Reply to Anonymous Referee #1

General comments

The authors have developed a new approach to retrieve CCN number concentrations. It is my opinion that this topic is one of the most important for tackling the uncertainties surrounding climate radiative forcing. The method here is unique in that it utilizes some machine learning along with lidar and in-situ measurements to derive CCN number concentration. I found the work to be interesting and of high quality, but the methods and figure presentation need some revision and/or clarification. After my comments and concerns have been addressed, I feel the manuscript will be suitable for publication in AMT. Therefore, I recommend acceptance after some minor revision.

Response:

Thanks for your encouraging comments. The manuscript has been revised according to your suggestions. Please see the responses to the specific comments.

Major comments

One major point I would like to see addressed is on the performance of the humidogram parameter estimates. The N_{CCN} retrieval heavily relies on calculated dry optical parameters (dry angstrom exponent, dry lidar ratio) which are determined from the fitted dry extinction/backscatter, and the κ parameter. I'd like to see a figure of example MWRL profiles that were used in this study with the model fit lines. A statistical summary plot would suffice if there is large scatter. The actual fits and profiles would be highly beneficial for me as the reader to visually assess the fit performance and also validate the layer selection that was mentioned in section 3.3.2.

Response:

Our paper presents the methodlogy based on theoretical simulation, and the methodlogy is not applied to real cases. All the results in our paper are only based on in situ measurements and there is no real measured lidar data. The MWRL backscatter and extinction data mentioned in the paper is all simulated using in situ data and Mie model. In case of more misunderstandings, we have added an introduction at the beginning of Sect. 2:

'Since it is not easy to accumulate large datasets of simultaneous measurements of lidars and aircrafts, ground-measured aerosol microphysical and chemical data are used to simulate lidar-derived backscatter and extinction coefficients and corresponding CCN number concentrations. The simulations are based on κ-Köhler theory and Mie theory. The required datasets include: particle number size distribution (PNSD), black carbon (BC) mass concentrations (m_{BC}), mixing state of BC containing particles, and size-resolved hygroscopicity. The simulation results are used to establish and validate the new retrieval method.'

Our method is based on many previous studies about aerosol hygroscopicity using lidar techniques (Wulfmeyer and Feingold, 2000;Feingold and Morley, 2003;Pahlow et al., 2006;Fernández et al., 2015;Rosati et al., 2016;Fernández et al., 2017;Haarig et al., 2017;Lv et al., 2017;Bedoya-Velásquez et al., 2018). An example from Bedoya-Velásquez et al. (2018) is shown below. Figure 3 from Bedoya-Velásquez et al. (2018) (Fig. R1) explains how to derive backscatter enhancement factors using lidar-retrieved backscatter profiles and RH profiles with the selection criteria in Sect. 3.3.2. The method in our paper mainly focus on the procedure start from the variations of backscatter and extinction with RH. We assume that Mie model simulated dataset can represent actual lidar measurements. Figure 7 from Bedoya-Velásquez et al. (2018) (Fig. R2) shows lidar observation and model simulation are in good agreements, especially for RH below 90%. Therefore, we think it is reasonable to use Mie model simulated backscatter and extinction at different RH in this study.

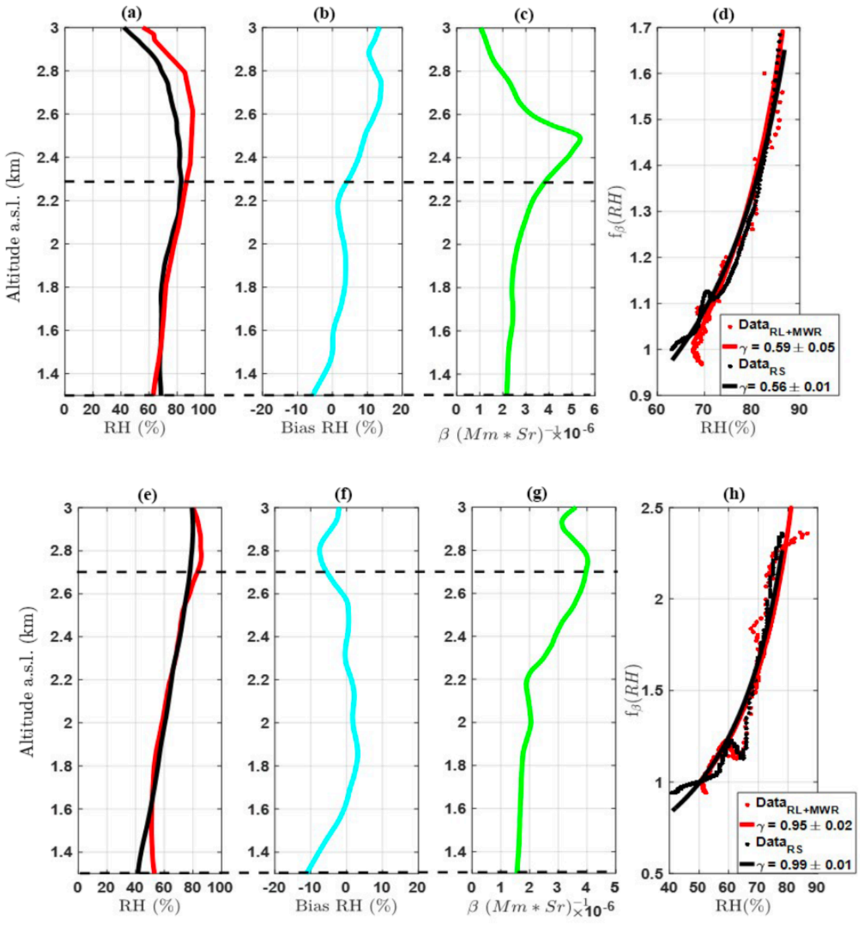


Figure 3. **(a, e)** Profiles of RH retrieved from RS (black line) and by the synergy RL + MWR (red line), **(b, f)** RH bias profiles (cyan line), **(c, g)** β_{par} retrieved by using the Klett–Fernald algorithm and lidar ratio of 65 Sr (green line), and **(d, f)** f_β (RH) calculated for RS (black dots) and by the synergy RL + MWR (red dots) and the corresponding Hänel parameterizations (solid lines), where the red line refers to the RL + MWR method (case I: $\gamma = 0.59 \pm 0.05$, case II: $\gamma = 0.95 \pm 0.02$) and the black line refers to the RS method (case I: $\gamma = 0.56 \pm 0.01$, case II: $\gamma = 0.99 \pm 0.01$). The top row corresponds to case I (22 July 2011, 20:30–21:00 UTC) and the bottom row to case II (22 July 2013, 20:00–20:30 UTC). Horizontal dashed lines indicate the altitude range analyzed for each case (1.3 to 2.3 km for case I and 1.3 to 2.7 km for case II). All these profiles were measured at the EARLINET IISTA-CEAMA station.

Figure R1. Example of deriving backscatter enhancement factors using lidar-retrieved backscatter profiles and RH profiles. (Figure 3 in Bedoya-Velásquez et al. (2018))

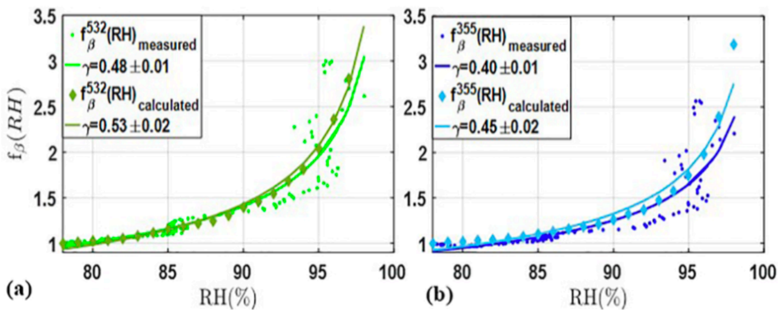


Figure 7. Humidograms calculated **(a)** at 532 nm and **(b)** at 355 nm, within the 1.5 to 2.4 km a.s.l. aerosol layer from the RL + MWR measurements and calculated using Mie theory and measured chemical composition and size distribution at 2.5 km a.s.l. RH_{ref} = 78 % was used for both methods.

Figure R2. Comparison of lidar-derived humidogram curves and modelled humidogram curves. (Figure 7 in Bedoya-Velásquez et al. (2018))

Here we use Mie model simulation to show the performance of the two humidogram parameterization. Figure R3 gives an example of humidogram fitting. All the dots represent Mie model simulations, and the dots in red (within RH range of 60-90%) are used to fit parameterization lines. The blue line is the result of γ-equation, and the green line represents the result of κ-equation. Both equations fit quite well for RH range of 60-90%. However, κ-equation has a better performance on estimating optical properties at dry condition. The figure has been added to supplement file of the paper.

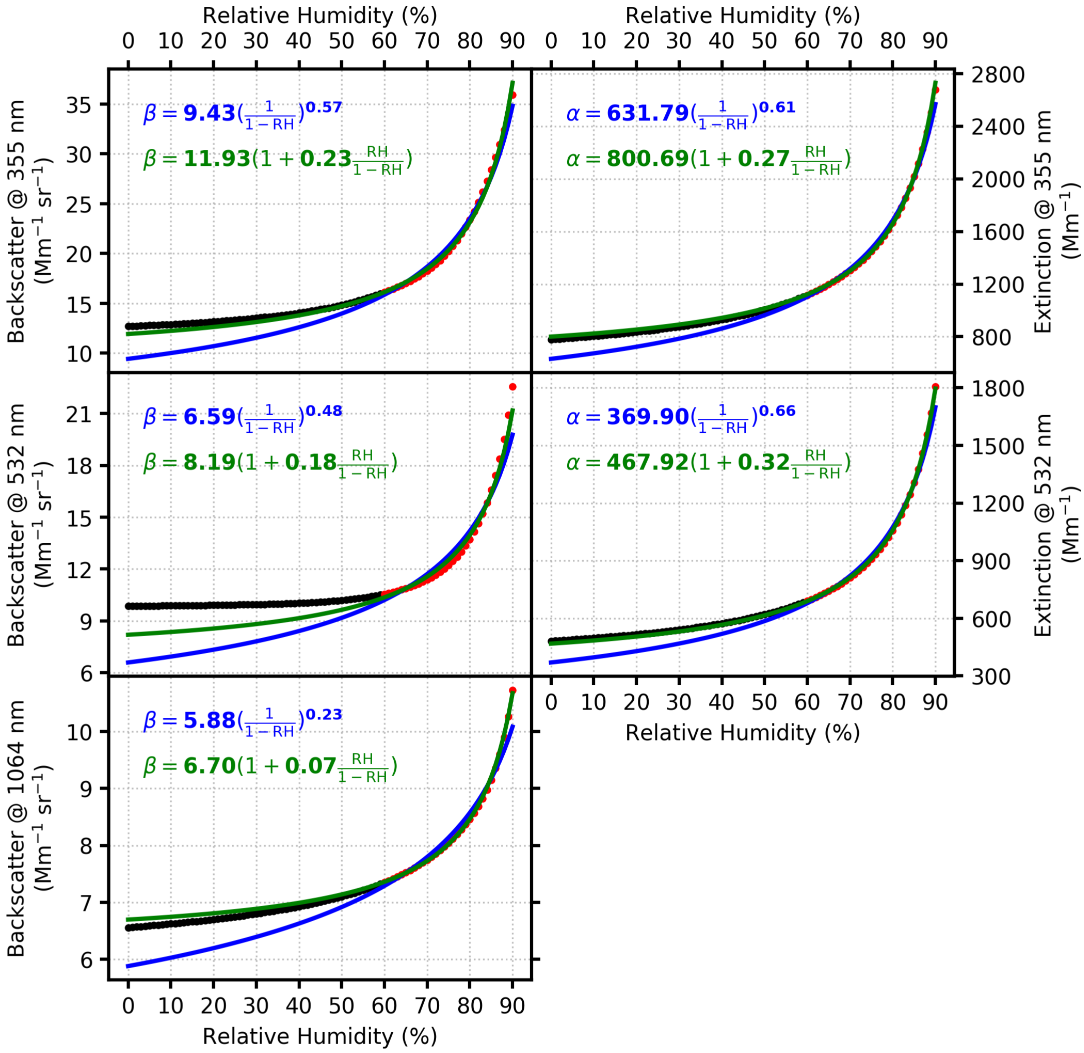


Figure R3. Example of humidogram fitting using different functions. The example is calculated with one set of PNSD, BC, r_{ext}, and size-resolved κ distribution.

Figure R4 gives the performance on the estimation of dry backscatter and extinction (Same as Table 2). The figure only shows the results of RH range 60-90%. Dry optical properties fitted with κ-equation agree better with Mie model simulations, especially for extinction. The figure has been added to supplement file of the paper.

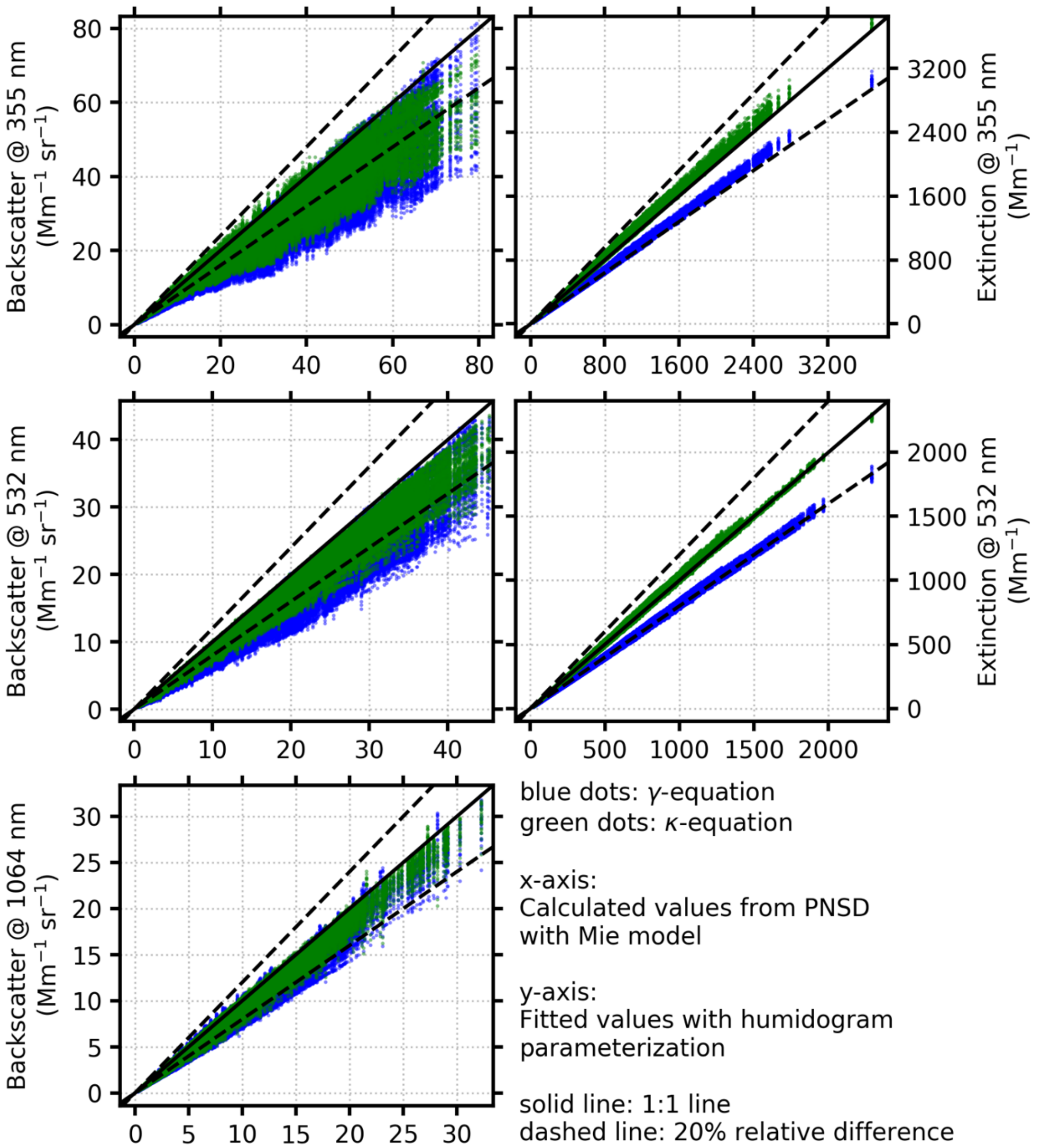


Figure R4. Comparison between Mie model calculated dry particle backscatter and extinction and those fitted from humidograms.

Also in reference to the above comment, Table 3 is confusing to me. I'm not sure what information is gained by partitioning the humidogram curve in a way that ignores the high or the low end and then comparing the "partial" fit coefficient to the full range fit coefficient, since it is the entirety of the curve that describes the aerosol chemistry/size distribution properties. I recommend instead, that the authors show the parameterization statistics to the fitted data (e.g. RMSE or another metric) rather than what is shown now in Table 3.

Response:

Yes, it is the whole curve that describes the aerosol properties. The motivation of showing Table 3 is based on practical situation of lidar observations. Unlike in situ measurements, hygroscopicity studies based on lidar measurements are facing a problem that we cannot manually control the RH we measure. The RH range that lidar observation can get is limited by the actual RH profile and the well-mixed assumption. Therefore, for every hygroscopic case of lidar, the RH ranges can be limited and very different. Humidogram parameters fitted from backscatter and extinction enhancements with RH (i.e. γ_ξ or κ_ξ) are often used to represent the hygroscopicity of particles. However, if large difference in γ_ξ or κ_ξ for different RH range accrues for the same group of particles, then the humidogram parameters may not be comparable for different cases that have different RH ranges. That is what we want to stress through the comparison in Table 3.

For our study, the values of κ_ξ are important for retrieving CCN. The RH range is also important. For example, the random forest model is now trained by κ_ξ fitted from RH range of 60-90%, and the data collect from lidar measurements only contains RH within 80-90%. The results are presented in Fig. R5. Compared to the results in Fig. 4 in the paper, more uncertainties will arise if RH range is different between the training and test data.

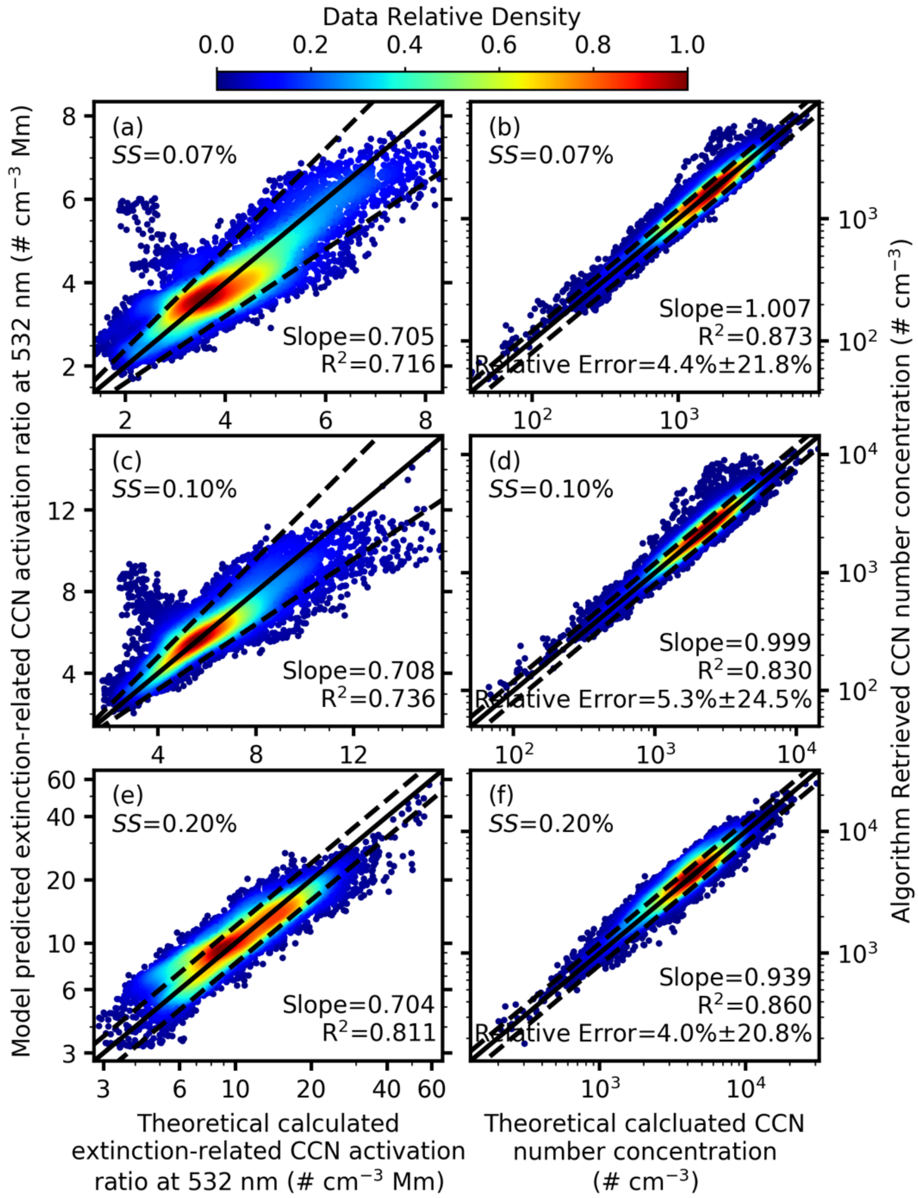


Figure R5. Same as Fig. 4 in the paper. The training data is the same as Fig. 4 which is derived from RH range of 60-90%, but the test data is derived from RH range of 80-90%.

I can appreciate investigating the performance of the parameterizations as a function of the RH range. But I don't think the breakdown here is important. The κ parameterization has been well compared to the γ parameterization as the authors note with the Brock et al. (2016) reference. Of more interest to me is the visual performance of these fits. I'd still like to see these in addition to the discussion and the Table entries that are already in the manuscript.

Response:

As you addressed, the performance of the two parameterization has been well evaluated by Brock et al. (2016). Most of the comparisons focus on the performance on describing enhancement values at different RH, especially high RH, so we did not show many evaluation results on this performance. You can see both parameterizations fit quite well for high RH ranges from the figure in Fig. R3 and Fig. R4. Compared to the accuracy at high RH values, the estimated dry optical properties are much more important to our method. Therefore, we paid more attention to the performance on estimating dry backscatter and extinction. Through the investigation of the performance as a function of RH range, we found that if the data has lower RH, the estimated dry backscatter and extinction will be closer to the simulated values (Table 2). As for the visual performance, we have put Fig. R3 and Fig. R4 to the supplement file.

