Optimizing CCN predictions through inferred modal aerosol composition – a boreal forest case study

Rahul Ranjan[1, 2], Liine Heikkinen [1, 2], Lauri R. Ahonen[3], Krista Luoma[3, 6], Paul Bowen[5], Tuukka Petäjä[3], Annica M. L. Ekman[4, 2], Daniel G. Partridge[5] and Ilona Riipinen[1, 2]

[1]Department of Environmental Science (ACES), Stockholm University, 10691, Stockholm, Sweden
[2]Bolin Centre for Climate Research, Stockholm University, 10691, Stockholm, Sweden
[3]Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, 00014, Helsinki, Finland
[4]Department of Meteorology, Stockholm University, Stockholm, Sweden
[5]Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, EX4 4QF, Exeter, United Kingdom
[6]Atmospheric Composition Research, Finnish Meteorological Institute, Helsinki, 00560, Finland

Correspondence to: Rahul Ranjan (rahul.ranjan@aces.su.se) and Ilona Riipinen (ilona.riipinen@aces.su.se)

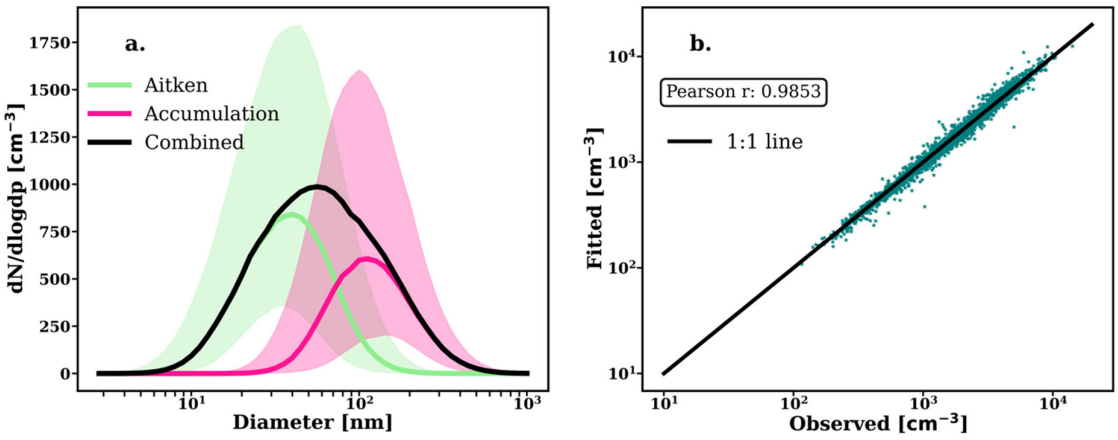

Figure S1 (a) Lognormal size distributions calculated from the fit parameters obtained after applying the algorithm by Hussein et al. (2005) on observed size distributions. The solid lines represent the median values, while the shading indicates the percentiles. (b) Fitted versus observed total particle number concentration.

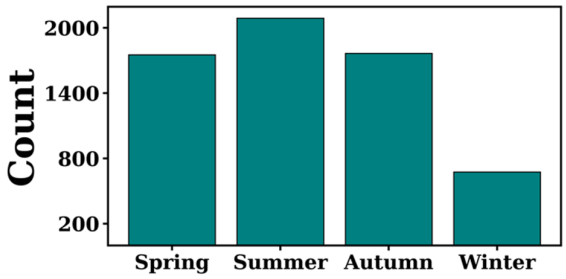

Figure S2 Number of data points for each season.

