



Supplement of

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

Rahul Ranjan et al.

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

The copyright of individual parts of the supplement might differ from the article licence.

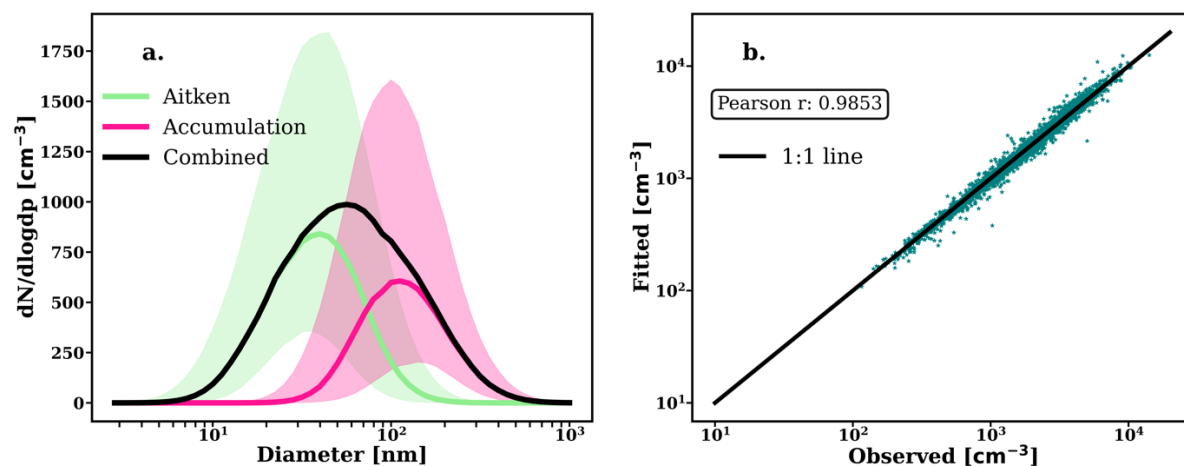


Figure S1 (a) Lognormal size distributions calculated from the fit parameters obtained after applying the algorithm by Hussein et al. (2005) on observed size distributions. The solid lines represent the median values, while the shading indicates the percentiles. (b) Fitted versus observed total particle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

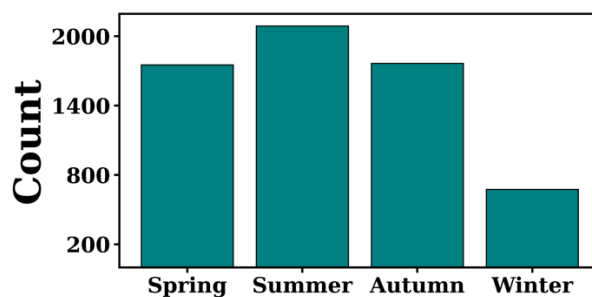


Figure S2 Number of data points for each season.

