S1 Site information

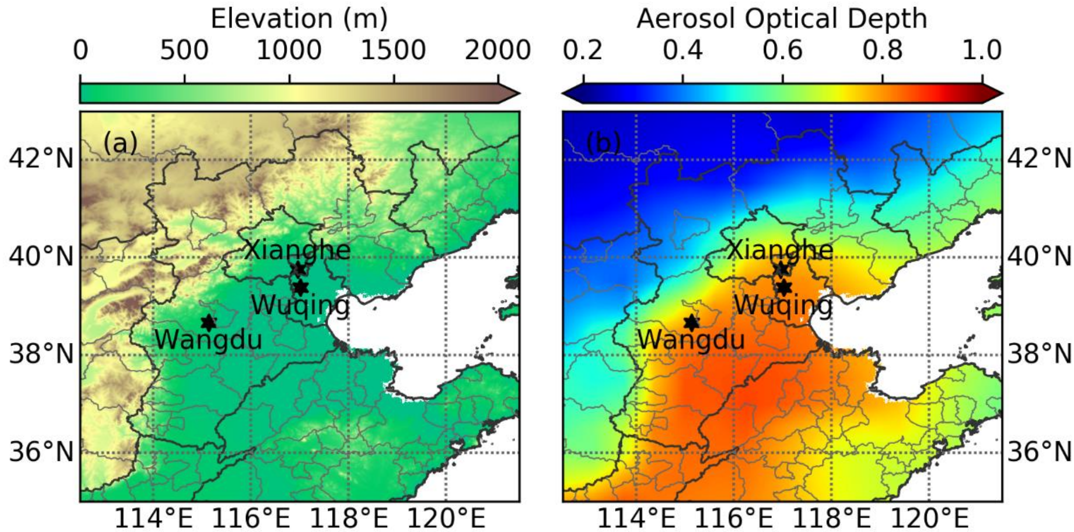

Figure S1. Site locations of Wuqing (39°23′N, 117°01′E, 7.4 m a.s.l), Xianghe (39°45′N, 116°58′E, 36 m a.s.l), and Wangdu (38°40′N, 115°08′E, 51 m a.s.l). Filled colors represents **(a)** elevation and **(b)** averaged aerosol optical depth (AOD). The AOD data is from reanalysis datasets of the Modern-Era 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 Black carbon size distribution used in theoretical simulations of lidar backscatter and extinction

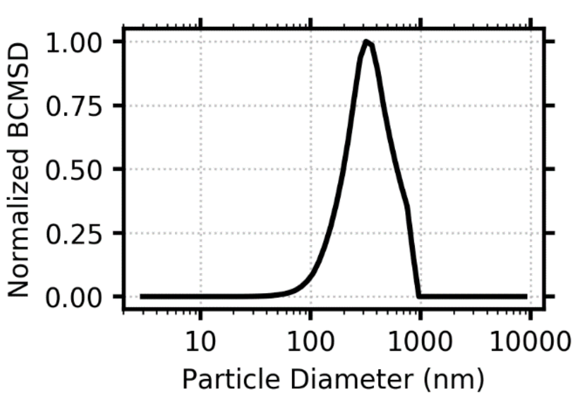

Figure S2. Normalized size distribution of black carbon (BC) mass/volume concentration. The distribution is the average black carbon mass concentration obtained from Berner impactor measurements (Ma et al., 2012) and is normalized by the maximum value of the distribution.

S3 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. 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 S3. The accuracy of predictions increase 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 S1. 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

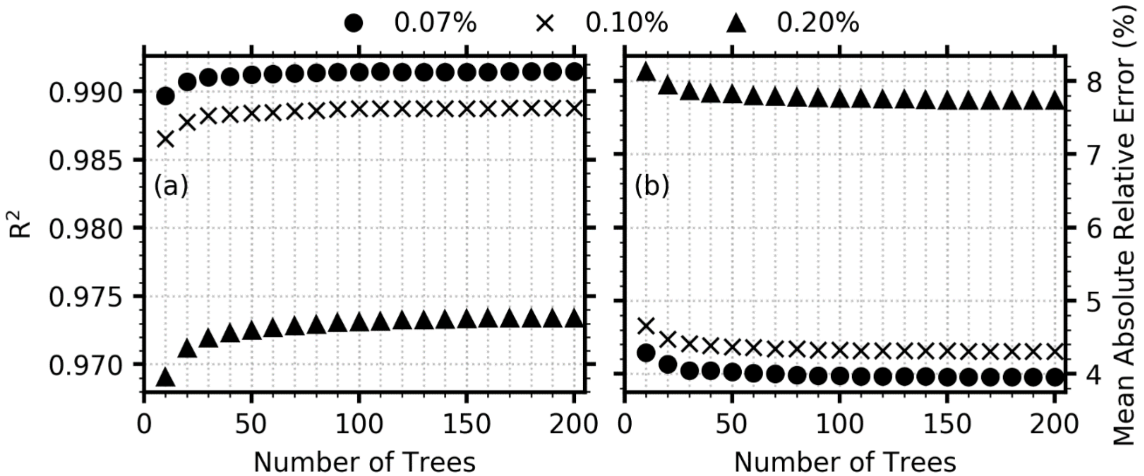

Figure S3. Influence of the number of trees in RF model on retrieving CCN number concentrations. Dependencies of tree numbers on **(a)** R^2 and **(b)** MARE between theoretical calculated CCN number concentrations and retrieved CCN number concentrations under different supersaturations.