Figure 1. I think this figure could be constructed instead by normalizing by the maximum value at the peak diameter instead of total number. As it is constructed now, the range in the y-axis values makes this figure hard to interpret. If normalizing by the maximum value at the peak diameter (so that each distribution peaks at 1 rather than something less than 1) doesn't result in much change, you could also consider a time-series with diameter on the y-axis and colors representing normalized PNSD.

Response:

Figure 1a is now reconstructed by normalizing the PNSD by the maximum value at the peak diameter (Figure R6). The time series of the normalized PNSD is shown in Fig. R7. The figure has been added to the Supplement.

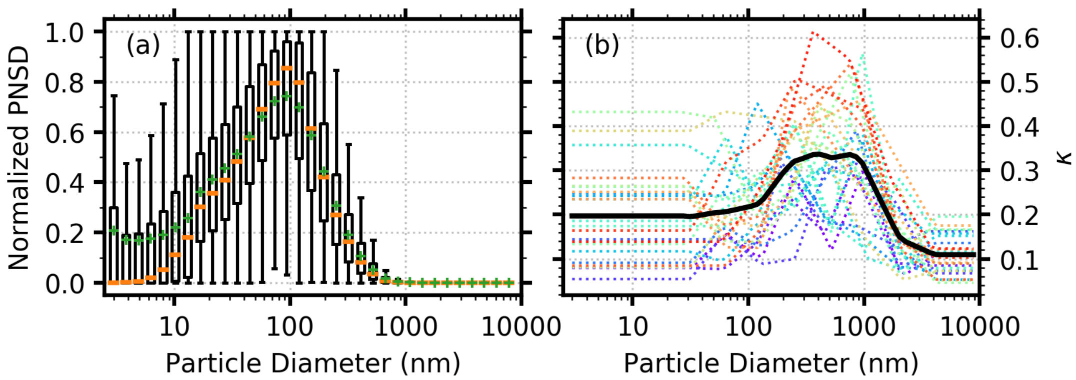


Figure R6. Reconstructed Fig. 1.

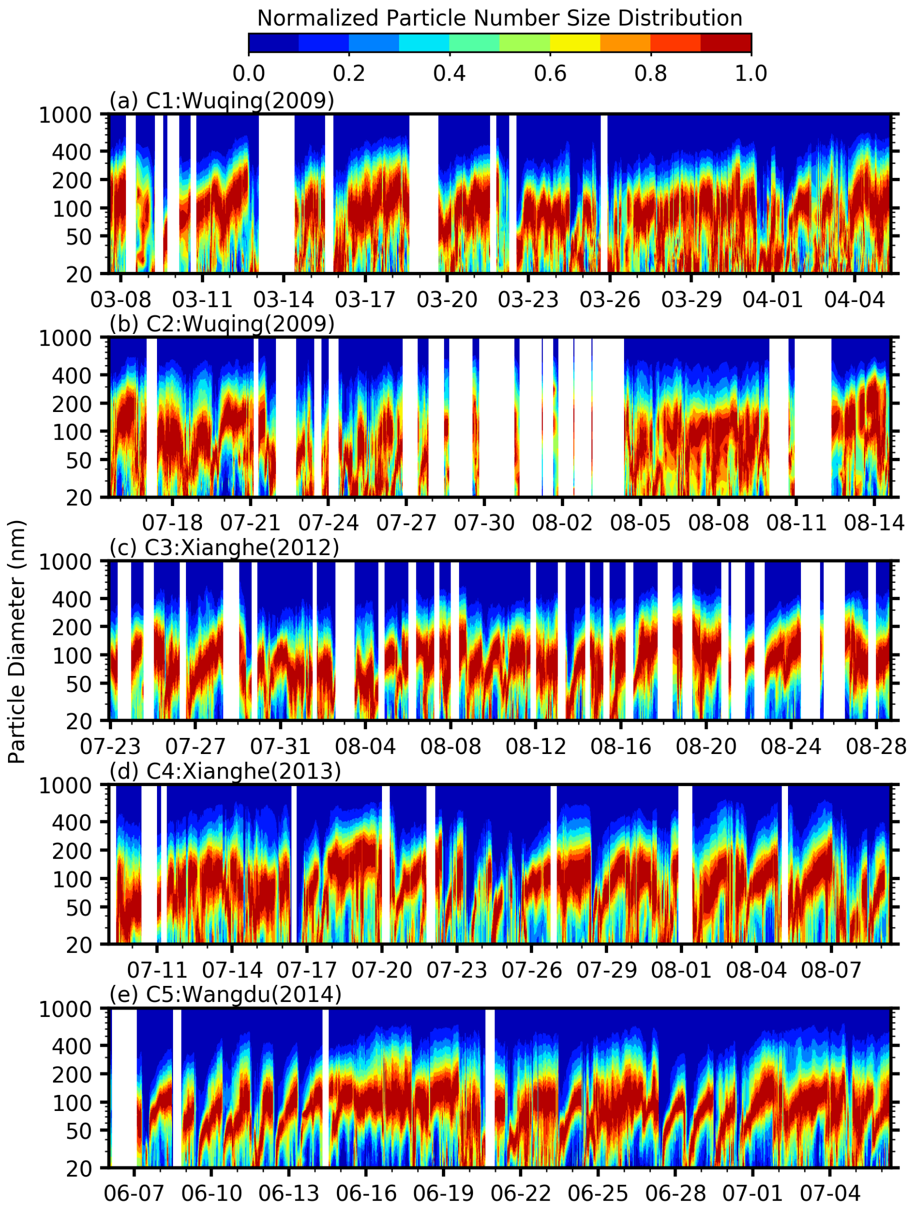


Figure R7. Time series of the normalized PNSD.

Minor comments

Table 4 and Text The 9 parameters that are selected to determine AR_{ξ} are declared to have no explicit expressions between them and have highly non-linear relationships (pg. 8 line 23-24). It's not clear that the normalized extinction at 532 nm and at 355 nm, for example, would have different enough κ fit parameters to yield information for AR_{ξ}. Could you comment on this or possibly add a supplementary figure that would give support to the statement that there is no explicit expression between the 9 selected parameters? Worded another way, what information content can be gained from a spectrally dependent hygroscopicity fit parameter?

<u>Response:</u>

Thanks for your comment and suggestion. There should be some linear relationship between some parameters. Figure R8 shows the relationship between fitted humidogram parameter of extinction ($\kappa_{\alpha 355}$) and other 8 parameters.

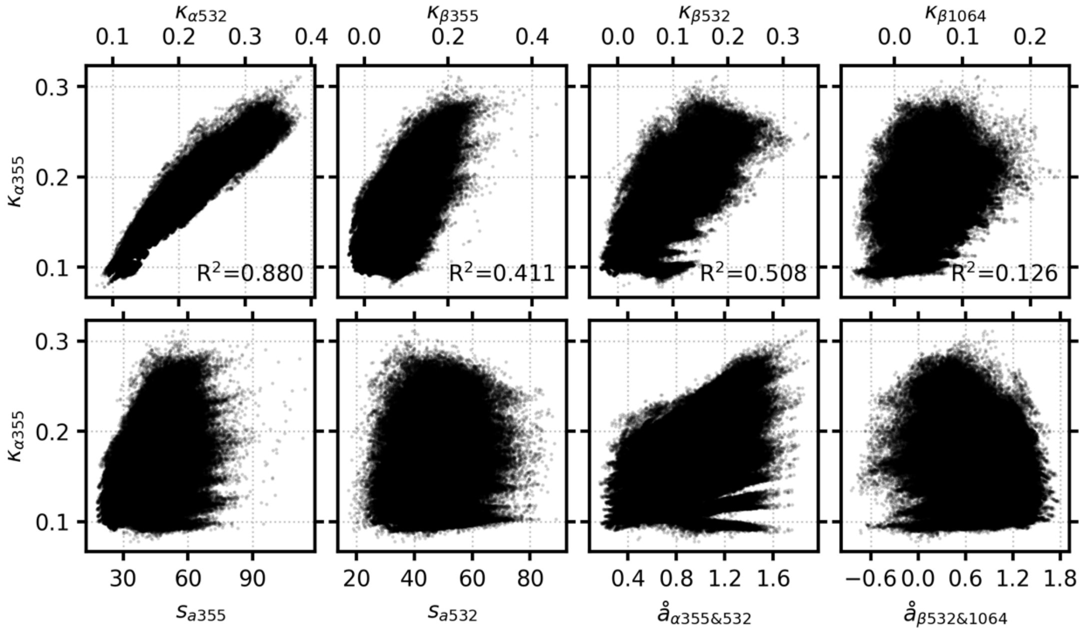


Figure R8. Relationship between $\kappa_{\alpha 355}$ and other 8 parameters.

High correlation (R^2=0.880) is found between humidogram parameters fitted from extinction at 355 nm and 532 nm. R^2 between $\kappa_{\alpha 355}$ and $\kappa_{\beta 1064}$ is only 0.126 which quite fits the results from Fig. 5 in the paper. The two parameters with low correlation are both important to predict AR_{ξ}.

We also present Table R1 to show the relationship between every two of the nine parameters. The table indicates that some parameters are linear correlated to some extent. Accordingly, we remove the sentence you mention.

Table R1. Determine coefficients (R^2) between input parameters. R^2 larger than 0.5 are marked in red.

R^2	$\kappa_{\alpha 355}$	$\kappa_{\alpha 532}$	$\kappa_{\beta 355}$	$\kappa_{\beta 532}$	$\kappa_{\beta 1064}$	S_{a355}	S_{a532}	$\mathring{a}_{\alpha 355\&532}$
$\kappa_{\alpha 355}$	–	–	–	–	–	–	–	–
$\kappa_{\alpha 532}$	0.880	–	–	–	–	–	–	–
$\kappa_{\beta 355}$	0.411	0.321	–	–	–	–	–	–
$\kappa_{\beta 532}$	0.508	0.644	0.450	–	–	–	–	–
$\kappa_{\beta 1064}$	0.126	0.222	0.016	0.056	–	–	–	–
S_{a355}	0.085	0.073	0.680	0.292	0.019	–	–	–
S_{a532}	0.026	0.070	0.117	0.423	0.070	0.360	–	–
$\mathring{a}_{\alpha 355\&532}$	0.149	0.135	0.505	0.267	0.027	0.627	0.089	–
$\mathring{a}_{\beta 532\&1064}$	0.062	0.023	0.550	0.169	0.464	0.409	0.023	0.317

As for the information content of spectrally dependent κ_{ξ}, we have tried to explain it with Fig. 2b in the paper. The enhancement contribution is influenced by both hygroscopicity and number concentration of each size. Still, from the figure, we can see different κ_{ξ} is sensitive to the hygroscopicity of different size. We want to obtain size-dependent hygroscopicity information with the use of spectrally dependent κ_{ξ}, because size-dependent hygroscopicity is important to estimate CCN rather than a bulk hygroscopicity information, especially for different supersaturation conditions. One humidogram may indicate the bulk hygroscopicity, but it is the hygroscopicity of small particles that influences CCN number concentrations most. Spectrally dependent κ_{ξ} can provide some information about the hygroscopicity of small particles. Also from Fig. 5 in the paper which show the relative importance of the input parameters for random forest model, humidogram parameters of extinction at 355 nm and backscatter at 1064 nm are rather important. Therefore, at least these two parameters are needed for the retrieval method. We have added the explanation above to the corresponding text.

Training set I think the exercise would be more convincing if you divided your entire dataset into a training and test class where, for example, 70% of the data is randomly chosen for training and the other 30% is used for the test performance.

Response:

It is true that many studies randomly select their training and test class. One thing I am concerning here is that our dataset is from continuously in situ measurements. Two observations of very close time are likely to be similar. If the dataset is divided randomly, the random forest model can possibly be trained and tested with two very similar dataset. We have tried random selection strategy and found that the test results are better than the strategy used now in the paper. Good agreements are found even in high supersaturations if the data is divided randomly for training and testing. Thus, we think the random selection strategy is not appropriate for our data. We train the model with data from some sites and test with data from other sites, in order to test the feasibility of the method at different locations. There are some papers used the same strategy as we do (Kuang et al., 2018;Zhao et al., 2018).

The Mie model results are calculated for the entire range of 25 kappa size resolved distributions but how is the final result selected for comparison to the MWRL retrieval method? Could you more clearly state this somewhere appropriate? Could you also, before section 2.2, explain the significance of a “size-resolved" kappa distribution? Kappa is thought to be size-independent for particles of certain chemical composition. It might be important to include a sentence or two explaining that the size-resolved kappa distribution is for particles of changing chemical composition with the Liu et al. (2014) reference.

Response:

Thanks for your suggestion.

Since no concurrent measured microphysical and chemical properties, we conducted the simulation with all 25 size-resolved κ distributions for every PNSD and did not select them for application. We trained the random forest model with all the results of 25 size-resolved κ distributions. Actually, there should be some relationship between particle dry optical properties and their hygroscopicity (e.g. black carbon influences both lidar ratio and hygroscopicity very much). More details can be discussed if more concurrent measured data is available in the future. In the paper, we applied all 25 distributions to every PNSD in order to cover various situations. Training the random forest model without selection, to some extent, is reasonable because the simulations contain all situations.

We have added the following statement at the end of Sect. 2.2:

‘The new method and all the analyses in this paper are based on the Mie model simulated datasets, and all the simulations mentioned above are implemented.’

And we also added a sentence at the end of Sect. 2.1 to explain the size-resolved κ distribution:

‘The chemical compositions are found to be size dependent during the campaign C2, especially the mass fraction of organic matter (Liu et al., 2014).’

Technical corrections

Table 2. Can you clarify in the caption if the Mie model simulated dry parameters are from the measured PNSDs?

<u>Response:</u>

Thanks for your suggestion. We have rewritten the caption:

'Slopes of linear regressions, determination coefficients (R^2), and relative errors (RE) between Mie model simulated particle dry backscatter or extinction coefficients and those inferred from humidogram functions. 404575 pairs of the simulations from in situ dataset are used. The RE are given in the form of mean value ± one standard deviation (std).'

Pg. 1 Line 18 change "datasets" to "dataset"

Pg. 1 Line 25 change "a huge" to "an"

Pg. 2 Line 8 change "always" to "can"

Pg. 2 Line 10 add the word in brackets: "in [the] natural"

Pg. 2 Line 13 change "showing" to "suggesting"

Pg. 2 Line 18 add characters in brackets: "Existing approach[es]"

Pg. 2 Line 27 add the word in brackets: "...humidified in [the] ambient..."

Pg. 3 Line 5 add the word in brackets: "...hints [at] the ability..."

Pg. 3 Line 10 remove sentence beginning "Enhancements of ..."

Pg. 3 Line 13 add an "s" to scheme and humidogram

<u>Response:</u>

Thanks! Corrections have been made according to your suggestions.

Pg. 7 Line 17 can you provide a reference for the backscatter angstrom exponent relationship?

<u>Response:</u>

The following reference (Komppula et al. 2012) has been added to the corresponding text:

'Komppula, M., Mielonen, T., Arola, A., Korhonen, K., Lihavainen, H., Hyvärinen, A. P., Baars, H., Engelmann, R., Althausen, D., Ansmann, A., Müller, D., Panwar, T. S., Hooda, R. K., Sharma, V. P., Kerminen, V. M., Lehtinen, K. E. J., and Viisanen, Y.: Technical Note: One year of Raman-lidar measurements in Gual Pahari EUCAARI site close to New Delhi in India – Seasonal characteristics of the aerosol vertical structure, Atmos. Chem. Phys., 12, 4513-4524, 10.5194/acp-12-4513-2012, 2012.'

Pg. 7 Line 28-29 Rewrite the sentence beginning "Particle type information..." as follows: "The lidar ratio can provide information on particle type and also serve as a proxy for particle hygroscopicity. Therefore, the lidar ratio of dry particles could be a reliable parameter to estimate AR_{ξ}." or something like this.

Pg. 8 Line 1 change "huge" to "large".

Pg. 8 Line 26 add characters in brackets: "...been a field that [has] develop[ed] rapidly..."

Pg. 10 Line 8 add characters in brackets: "...lidars may not [be] sufficient enough..." Pg. 10 Line 14 change "huge" to "the"

Pg. 10 Line 30 change "In average" to "On average"

<u>Response:</u>

Thanks! Corrections have been made according to your suggestions.

Pg. 11 Line 3 rewrite the sentence begging with "Bigger...." This needs to clearly state that smaller particles have larger angstrom exponents and bigger particles have smaller angstrom exponents (or more compactly, extinction angstrom exponents are inversely proportional to particle size). Do you mean here that $\alpha_{355:532} > \alpha_{532:1064}$ means smaller particles? I'm not sure if that's true since the relationship is complex (e.g. Schuster et al., 2006; doi:10.1029/2005JD006328)

Response:

Thanks for your suggestion. The symbol 'α_{355}' here represents extinction coefficient at 355 nm but not Ångström exponent. What I mean here is just that if α_{355} is overestimated and backscatter and extinction at other wavelengths remain unchanged, the corresponding Ångström exponent will become bigger. The sentence has been removed since the result here is obvious and needs no more detailed discussion.

Pg. 12 Line 20 change to "It should be [noted]...."

Response:

Thanks! Correction has been made.

Reference

Bedoya-Velásquez, A. E., Navas-Guzmán, F., Granados-Muñoz, M. J., Titos, G., Román, R., Casquero-Vera, J. A., Ortiz-Amezcua, P., Benavent-Oltra, J. A., de Arruda Moreira, G., Montilla-Rosero, E., Hoyos, C. D., Artiñano, B., Coz, E., Olmo-Reyes, F. J., Alados-Arboledas, L., and Guerrero-Rascado, J. L.: Hygroscopic growth study in the framework of EARLINET during the SLOPE I campaign: synergy of remote sensing and in situ instrumentation, Atmos. Chem. Phys., 18, 7001-7017, 10.5194/acp-18-7001-2018, 2018.

Feingold, G., and Morley, B.: Aerosol hygroscopic properties as measured by lidar and comparison with in situ measurements, J. Geophys. Res. Atmos., 108, n/a-n/a, 10.1029/2002JD002842, 2003.

Fernández, A. J., Apituley, A., Veselovskii, I., Suvorina, A., Henzing, J., Pujadas, M., and Artíñano, B.: Study of aerosol hygroscopic events over the Cabauw experimental site for atmospheric research (CESAR) using the multi-wavelength Raman lidar Caeli, Atmos. Environ., 120, 484-498, http://dx.doi.org/10.1016/j.atmosenv.2015.08.079, 2015.

Fernández, A. J., Molero, F., Becerril-Valle, M., Coz, E., Salvador, P., Artíñano, B., and Pujadas, M.: Application of remote sensing techniques to study aerosol water vapour uptake in a real atmosphere, Atmos. Res., https://doi.org/10.1016/j.atmosres.2017.11.020, 2017.

Haarig, M., Ansmann, A., Gasteiger, J., Kandler, K., Althausen, D., Baars, H., Radenz, M., and Farrell, D. A.: Dry versus wet marine particle optical properties: RH dependence of depolarization ratio, backscatter, and extinction from multiwavelength lidar measurements during SALTRACE, Atmos. Chem. Phys., 17, 14199-14217, 10.5194/acp-17-14199-2017, 2017.

Kuang, Y., Zhao, C. S., Zhao, G., Tao, J. C., Xu, W., Ma, N., and Bian, Y. X.: A novel method for calculating ambient aerosol liquid water content based on measurements of a humidified nephelometer system, Atmos. Meas. Tech., 11, 2967-2982, 10.5194/amt-11-2967-2018, 2018.

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

2014, 2014.

Lv, M., Liu, D., Li, Z., Mao, J., Sun, Y., Wang, Z., Wang, Y., and Xie, C.: Hygroscopic growth of atmospheric aerosol particles based on lidar, radiosonde, and in situ measurements: Case studies from the Xinzhou field campaign, J. Quant. Spectrosc. Ra., 188, 60-70, 10.1016/j.jqsrt.2015.12.029, 2017.

Pahlow, M., Feingold, G., Jefferson, A., Andrews, E., Ogren, J. A., Wang, J., Lee, Y. N., Ferrare, R. A., and Turner, D. D.: Comparison between lidar and nephelometer measurements of aerosol hygroscopicity at the Southern Great Plains Atmospheric Radiation Measurement site, J. Geophys. Res. Atmos., 111, n/a-n/a, 10.1029/2004JD005646, 2006.

Rosati, B., Herrmann, E., Bucci, S., Fierli, F., Cairo, F., Gysel, M., Tillmann, R., Größ, J., Gobbi, G. P., Di Liberto, L., Di Donfrancesco, G., Wiedensohler, A., Weingartner, E., Virtanen, A., Mentel, T. F., and Baltensperger, U.: Studying the vertical aerosol extinction coefficient by comparing in situ airborne data and elastic backscatter lidar, Atmos. Chem. Phys., 16, 4539-4554, 10.5194/acp-16-4539-2016, 2016.

Wulfmeyer, V., and Feingold, G.: On the relationship between relative humidity and particle backscattering coefficient in the marine boundary layer determined with differential absorption lidar, J. Geophys. Res. Atmos., 105, 4729-4741, 10.1029/1999JD901030, 2000.

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

Method to retrieve cloud condensation nuclei number concentrations using lidar measurements

Wangshu Tan[1], Gang Zhao[1], Yingli Yu[1], Chengcai Li[1], Jian Li[2], Ling Kang[3], Tong Zhu[3], Chunsheng Zhao[1]

[1]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China
[2]State Key Laboratory of Severe Weather & Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing 100081, China
[3]State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Science & Engineering, Peking University, Beijing 100871, China

Correspondence to: Chengcai Li (ccli@pku.edu.cn)

Abstract. Determination of cloud condensation nuclei (CCN) number concentrations at cloud base is important to constrain aerosol-cloud interactions. A new method to retrieve CCN number concentrations using backscatter and extinction profiles from multiwavelength Raman lidars is proposed. The method implements hygroscopic enhancements of ~~backscatter/extinction~~backscatter and extinction with relative humidity to derive dry ~~backscatter/extinction~~backscatter and extinction and humidogram parameters. Humidogram parameters, Ångström exponents, and lidar extinction-to-backscatter ratios are then linked to the ratio of CCN number concentration to dry ~~backscatter/extinction~~backscatter and extinction coefficient (AR_{ξ}). This linkage is established based on the datasets simulated by Mie theory and κ-Köhler theory with in situ measured particle size distributions and chemical compositions. CCN number concentration can thus be calculated with AR_{ξ} and dry ~~backscatter/extinction~~backscatter and extinction. An independent theoretical simulated dataset~~s~~ is used to validate this new method and results show that the retrieved CCN number concentrations at supersaturations of 0.07%, 0.10%, and 0.20% are in good agreement with theoretical calculated values. Sensitivity tests indicate that retrieval error in CCN arise mostly from uncertainties in extinction coefficients and RH profiles. The proposed method improves CCN retrieval from lidar measurements and has great potential in deriving scarce long-term CCN data at cloud base which benefits aerosol-cloud interaction studies.