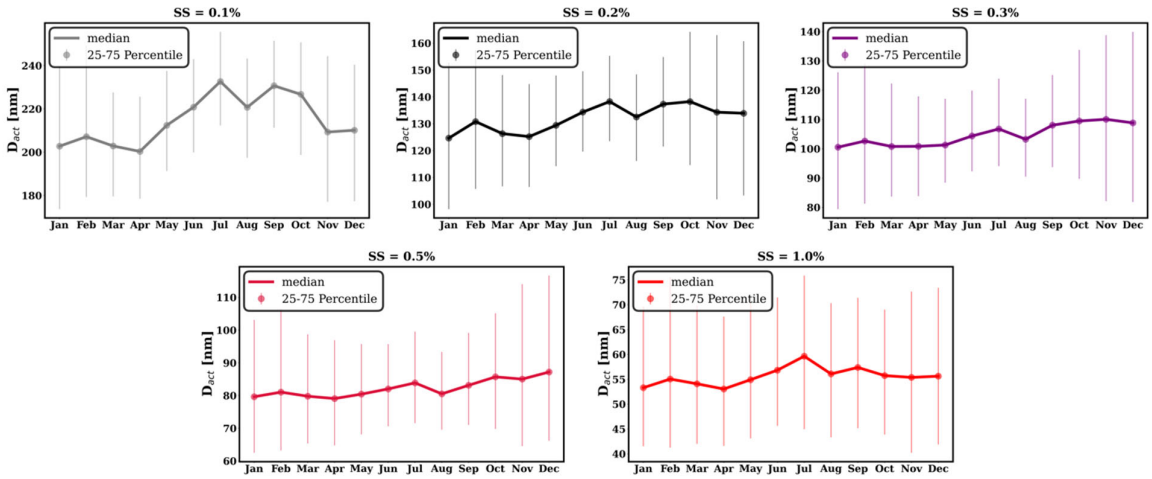

Figure S3 Monthly variation of activation diameter (D_{act}) for different supersaturation levels (SS) derived from aerosol number size distributions and observed CCN concentration. The plots display the median values (solid lines) and the corresponding 25th to 75th percentile range (error bars). Each panel represents a different supersaturation level ranging from 0.1% to 1.0%. The x-axis shows the months of the year, while the y-axis indicates D_{act}. Error bars represent the variability within each month.

Table S1. Median activation diameters (D_{act}) in nanometers by season:

Seasons	D_{act} at SS = 0.1%	D_{act} at SS = 0.2%	D_{act} at SS = 0.3%	D_{act} at SS = 0.5%	D_{act} at SS = 1.0%
Spring	206	127	100	80	54
Summer	224	135	105	82	57
Autumn	224	137	109	84	56
Winter	206	129	104	82	55

Supplementary note 1

We have used the fragmentation pattern of nitrate in the ACSM to determine whether the detected nitrate is from ammonium nitrate or organic nitrate. The idea is based on Farmer et al. (2010) and in a nutshell, ammonium nitrate will yield a higher fraction of NO_2^+ as opposed to NO^+ compared to organic nitrate. NO_2^+ and NO^+ are the main peaks where nitrate will be distributed in. the fraction of nitrate that would be from organic nitrates (f_{ON}). The formula as mentioned in Farmer et al., 2010 is:

$$f_{ON} = \frac{(R_{obs} - R_{AN}) \times (1 + R_{ON})}{(R_{ON} - R_{AN}) \times (1 + R_{obs})}, \quad \text{(S1)}$$

- f_{ON} is the fraction of nitrate that is from organic nitrates

- R_{obs} is the NO^+: NO_2^+ from ACSM measurements (nitrate fractions of *m/z* 30 and *m/z* 46)
- R_{AN} is the NO^+: NO_2^+ that the AN calibration would yield
- R_{ON} is the NO^+: NO_2^+ for pure organic nitrate

Here, it is assumed that R_{AN} = 2 since this would limit the number of negative values for f_{ON}, when R_{ON} = 10. The R_{ON} = 10 is to be expected from the NO_3 oxidation of **α**-pinene, which is assumed as a major source of ON at SMEAR II. Finally, using Eq. S1 with the R_{AN} and R_{ON} constants for 2016 onwards, a conservative guess for the time series for f_{ON} can be derived (Fig. S4). As f_{ON} exhibits a clear, and rather consistent seasonality, a seasonal mask could be retrieved using day-of-the year -based 3-month running median (also shown in Fig. S3). This day-of-the-year based mask is used further to estimate how much of the measured nitrate was present as ON and AN (the ammonium nitrate mass fraction $f_{AN} = 1 - f_{ON}$).