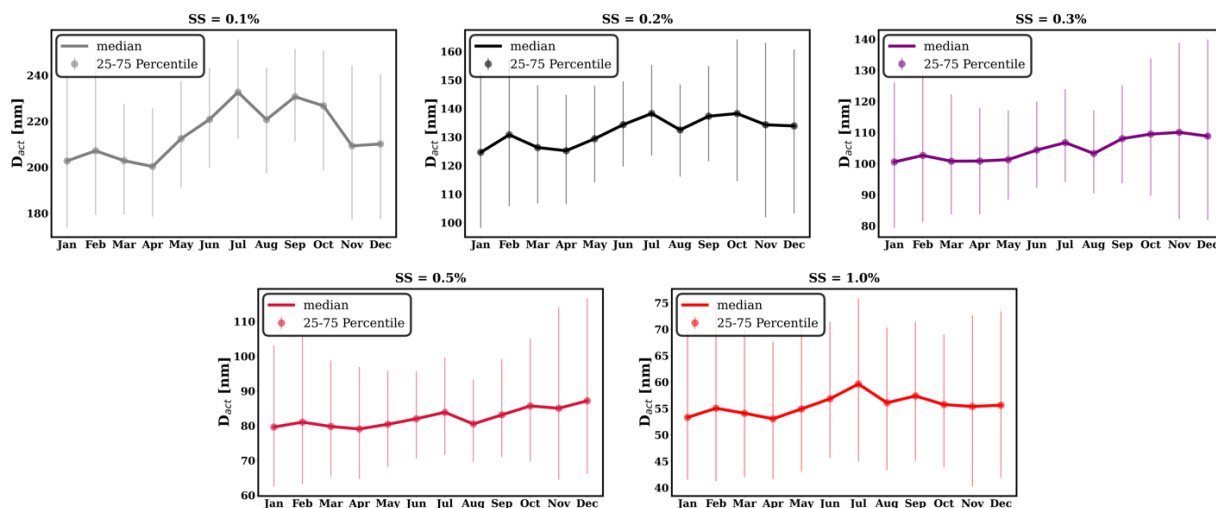


Figure S3 Monthly variation of activation diameter (D_{act}) for different supersaturation levels (SS) derived from aerosol number size distributions and observed CCN concentration. The plots display the median values (solid lines) and the corresponding

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

S1. Calculation of mass fraction of organic nitrate

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:

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

- f_{ON} is the fraction of nitrate that is from organic nitrates
- R_{obs} is the $\text{NO}^+ : \text{NO}_2^+$ from ACSM measurements (nitrate fractions of m/z 30 and m/z 46)
- R_{AN} is the $\text{NO}^+ : \text{NO}_2^+$ that the AN calibration would yield
- R_{ON} is the $\text{NO}^+ : \text{NO}_2^+$ for pure organic nitrate

Here, it is assumed that $R_{\text{AN}} = 2$ since this would limit the number of negative values for f_{ON} , when $R_{\text{ON}} = 10$. The $R_{\text{ON}} = 10$ 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 (Fig. S4). 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 $f_{\text{AN}} = 1 - f_{\text{ON}}$).

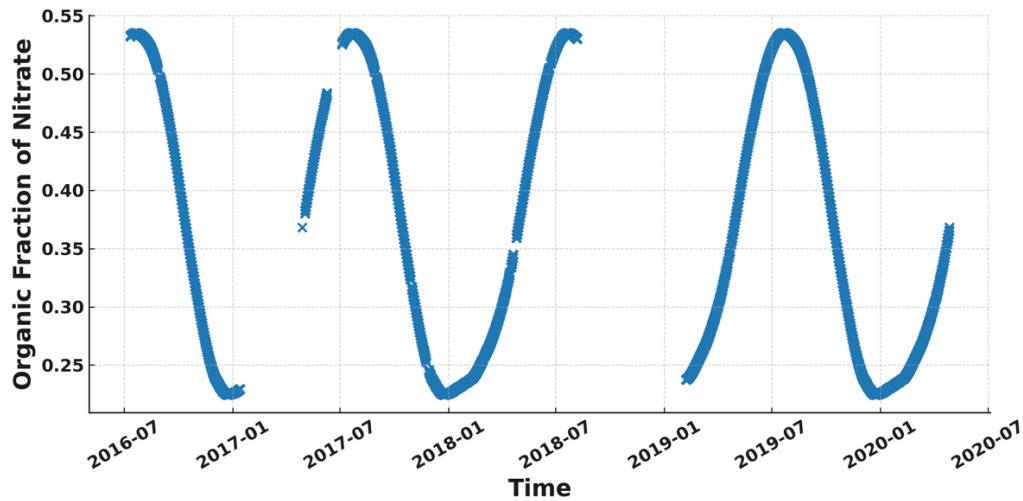


Figure S4 Mask determining how much of the NO_3 ion is organic nitrate, and how much ammonium nitrate.