1 Introduction

Anthropogenic activities have caused ~~a huge~~an increase in atmospheric aerosols, and some of the aerosol particles affect the climate by serving as cloud condensation nuclei (CCN). CCN in clouds can modify cloud forming processes and cloud microphysical properties (Rosenfeld et al., 2014). Although numerous impacts of aerosol-cloud interactions on radiative forcing (McCoy et al., 2017;Zhou et al., 2017), precipitation (Xu et al., 2017;Fan et al., 2018), cloud electrification (Wang et al., 2018), and severe weathers or hazards (Fu et al., 2017) have been discovered, constraining the relationships between aerosols and clouds is still a big challenge (Seinfeld et al., 2016). Lacking the knowledge of aerosol-cloud interactions limits our ability to estimate climate forcing caused by aerosols (Boucher et al., 2013).

Aerosol CCN supersaturation activation spectrum is one of the most critical parameters to quantify aerosol-cloud interactions (Schmale et al., 2018). Despite that a large amount of CCN number concentrations near ground have been measured worldwide (Tao et al., 2017), ground-measured CCN may not represent CCN at cloud base that alter clouds directly. Obtaining CCN near cloud base becomes a crucial issue. Cloud base CCN can be measured in situ on aircraft platforms, but airborne measurements have the limitations of huge costs and discontinuity. Satellites are difficult to observe CCN at cloud base, because clouds ~~always~~can obscure aerosol signals beneath them. Rosenfeld et al. (2016) have proposed an alternative approach for satellites to retrieve CCN concentrations using clouds as CCN chambers, however, employing CCN concentrations derived with this strategy limits our exploration of the relationship between CCN concentrations and cloud droplet concentrations in the natural environment. So far, CCN concentrations at cloud base are scarce for aerosol-cloud interaction studies.

Ground-based lidars can continuously provide optical properties of aerosol particles from ground up to cloud base (Mattis et al., 2016;Barreto et al., 2019), ~~showing~~suggesting great potential in deriving CCN concentrations near cloud base. Ghan and Collins (2004) propose a simple method to infer CCN profiles with the combination of surface in situ CCN and aerosol optical measurements. The method is only applicable when boundary layer is well mixed from surface to cloud base (Ghan et al., 2006). Multiwavelength Raman lidars (MWRLs) are increasingly used to detect aerosol vertical distributions in recent years. The principle of MWRLs allows independent retrieval of particle backscatter (β) and extinction coefficients (α), which provides more information about particle microphysical properties (Müller et al., 2016). The $3\beta+2\alpha$ MWRL systems (backscatter coefficients at 355, 532, and 1064 nm and extinction coefficients at 355 and 532 nm) have been widely recommended to derive particle microphysical properties (Burton et al., 2016). Existing approaches to retrieve CCN using MWRLs ~~is~~are based on microphysical inversion techniques (Mamouri and Ansmann, 2016). Retrieved optical-equivalent particle size distributions together with assumed activation critical diameters are utilized to calculate CCN concentrations (Lv et al., 2018).

There are three major challenges in CCN concentration retrieval with lidars. The first is the conversion of lidar-derived optical properties into particle number concentrations. High uncertainties of retrieved particle number concentrations could be an important source of CCN retrieval error. The second one is the determination of particle hygroscopicity in order to evaluate the ability of particles to participate as CCN. Particle hygroscopicity, which is highly related to chemical composition and aging/coating effect, is found to cause nonnegligible variations in cloud droplet activation (Hudson, 2007;Zhang et al., 2017). The last is the influence of high relative humidity (RH) near clouds. Aerosol particles are likely to be humidified in the ambient environment, and the consequent changes in optical properties make CCN retrieval more ~~complicated~~challenging. Most studies working on CCN retrieval with MWRLs mainly focus on deriving particle number concentrations, but seldom commence to solve the issue of hygroscopicity.

In recent years, several aerosol hygroscopic studies based on lidar measurements have been carried out (Fernández et al., 2017;Lv et al., 2017;Bedoya-Velásquez et al., 2018). Backscatter and extinction enhancement factors can be derived with lidar measurements and RH profiles. The enhancement factor, which is associated with both particle size and hygroscopicity (Kuang et al., 2017), is defined as:

$$f_\xi(\mathrm{RH},\lambda) = \frac{\xi(\mathrm{RH},\lambda)}{\xi(\mathrm{RH_{ref}},\lambda)}, \quad (1)$$

where f_ξ is the enhancement factor of the optical property ξ (backscatter or extinction) at a specific light wavelength λ and RH, and $\mathrm{RH_{ref}}$ is the reference RH value. Many studies manifest that lidar-derived enhancement factors are in good agreement with in situ measurements (Wulfmeyer and Feingold, 2000;Pahlow et al., 2006;Fernández et al., 2015;Rosati et al., 2016). Feingold and Morley (2003) demonstrate that the extent of backscatter and extinction enhancements hints at the ability of particles to serve as CCN. Tao et al. (2018) use in situ measured light scattering enhancement factors to predict N_{CCN} at 0.07% supersaturation, and the result shows strong consistency with CCN counter.

~~In this paper, a new method to retrieve CCN number concentrations for 3β+2α MWRL systems (backscatter coefficients at 355, 532, and 1064 nm and extinction coefficients at 355 and 532 nm) is proposed based on κ-Köhler theory (Petters and Kreidenweis, 2007) and Mie theory (Bohren and Huffman, 2007). Enhancements of backscatter and extinction with RH are first implemented in CCN retrieval using MWRLs.~~ In this paper, a new method to retrieve CCN number concentrations for 3β+2α MWRL systems is proposed. Theoretical simulations are carried out to seek the relationship between CCN number concentrations and lidar-derived optical properties. The simulation implements κ-Köhler theory (Petters and Kreidenweis, 2007) to describe particle hygroscopic growth and activation process. Mie theory (Bohren and Huffman, 2007) is utilized to calculate particle backscatter and extinction coefficients from in situ measured aerosol microphysical and chemical properties. The paper is structured as follows. Section 2 introduces the measured and simulated datasets used in this paper. ~~Section 3 presents the methodology. Firstly, suitable supersaturation conditions for lidar retrieval are discussed in Sect 3.1. Performances of two parameterization scheme for backscatter and extinction humidogram are evaluated in Sect 3.2. In Sect. 3.3, the~~The new CCN retrieval method for MWRLs and the sensitivity tests are respectively described in ~~detail~~ Sect. 3.1 and Sect. 3.2. Sensitivity tests are carried out in Sect. ~~3.4~~3.2. Results and ~~summary~~ discussions are given in Sect. 4 ~~and Sect. 5, respectively~~. Section 5 summarizes the paper.

2 Data

Since it is not easy to accumulate large datasets of simultaneous measurements of lidars and aircrafts, ground-measured aerosol microphysical and chemical data are used to simulate lidar-derived backscatter and extinction coefficients and corresponding CCN number concentrations. The simulations are based on κ-Köhler theory and Mie theory. The required datasets include: particle number size distribution (PNSD), black carbon (BC) mass concentrations (m_{BC}), mixing state of BC containing particles, and size-resolved hygroscopicity. The simulation results are used to establish and validate the new retrieval method.

2.1 Datasets of aerosol microphysical and chemical properties

In situ measured aerosol properties were collected from five field campaigns at three different measurement sites in the North China Plain (NCP). The measurement sites are located at Wuqing (39°23′ N, 117°01′ E, 7.4 m a.s.l) in Tianjin, Xianghe (39°45′ N, 116°58′ E, 36 m a.s.l) and Wangdu (38°40′ N, 115°08′ E, 51 m a.s.l) in Hebei province. The specific locations, topographical information, and pollution status of these measurement~~s~~ sites are shown in Fig. S1 in the Supplement. These three sites all lie inside the polluted NCP region and are highly representative of the polluted background (Xu et al., 2011;Bian et al., 2018;Sun et al., 2018). Time periods, measured parameters, and corresponding instruments of individual campaign are listed in Table 1. During these field campaigns, except measurement for size-resolved chemical compositions, ambient particles were drawn in through a PM10 inlet (16.67 L/min), passed through a silica gel diffusion drier, and then were split into different instruments. All instruments were operated at RH less than 30%.

The particle number size distributions (PNSDs) were measured with the combination of a twin differential mobility particle sizer (TDMPS, IfT, Leipzig, Germany) or a scanning mobility particle size spectrometer (SMPS) and an aerodynamic particle sizer (APS, TSI, Inc., Shoreview, MN USA, Model 3320 or Model 3321). The statistical information about the measured PNSDs is shown in Fig. 1a. The peaks of the PNSDs are at about 100 nm (diameter in log-scale), which shows strong characteristics of continental aerosols.

The black carbon (BC) mass concentrations (m_{BC}) were measured by a multi-angle absorption photometer (MAAP, Thermo, Inc., Waltham, MA USA, Model 5012). As for mixing states of BC, BC and other non-absorbing compositions were found to be both externally mixed and core-shell mixed during the campaigns (Ma et al., 2012). The mass fraction of externally-mixed BC (r_{ext}) is defined to quantify the mixing states of BC:

$$r_{\mathrm{ext}} = \frac{m_{\mathrm{ext\_BC}}}{m_{\mathrm{BC}}}, \qquad (2)$$

where $m_{\mathrm{ext\_BC}}$ is the mass concentration of externally mixed BC. According to Ma et al. (2012), r_{ext} can be retrieved from hemispheric backscattering fractions (HBFs) measured by an integrating nephelometer (TSI, Inc., Shoreview, MN USA, Model 3563).

Size-resolved chemical compositions all come from campaign C2. The size-resolved aerosol sampling was carried out with a ten-stage Berner low pressure impactor (BLPI). Chemical species including inorganic ions (NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, NO_3^-, SO_4^{2-}, Cl^-), elemental carbon, organic carbon, water-soluble organic carbon and some other species such as dicarboxylic acids were analyzed from sample substrates. After transforming the ambient wet aerodynamic diameters into dry volume-equivalent diameters, size-resolved κ distributions were derived from measured size-resolved chemical compositions. The chemical compositions are found to be size dependent during the campaign C2, especially the mass fraction of organic matter (Liu et al., 2014). Twenty-five typical size-resolved κ distributions in the NCP are given in Fig. 1b. The measured size-resolved κ distributions vary a lot and cover a wide range of aerosol hygroscopicity (Kuang et al., 2018). More details about the measurements can be found in Liu et al. (2014).

2.2 Datasets of CCN number concentrations and lidar-derived optical properties

In situ measured aerosol properties mentioned above are utilized to calculate CCN number concentrations and particle backscatter and extinction coefficients base on κ-Köhler theory and Mie theory. For each simultaneously measured PNSD, m_{BC}, and r_{ext} (16183 sets of data), simulations are carried out with every one of the twenty-five size-resolved κ distributions. CCN number concentrations can be calculated with PNSD and size-resolved κ distributions based on κ-Köhler equation. Petters and Kreidenweis (2007) introduce the κ-Köhler equation to describe the relationship between ~~particle/droplet~~particle or droplet diameter D and critical supersaturation ratio (SS) or RH with a single hygroscopic parameter κ:

$$\mathrm{RH}(D) = 1 + \mathrm{SS}(D) = \frac{D^3 - D_{\mathrm{dry}}^3}{D^3 - D_{\mathrm{dry}}^3(1-\kappa)} \exp\left(\frac{4\sigma_{\mathrm{s/a}} M_{\mathrm{w}}}{RT\rho_{\mathrm{w}} D}\right), \tag{3}$$

where D_{dry} is particle dry diameter, $\sigma_{\mathrm{s/a}}$ is the surface tension of the ~~solution/air~~solution-air interface, M_{w} is the molecular weight of water, R is the universal gas constant, T is temperature, and ρ_{w} is the density of water. For a specific supersaturation, critical activation diameter can be derived with κ-Köhler equation using size-resolved κ distributions. CCN number concentrations thereby can be calculated by integrating number concentrations of particles larger than the critical diameter. ~~Given PNSD at dry condition, SRKD, and r_{ext}, κ-Köhler equation can be used to estimate CCN number concentrations by calculating critical diameter.~~ CCN number concentrations at the supersaturations of 0.07%, 0.10%, 0.20%, 0.40%, and 0.80% are accordingly simulated. The selected supersaturation ratios are widely used in CCN measurements.

Particle backscatter and extinction can be calculated with PNSD, m_{BC}, and r_{ext} using Mie models. Mie theory can solve light scattering problems of homogeneous and coated spherical particles. Without the consideration of mineral dust, using Mie model is quite reasonable because particles are likely to be spherical near clouds where the RH ~~cloud~~ could be relatively high. When simulating particle backscatter and coefficients, PNSD, m_{BC}, r_{ext}, and complex refractive index are ~~essential~~needed. PNSD at different RH can be calculated with κ-Köhler equation as well. The refractive indices of BC, non-absorbing component, and pure water are set to be 1.8+0.54i (Ma et al., 2012), 1.53+$10^{-7}i$ (Wex et al., 2002), and 1.33+$10^{-7}i$ respectively. ~~More detail about calculations of lidar-derived optical properties can be found in (Zhao et al., 2017).~~ Backscatter coefficients (355, 532, and 1064 nm) and extinction coefficients (355 and 532 nm) at dry condition and RH from 60-90% are simulated with an interval of 1%.

The simulations are introduced in detail in Sect. S3 in the Supplement. The new method and all the analyses in this paper are based on the Mie model simulated datasets, and all the simulations mentioned above are implemented.

3 Methodology

~~3.1 Supersaturations for lidar CCN retrieval~~

~~CCN number concentrations are related with supersaturations. Critical diameters of each supersaturations calculated with twenty-five size-resolved κ distributions are shown in Fig. 2a. Most of the critical diameters at supersaturation of 0.07% are~~

~~larger than 200 nm, while critical diameters at supersaturation of 0.80% are around 50 nm. Suitable supersaturations for lidar CCN retrieval depend on the ability of lidar optical properties to provide information about number and hygroscopicity of CCN-related sizes.~~

~~Size-cumulative contributions of particle number of all measured particle size distribution and corresponding calculated backscatter and extinction at dry condition are also displayed in Fig. 2a. As the cumulative contributions of particle number suggest, particles with diameter less than 100 nm dominate particle number concentrations (over 65%). However, most backscatter and extinction come from particles larger than 200 nm (around 90%) and almost 100% come from particles larger than 100 nm. If critical diameter is small, dry backscatter and extinction are insensitive to particles diameters that contribute to most CCN concentrations.~~

~~Size-resolved enhancement contributions of backscatter and extinction are calculated to discuss hygroscopicity sensitive size of optical enhancement factor measurement. The enhancement contribution is defined as the difference between optical cross-sections of RH at 90% and 60%, and represents the proportion of each size to the enhancement in backscatter or extinction. As is shown in Fig. 2b, the contributions of the extinction enhancements are concentrated in the diameters within 200 nm to 700 nm, and extinction enhancement at 355 nm is related to smaller particles than that at 532 nm. Strong oscillations are found in size enhancement contributions of backscatter coefficients. Similar to particle number, particles with diameters smaller than 100nm contributes little to the enhancements of both backscatter and extinction.~~

~~Comparing sensitive size of optical properties and critical diameters at different supersaturations. 3*β*+2*α* MWRL systems have potential to retrieve CCN number concentrations at supersaturations smaller than 0.20%. It is not recommended to estimate CCN concentrations using lidar data at superstations larger than 0.40%.~~

~~3.2 Humidogram parameterization for backscatter and extinction enhancements~~

~~Humidogram parameterization is needed to find a representative parameter for the relationship between enhancement factor and RH. Unlike in situ controlled-RH measurements, there is no such a generic reference RH as dry condition for lidar measurements to derive enhancement factor. Inferring backscatter and extinction coefficients at dry condition~~ (ξ_{dry}) ~~is also an important issue in CCN retrieval. Therefore, humidogram parameterization of lidar-derived optical properties should combine~~ ξ_{dry} ~~and~~ $f_\xi(\mathrm{RH},\lambda)$ ~~together.~~

~~Many equations to parameterize enhancement factors have been proposed by previous studies (Titos et al., 2016). Two one-parameter equations are selected to test their performance on estimating~~ ξ_{dry} ~~and representing particle hygroscopic growth characteristics. The first equation is the most commonly used one initially introduced by Kasten (1969):~~

$$\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot f_\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot (1-\mathrm{RH})^{-\gamma_\xi(\lambda)}, \tag{4}$$

~~where the exponent~~ γ_ξ ~~is the fitting parameter and describes the hygroscopic behavior of the particles; the other equation is proposed based on physical understanding by Brock et al. (2016), which has been reported to have better performance in describing light scattering enhancement factor than Eq. (4) (Shin et al., 2018):~~

$$\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda)\cdot f_{\xi}(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda)\cdot\left[1+\kappa_{\xi}(\lambda)\frac{\mathrm{RH}}{1-\mathrm{RH}}\right], \quad (5)$$

~~where κ_{ξ} is the fitting parameter and shows significant correlation with bulk hygroscopic parameter κ (Kuang et al., 2017). Here, Eq. (4) and Eq. (5) are denoted as γ-equation and κ-equation respectively. With given backscatter and extinction at different RH, ξ_{dry} and γ_{ξ} or κ_{ξ} can be fitted simultaneously by means of least squares.~~

~~Comparisons between the performances of γ-equation and κ-equation on inferring backscatter and extinction at dry condition are carried out to select a better parameterization. Four RH ranges (60%-90%, 60%-70%, 70%-80%, and 80%-90%) are selected. The fitted ξ_{dry} are compared with the ξ_{dry} calculated by Mie model. The slopes of linear regressions, determination coefficients (R^2), and relative errors are listed in Table 2. Apparently, κ-equation has a better performance than γ-equation for all RH ranges. Inferring ξ_{dry} with γ-equation will underestimate about 10%-30%. It is consistent with the finding of Haarig et al. (2017) that γ-equation does not hold for RH lower than 40%. The bias of backscatter is found to be larger than the bias of extinction.~~

~~The RH range of humidogram equations also influences the fitting results. Table 2 shows the fitted ξ_{dry} have larger bias when the value of RH increase. The fitted humidogram parameters γ_{ξ} and κ_{ξ} from different RH ranges are compared to each other, and the results are displayed in Table 3. Parameterization equations are not always perfect for the whole RH ranges, so humidogram parameters fitted with various RH ranges can be different. If γ_{ξ} and κ_{ξ} are used to represent hygroscopic behavior of particles, more careful attention should be paid to the RH ranges.~~

~~Based on the comparisons above, Eq. (5) (κ-equation) is selected as our humidogram equation to derive ξ_{dry} and κ_{ξ}. The RH range for parameter fitting used is fixed to 60%-90% in the following method.~~

~~3.3~~3.1 Method to retrieve CCN number concentrations using MWRL

~~3.3.1~~3.1.1 Overview

An optical-related CCN activation ratio, AR_{ξ}, is introduced to bridge the gap between CCN and lidar-derived optical properties. AR_{ξ} is the ratio between CCN number concentration and ~~backscatter/extinction~~backscatter or extinction coefficient, which can be expressed as:

$$\mathrm{AR}_{\xi}(\mathrm{SS},\lambda) = \frac{N_{\mathrm{CCN}}(\mathrm{SS})}{\xi_{\mathrm{dry}}(\lambda)} = \frac{N_{\mathrm{CCN}}(\mathrm{SS})}{N_{\mathrm{aerosol}}}\cdot\frac{N_{\mathrm{aerosol}}}{\xi_{\mathrm{dry}}(\lambda)}, \quad (4~~46~~)$$

where N_{CCN} is the CCN number concentration, and N_{aerosol} is the total number concentration of aerosol particles. AR_{ξ} can be divided into two parts: one is the ratio of CCN to the total particles, which is the origin definition of CCN activation ratio; the other is the ratio of total number concentration to backscatter or extinction at dry condition. Bulk CCN activation ratio is related with particle size distribution and hygroscopicity, and the relationship between particle number concentration and optical properties is mainly controlled by size distribution. Therefore, AR_{ξ} could be quantified with size and hygroscopicity information. The key point of our method is to seek parameters that can indicate size and hygroscopicity of particles from lidar measurement and use these parameters to estimate AR_ξ. Besides, deriving backscatter and extinction coefficients at dry condition is also important.

~~The~~A schematic diagram of the ~~whole algorithm~~method to retrieve CCN number concentration is shown in Fig. ~~3~~2. Firstly, enhancement of backscatter and extinction coefficients with RH (also called humidogram) is derived from lidar measurements and additional ancillary data (i.e. pressure, temperature, RH profiles). Humidogram parameter which can indicate particle hygroscopicity can be fitted from humidograms with parameterization equation. Particle dry backscatter and extinction can also be inferred from the humidograms. This step is applied to all the $3\beta+2\alpha$ parameters. The approaches to select appropriate hygroscopic layers and fit humidogram parameters, dry backscatter, and dry extinction are described in Sect. 3.1.2.

~~With the method in Sect. 3.2, ξ_{dry} and κ_ξ can be derived with backscatter and extinction enhancements. Optical humidogram parameters κ_ξ can be regarded as parameters indicating hygroscopicity.~~ Then, Ångström exponent ($\mathring{a}$) and lidar extinction-to-backscatter ratio (lidar ratio, s_a) are calculated from inferred dry backscatter and extinction coefficients. Extinction-related Ångström exponent ($\mathring{a}_\alpha$) is the most commonly used parameter to reveal information about the predominant size of aerosols. Generally speaking, a smaller $\mathring{a}_\alpha$ represents there are more large particles. Similarly, backscatter-related Ångström exponent ($\mathring{a}_\beta$) are often employed in lidar analysis (Fernández et al., 2015), and particle backscatter coefficients of different wavelengths also have been proved to have a valid Ångström exponent relationship (Komppula et al., 2012). Ångström exponent of dry backscatter and extinction coefficients ($\mathring{a}_\xi$) between two wavelengths can be derived using Eq. (5~~7~~):

$$\mathring{a}_\xi(\lambda_1,\lambda_2) = -\frac{\log(\xi_1/\xi_2)}{\log(\lambda_1/\lambda_2)}, \quad (5\cancel{57})$$

where the subscript 1 and 2 represents different wavelengths. Another widely used parameter to express aerosol characteristics in lidar studies is the particle lidar extinction-to-backscatter ratio (lidar ratio, s_a), which is defined as the ratio of extinction coefficient to backscatter coefficient at a specific light wavelength:

$$s_a(\lambda) = \frac{\alpha(\lambda)}{\beta(\lambda)} = \frac{4\pi}{P(\pi)\cdot\omega}. \quad (6\cancel{68})$$

As is shown in Eq. (~~7~~6), lidar ratio is determined by the scattering phase function at 180° $P(\pi)$ and the single scattering albedo ω. $P(\pi)$ is mainly influenced by particle size and ω indicates the content and mixing state of light absorbing components. Lidar ratio is often utilized in aerosol type classification and is proved to be very sensitive to particle sizes (Zhao et al., 2017). ~~Particle type information can also be regard as an alternative representative of hygroscopicity.~~The lidar ratio can provide information on particle type and also serve as a proxy for particle hygroscopicity. Therefore, lidar ratio of dry particles could be a reliable parameter to estimate AR_ξ.