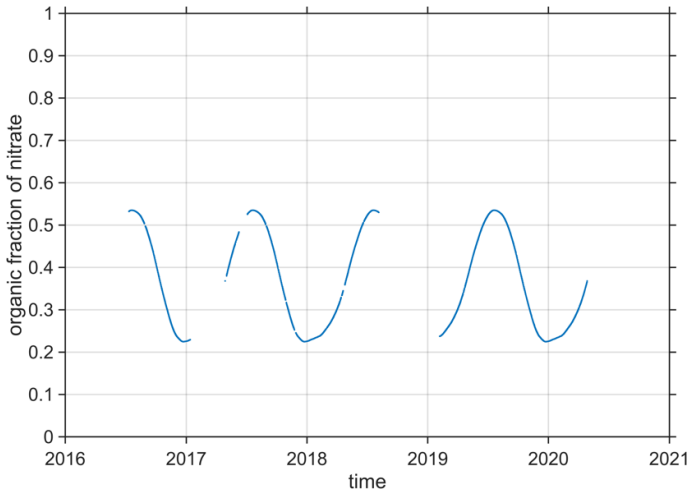

Figure S4 Final mask determining how much of the NO_3 ion is organic nitrate, and how much ammonium nitrate.

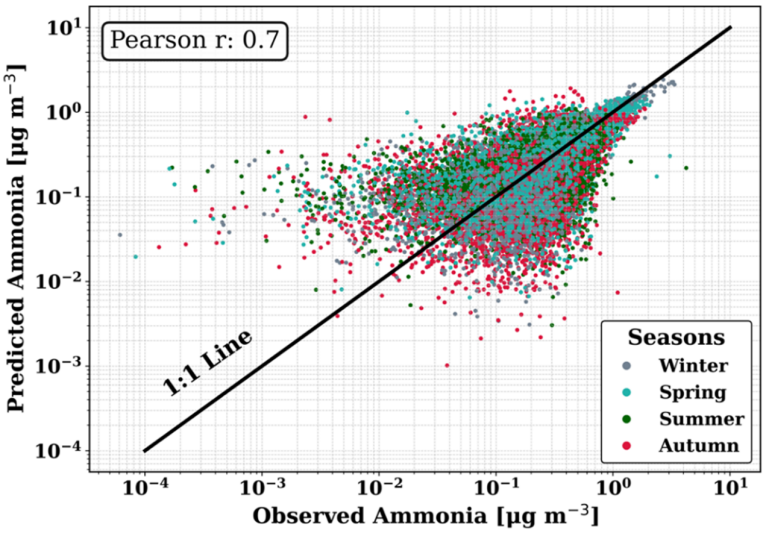

Figure S5 Predicted ammonium ion (concentration necessary to achieve ion balance within the particles) and observed mass concentration of ammonium ion.

Supplementary note 2

In this section, we describe the method (also referenced in Sect. 2.2.3 of the main article) used to scale the fitted size distribution to match the observations. This ensures that the total particle concentration, as well as the number of particles in each bin of the reconstructed size distribution (derived from fitted bimodal lognormal parameters), remains consistent with the observed size distribution. As a result, whether we use the original observed size distribution or the fitted bimodal size distribution when applying bulk chemical composition, the CCN spectra remain the same. During scaling process, for each bin, the observed concentration is compared to the sum of the two fitted modes. If only one mode contributes, its concentration is adjusted directly to match the observed value. When both modes contribute, the difference between observed and fitted totals is distributed proportionally based on each mode's relative contribution.

For scaling the steps are as follows:

We have (in cm^{-3}):

$N_{\mathrm{obs,i}}$: Observed concentration in bin *i*

$N_{\mathrm{fit,Ait,i}}$: Fitted concentration for Aitken mode in bin *i*

$N_{\mathrm{fit,acc,i}}$: Fitted concentration for accumulation mode in bin *i*

$N_{\mathrm{scaled,Ait,i}}$: Scaled concentration for Aitken mode in bin *i*

$N_{\mathrm{scaled,acc,i}}$: Scaled concentration for accumulation in bin *i*

The scaling follows these steps:

Step 1: Scaling in the size distribution where only the Aitken mode has particles ($N_{\mathrm{fit,acc,i}} = 0$)

$$N_{\mathrm{scaled,Ait,i}} = N_{\mathrm{fit,Ait,i}} + (N_{\mathrm{obs,i}} - N_{\mathrm{fit,Ait,i}})$$
$$N_{\mathrm{scaled,acc,i}} = 0$$

Step 2: Scaling in the size distribution where only the accumulation mode has particles ($N_{\mathrm{fit,Ait,i}} = 0$)

$$N_{\mathrm{scaled,acc,i}} = N_{\mathrm{fit,acc,i}} + (N_{\mathrm{obs,i}} - N_{\mathrm{fit,acc,i}})$$
$$N_{\mathrm{scaled,Ait,i}} = 0$$

Step 3: Scaling where both modes have contribution to the number of particles ($N_{\mathrm{fit,Ait,i}} \neq 0$, $N_{\mathrm{fit,acc,i}} \neq 0$); Compute the fractional contributions of each mode

fractional contribution of Aitken mode, $$\mathcal{X}1_i = \frac{N_{\text{fit,Ait,i}}}{N_{\text{fit,Ait,i}} + N_{\text{fit,acc,i}}}$$

fractional contribution of accumulation mode, $$\mathcal{X}2_i = \frac{N_{\text{fit,acc,i}}}{N_{\text{fit,Ait,i}} + N_{\text{fit,acc,i}}}$$

scaling factor, $$\mathcal{X}_i = N_{\text{obs,i}} - (N_{\text{fit,Ait,i}} + N_{\text{fit,acc,i}})$$

apply scaling, $$N_{\text{scaled,Ait,i}} = N_{\text{fit,Ait,i}} + \mathcal{X}_i.\mathcal{X}1_i$$
$$N_{\text{scaled,acc,i}} = N_{\text{fit,acc,i}} + \mathcal{X}_i.\mathcal{X}2_i$$

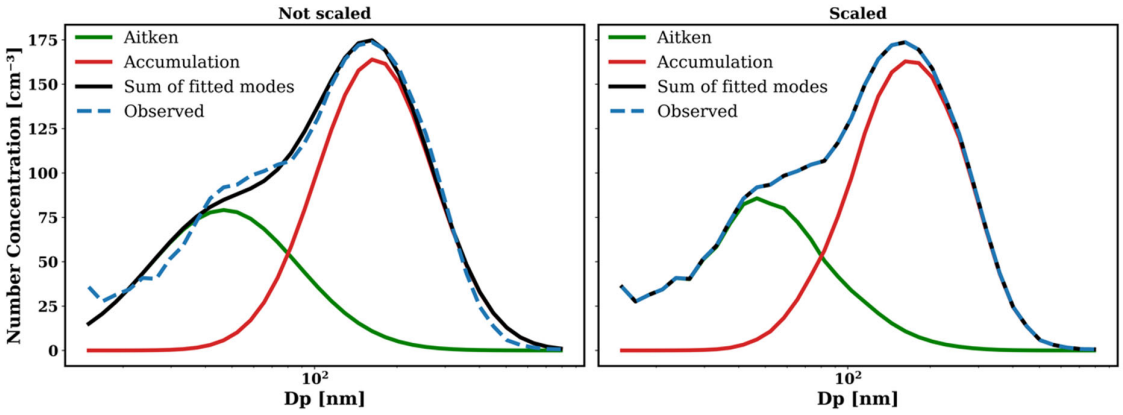

Figure S6 An example of how scaling works. Bimodal fitting is done first and then the scaling aligns the reconstructed lognormal size distributions (black line in the left panel i.e. sum of fitted modes) with the observed distribution across binned diameters (Dp).

Supplementary note 3

To investigate how different combinations of mass fractions of different species in the Aitken and accumulation modes influence the model performance, we employed a brute-force sampling approach. This allowed us to explore the full parameter space and examine the variation in normalized root mean square error (NRMSE) between measured and modelled CCN spectra across the entire dataset (2016-2020). Figure S7 illustrates the distribution of NRMSE values over this parameter space with respect to all samples of mass fraction of organics in Aitken mode ($X_{\text{org, Aitken}}$) and highlights the regions where model performance is optimal (i.e., where NRMSE

is minimized). A small reduction in NRMSE spread and median NRMSE is visible with increasing $X_{org,\ Aitken}$ suggesting that for a better CCN closure, Aitken mode has to be enriched with organics likewise in the optimized mass fractions in the main text (Fig. 6b).