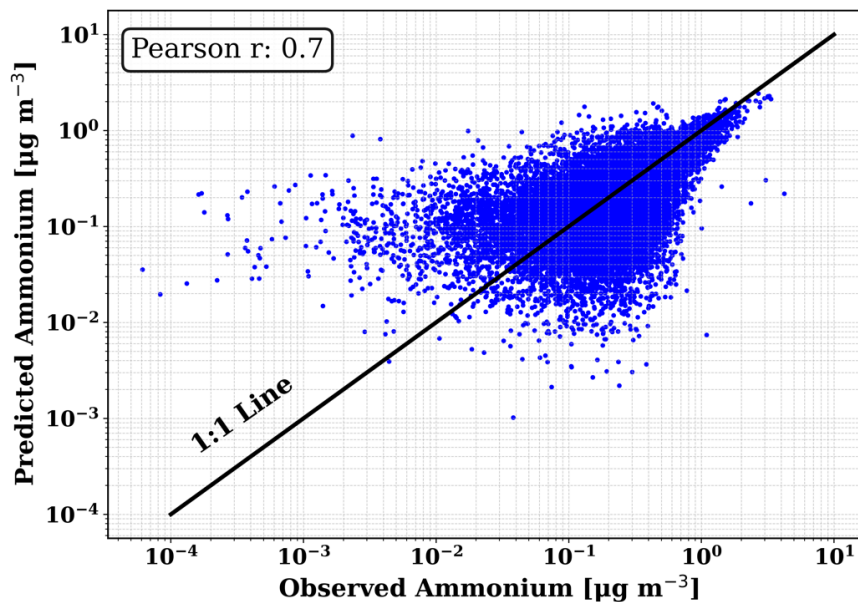


Figure S5 Predicted ammonium ion (concentration necessary to achieve ion balance within the particles) and observed mass concentration of ammonium ion.

S2. Demonstration of how scaling works

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 (see Fig. S6). As a result, whether we use the original observed size distribution or the fitted bimodal size distribution when applying bulk chemical composition, the predicted 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

59 adjusted directly to match the observed value. When both modes contribute, the difference between observed and
 60 fitted totals is distributed proportionally based on each mode's relative contribution.

61 To better understand the process, scaling is explained in the steps as follows:

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

63 $N_{\text{obs},i}$: Observed concentration in bin i

64 $N_{\text{fit,Ait},i}$: Fitted concentration for Aitken mode in bin i

65 $N_{\text{fit,acc},i}$: Fitted concentration for accumulation mode in bin i

66 $N_{\text{scaled,Ait},i}$: Scaled concentration for Aitken mode in bin i

67 $N_{\text{scaled,acc},i}$: Scaled concentration for accumulation in bin i

68 The scaling follows these steps:

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

$$N_{\text{scaled,Ait},i} = N_{\text{fit,Ait},i} + (N_{\text{obs},i} - N_{\text{fit,Ait},i}) \quad (\text{S2})$$

$$N_{\text{scaled,acc},i} = 0 \quad (\text{S3})$$

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

$$N_{\text{scaled,acc},i} = N_{\text{fit,acc},i} + (N_{\text{obs},i} - N_{\text{fit,acc},i}) \quad (\text{S4})$$

$$N_{\text{scaled,Ait},i} = 0$$

72

73 Step 3: Scaling where both modes have contributions to the number of particles ($N_{\text{fit,Ait},i} \neq 0, N_{\text{fit,acc},i} \neq 0$);

74 Compute the fractional contributions of each mode

$$\text{fractional contribution of Aitken mode,} \quad \mathcal{X}1_i = \frac{N_{\text{fit,Ait},i}}{N_{\text{fit,Ait},i} + N_{\text{fit,acc},i}} \quad (\text{S5})$$

$$\text{fractional contribution of accumulation mode,} \quad \mathcal{X}2_i = \frac{N_{\text{fit,acc},i}}{N_{\text{fit,Ait},i} + N_{\text{fit,acc},i}} \quad (\text{S6})$$

$$\text{scaling factor, } \mathcal{X}_i = N_{\text{obs},i} - (N_{\text{fit,Ait},i} + N_{\text{fit,acc},i}) \quad (\text{S7})$$

$$\text{apply scaling, } N_{\text{scaled,Ait},i} = N_{\text{fit,Ait},i} + \mathcal{X}_i \cdot \mathcal{X}1_i \quad (\text{S8})$$

$$N_{\text{scaled,acc},i} = N_{\text{fit,acc},i} + \mathcal{X}_i \cdot \mathcal{X}2_i \quad (\text{S9})$$