Next, $\mathring{a}_\xi$, s_a, and humidogram parameters are utilized to estimate AR_ξ. AR_ξ of all the $3\beta+2\alpha$ parameters is calculated. Statistical relationship among humidogram parameters, ~~-~~$\mathring{a}_\xi$, s_a~~, κ_ξ~~, and AR_ξ are used in our new method. ~~Based on the statistical relationship, AR_ξ can be estimated by $\mathring{a}_\xi$, s_a, and κ_ξ.~~The estimation of AR_ξ is introduced in Sect. 3.1.3 in detail. The implement of $\mathring{a}_\xi$ and s_a is quite similar to the microphysical inversion process for particle size distribution retrieval.

Microphysical inversion is a physics-based approach but will bring ~~huge~~large uncertainties in retrieving particle number concentrations. Constraining AR_ξ directly with statistical relationship is a much more simple and straightforward way.

Finally, ~~After~~after AR_ξ of backscatter and extinction at different wavelengths are derived, CCN number concentration can be calculated by multiplying AR_ξ by the corresponding ξ_{dry}. The average value of CCN concentrations calculated by different ξ_{dry} is the final retrieval result.

~~The schematic diagram of the whole algorithm is shown in Fig. 3.~~

~~3.3.2~~3.1.2 ~~Appropriate retrieval layers~~Derivation of humidogram parameters, dry backscatter, and dry extinction from lidar measurement

A constraint needs to be satisfied when quantifying the enhancements of backscatter and extinction coefficients with lidar measurements. The selected vertical layers must be well-mixed, so we can guarantee that the variations of particle ~~backscatter/extinction~~backscatter and extinction coefficients are caused by different RH and not by various aerosol types or loads. Atmospheric vertical homogeneity is fulfilled if the layer has little variability of virtual potential temperature profile and water vapor mixing ratio profile (Lv et al., 2017). Additional analyses can also be considered to evaluate vertical mixing of air masses, such as backward trajectory, horizontal wind velocities at different altitude, or the third moment of the frequency distribution of vertical wind velocities (Bedoya-Velásquez et al., 2018).

Once vertical homogeneity is ensured, physical and chemical properties at dry condition can be assumed to be uniform in the selected layer, and the number concentrations are proportional to air molecule number density. Accordingly, the relative variations of particle ~~backscatter/extinction~~backscatter and extinction coefficients against different RH can be achieved after normalizing the backscatter and extinction coefficients with air molecule number density.

Humidogram parameterization is needed to find a representative parameter for the relationship between enhancement factor and RH. Unlike in situ controlled-RH measurements, there is no such a generic reference RH as dry condition for lidar measurements to derive enhancement factor. Inferring backscatter and extinction coefficients at dry condition (ξ_{dry}) is also an important issue in CCN retrieval. Therefore, humidogram parameterization of lidar-derived optical properties should combine ξ_{dry} and $f_\xi(\mathrm{RH},\lambda)$ together.

Many equations to parameterize enhancement factors have been proposed by previous studies (Titos et al., 2016). Two one-parameter equations are selected to test their performance on estimating ξ_{dry} and representing particle hygroscopic growth characteristics. The first equation is the most commonly used one initially introduced by Kasten (1969):

$$\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot f_\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot (1-\mathrm{RH})^{-\gamma_\xi(\lambda)}, \qquad (7\cancel{7}4)$$

where the exponent γ_ξ is the fitting parameter and describes the hygroscopic behavior of the particles; the other equation is proposed based on physical understanding by Brock et al. (2016), which has been reported to have better performance in describing light scattering enhancement factor than Eq. (47) (Shin et al., 2018):

$$\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot f_\xi(\mathrm{RH},\lambda) = \xi_{\mathrm{dry}}(\lambda) \cdot \left[1 + \kappa_\xi(\lambda)\frac{\mathrm{RH}}{1-\mathrm{RH}}\right], \qquad (8\cancel{8}5)$$

where κ_ξ is the fitting parameter and shows significant correlation with bulk hygroscopic parameter κ (Kuang et al., 2017). Here, Eq. (47) and Eq. (~~5~~8) are denoted as γ-equation and κ-equation respectively. With given backscatter and extinction at different RH, ξ_{dry} and γ_ξ or κ_ξ can be fitted simultaneously by means of least squares.

Comparisons between the performances of γ-equation and κ-equation on inferring backscatter and extinction at dry condition are carried out to select a better parameterization. Four RH ranges (60%-90%, 60%-70%, 70%-80%, and 80%-90%) are selected. The fitted ξ_{dry} are compared with the ξ_{dry} calculated by Mie model. The slopes of linear regressions, determination coefficients (R^2), and relative errors are listed in Table 2. Apparently, κ-equation has a better performance than γ-equation for all RH ranges. Inferring ξ_{dry} with γ-equation will underestimate about 10%-30%. It is consistent with the finding of Haarig et al. (2017) that γ-equation does not hold for RH lower than 40%. The bias of backscatter is found to be larger than the bias of extinction.

The RH range of humidogram equations also influences the fitting results. Table 2 shows the fitted ξ_{dry} have larger bias when the value of RH increase. The fitted humidogram parameters γ_ξ and κ_ξ from different RH ranges are compared to each other, and the results are displayed in Table 3. Parameterization equations are not always perfect for the whole RH ranges, so humidogram parameters fitted with various RH ranges can be different. If γ_ξ and κ_ξ are used to represent hygroscopic behavior of particles, more careful attention should be paid to the RH ranges.

Based on the comparisons above, Eq. (~~5~~8) (κ-equation) is selected as our humidogram equation to derive ξ_{dry} and κ_ξ. The RH range for parameter fitting used is fixed to 60%-90% in the following method.

~~3.3.3~~3.1.3 Estimation of AR_ξ

Ångström exponents, lidar ratios, and optical humidogram parameters κ_ξ are used to estimate optical-related activation ratio AR_ξ. Concerning the Ångström exponents and lidar ratios are not independent to each other (any parameter can be calculated from other parameters), we reduce the number of parameters to a sufficient number to represent all the information. The selected nine parameters are listed in Table 4. ~~There are no explicit expressions between these parameters and AR_ξ, and the relationships between them are highly nonlinear.~~ One possible way to seek the relationship between the nine parameters and AR_ξ ~~solve this problem~~ is to build a lookup table, but too many input parameters would make the lookup table so large to build and operate.

In the past few decades, machine learning has been a field that has develop~~eds~~ rapidly, which experiences a very wide range of applications (Grange et al., 2018). Compared to traditional statistical methods, many machine learning techniques are nonparametric and do not need to fulfill many assumptions required for statistical methods (Immitzer et al., 2012). Random forest (RF) is an ensemble decision tree machine learning method that can be used for regression. (Breiman, 2001;Tong et al., 2003). Beside the free restraints on input parameters and assumptions, RF also has the advantage of being able to explain and investigate the learning process (Kotsiantis, 2013). The Python module *RandomForestRegressor* from the Python Scikit-Learn library (http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html, last access: 18 December 2018) are utilized as the RF model. The nine parameters in Table 4 are the input parameters for the RF model, and the AR_{ξ} of the $3\beta+2\alpha$ are the output parameters.

Some tuning parameters required by RF model need to be specified by users. Experiments are made to determine the optimal values of the tuning parameters. Experiment results are showed in Fig. ~~S3~~ S7 in the Supplement and the detailed settings of the RF model are listed in Table ~~S1~~ S2 in the Supplement. In this case, the results are rather insensitive to the tuning parameters. Data simulated with datasets measured from campaign C1-C4 are utilized as the training data, and those from C5 are used as test data.

~~3.4~~3.2 Sensitivity test

Both systematic and random errors exist in lidar-retrieved backscatter and extinction coefficients (Mattis et al., 2016). Systematic errors in backscatter and extinction can come from instrumentation setup, data processing method, and retrieval algorithm. Sensitivity test is carried out to test the impact of systematic errors of backscatter and extinction on CCN retrieval. Errors in backscatter or extinction influence the value of Ångström exponents and lidar ratios~~ but have no impact on κ_{ξ}~~. The errors of individual backscatter or extinction are considered to be independent, though systematic errors of different parameters are related. The systematic errors are given in the range of -20% to 20% with an interval of 2%. In each test, the error is only applied to one parameter, and other parameters are error-free.

RH is another crucial factor in this new method to retrieve CCN. Profiles of RH derived by remote sensing techniques are also influenced by errors. At present, RH profiles are usually obtained with the combination of temperature from microwave radiometer and water vapor mixing ratio from MWRL. Both measurements can cause systematic and random errors in RH (Bedoya-Velásquez et al., 2018). Errors in RH will influence the values of ξ_{dry} and κ_{ξ}, which in turn influence all the nine input parameters. Systematic errors ranging from -10% to 10% in intervals of 1% are considered for RH.

Random errors in observations can be reduced by temporal averaging but cannot be eliminated. The influence of random errors in backscatter, extinction, and RH on CCN retrieval are investigated with Monte Carlo method. Errors obeying Gaussian distribution are generated randomly with the mean value of zero. The standard deviation of Gaussian distribution is 10% for ~~backscatter/extinction~~backscatter and extinction and 5% for RH. The procedure is repeated for 2000 times. All the 80575 sets of data from campaign C5 are used for sensitivity test.

4 Results and discussions

~~3.1~~4.1 Supersaturations for lidar CCN retrieval

CCN number concentrations are related with supersaturations. ~~-~~Critical diameters of each supersaturations calculated with twenty-five size-resolved κ distributions are shown in Fig. ~~2~~3a. Most of the critical diameters at supersaturation of 0.07% are larger than 200 nm, while critical diameters at supersaturation of 0.80% are around 50 nm. Suitable supersaturations for lidar CCN retrieval depend on the ability of lidar optical properties to provide information about number and hygroscopicity of CCN-related sizes.

Size cumulative contributions of particle number of all measured particle size distribution and corresponding calculated backscatter and extinction at dry condition are also displayed in Fig. ~~2~~3a. As the cumulative contributions of particle number suggest, particles with diameter less than 100 nm dominate particle number concentrations (over 65%). However, most backscatter and extinction come from particles larger than 200 nm (around 90%) and almost 100% come from particles larger than 100 nm. If critical diameter is small, dry backscatter and extinction are insensitive to particles diameters that contribute to most CCN concentrations.

Size-resolved enhancement contributions of backscatter and extinction are calculated to discuss hygroscopicity sensitive size of optical enhancement factor measurement. The enhancement contribution is defined as the difference between optical cross-sections of RH at 90% and 60%, and represents the proportion of each size to the enhancement in backscatter or extinction. As is shown in Fig. ~~2~~3b, the contributions of the extinction enhancements are concentrated in the diameters within 200 nm to 700 nm, and extinction enhancement at 355 nm is related to smaller particles than that at 532 nm. ~~Strong oscillations are found in size enhancement contributions of backscatter coefficients.~~ Similar to particle number, particles with diameters smaller than 100nm contributes little to the enhancements of both backscatter and extinction.

Figure 3b also shows that different κ_{ξ} is sensitive to the hygroscopicity of different size. Size-dependent hygroscopicity is important to estimate CCN rather than a bulk hygroscopicity information, especially for different supersaturation conditions. One humidogram may indicate the bulk hygroscopicity, but it is the hygroscopicity of small particles that influences CCN number concentrations most. Using κ_{ξ} of all the 3β+2α can provide some information about the hygroscopicity of small particles.

Comparing sensitive size of optical properties and critical diameters at different supersaturations. 3β+2α MWRL systems have potential to retrieve CCN number concentrations at supersaturations smaller than 0.20%. It is not recommended to estimate CCN concentrations using lidar data at superstations larger than 0.40%.

~~4.1~~4.2 CCN number concentrations retrieved with error-free data

With error-free data as input, the model predicted extinction-related activation ratio at 532 nm ($AR_{\alpha 532}$) and the retrieved CCN number concentrations at supersaturations of 0.07%, 0.10%, and 0.20% are compared to the theoretical calculated values. A total of 80575 pairs of data calculated from campaign C5 are used for verification. The retrieval results are displayed in Fig. 4. The values $AR_{\alpha 532}$ at a specific supersaturation are distributed in a wide range and can span over an order of magnitude, indicating that the relationship between CCN and optical parameters is very complex. According to Fig. 4, all data points are distributed almost evenly on both sides of the 1:1 line and the relative errors of most points are within 20%. The determination coefficients (R^2) of CCN concentrations are all larger than 0.97, and the results do not show obvious systematic deviations.

The retrieval errors are found to grow with supersaturation. Retrieval results for higher supersaturations (i.e. 0.40% and 0.80%) is displayed in Fig. ~~S4~~ S8 in the Supplement. There are larger errors for supersaturations of 0.40% and 0.80%. Only 47.76% of the retrieved CCN number concentration at supersaturation of 0.80% have relative errors less than 20%. The results ~~are consistent with the previous analysis in Sect. 3.1, which means~~demonstrate again that lidars may not be sufficient enough to retrieve CCN number concentrations at supersaturations lager than 0.40%.

~~4.2~~4.3 Importance of size-related and hygroscopicity-related parameters

RF models can evaluate the importance of features (input parameters) by calculating the mean decrease impurity (MDI) for each feature among all the trees in the forest. The MDIs and corresponding standard deviations of each parameter at different supersaturations are shown in Fig. 5. Importance of the nine input parameters varies with supersaturations. For 0.07% and 0.10%, $\kappa_{\alpha355}$ and $\kappa_{\beta1064}$ are the two most important parameters, showing ~~huge~~the impact of hygroscopicity on the relationship between CCN and optical properties. For 0.20%, $\mathring{a}_{\alpha355\&532}$ becomes much more important. Among the nine input parameters, κ_{ξ} are denoted as hygroscopicity-related parameters, and $\mathring{a}_{\xi}$ are denoted as size-related parameters. Particularly, s_a can be regarded as both size- and hygroscopicity-related parameter. As is shown in Fig. 5, hygroscopicity-related parameters, especially $\kappa_{\alpha355}$, $\kappa_{\beta1064}$, and $s_{\alpha532}$, play crucial roles in retrieving CCN. Size-related parameters have already been proved to be vital in retrieving CCN, however, humidogram parameters κ_{ξ} have not been implemented in previous methods. CCN concentrations retrieved with and without κ_{ξ} are compared to show the importance of κ_{ξ}. When retrieving CCN without κ_{ξ}, the RF model is also trained with datasets from campaign C1-C4, but the input data only contains Ångström exponents and lidar ratios. The retrieved CCN concentrations are all compared with datasets from campaign C5, and the results are listed in Table 5. R^2 of retrieved CCN decreases from 0.991 to 0.887 for supersaturations of 0.07%, from 0.992 to 0.857 for 0.10%, and from 0.973 to 0.785 for 0.20%. Retrieval errors also increase overwhelmingly, and there are significant positive systematic biases. Parameters which are derived from backscatter and extinction enhancements, κ_{ξ}, are indispensable parameters in CCN retrieval.

~~4.3~~4.4 Impact of systematic and random error on CCN retrieval

Figure 6 shows the relative errors of CCN retrieved with systematic errors in backscatter and extinction. Errors of retrieved CCN increase as errors of backscatter and extinction increase, and higher supersaturations are more affected by errors of optical parameters. Errors in extinction coefficients at 355 nm (α_{355}) influence the retrieval results most. ~~In~~ On average, a positive relative error of 20% in α_{355} will cause about 20% overestimate in CCN number concentrations for supersaturation of 0.07%, about 40% overestimate for 0.10%, and about 60% overestimate for 0.20%. A negative error of 20% in α_{355} will underestimate CCN concentrations, and the degree of impact is slightly smaller than positive error. Errors in extinction coefficient at 532 nm (α_{532}) and at 355nm have opposite effect on retrieval error. ~~Bigger α_{355} means more small particles and higher number concentrations, and bigger α_{532} means more large particles.~~ Errors in α_{532} do not show significant impact at supersaturations of 0.07% and 0.10%, but an overwhelming effect is found at supersaturations of 0.20%. It is interesting to note that the errors in backscatter coefficients do not affect the results much. However, in practical applications of MWRLs, the errors in extinction are always much larger than the errors of backscatter. If the error of retrieved CCN concentrations needs to be limited to 20% at supersaturation of 0.20%, the errors of retrieved extinction coefficients should to be controlled within 5%.

The test result of systematic error in RH is shown in Fig. 7. When RH has a negative systematic error, CCN concentrations are overestimated, and the extent of overestimation increases as the error increase. A negative error of 10% in RH will overestimate CCN at supersaturations at 0.20% by about 60% in average, and the standard deviation is over 60%. Effects of positive errors in RH is much smaller than negative errors but more ~~complicated~~complex. The standard deviations of retrieval relative error increase with RH error, and the extreme value of the mean retrieval error appears at the RH error of 5%. Underestimating RH will cause~~s~~ much more errors than overestimation. Great care should be paid to RH profiles if enhancements of backscatter and extinction with RH are utilized.

The relative error of retrieved CCN with random errors are presented in Table 6. The mean values of relative error ~~is~~are -2.8%, -1.3%, and 1.3% for CCN at supersaturations of 0.07%, 0.10%, and 0.20%, respectively, and the corresponding standard deviations are 29.7%, 31.5%, and 42.9%. The impact of random errors on the nine input parameters is also evaluated and is shown in Fig. 8. Random errors (10% for backscatter and extinction, and 5% RH) underestimate κ_ξ by 30%-35% in average, and the standard deviations are about 40% or more. s_{a355} , s_{a532}, and $\mathring{a}_{\beta532\&1064}$ are overestimated by 5%-10%.

5 Summary

CCN number concentration at cloud base is a crucial and scarce parameter to constrain the relationship between aerosols and clouds. A new method to retrieve CCN number concentrations using backscatter and extinction coefficients from MWRL measurements is proposed. Enhancements of backscatter and extinction coefficients with RH are implemented to derive dry ~~backscatter/extinction~~backscatter and extinction ξ_{dry} and humidogram parameter κ_ξ. The ratio of CCN number concentration to dry backscatter or extinction coefficient AR_ξ, which is estimated by κ_ξ, Ångström exponents, and lidar ratios, is introduced to retrieve CCN number concentrations.

The method is established and verified by theoretical simulations using Mie theory and κ-Köhler theory with in situ measured particle size distributions, mixing states, and chemical compositions. The values of AR_ξ are found to have large variations due to different size distributions and hygroscopicity. Theoretical analyses show that optical properties provided by current 3β+2α MWRL systems basically contains size distribution and hygroscopicity information of particles with diameters larger than 100 nm, which only fits the critical diameters for supersaturations lower than 0.20%. Accordingly, CCN number concentrations at supersaturations of 0.07%, 0.10%, and 0.20% are retrieved. The performance of the new method is evaluated with error-free data, and CCN number concentrations at all three supersaturations are in good agreements with theoretical calculated values. Sensitivity tests are carried out to show the influence of systematic and random errors of lidar-derived optical properties and auxiliary RH profiles on CCN retrieval. Systematic errors in extinction coefficients and RH are found to have large impact on error in retrieved CCN. Parameters fitted from backscatter and extinction enhancements (i.e. ξ_{dry} and κ_{ξ}) is significantly influenced by RH. The uncertainty of RH profiles derived by remote sensing techniques is a major problem in CCN retrieval. Optical properties near cloud base from lidar measurements always influenced by high RH. Thus, transforming backscatter and extinction coefficients at ambient RH to dry conditions is a must for CCN retrieval, and accurate RH profiles are highly demanded.

The importance of humidogram parameters κ_{ξ} is demonstrated by comparing the error of CCN concentration retrieved with and without κ_{ξ}. Neglecting hygroscopicity information contained in backscatter and extinction enhancements will bring huge errors to CCN retrieval by lidars. The performance of two parameterization schemes for backscatter and extinction humidograms are evaluated. The κ-equation shows better performance on inferring dry backscatter and extinction than γ-equation. The κ-equation, therefore, is recommended to describe the hygroscopic behaviors of the backscatter and extinction coefficients from lidar measurements. The fitted hygroscopic parameter are found to be sensitive to fitting RH range when the RH range is limited and relatively high (between 60%-90%). This is an extreme essential problem for current research for aerosol hygroscopicity with lidar measurements. Great care should be paid to the RH range when evaluating the hygroscopic growth of the lidar-related optical properties.

It should be noted that the theoretical analyses in this paper are based on datasets of continental aerosols, and the implement of Mie theory also limits the scope of the results. The results can be applied in the North China Plain but are not fit for sea salts and mineral dust. Studies with datasets of other aerosol types should be carried out in the future. Although the applicability of this new method is limited by large uncertainties in RH profiles, comparison between real measured MWRL data and airborne in situ measurement should also be conducted.

This work furthers our understanding of the relationship between CCN and aerosol optical properties and providing an optional way to retrieve CCN number concentration profiles from lidar measurements. The newly proposed method has potential to provide long-term CCN at cloud base for aerosol-cloud-interaction studies.

Author contribution. C. Zhao and C. Li determined the main goal of this study. W. Tan and G. Zhao designed the methods. W. Tan carried them out and prepared the manuscript with contributions from all co-authors.

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

Acknowledgements. The study is supported by the National Key R&D Program of China (2016YFC0202000: Task 4) and the National Natural Science Foundation of China (41375008, 41590872, 9154400001, 41527807).

References

Barreto, A., Román, R., Cuevas, E., Pérez-Ramírez, D., Berjón, A. J., Kouremeti, N., Kazadzis, S., Gröbner, J., Mazzola, M., Toledano, C., Benavent-Oltra, J. A., Doppler, L., Juryšek, J., Almansa, A. F., Victori, S., Maupin, F., Guirado-Fuentes, C., González, R., Vitale, V., Goloub, P., Blarel, L., Alados-Arboledas, L., Woolliams, E., Taylor, S., Antuña, J. C., and Yela, M.: Evaluation of night-time aerosols measurements and lunar irradiance models in the frame of the first multi-instrument nocturnal intercomparison campaign, Atmos. Environ., 10.1016/j.atmosenv.2019.01.006, 2019.