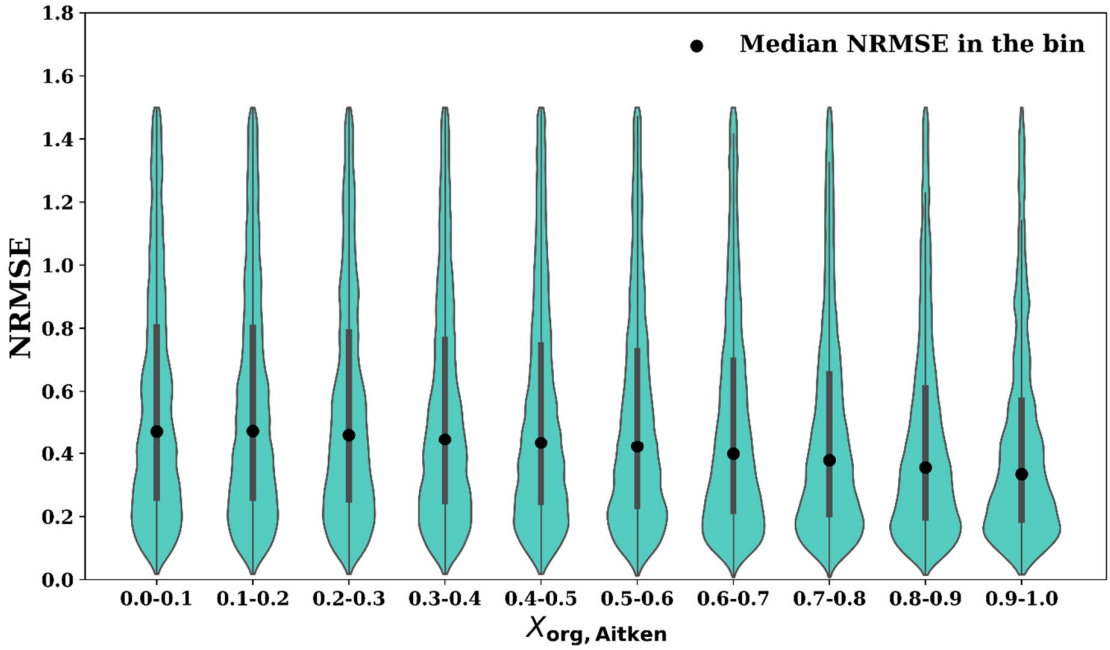

Figure S7 Violin plots of the normalized root mean square error (NRMSE) between measured and modelled CCN spectra during optimization, grouped by the mass fraction of organics in the Aitken mode ($X_{org,\ Aitken}$).

Supplementary note 4

The activation diameter, often referred to as the critical diameter, represents the minimum size of particles that can activate into cloud droplets at a given *SS*. This method estimates the activation diameter based on observed cloud condensation nuclei concentrations and particle number size distribution data.

Methodology

1. **For a given CCN concentration at a particular supersaturation level**:
    - Find the bin where the cumulative particle concentration is either equal to or exceeds the CCN concentration (upper bound).
    - Find the size bin where the cumulative particle concentration just falls below the CCN concentration.
2. **Linear interpolation between size bins:**
    - The activation diameter lies between the sizes associated with the lower and upper bounds. Assuming that particle concentration changes linearly between these bins:

$$D_{act} = D_0 + \frac{N_{CCN} - N_0}{N_1 - N_0} \times (D_1 - D_0)$$

Where:

- D_{act} : Activation diameter
- D_0, D_1: Diameters corresponding to the lower and upper cumulative concentrations, N_0 and N_1, respectively.
- N_{CCN}: Observed CCN concentration.
- If the cumulative concentrations at the lower and upper bounds are identical ($N_0 = N_1$), the D_{act} is assigned the diameter of the lower bound (D_0).