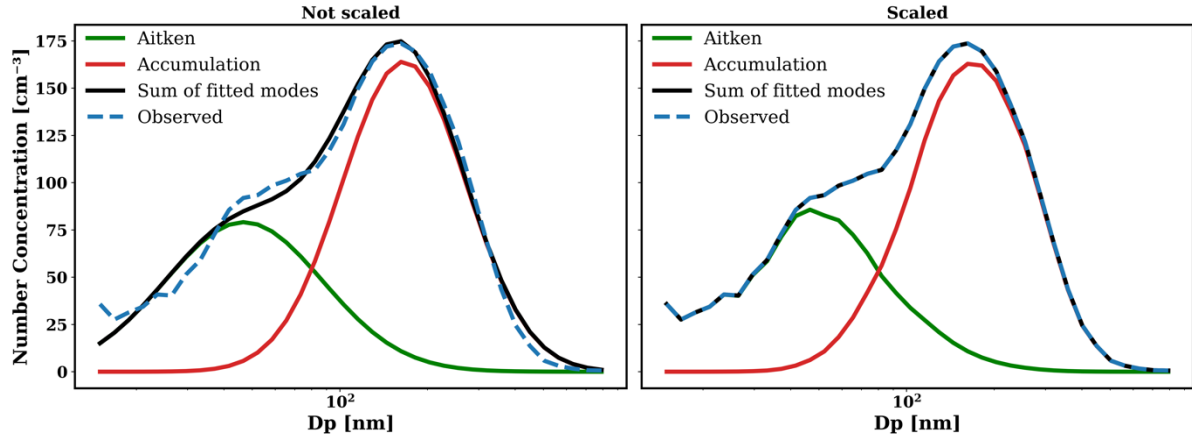


Figure S6 An example of the size distribution scaling procedure. Bimodal fitting is done first and then the scaling aligns the reconstructed lognormal size distributions (black line in the left panel i.e. sum of fitted modes) with the observed distribution across binned diameters (D_p).

S3. Conservation of mass

In both the Nelder–Mead and DREAM-MCMC optimizations, the total aerosol mass is conserved across all chemical species (organics, inorganics, and eBC). Additionally, the eBC mass is fixed in both the Aitken and accumulation modes. This constraint reduces the optimization problem: we only vary the mass of one species (e.g., organics in the Aitken mode), while the remaining components are derived from conservation laws. For a given time step in the time series, let, $m_{\text{Ait}}^{\text{tot}}$ and $m_{\text{acc}}^{\text{tot}}$ are the total masses in Aitken and accumulation mode, respectively, obtained from bimodal size distribution and mass fractions. Similarly, let $m_{\text{BC,Ait}}$ and $m_{\text{BC,acc}}$ are the fixed eBC masses, then for each $m_{\text{org,Ait}}$ the optimization algorithm explore:

$$m_{\text{inorg,Ait}} = m_{\text{Ait}}^{\text{tot}} - (m_{\text{org,Ait}} + m_{\text{BC,Ait}}) \quad (\text{S10})$$

$$m_{\text{org,acc}} = m_{\text{org}}^{\text{tot}} - m_{\text{org,Ait}} \quad (\text{S11})$$

$$m_{\text{inorg,acc}} = m_{\text{acc}}^{\text{tot}} - (m_{\text{org,acc}} + m_{\text{BC,acc}}) \quad (\text{S12})$$