Bedoya-Velásquez, A. E., Navas-Guzmán, F., Granados-Muñoz, M. J., Titos, G., Román, R., Casquero-Vera, J. A., Ortiz-Amezcua, P., Benavent-Oltra, J. A., de Arruda Moreira, G., Montilla-Rosero, E., Hoyos, C. D., Artiñano, B., Coz, E., Olmo-Reyes, F. J., Alados-Arboledas, L., and Guerrero-Rascado, J. L.: Hygroscopic growth study in the framework of EARLINET during the SLOPE I campaign: synergy of remote sensing and in situ instrumentation, Atmos. Chem. Phys., 18, 7001-7017, 10.5194/acp-18-7001-2018, 2018.

Bian, Y., Zhao, C., Xu, W., Kuang, Y., Tao, J., Wei, W., Ma, N., Zhao, G., Lian, S., Tan, W., and Barnes, J. E.: A novel method to retrieve the nocturnal boundary layer structure based on CCD laser aerosol detection system measurements, Remote Sens. Environ., 211, 38--47, 10.1016/j.rse.2018.04.007, 2018.

Bohren, C. F., and Huffman, D. R.: Absorption and Scattering by an Arbitrary Particle, in: Absorption and Scattering of Light by Small Particles, Wiley-VCH Verlag GmbH, 57-81, 2007.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B., and Zhang, X. Y.: Clouds and Aerosols, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 571–658, 2013.

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

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

Burton, S. P., Chemyakin, E., Liu, X., Knobelspiesse, K., Stamnes, S., Sawamura, P., Moore, R. H., Hostetler, C. A., and Ferrare, R. A.: Information content and sensitivity of the 3β + 2α lidar measurement system for aerosol microphysical retrievals, Atmos. Meas. Tech., 9, 5555-5574, 10.5194/amt-9-5555-2016, 2016.

Fan, J., Rosenfeld, D., Zhang, Y., Giangrande, S. E., Li, Z., Machado, L. A. T., Martin, S. T., Yang, Y., Wang, J., Artaxo, P., Barbosa, H. M. J., Braga, R. C., Comstock, J. M., Feng, Z., Gao, W., Gomes, H. B., Mei, F., Pöhlker, C., Pöhlker, M. L., Pöschl, U., and de Souza, R. A. F.: Substantial convection and precipitation enhancements by ultrafine aerosol particles, Science, 359, 411-418, 10.1126/science.aan8461, 2018.

Feingold, G., and Morley, B.: Aerosol hygroscopic properties as measured by lidar and comparison with in situ measurements, J. Geophys. Res. Atmos., 108, n/a-n/a, 10.1029/2002JD002842, 2003.

Fernández, A. J., Apituley, A., Veselovskii, I., Suvorina, A., Henzing, J., Pujadas, M., and Artíñano, B.: Study of aerosol hygroscopic events over the Cabauw experimental site for atmospheric research (CESAR) using the multi-wavelength Raman lidar Caeli, Atmos. Environ., 120, 484-498, http://dx.doi.org/10.1016/j.atmosenv.2015.08.079, 2015.

Fernández, A. J., Molero, F., Becerril-Valle, M., Coz, E., Salvador, P., Artíñano, B., and Pujadas, M.: Application of remote sensing techniques to study aerosol water vapour uptake in a real atmosphere, Atmos. Res., https://doi.org/10.1016/j.atmosres.2017.11.020, 2017.

Fu, S., Deng, X., Li, Z., and Xue, H.: Radiative effect of black carbon aerosol on a squall line case in North China, Atmos. Res., 197, 407-414, 10.1016/j.atmosres.2017.07.026, 2017.

Ghan, S. J., and Collins, D. R.: Use of In Situ Data to Test a Raman Lidar–Based Cloud Condensation Nuclei Remote Sensing Method, J. Atmos. Ocean. Tech., 21, 387-394, 10.1175/1520-0426(2004)021<0387:uoisdt>2.0.co;2, 2004.

Ghan, S. J., Rissman, T. A., Elleman, R., Ferrare, R. A., Turner, D., Flynn, C., Wang, J., Ogren, J., Hudson, J., Jonsson, H. H., VanReken, T., Flagan, R. C., and Seinfeld, J. H.: Use of in situ cloud condensation nuclei, extinction, and aerosol size distribution measurements to test a method for retrieving cloud condensation nuclei profiles from surface measurements, J. Geophys. Res. Atmos., 111, 10.1029/2004jd005752, 2006.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM_{10} trend analysis, Atmos. Chem. Phys., 18, 6223-6239, 10.5194/acp-18-6223-2018, 2018.

Haarig, M., Ansmann, A., Gasteiger, J., Kandler, K., Althausen, D., Baars, H., Radenz, M., and Farrell, D. A.: Dry versus wet marine particle optical properties: RH dependence of depolarization ratio, backscatter, and extinction from multiwavelength lidar measurements during SALTRACE, Atmos. Chem. Phys., 17, 14199-14217, 10.5194/acp-17-14199-2017, 2017.

Hudson, J. G.: Variability of the relationship between particle size and cloud-nucleating ability, Geophys. Res. Lett., 34, n/a-n/a, 10.1029/2006GL028850, 2007.

Immitzer, M., Atzberger, C., and Koukal, T.: Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data, Remote Sens., 4, 2661, 2012.

Kasten, F.: Visibility forecast in the phase of pre-condensation, Tellus, 21, 631-635, 10.3402/tellusa.v21i5.10112, 1969.

Komppula, M., Mielonen, T., Arola, A., Korhonen, K., Lihavainen, H., Hyvärinen, A. P., Baars, H., Engelmann, R., Althausen, D., Ansmann, A., Müller, D., Panwar, T. S., Hooda, R. K., Sharma, V. P., Kerminen, V. M., Lehtinen, K. E. J., and Viisanen, Y.: Technical Note: One year of Raman-lidar measurements in Gual Pahari EUCAARI site close to New Delhi in India – Seasonal characteristics of the aerosol vertical structure, Atmos. Chem. Phys., 12, 4513-4524, 10.5194/acp-12-4513-2012, 2012.

Kotsiantis, S. B.: Decision trees: a recent overview, Artif. Intell. Rev., 39, 261-283, 10.1007/s10462-011-9272-4, 2013.

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

Kuang, Y., Zhao, C. S., Zhao, G., Tao, J. C., Xu, W., Ma, N., and Bian, Y. X.: A novel method for calculating ambient aerosol liquid water content based on measurements of a humidified nephelometer system, Atmos. Meas. Tech., 11, 2967-2982, 10.5194/amt-11-2967-2018, 2018.

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

Lv, M., Liu, D., Li, Z., Mao, J., Sun, Y., Wang, Z., Wang, Y., and Xie, C.: Hygroscopic growth of atmospheric aerosol particles based on lidar, radiosonde, and in situ measurements: Case studies from the Xinzhou field campaign, J. Quant. Spectrosc. Ra., 188, 60-70, 10.1016/j.jqsrt.2015.12.029, 2017.

Lv, M., Wang, Z., Li, Z., Luo, T., Ferrare, R., Liu, D., Wu, D., Mao, J., Wan, B., Zhang, F., and Wang, Y.: Retrieval of Cloud Condensation Nuclei Number Concentration Profiles From Lidar Extinction and Backscatter Data, J. Geophys. Res. Atmos., 123, 6082-6098, 10.1029/2017jd028102, 2018.

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

Mamouri, R.-E., and Ansmann, A.: Potential of polarization lidar to provide profiles of CCN- and INP-relevant aerosol parameters, Atmos. Chem. Phys., 16, 5905-5931, 10.5194/acp-16-5905-2016, 2016.

Mattis, I., apos, Amico, G., Baars, H., Amodeo, A., Madonna, F., and Iarlori, M.: EARLINET Single Calculus Chain – technical – Part 2: Calculation of optical products, Atmos. Meas. Tech., 9, 3009-3029, 10.5194/amt-9-3009-2016, 2016.

McCoy, D. T., Bender, F. A. M., Mohrmann, J. K. C., Hartmann, D. L., Wood, R., and Grosvenor, D. P.: The global aerosol-cloud first indirect effect estimated using MODIS, MERRA, and AeroCom, J. Geophys. Res. Atmos., 122, 1779-1796, 10.1002/2016JD026141, 2017.

Müller, D., Böckmann, C., Kolgotin, A., Schneidenbach, L., Chemyakin, E., Rosemann, J., Znak, P., and Romanov, A.: Microphysical particle properties derived from inversion algorithms developed in the framework of EARLINET, Atmos. Meas. Tech., 9, 5007-5035, 10.5194/amt-9-5007-2016, 2016.

Pahlow, M., Feingold, G., Jefferson, A., Andrews, E., Ogren, J. A., Wang, J., Lee, Y. N., Ferrare, R. A., and Turner, D. D.: Comparison between lidar and nephelometer measurements of aerosol hygroscopicity at the Southern Great Plains Atmospheric Radiation Measurement site, J. Geophys. Res. Atmos., 111, n/a-n/a, 10.1029/2004JD005646, 2006.

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, 10.5194/acp-7-1961-2007, 2007.

Rosati, B., Herrmann, E., Bucci, S., Fierli, F., Cairo, F., Gysel, M., Tillmann, R., Größ, J., Gobbi, G. P., Di Liberto, L., Di Donfrancesco, G., Wiedensohler, A., Weingartner, E., Virtanen, A., Mentel, T. F., and Baltensperger, U.: Studying the vertical aerosol extinction coefficient by comparing in situ airborne data and elastic backscatter lidar, Atmos. Chem. Phys., 16, 4539-4554, 10.5194/acp-16-4539-2016, 2016.

Rosenfeld, D., Andreae, M. O., Asmi, A., Chin, M., de Leeuw, G., Donovan, D. P., Kahn, R., Kinne, S., Kivekäs, N., Kulmala, M., Lau, W., Schmidt, K. S., Suni, T., Wagner, T., Wild, M., and Quaas, J.: Global observations of aerosol-cloud-precipitation-climate interactions, Rev. Geophys., 52, 750-808, 10.1002/2013RG000441, 2014.

Rosenfeld, D., Zheng, Y., Hashimshoni, E., Pohlker, M. L., Jefferson, A., Pohlker, C., Yu, X., Zhu, Y., Liu, G., Yue, Z., Fischman, B., Li, Z., Giguzin, D., Goren, T., Artaxo, P., Barbosa, H. M., Poschl, U., and Andreae, M. O.: Satellite retrieval of cloud condensation nuclei concentrations by using clouds as CCN chambers, P. Natl. Acad. Sci. USA, 113, 5828-5834, 10.1073/pnas.1514044113, 2016.

Schmale, J., Henning, S., Decesari, S., Henzing, B., Keskinen, H., Sellegri, K., Ovadnevaite, J., Pöhlker, M. L., Brito, J., Bougiatioti, A., Kristensson, A., Kalivitis, N., Stavroulas, I., Carbone, S., Jefferson, A., Park, M., Schlag, P., Iwamoto, Y., Aalto, P., Äijälä, M., Bukowiecki, N., Ehn, M., Frank, G., Fröhlich, R., Frumau, A., Herrmann, E., Herrmann, H., Holzinger, R., Kos, G., Kulmala, M., Mihalopoulos, N., Nenes, A., amp, apos, Dowd, C., Petäjä, T., Picard, D., Pöhlker, C., Pöschl, U., Poulain, L., Prévôt, A. S. H., Swietlicki, E., Andreae, M. O., Artaxo, P., Wiedensohler, A., Ogren, J., Matsuki, A., Yum, S. S., Stratmann, F., Baltensperger, U., and Gysel, M.: Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories, Atmos. Chem. Phys., 18, 2853-2881, 10.5194/acp-18-2853-2018, 2018.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R., Kraucunas, I., Kreidenweis, S. M., Molina, M. J., Nenes, A., Penner, J. E., Prather, K. A., Ramanathan, V., Ramaswamy, V., Rasch, P. J., Ravishankara, A. R., Rosenfeld, D., Stephens, G., and Wood, R.: Improving our fundamental understanding of the role of aerosol–cloud interactions in the climate system, P. Natl. Acad. Sci. USA, 113, 5781-5790, 10.1073/pnas.1514043113, 2016.

Shin, S.-K., Tesche, M., Kim, K., Kezoudi, M., Tatarov, B., Müller, D., and Noh, Y.: On the spectral depolarisation and lidar ratio of mineral dust provided in the AERONET version 3 inversion product, Atmos. Chem. Phys., 18, 12735-12746, 10.5194/acp-18-12735-2018, 2018.

Sun, T., Che, H., Qi, B., Wang, Y., Dong, Y., Xia, X., Wang, H., Gui, K., Zheng, Y., Zhao, H., Ma, Q., Du, R., and Zhang, X.: Aerosol optical characteristics and their vertical distributions under enhanced haze pollution events: effect of the regional transport of different aerosol types over eastern China, Atmos. Chem. Phys., 18, 2949-2971, 10.5194/acp-18-2949-2018, 2018.

Tao, J., Zhao, C., Ma, N., and Kuang, Y.: Consistency and applicability of parameterization schemes for the size-resolved aerosol activation ratio based on field measurements in the North China Plain, Atmos. Environ., https://doi.org/10.1016/j.atmosenv.2017.11.021, 2017.

Tao, J., Zhao, C., Kuang, Y., Zhao, G., Shen, C., Yu, Y., Bian, Y., and Xu, W.: A new method for calculating number concentrations of cloud condensation nuclei based on measurements of a three-wavelength humidified nephelometer system, Atmos. Meas. Tech., 11, 895-906, 10.5194/amt-11-895-2018, 2018.

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

Tong, W., Hong, H., Fang, H., Xie, Q., and Perkins, R.: Decision Forest: Combining the Predictions of Multiple Independent Decision Tree Models, J. Chem. Inf. Comp. Sci., 43, 525-531, 10.1021/ci020058s, 2003.

Wang, Q., Li, Z., Guo, J., Zhao, C., and Cribb, M.: The climate impact of aerosols on the lightning flash rate: is it detectable from long-term measurements?, Atmos. Chem. Phys., 18, 12797-12816, 10.5194/acp-18-12797-2018, 2018.

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, J. Geophys. Res. Atmos., 107, LAC 4-1-LAC 4-18, 10.1029/2000JD000234, 2002.

Wulfmeyer, V., and Feingold, G.: On the relationship between relative humidity and particle backscattering coefficient in the marine boundary layer determined with differential absorption lidar, J. Geophys. Res. Atmos., 105, 4729-4741, 10.1029/1999JD901030, 2000.

Xu, W. Y., Zhao, C. S., Ran, L., Deng, Z. Z., Liu, P. F., Ma, N., Lin, W. L., Xu, X. B., Yan, P., He, X., Yu, J., Liang, W. D., and Chen, L. L.: Characteristics of pollutants and their correlation to meteorological conditions at a suburban site in the North China Plain, Atmos. Chem. Phys., 11, 4353-4369, 10.5194/acp-11-4353-2011, 2011.

Xu, X., Guo, X., Zhao, T., An, X., Zhao, Y., Quan, J., Mao, F., Gao, Y., Cheng, X., Zhu, W., and Wang, Y.: Are precipitation anomalies associated with aerosol variations over eastern China?, Atmos. Chem. Phys., 17, 8011-8019, 10.5194/acp-17-8011-2017, 2017.

Zhang, F., Wang, Y., Peng, J., Ren, J., Collins, D., Zhang, R., Sun, Y., Yang, X., and Li, Z.: Uncertainty in Predicting CCN Activity of Aged and Primary Aerosols, J. Geophys. Res. Atmos., 10.1002/2017jd027058, 2017.

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., 17, 12133-12143, 10.5194/acp-17-12133-2017, 2017.

Zhou, X., Ackerman, A. S., Fridlind, A. M., Wood, R., and Kollias, P.: Impacts of solar-absorbing aerosol layers on the transition of stratocumulus to trade cumulus clouds, Atmos. Chem. Phys., 17, 12725-12742, 10.5194/acp-17-12725-2017, 2017.

Table 1. Locations, time periods, parameters, and instruments of five field campaigns

~~Location~~	~~Wuqing~~	~~Wuqing~~	~~Xianghe~~	~~Xianghe~~	~~Wangdu~~
~~Campaign name~~	~~C1~~	~~C2~~	~~C3~~	~~C4~~	~~C5~~
~~Time period~~	~~7 March to 4 April, 2009~~	~~12 July to 14 August, 2009~~	~~22 July to 30 August, 2012~~	~~9 July to 30 August, 2013~~	~~4 June to 14 July, 2014~~
~~PNSD~~	~~TSMPS+APS~~	~~TSMPS+APS~~	~~SMPS+APS~~	~~TSMPS+APS~~	~~TSMPS+APS~~
~~m_{BC}~~	~~MAAP~~				
~~HBF~~	~~TSI 3563 nephelometer~~				
~~Size-resolved chemical composition~~	~~–~~	~~Substrates sampled by BLPI~~	~~–~~		

Location **Campaign name**	Wuqing C1	Wuqing C2	Xianghe C3	Xianghe C4	Wangdu C5
Time period	7 March to 4 April, 2009	12 July to 14 August, 2009	22 July to 30 August, 2012	9 July to 30 August, 2013	4 June to 14 July, 2014
PNSD	TSMPS+APS	TSMPS+APS	SMPS+APS	TSMPS+APS	TSMPS+APS
m_{BC}	MAAP	MAAP	MAAP	MAAP	MAAP
HBF	TSI 3563	TSI 3563	TSI 3563	TSI 3563	TSI 3563
Size-resolved chemical composition	–	Substrates sampled by BLPI	–	–	–

Table 2. Slopes of linear regressions, determination coefficients (R^2), and relative errors (RE) between Mie model simulated particle dry ~~backscatter/extinction~~backscatter or extinction coefficients and those inferred from humidogram functions. 404575 pairs of the simulations from in situ dataset ~~is~~ are used. The RE are given in the form of mean value ± one standard deviation (std).

RH (%)	ξ	***γ*-equation** slope	***γ*-equation** R^2	***γ*-equation** RE(%)	***κ*-equation** slope	***κ*-equation** R^2	***κ*-equation** RE(%)
	$\alpha_{355,dry}$	0.850	0.998	-16.2 ± 2.1	1.045	0.998	3.4 ± 2.4
	$\alpha_{532,dry}$	0.820	0.998	-19.2 ± 2.0	1.017	0.999	0.5 ± 1.8
60-90	$\beta_{355,dry}$	0.784	0.960	-20.8 ± 7.2	0.817	0.971	-9.6 ± 7.5
	$\beta_{532,dry}$	0.812	0.972	-22.7 ± 7.6	0.874	0.988	-11.7 ± 5.6
	$\beta_{1064,dry}$	0.878	0.986	-12.9 ± 5.7	0.935	0.994	-5.4 ± 4.4
	$\alpha_{355,dry}$	0.913	1.000	-9.2 ± 1.1	1.016	1.000	1.1 ± 0.9
	$\alpha_{532,dry}$	0.900	0.999	-10.4 ± 1.3	1.005	1.000	0.0 ± 0.7
60-70	$\beta_{355,dry}$	0.939	0.989	-9.1 ± 6.0	0.906	0.991	-5.6 ± 4.9
	$\beta_{532,dry}$	0.939	0.990	-9.9 ± 5.6	0.939	0.996	-6.4 ± 3.9
	$\beta_{1064,dry}$	0.966	0.997	-3.9 ± 2.9	0.974	0.999	-1.9 ± 2.0
	$\alpha_{355,dry}$	0.852	0.999	-15.8 ± 1.9	1.037	0.999	2.7 ± 2.1
	$\alpha_{532,dry}$	0.827	0.998	-18.3 ± 1.9	1.012	0.999	0.3 ± 1.5
70-80	$\beta_{355,dry}$	0.799	0.950	-20.5 ± 8.9	0.818	0.968	-10.5 ± 8.1
	$\beta_{532,dry}$	0.833	0.966	-21.4 ± 9.0	0.880	0.986	-11.7 ± 6.6
	$\beta_{1064,dry}$	0.898	0.987	-10.8 ± 5.7	0.942	0.995	-4.6 ± 4.1
	$\alpha_{355,dry}$	0.756	0.922	-26.5 ± 3.8	1.110	0.991	8.5 ± 5.5
	$\alpha_{532,dry}$	0.702	0.994	-31.9 ± 3.1	1.047	0.995	1.9 ± 4.2
80-90	$\beta_{355,dry}$	0.547	0.848	-37.0 ± 11.1	0.695	0.892	-13.4 ± 14.1
	$\beta_{532,dry}$	0.593	0.925	-42.1 ± 8.7	0.775	0.961	-19.2 ± 8.7
	$\beta_{1064,dry}$	0.702	0.934	-30.4 ± 10.3	0.867	0.971	-11.5 ± 8.8

Table 3. Slopes of linear regressions and determination coefficients (R^2) between γ_ξ or κ_ξ fitted from RH range 60%-90% and those fitted from limited RH ranges (60%-70%, 70%-80%, and 80%-90%).

RH (%)	ξ	γ_ξ		κ_ξ	
		slope	$\mathbf{R^2}$	**slope**	$\mathbf{R^2}$
60-70	α_{355}	0.992	0.958	1.113	0.955
	α_{532}	0.969	0.978	1.007	0.977
	β_{355}	1.019	0.814	1.213	0.819
	β_{532}	0.790	0.797	0.891	0.799
	β_{1064}	0.806	0.834	1.011	0.812
70-80	α_{355}	1.021	0.996	1.045	0.995
	α_{532}	1.015	0.997	1.014	0.997
	β_{355}	1.115	0.968	1.195	0.958
	β_{532}	1.078	0.973	1.128	0.969
	β_{1064}	0.999	0.979	1.034	0.972
80-90	α_{355}	0.941	0.939	0.847	0.934
	α_{532}	0.957	0.969	0.969	0.967
	β_{355}	0.741	0.679	0.684	0.626
	β_{532}	0.970	0.851	1.002	0.827
	β_{1064}	1.090	0.816	1.036	0.818

Table 4. Lidar derived parameters for predicting optical-related CCN activation ratio ~~AR_{ξ}~~ AR_{ξ}

Parameter	Description
$\kappa_{\alpha 355}$	Fitted parameter of extinction humidogram at 355 nm in κ-equation form
$\kappa_{\alpha 532}$	Fitted parameter of extinction humidogram at 532 nm in κ-equation form
$\kappa_{\beta 355}$	Fitted parameter of backscatter humidogram at 355 nm in κ-equation form
$\kappa_{\beta 532}$	Fitted parameter of backscatter humidogram at 532 nm in κ-equation form
$\kappa_{\beta 1064}$	Fitted parameter of backscatter humidogram at 1064 nm in κ-equation form
$s_{\alpha 355}$	Particle dry lidar extinction-to-backscatter ratio at 355 nm
$s_{\alpha 532}$	Particle dry lidar extinction-to-backscatter ratio at 532 nm
$\mathring{a}_{\alpha 355\&532}$	Ångström exponent of particle dry extinction coefficients between 355 and 532 nm
$\mathring{a}_{\beta 532\&1064}$	Ångström exponent of particle dry backscatter coefficients between 532 and 1064 nm

Table 5. Slopes of linear regressions, determination coefficients (R^2), and relative errors (RE) between theoretical calculated CCN number concentrations and CCN number concentrations retrieved with/without κ_ξ as input parameter. The relative errors are given in the form of mean value ± one standard deviation (std).