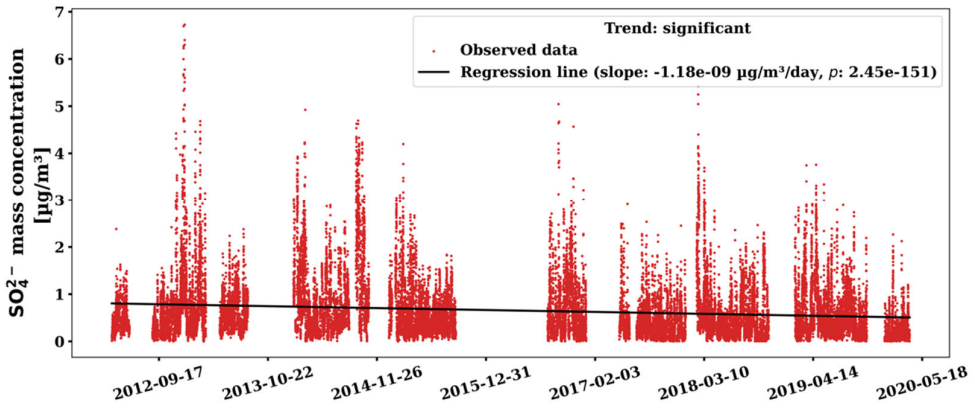

Figure S8 Time series of sulfate ion mass concentration measured by the ACSM. Scatter points represent individual sulfate ion concentration values, while the black solid line indicates the linear regression fit. The slope and *p*-value of the regression are provided in the legend. The statistical significance of the trend is assessed at the 95% confidence level ($p < 0.05$).

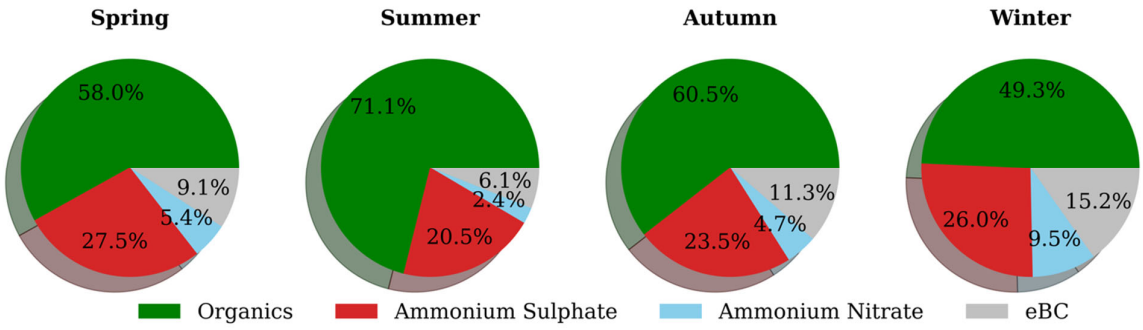

Figure S9 Mass fractions of various chemical species during different seasons.

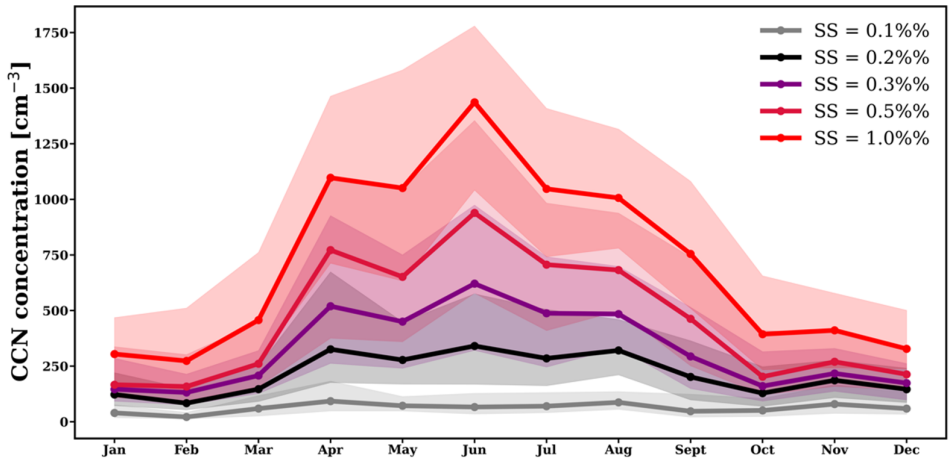

Figure S10 Monthly median CCN concentrations at different supersaturations with shaded areas indicating the interquartile range (25th to 75th percentiles).