S4 Retrieving CCN number concentrations at high supersaturations

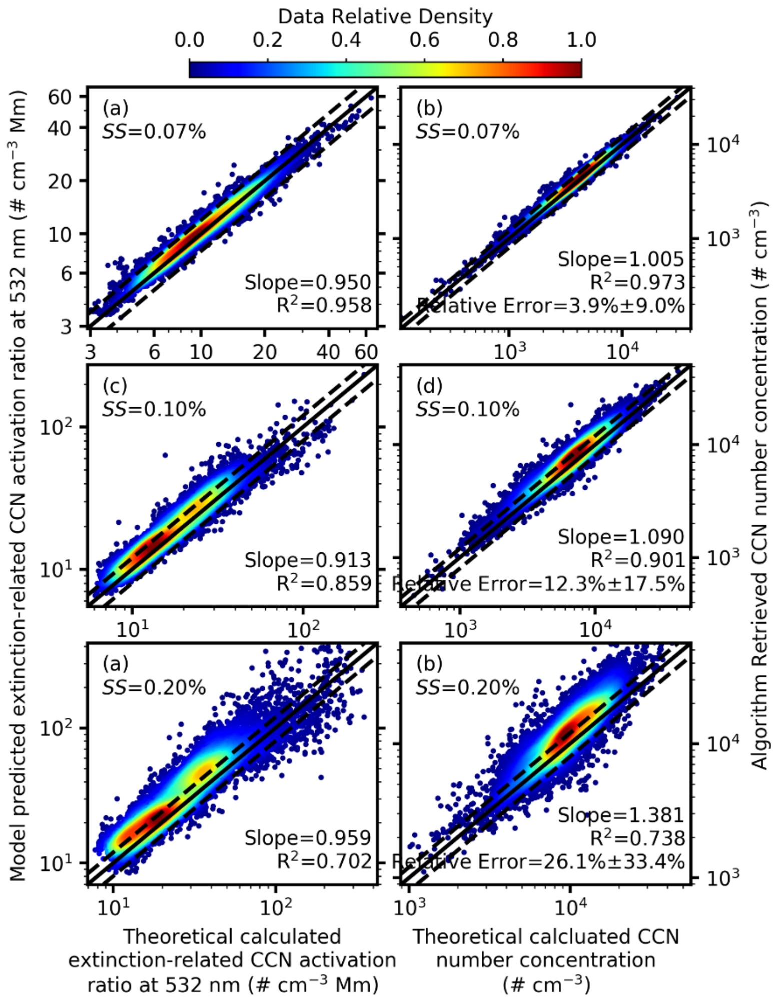

Figure S4. Comparison of the theoretical calculated extinction-related CCN activation ratio at 532 nm and the model predicted extinction-related CCN activation ratios at 532 nm at supersaturations of **(a)** 0.20%, **(c)** 0.40%, and **(e)** 0.80%, and of the theoretical calculated CCN number concentrations 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 Retrieving CCN number concentrations without hygroscopicity information

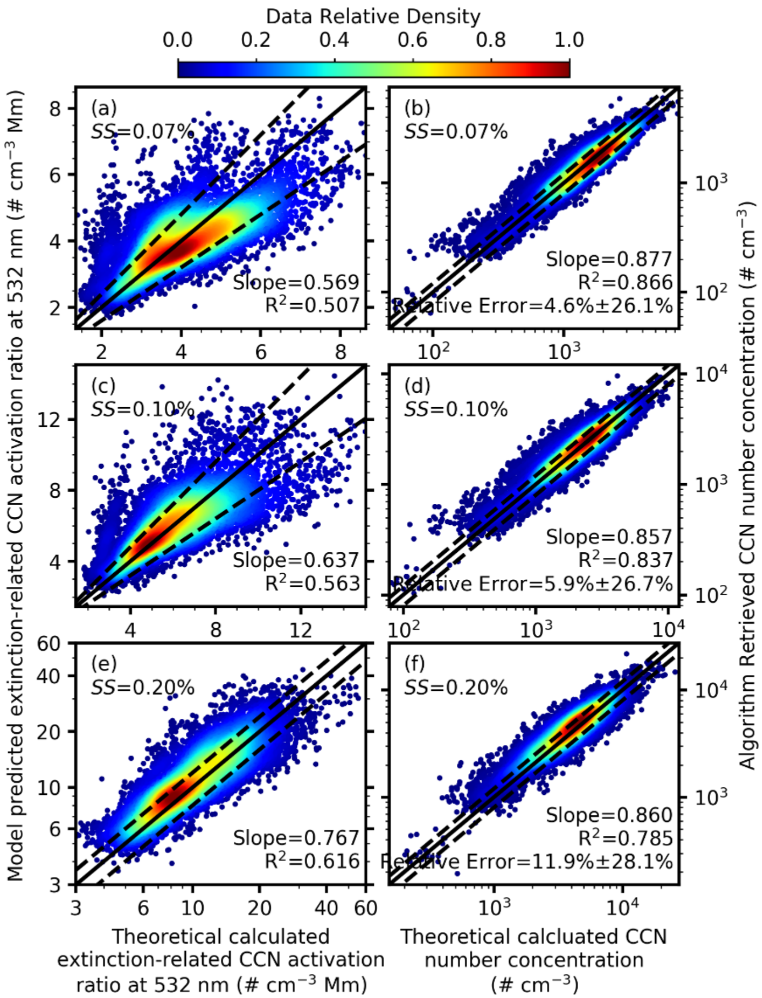

Figure S5. Comparison of the theoretical calculated extinction-related CCN activation ratio at 532 nm and the model predicted extinction-related CCN activation ratios at 532 nm at supersaturations of **(a)** 0.07%, **(c)** 0.10%, and **(e)** 0.20%, and of the theoretical calculated CCN number concentrations 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

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.