Supersaturation Ratio	With κ_ξ			Without κ_ξ		
	slope	R^2	RE(%)	slope	R^2	RE(%)
0.07%	0.991	0.991	-0.8 ± 6.0	0.877	0.866	4.6 ± 26.1
0.10%	0.992	0.989	0.1 ± 6.3	0.857	0.837	5.9 ± 26.7
0.20%	1.005	0.973	3.9 ± 9.0	0.860	0.785	11.9 ± 28.1

Table 6. Mean and one standard deviation (std) values of relative errors in retrieved CCN number concentrations at different supersaturations with error-free and random error (10% for ~~backscatter/extinction~~backscatter and extinction and 5% for relative humidity) conditions.

Supersaturation Ratio	**Error-free (mean ± std)**	**Random Error (mean ± std)**
0.07%	-0.8% ± 6.0%	-2.8% ± 29.7%
0.10%	0.1% ± 6.3%	-1.3% ± 31.5%
0.20%	3.9% ± 9.0%	1.3% ± 42.9%

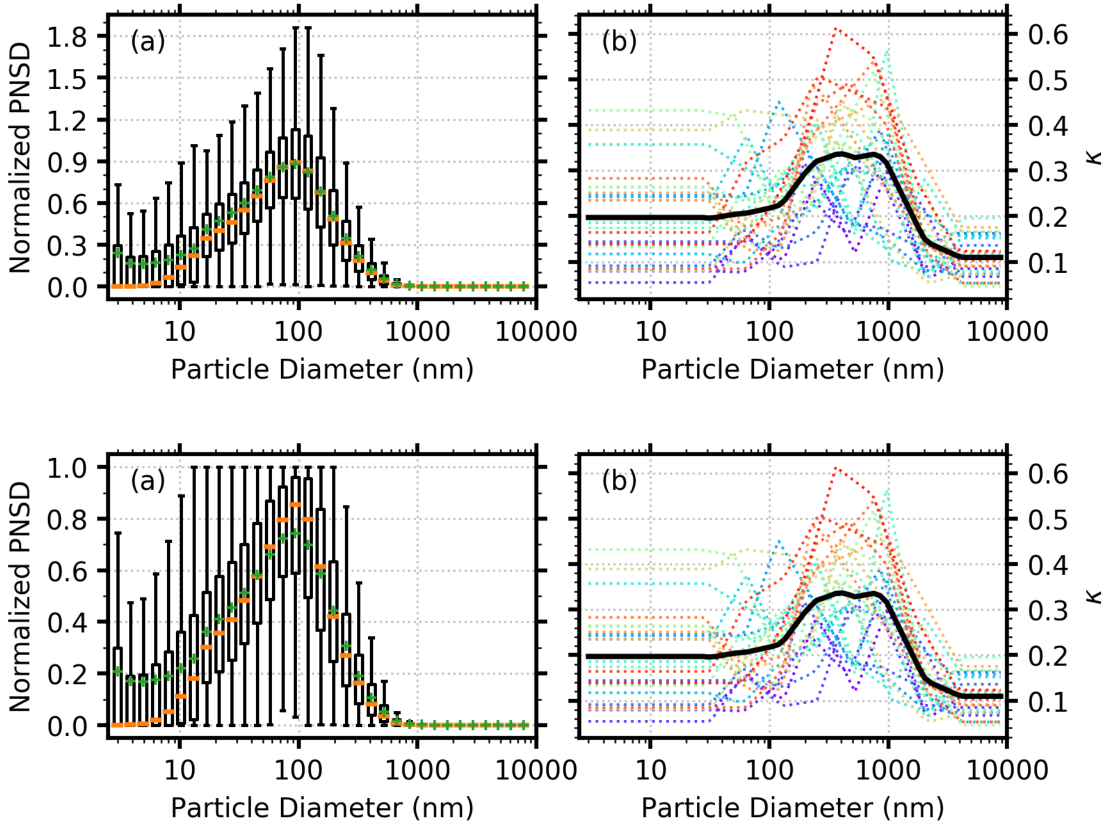
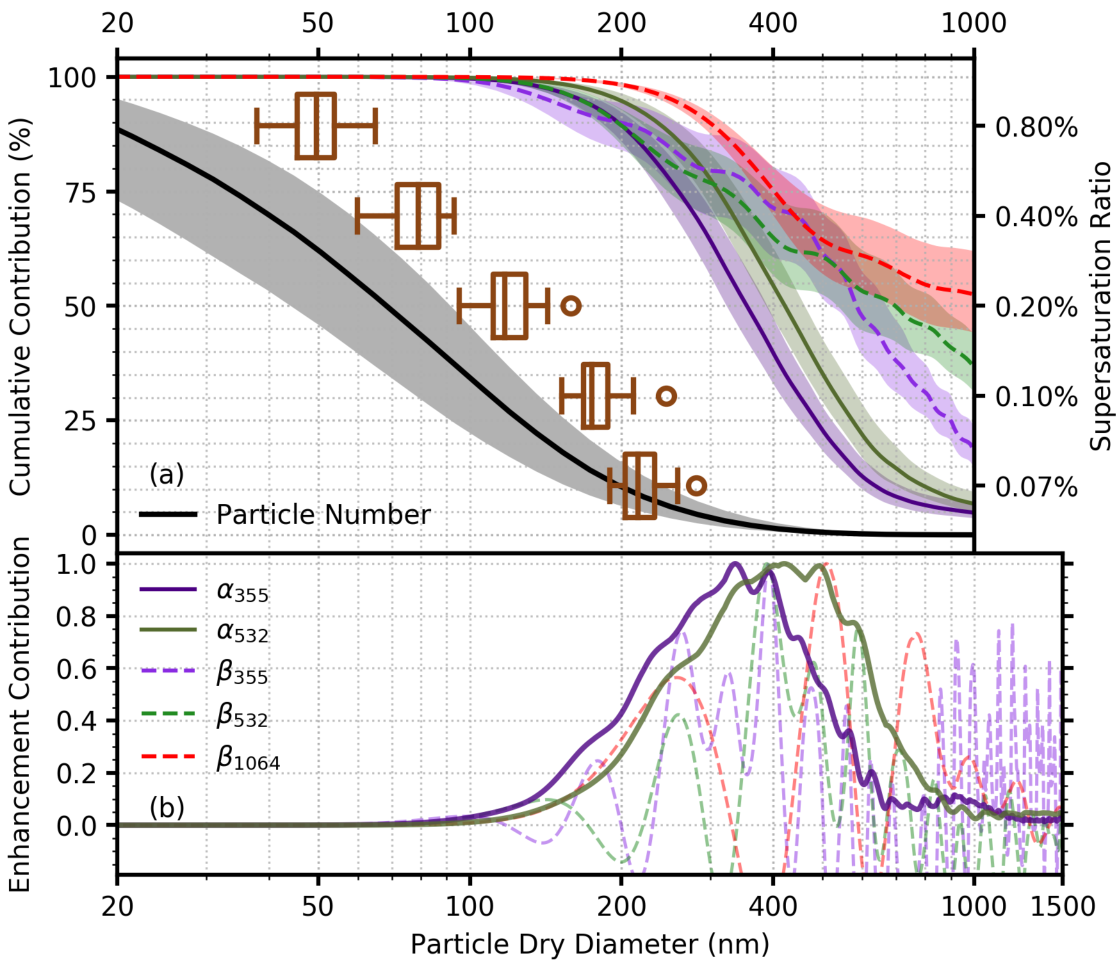


Figure 1. (a) Boxplot of particle number size distributions (PNSDs) in the datasets from five field campaigns. Each PNSD is normalized by its ~~number concentration of total particles~~maximum value at the peak diameter. Green markers "+" represent the mean value of each diameter. The boxes extend from the lower to upper quartile values, with orange lines at the median. The whiskers extend from the box to the minimum/maximum values or extend from the box by 1.5 times of interquartile range. The flyers are not shown in the plot. **(b)** Twenty-five typical size-resolved κ distributions. Each dotted line with color represents one size-resolved κ distribution. The solid black line represents the mean value of the size-resolved κ distributions.



~~**Figure 2. (a)** Cumulative contributions (accumulate from large particle size to small particle size) of particle number concentrations (measured), dry particle backscatter coefficients (simulated), and dry particle extinction coefficients (simulated). The solid and dashed lines represent the median values of five field campaigns, and the shadows cover from the lower to upper quartile values. The box plots in brown contain statistical information about critical diameter of each supersaturation condition (right y-axis). The boxes extend from the lower to upper quartile values, with lines at the median. The whiskers extend from the box to the minimum/maximum values or extend from the box by 1.5 times of interquartile range. The markers "o" are the flyers. **(b)** Normalized size-resolved enhancement contributions when relative humidity increase from 60% to 90%, which are theoretically calculated by the mean particle number size distribution, the mean black carbon mass concentration (4.717 μg m^{-3}), the mean mass ratio of externally mixed black carbon (0.664%), and the mean size-resolved κ distribution.~~

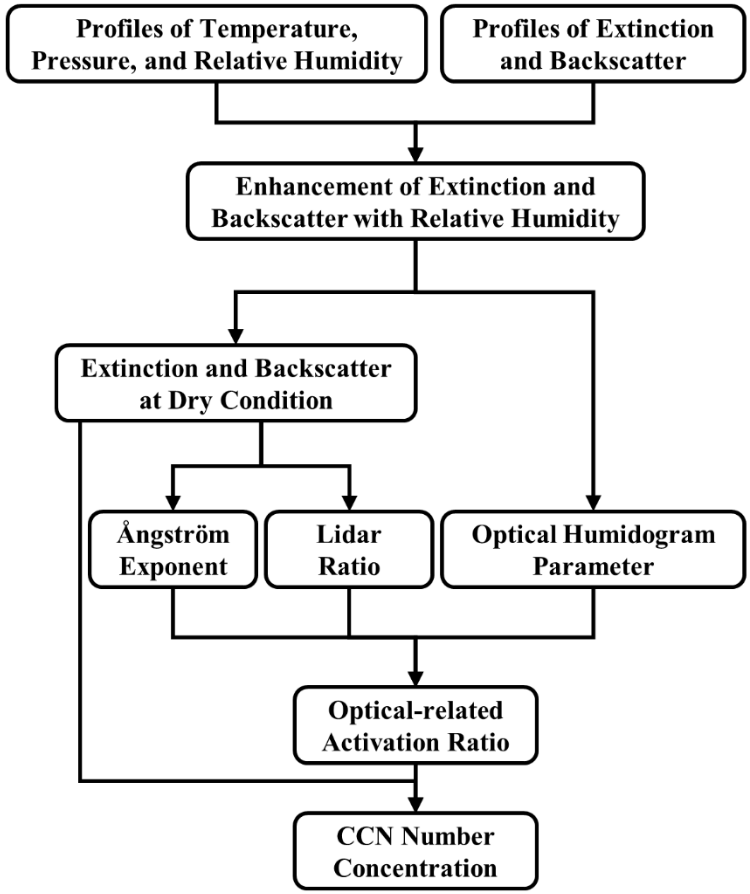


Figure 22~~3~~. Schematic diagram of newly proposed method to retrieve cloud condensation nuclei number concentrations using multiwavelength Raman lidar.

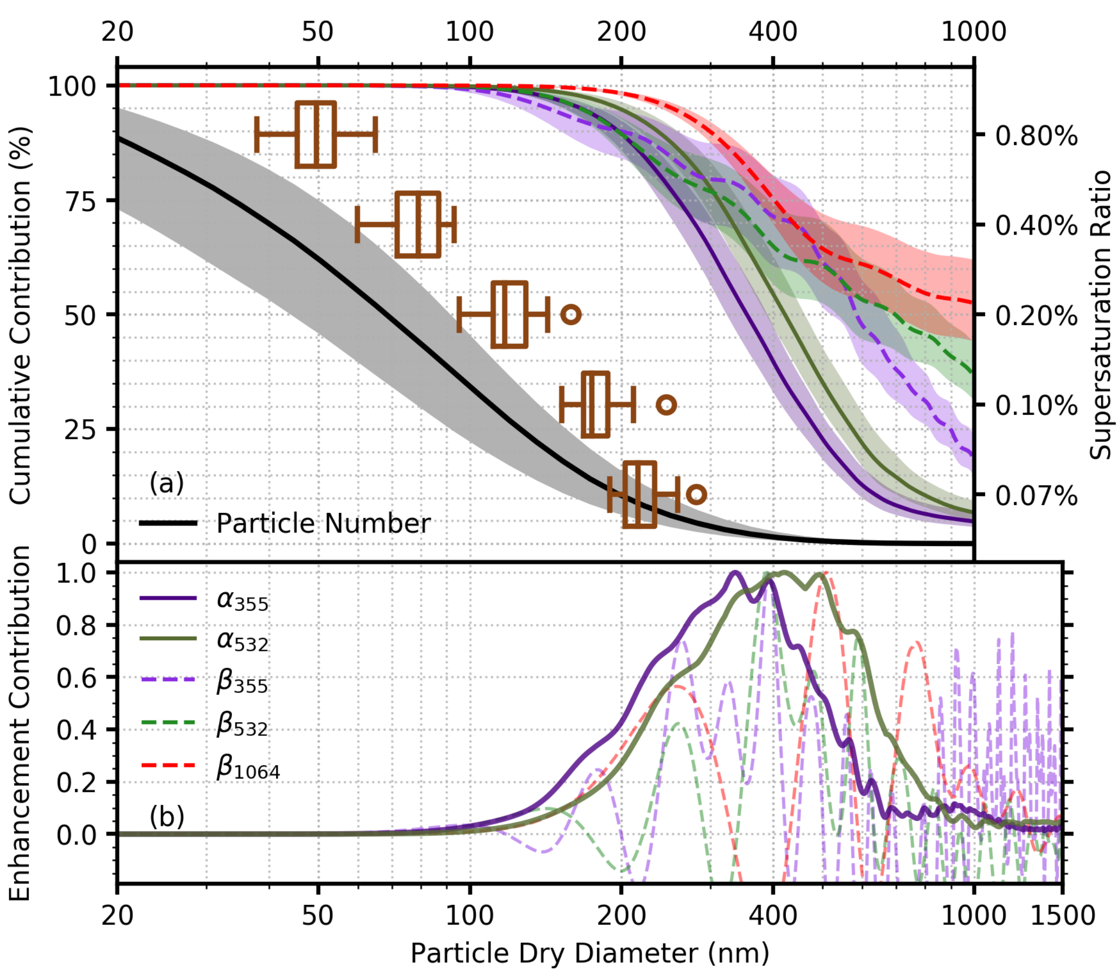


Figure 3~~32~~. (a) Cumulative contributions (accumulate from large particle size to small particle size) of particle number concentrations (measured), dry particle backscatter coefficients (simulated), and dry particle extinction coefficients (simulated). The solid and dashed lines represent the median values of five field campaigns, and the shadows cover from the lower to upper quartile values. The box plots in brown contain statistical information about critical diameter of each supersaturation condition (right y-axis). The boxes extend from the lower to upper quartile values, with lines at the median. The whiskers extend from the box to the minimum/maximum values or extend from the box by 1.5 times of interquartile range. The markers "o" are the flyers. **(b)** Normalized size-resolved enhancement contributions when relative humidity increases from 60% to 90%, which are theoretically calculated by the mean particle number size distribution, the mean black carbon mass concentration (4.717 μg m^{-3}), the mean mass ratio of externally mixed black carbon (0.664%), and the mean size-resolved κ distribution.

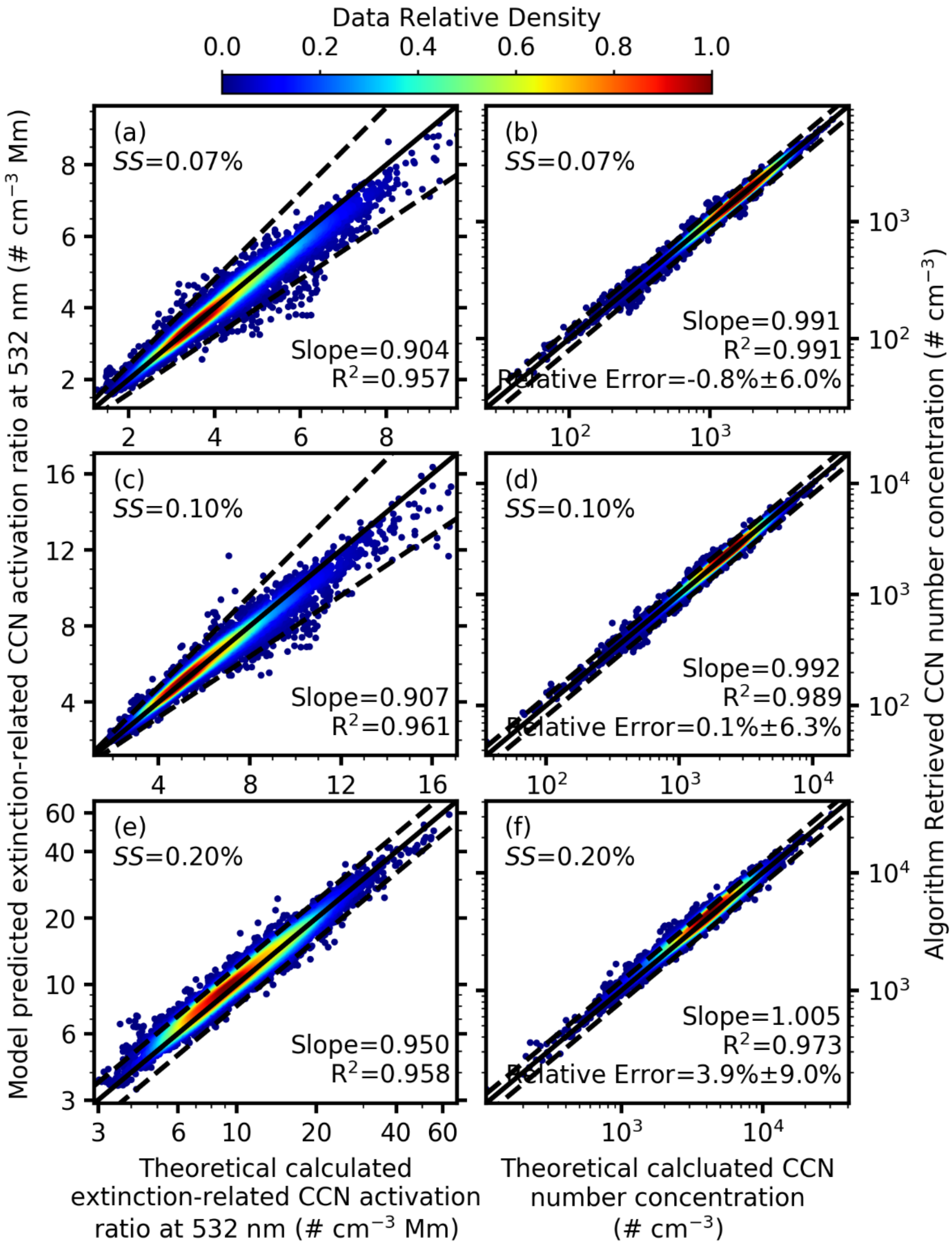


Figure 4. Comparison of the theoretical calculated extinction-related CCN activation ratio at 532 nm and the model predicted extinction-related CCN activation ratios at 532 nm at supersaturations of **(a)** 0.07%, **(c)** 0.10%, and **(e)** 0.20%, and of the theoretical calculated CCN number concentrations and the retrieved CCN number concentrations at supersaturations of **(b)**

0.07%, **(d)** 0.10%, and **(f)** 0.20%. A total of 80575 pairs of data calculated from campaign C5 are used. The solid line is 1:1 line, and the dashed lines are 20% relative difference lines. Colors represent the relative density of the data points normalized by the maximum data density of each panel. The relative error showed in the figure is mean value ± one standard deviation.

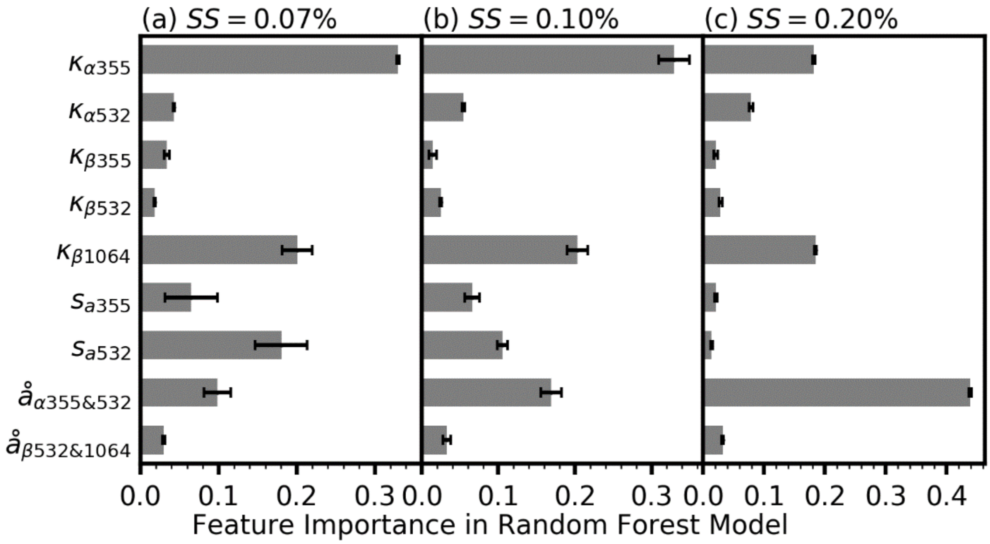


Figure 5. Importance of each feature (input parameter) output by the Random Forest model for predicting optical-related CCN activation ratios at supersaturations of **(a)** 0.07%, **(b)** 0.10%, and **(c)** 0.20%. The values of feature importance indicate the decrease in impurity for each feature. The length of the bar~~s~~ represents the mean values among all trees and the error bars give the standard deviations.