Table S2. Geometric mean bias (GMB) and Pearson's correlation coefficient (in brackets) corresponding to different methods and supersaturations

Methods	Bias (R) at SS = 0.1%	Bias (R) at SS = 0.2%	Bias (R) at SS = 0.3%	Bias (R) at SS = 0.5%	Bias (R) at SS = 1.0%
κ_{bulk}	1.56 (0.89)	1.19 (0.93)	1.19 (0.93)	1.34 (0.92)	1.34 (0.89)
$\kappa_{0.18}$	1.35 (0.85)	1.11 (0.93)	1.11 (0.93)	1.24 (0.93)	1.26 (0.90)
κ_{opt}	1.53 (0.89)	1.13 (0.93)	1.12 (0.94)	1.21 (0.94)	1.20 (0.93)
$\kappa_{org} = 0$	0.92 (0.83)	0.87 (0.87)	0.94(0.87)	1.086 (0.89)	1.938 (0.87)

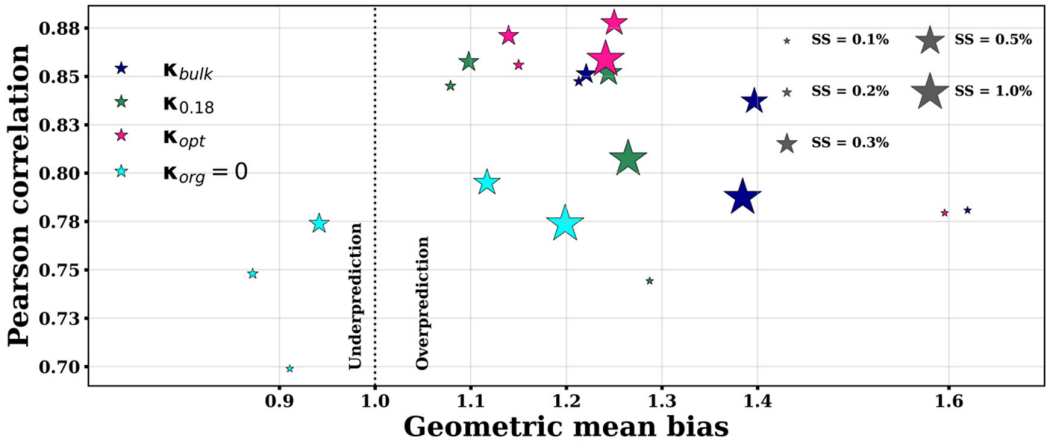

Figure S11 Scatter plot illustrating the geometric mean bias (GMB) and Pearson correlation for different supersaturation (*SS*) levels, comparing four methods: κ_{bulk}, $\kappa_{0.18}$, and κ_{opt}, κ_{org} = 0. Each marker represents a different SS level, with size proportional to SS level, ranging from 0.1% to 1.0%. Bias is presented on the x-axis and correlation on the y-axis, with vertical lines indicating perfect prediction (bias = 1). The plot shows the performance of each method across various *SS* levels, with annotations indicating underprediction (bias < 1) and overprediction (bias > 1).

Table S3. Median of optimized mass fractions of various chemical species in Aitken mode in different seasons

	Spring	Summer	Autumn	Winter
Organics	0.89	0.92	0.87	0.83
Inorganics	0.0002	0.0001	0.0001	0.0001
eBC	0.10	0.07	0.13	0.17

Table S4. Median of optimized mass fractions of various chemical species in accumulation mode in different seasons

	Spring	Summer	Autumn	Winter
Organics	0.54	0.71	0.57	0.46
Inorganics	0.37	0.23	0.32	0.40
eBC	0.09	0.06	0.10	0.14