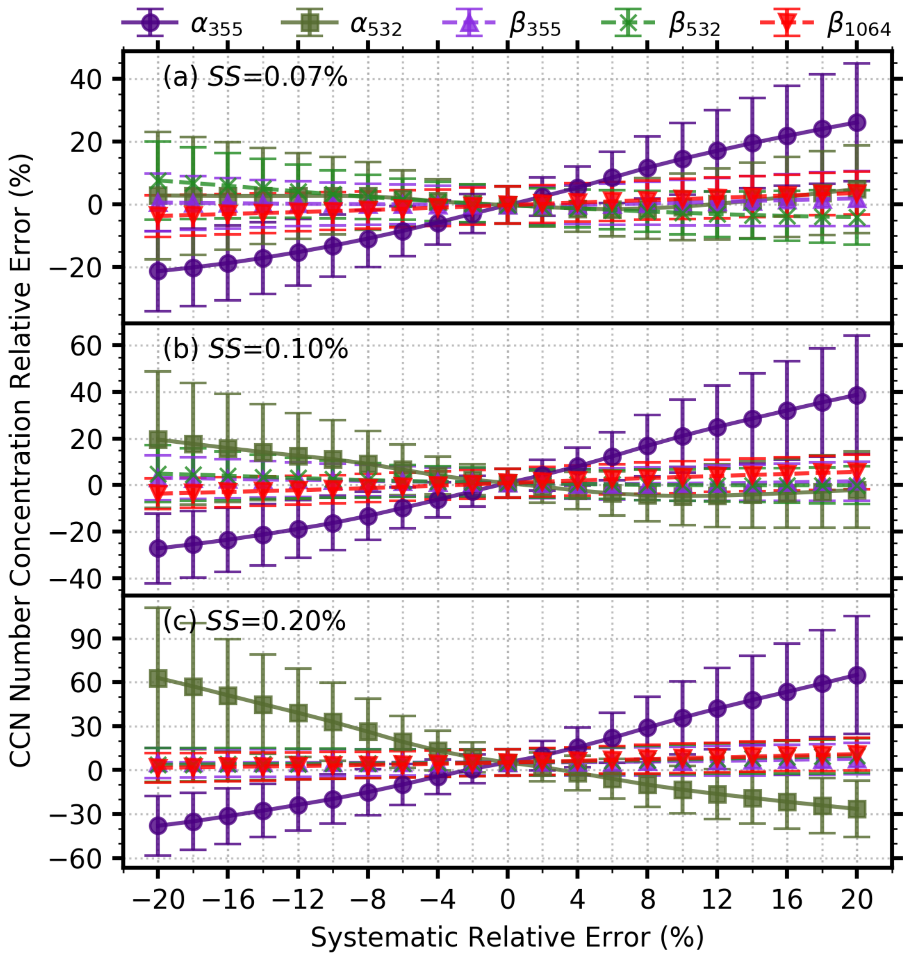


Figure 6. Relative errors in retrieved CCN number concentrations at supersaturations of **(a)** 0.07%, **(b)** 0.10%, and **(c)** 0.20% as a function of systematic errors in backscatter or extinction. The markers are the mean values, and the error bars denote the standard deviations.

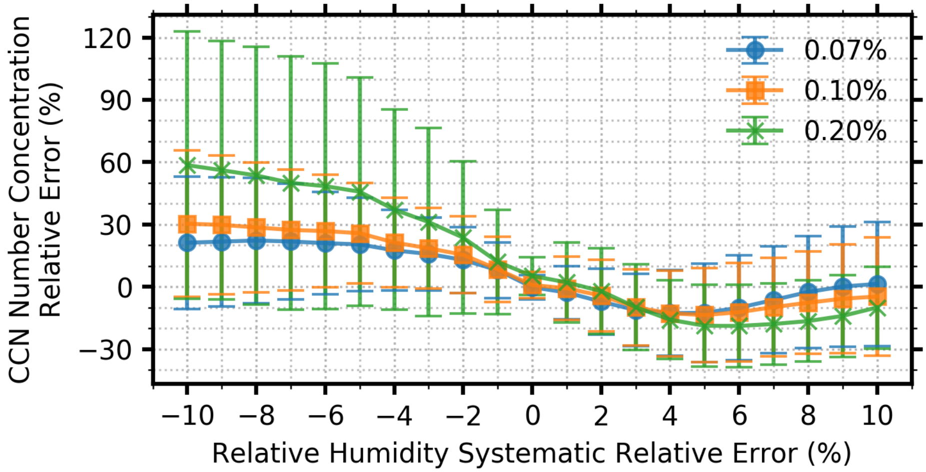


Figure 7. Relative errors in retrieved CCN number concentrations at supersaturations of 0.07%, 0.10%, and 0.20% as a function of systematic errors in relative humidity. The markers are the mean values, and the error bars denote the standard deviations.

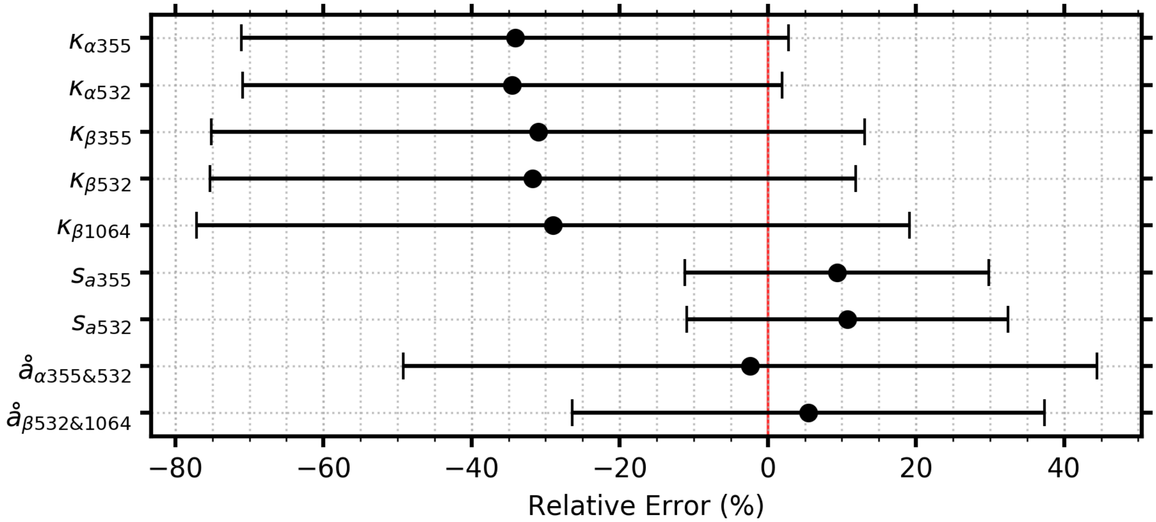


Figure 8. Relative errors in fitted and calculated parameters with 10% random errors for ~~backscatter/extinction~~backscatter and extinction and 5% random error for relative humidity. The dots are the mean values, and the error bars denote the standard deviations.

S1 Site information

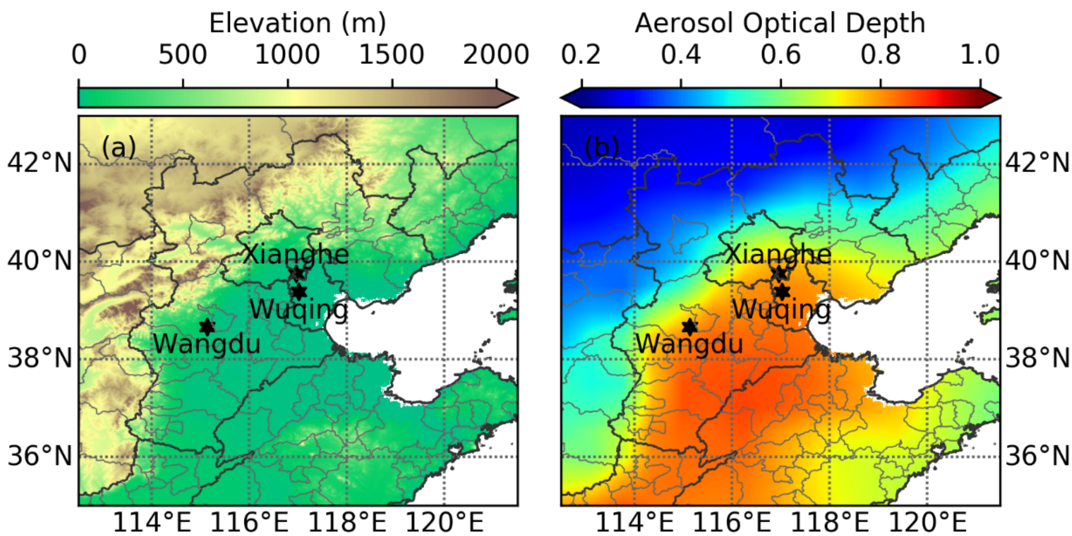


Figure S1. Site locations of Wuqing (39°23′N, 117°01′E, 7.4 m a.s.l), Xianghe (39°45′N, 116°58′E, 36 m a.s.l), and Wangdu (38°40′N, 115°08′E, 51 m a.s.l). Filled colors represents **(a)** elevation and **(b)** averaged aerosol optical depth (AOD). The AOD data is from reanalysis datasets of the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2, Global Modeling and Assimilation Office (GMAO) (2015), MERRA-2 instM_2d_gas_Nx: 2d,Monthly mean,Instantaneous,Single-Level,Assimilation,Aerosol Optical Depth Analysis V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [11 September 2018], 10.5067/XOGNBQEPLUC5). The averaged AOD is calculated from monthly mean values of all months during the five field campaigns shown in Table 1.

S2 Time series of the normalized particle number size distribution (PNSD)

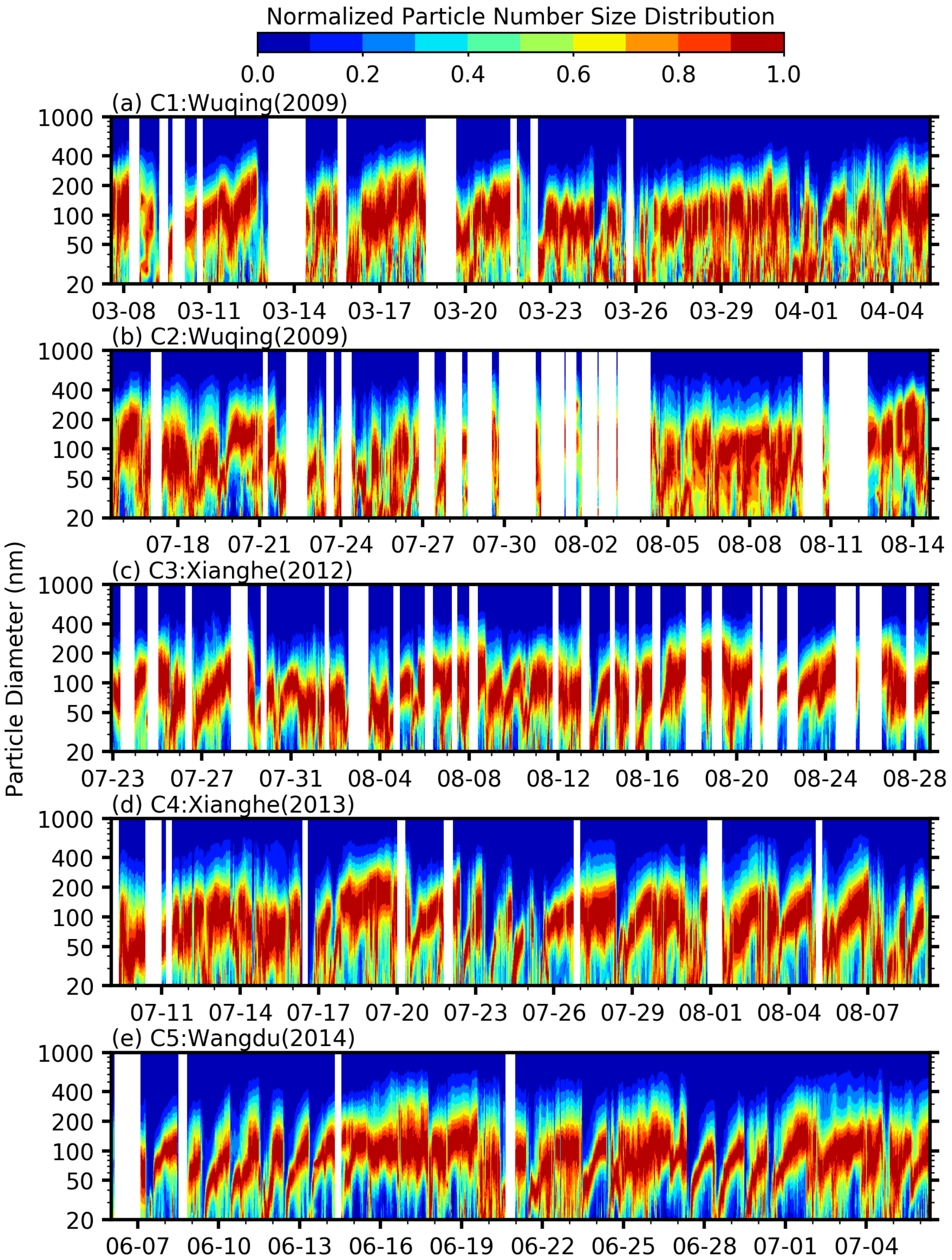


Figure S2. Time series of the normalized particle number size distribution from the five field campiangs. (a)-(e) represents campaign C1-C5, respectively.

~~S2~~ S3 ~~Black carbon size distribution used in t~~Theoretical simulations of CCN, lidar backscatter, and extinction

S3.1 Aerosol model assumptions

We assume the aerosol particles act as follows:

(1) Aerosol particles are spherical, which means the simulation results are not appropriate for mineral dust.

(2) Particles are partially externally mixed and partially core-shell mixed. Only two kinds of aerosols are contained: pure black carbon (BC) and BC coated by non-light-absorbing components. Note that if $r_{\mathrm{ext}} = 1$, there exists pure BC and pure non-light-absorbing particles. r_{ext} is defined with Eq. (2) in the paper.

(3) The shape of BC mass size distribution (BCMSD) remains unchanged and the amount is related to the total BC mass concentration (m_{BC}). The fixed distribution comes from the average BCMSD obtained from Berner impactor measurements (Ma et al., 2012) and is shown in Fig. S3.

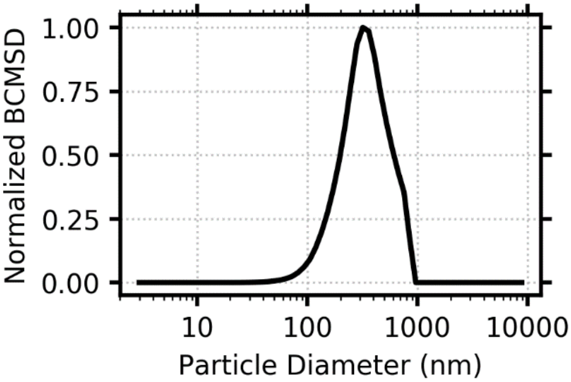


Figure S3~~2~~. Normalized size distribution of black carbon (BC) mass/volume concentration. The distribution is the average black carbon mass concentration obtained from Berner impactor measurements (Ma et al., 2012) and is normalized by the maximum value of the distribution.

(4) r_{ext} is uniform among different particle sizes. Accordingly, number concentrations of externally mixed BC (N_{ext}) can easily be calculated from BCMSD and r_{ext}:

$$N_{\mathrm{ext}}(D) = \frac{r_{\mathrm{ext}} \cdot m_{\mathrm{BC}}(D)}{\frac{\pi}{6} D^3 \cdot \rho_{\mathrm{BC}}}, \qquad \text{(S1)}$$

where D is diameter of the particle, ρ_{BC} is the density of BC. In this study, ρ_{BC} is assumed to be 1.5 g/cm^3, which is also the density when retrieving r_{ext}.

(5) For each particle size, the diameters of BC cores (D_{core}) are the same and can be derived using the following equation:

$$D_{\mathrm{core}}(D) = \sqrt[3]{\frac{6(1-r_{\mathrm{ext}}) \cdot m_{\mathrm{BC}}(D)}{\pi \rho_{\mathrm{BC}} \cdot [N(D) - N_{\mathrm{ext}}(D)]}}, \qquad \text{(S2)}$$

where $N(D)$ represents particle number size distribution (PNSD).

(6) The size-resolved κ distributions represent the bulk hygroscopicity of core-shell mixed particles, and externally mixed BC particles do not take up water.

All these assumptions above are strong and, to some extent, inconsistent with the reality, but are certified to be reasonable for calculating aerosol optical properties. Plenty of works on aerosol optical closure studies (Ma et al., 2011;Ma et al., 2012) and aerosol optical simulations (Kuang et al., 2017;Kuang et al., 2018;Zhao et al., 2018) have been carried out with these aerosol model assumptions. In particular, Zhao et al. (2017) use the aerosol model to simulate lidar backscatter and extinction under different relative humidity (RH) conditions.

S3.2 Calculations of CCN number concentrations using κ-Köhler theroy

According to κ-Köhler theory, CCN number concentrations at a specific supersaturation level can be calculated by PNSD and size-resolved κ distribution. Based on Eq. (3), the critical supersaturation ratio required to activate a particle is decided by corresponding κ and D_{dry}. In other word, we can get a critical activation dry diameter D_c with a given supersaturation ratio and size-resolved κ distribution. Then CCN number concentration $N_{\mathrm{CCN}}(SS)$ thereby can be calculated with Eq. (S3):

$$N_{\mathrm{CCN}}(SS) = \int_{D_c(SS)}^{D_{\max}} [N(D) - N_{\mathrm{ext}}(D)]\, dD \,, \tag{S3}$$

where $D_{\max}$ correspond to the upper bounds of the measured PNSD. Note that we regard externally mixed pure BC as non-hygroscopic particles, so r_{ext} and m_{BC} should also be involved to calculate N_{ext}, which needs to be subtracted. Otherwise, N_{CCN} will be overestimated.

S3.3 Calculations of particle backscatter and extinction coefficients at different RH using κ-Köhler theroy and Mie theory

We use a modified BHMIE Fortran code and a modified BHCOAT Fortran code to calculate optical properties of homogeneous spherical particles and coated spherical particles, respectively. For a homogeneous spherical particle (i.e. externally mixed BC in this study), BHMIE can calculate particle scattering and extinction efficiency (Q_{sca} and Q_{ext}) and scattering phase function with given light wavelength, particle diameter, and complex refractive index. For a coated spherical particle (i.e. core-shell mixed particle), diameters and complex refractive indices of both core and shell are needed in BHCOAT. The particle backscatter and extinction coefficients we need are derived from Q_{sca}, Q_{ext}, and scattering phase function at 180° $P(\pi)$:

$$\alpha = \sum_i \left[\int_{D_{\min}}^{D_{\max}} \frac{1}{4} D^2 Q_{\mathrm{ext}}(D,i) N(D,i)\, dD\right], \tag{S4}$$

$$\beta = \sum_i \left[\int_{D_{\min}}^{D_{\max}} \frac{1}{16} D^2 Q_{\mathrm{sca}}(D,i) P(\pi,D,i) N(D,i)\, dD\right], \tag{S5}$$

where $D_{\min}$ and $D_{\max}$ correspond respectively to the lower and upper bounds of the measured PNSD, and the index i indicates the mixing state of particles, i.e. external or core-shell in this paper. Polarization of lidar emitted laser is neglected.

Complex refractive indices are essential for Mie scattering calculation. Aerosol complex refractive indices are related to chemical components, morphology, and wavelengths of light (Cotterell et al., 2017). Both real part and imaginary part of refractive indices vary a lot in real ambient environment (Shettle and Fenn, 1979). Wavelength dependency of refractive indices at wavelengths of 355 nm, 532 nm, and 1064 nm are not significant except for brown carbon (Shettle and Fenn, 1979;Bond et al., 2013). Neglecting the effect of brown carbon in this study, we simply assume that complex refractive indices of corresponding components do not change with wavelengths. The refractive indices of BC and non-light-absorbing component (shell) are set to be $1.8+0.54i$ (Ma et al., 2012) and $1.53+10^{-7}i$ (Wex et al., 2002), respectively.

Concerning aerosol hygroscopicity for core-shell mixed particles, diameters of BC cores D_{core} remain unchanged, and particle diameters D at different RH can be calculated with Eq. (3). The refractive index of the swelling shell ($\widetilde{m}_{\mathrm{shell}}$) is calculated following the volume mixing law (Hanel, 1968):

$$\widetilde{m}_{\mathrm{shell}} = f_{\mathrm{solute}} \cdot \widetilde{m}_{\mathrm{solute}} + (1 - f_{\mathrm{solute}}) \cdot \widetilde{m}_{\mathrm{water}}\ , \tag{S6}$$

where $\widetilde{m}_{\mathrm{solute}}$ is the refractive index of solute (i.e. $1.53+10^{-7}i$ in this study), $\widetilde{m}_{\mathrm{water}}$ is the refractive index of pure water ($1.33+10^{-7}i$), and f_{solute} is the solute volume fraction of the in solution (shell), which is determined by Eq. (S7):

$$f_{\mathrm{solute}} = \frac{D_{\mathrm{dry}}^3 - D_{\mathrm{core}}^3}{D^3 - D_{\mathrm{core}}^3}\ . \tag{S7}$$

S4 Performance of fitting humidogram functions with parameterization equations

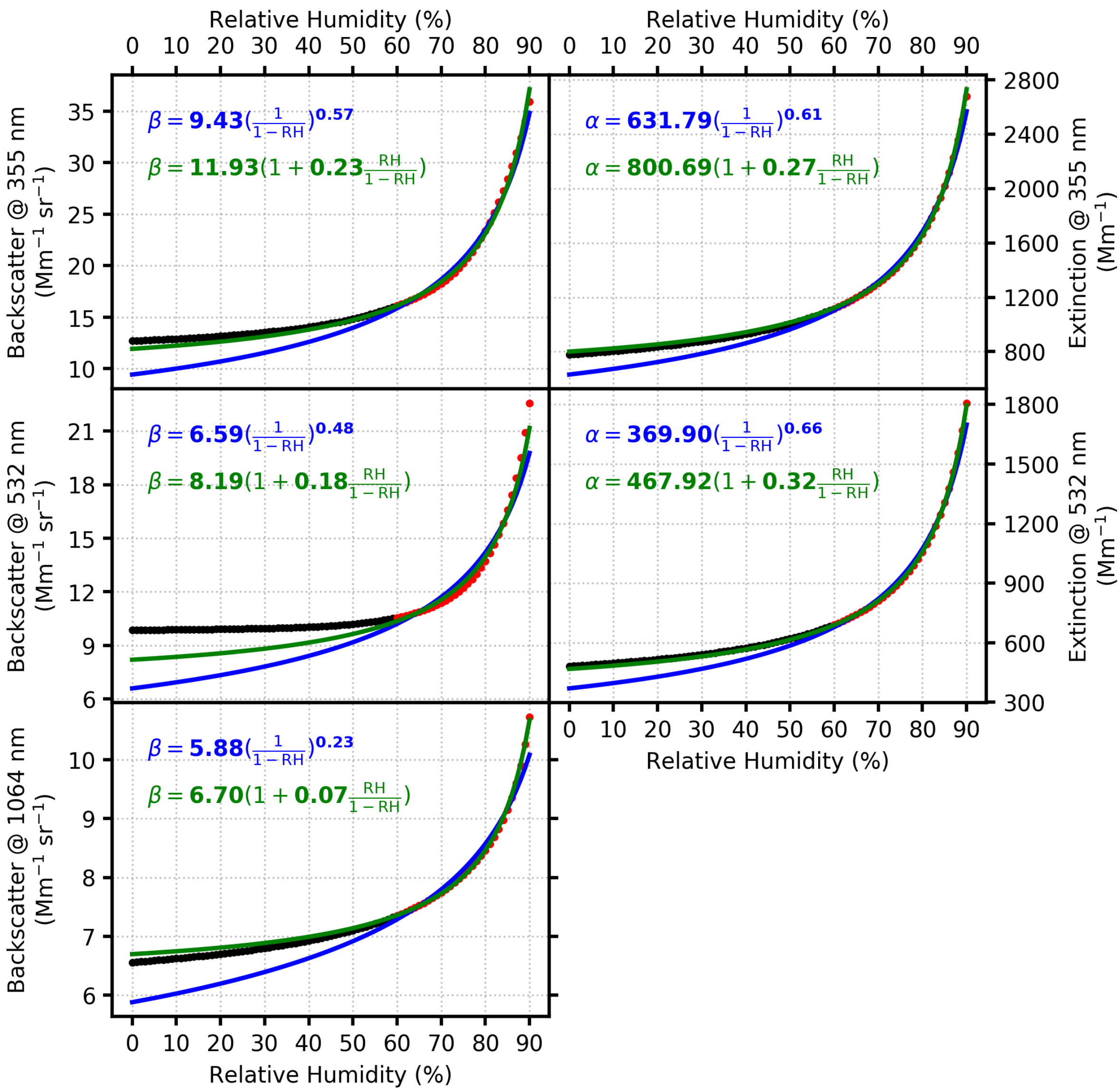


Figure S4. Example of humidogram fitting using different functions. The example is calculated with one set of PNSD, BC, r_{ext}, and size-resolved κ distribution. The dots represent Mie model simulations, and the dots in red (within RH range of 60-90%) are used to fit parameterization lines. The blue line is the result of γ-equation, and the green line represents the result of κ-equation.

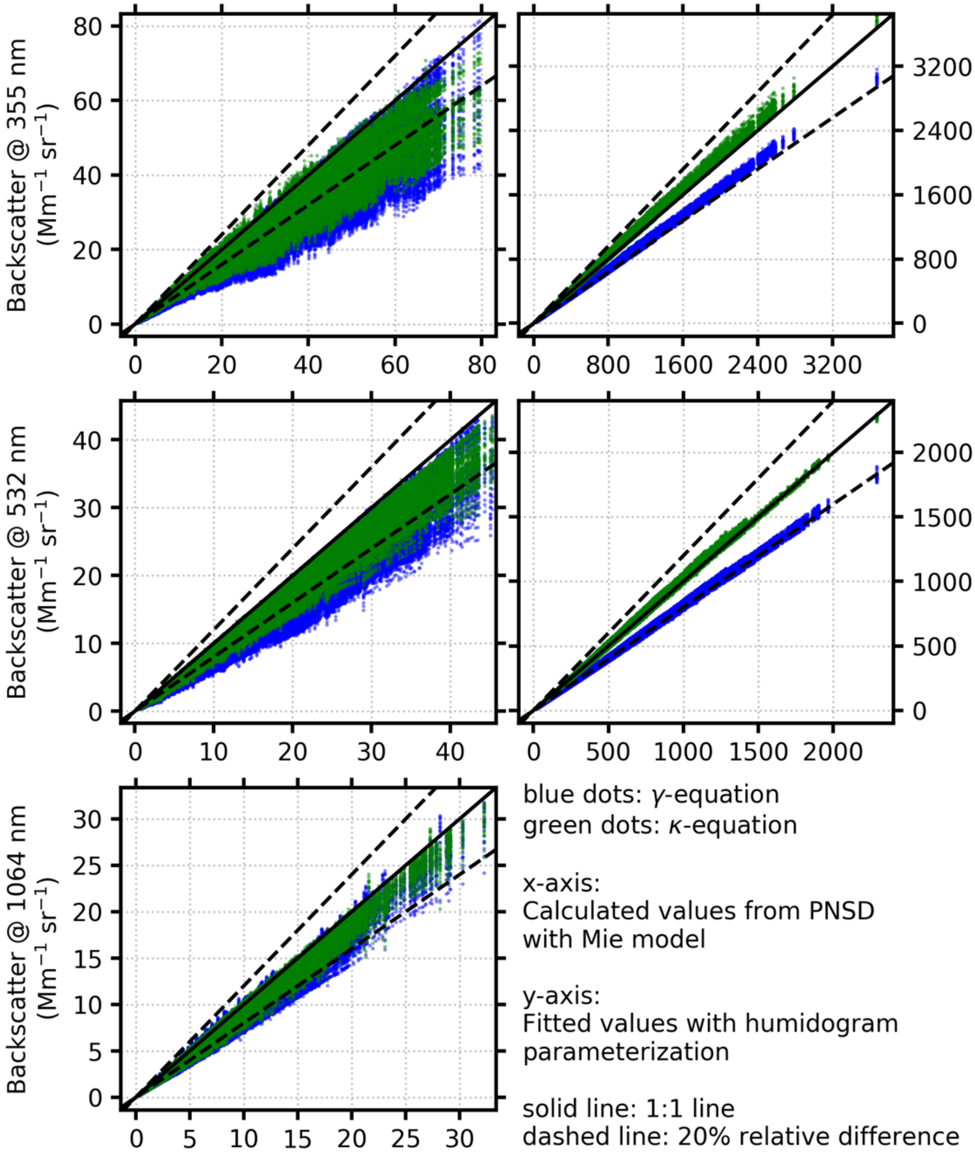


Figure S5. Comparison between Mie model calculated dry particle backscatter or extinction and those fitted from humidograms.

S5 Relationship between the nine input parameters in Table 4

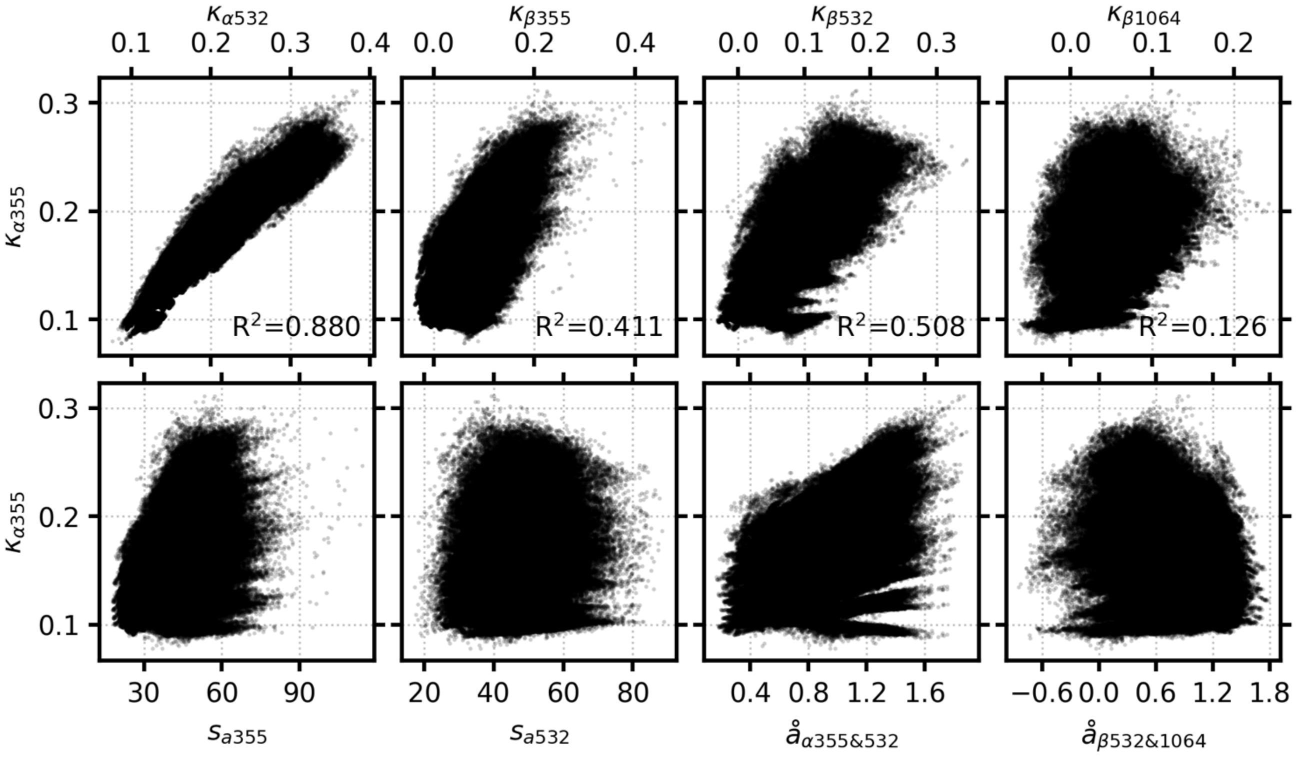


Figure S6. Relationship between $\kappa_{\alpha 355}$ and other 8 parameters.

Table S1. Determine coefficients (R2) between the 9 input parameters in Table 4.

R^2	$\kappa_{\alpha 355}$	$\kappa_{\alpha 532}$	$\kappa_{\beta 355}$	$\kappa_{\beta 532}$	$\kappa_{\beta 1064}$	S_{a355}	S_{a532}	$å_{\alpha 355\&532}$
$\kappa_{\alpha 355}$	–	–	–	–	–	–	–	–
$\kappa_{\alpha 532}$	0.880	–	–	–	–	–	–	–
$\kappa_{\beta 355}$	0.411	0.321	–	–	–	–	–	–
$\kappa_{\beta 532}$	0.508	0.644	0.450	–	–	–	–	–
$\kappa_{\beta 1064}$	0.126	0.222	0.016	0.056	–	–	–	–
S_{a355}	0.085	0.073	0.680	0.292	0.019	–	–	–
S_{a532}	0.026	0.070	0.117	0.423	0.070	0.360	–	–
$å_{\alpha 355\&532}$	0.149	0.135	0.505	0.267	0.027	0.627	0.089	–
$å_{\beta 532\&1064}$	0.062	0.023	0.550	0.169	0.464	0.409	0.023	0.317

~~S3~~ S6 Determine the tuning parameters for Random Forest model

In this study, we use the Python module *RandomForestRegressor* from the Python Scikit-Learn library (http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html, last access: 18 December 2018) as the Random Forest (RF) model tool. The tuning parameters of the model are listed in Table ~~S1~~S2. More detailed meanings about the setting values please refer to the user guide provided by the website.

The most import tuning parameter in the model is the number of trees in the forest (*n_estimators*). The influence of *n_estimators* on the accuracy of retrieved CCN number concentrations is tested. Here we use the same test method as introduced in Section 4.2 in the paper. The determination coefficients (R^2) and the mean absolute relative error (MARE) between theoretical calculated and retrieved CCN number concentrations with different *n_estimators* are shown in ~~Figure~~Fig. S~~3~~7. The accuracy of the predictions increases as *n_estimators* grows bigger and are insensitive when *n_estimators* is bigger than 60. Considering computational and time cost, we finally set *n_estimators* to 100.

Table S2~~1~~. Tuning parameters and their setting values of the Python module *RandomForestRegressor*.

Parameter	**Description**	**Values**
n_estimators	The number of trees in the forest	100
criterion	The function to measure the quality of a split	“mse”
max_features	The number of features to consider when looking for the best split	“auto”
max_depth	The maximum depth of the tree	None
min_samples_split	The minimum number of samples required to split an internal node	2
min_samples_leaf	The minimum number of samples required to be at a leaf node	1
min_weight_fraction_leaf	The minimum weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node	0
max_leaf_nodes	Grow trees with *max_leaf_nodes* in best-first fashion	None
min_impurity_decrease	A node will be split if this split induces a decrease of the impurity greater than or equal to this value	0

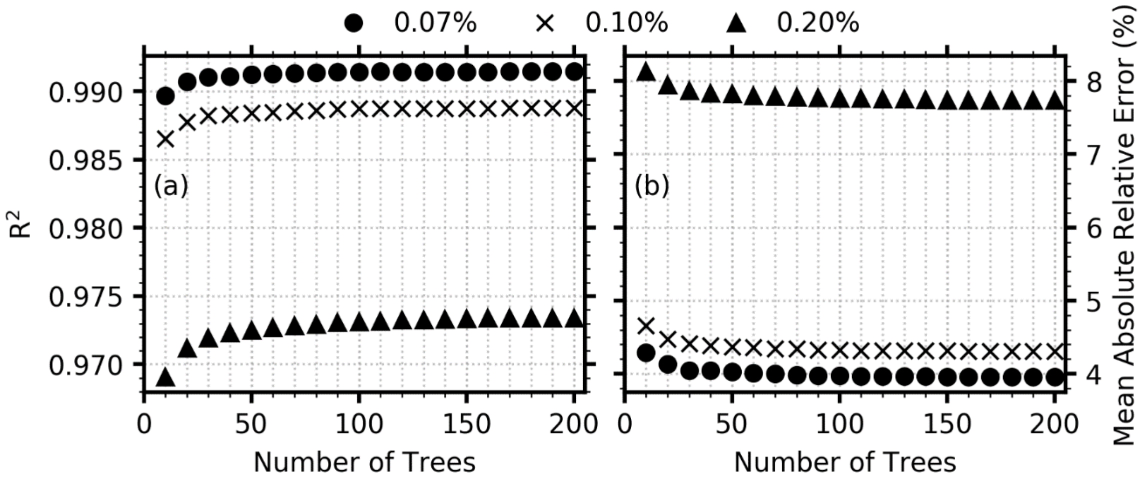


Figure S7~~3~~. Influence of the number of trees in RF model on retrieving CCN number concentrations. Dependencies of tree numbers on **(a)** R^2 and **(b)** MARE between theoretical calculated CCN number concentrations and retrieved CCN number concentrations under different supersaturations.

~~S4~~ S7 Retrieving CCN number concentrations at high supersaturations

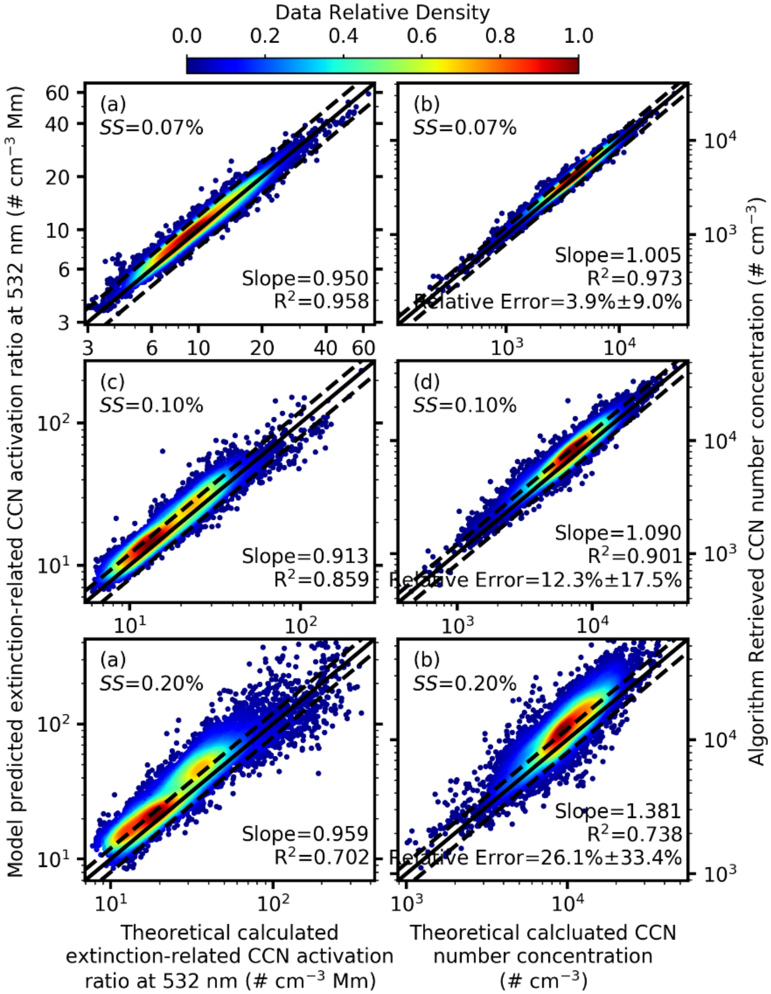


Figure S84. Comparison of the theoretical calculated extinction-related CCN activation ratio at 532 nm and the model predicted extinction-related CCN activation ratios at 532 nm at supersaturations of **(a)** 0.20%, **(c)** 0.40%, and **(e)** 0.80%, and of the theoretical calculated CCN number concentrations and the retrieved CCN number concentrations at supersaturations of **(b)** 0.20%, **(d)** 0.40%, and **(f)** 0.80%. A total of 80575 pairs of data calculated from campaign C5 are used. The solid line is 1:1 line, and the dashed lines are 20% relative difference lines. Colors represent the relative density of the data points normalized by the maximum data density of each panel. The relative error showed in the figure is mean value ± one standard deviation.

~~S5~~ S8 Retrieving CCN number concentrations without hygroscopicity information

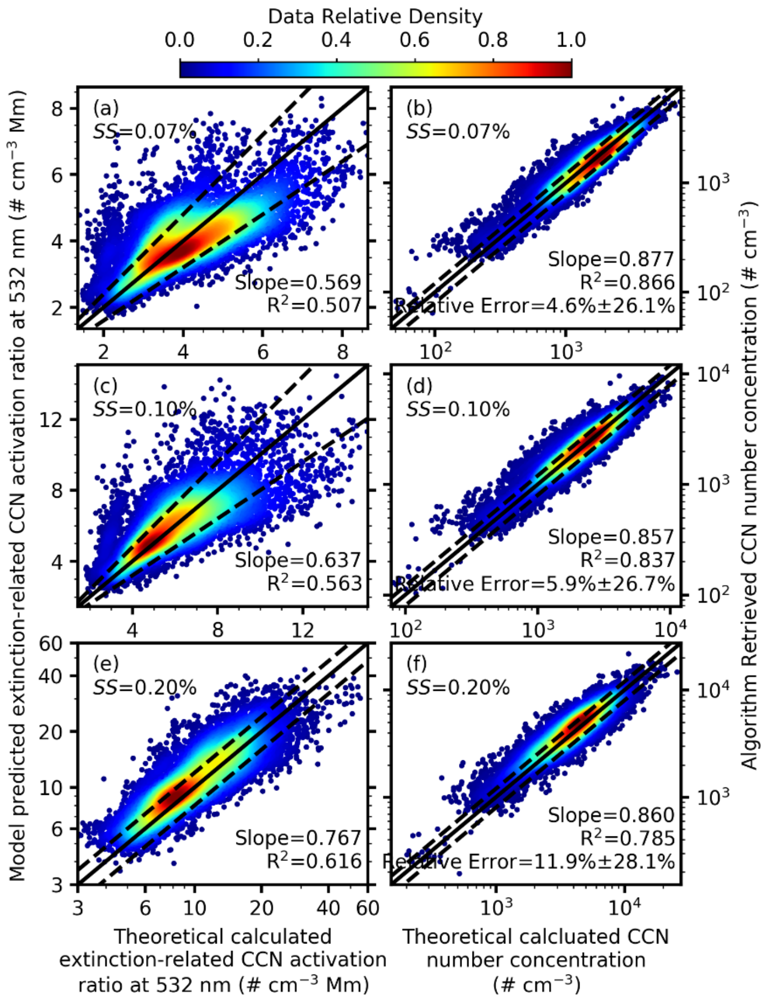


Figure S~~5~~9. Comparison of the theoretical calculated extinction-related CCN activation ratio at 532 nm and the model predicted extinction-related CCN activation ratios at 532 nm at supersaturations of **(a)** 0.07%, **(c)** 0.10%, and **(e)** 0.20%, and of the theoretical calculated CCN number concentrations and the retrieved CCN number concentrations at supersaturations of **(b)** 0.07%, **(d)** 0.10%, and **(f)** 0.20%. A total of 80575 pairs of data calculated from campaign C5 are used. The solid line is 1:1 line, and the dashed lines are 20% relative difference lines. Colors represent the relative density of the data points normalized by the maximum data density of each panel. The relative error showed in the figure is mean value ± one standard deviation.

References

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., K?rcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, J. Geophys. Res. Atmos., 118, 5380-5552, 10.1002/jgrd.50171, 2013.

Cotterell, M. I., Willoughby, R. E., Bzdek, B. R., Orr-Ewing, A. J., and Reid, J. P.: A complete parameterisation of the relative humidity and wavelength dependence of the refractive index of hygroscopic inorganic aerosol particles, Atmos. Chem. Phys., 17, 9837-9851, 10.5194/acp-17-9837-2017, 2017.

Hanel, G.: The real part of the mean complex refractive index and the mean density of samples of atmospheric aerosol particles, Tellus, 20, 371-379, 10.3402/tellusa.v20i3.10016, 1968.

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

Kuang, Y., Zhao, C. S., Zhao, G., Tao, J. C., Xu, W., Ma, N., and Bian, Y. X.: A novel method for calculating ambient aerosol liquid water content based on measurements of a humidified nephelometer system, Atmos. Meas. Tech., 11, 2967-2982, 10.5194/amt-11-2967-2018, 2018.

Ma, N., Zhao, C. S., Nowak, A., Müller, T., Pfeifer, S., Cheng, Y. F., Deng, Z. Z., Liu, P. F., Xu, W. Y., Ran, L., Yan, P., Göbel, T., Hallbauer, E., Mildenberger, K., Henning, S., Yu, J., Chen, L. L., Zhou, X. J., Stratmann, F., and Wiedensohler, A.: Aerosol optical properties in the North China Plain during HaChi campaign: an in-situ optical closure study, Atmos. Chem. Phys., 11, 5959-5973, 10.5194/acp-11-5959-2011, 2011.

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

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, Environmental Research Papers, 676, 89, 10.1109/TR.1987.5222381, 1979.

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, J. Geophys. Res. Atmos., 107, LAC 4-1-LAC 4-18, 10.1029/2000JD000234, 2002.

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., 17, 12133-12143, 10.5194/acp-17-12133-2017, 2017.

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