S4. Bayesian inference and MCMC

Bayesian inference is a rigorous method for quantifying uncertainty in model parameters (see also see Gelman et al., 2013), using probability statements. Unknown parameters are treated as random variables with some joint posterior probability distribution, which can be written using Bayes law as:

$$p(\theta|Y) = \frac{p(\theta)L(Y|\theta)}{p(Y)} \quad (\text{S13})$$

where $p(\theta)$ is the prior distribution which encompasses what is known about the parameters prior to observing any data, $L(Y|\theta)$ is the likelihood function which measures how well the model fits observed data, and

$$p(Y) = \int p(\theta)p(Y|\theta)d\theta \quad (\text{S14})$$

is the marginal distribution of Y , which represents the probability of observing Y given all possible parameter values θ . The result of conditioning the prior distribution with some observations is the posterior probability distribution $p(\theta|Y)$, which represents the updated probability of the model parameters. Bayesian inference is therefore a process of creating a probability model and iteratively updating that model based on some observations, resulting in a best estimate of the parameters and knowledge about their uncertainty, sensitivity, and correlation (in the case where θ is vector-valued).

Monte Carlo Markov Chain (MCMC) simulations are a methodology for sampling from posterior distributions. Generally, they involve repeatedly and sequentially sampling θ such that each new draw depends only on the previous sample and therefore forms a Markov chain, and correcting those draws so that the chain converges to the target distribution. Many different algorithms have been proposed for generating and correcting chain samples. Here we use the DiffereNTial Evolution Adaptive Metropolis Markov Chain Monte Carlo (DREAM-MCMC) algorithm (Vrugt et al. 2009). This algorithm runs multiple chains simultaneously and adaptively updates the proposal distribution using a randomized subset of the chains' joint history. It also supports large proposal jumps and outlier rejection during the initial burn-in phase, which accelerate convergence. This type of self-adaptive evolutionary strategy is particularly well suited to heavy-tailed or multi-modal posteriors, such as in this study where different combinations of aerosol chemical composition and size distributions parameters could result in similar CCN spectra.

S5. Chain evolution in DREAM-MCMC

As mentioned in sect. 2.2.3.1, we ran five DREAM-MCMC chains with 40,000 sampling steps each for each CCN observation window. Figure S7 shows that for both windows there is an overprediction of the CCN spectra when bulk composition is used, which is mostly corrected with the DREAM-MCMC optimized parameters, showing that when size-segregated composition is combined with variability in the lognormal size distribution parameters, most of the discrepancy between observed and predicted CCN can be resolved. Figures S8 and S9 show examples of the chain evolution and the resulting posterior distributions for two distinct cases.

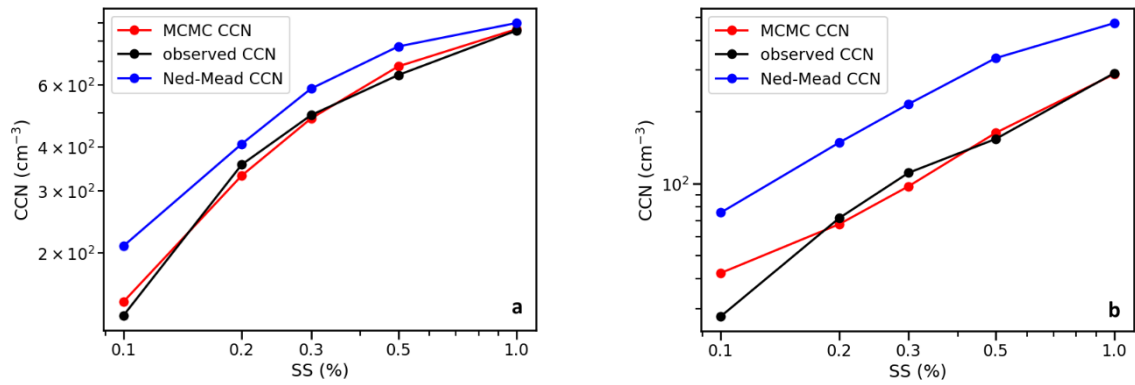


Figure S7. Observed and optimized CCN spectra. The black line shows the observed spectra, the blue line represents the optimized spectra using the Nelder–Mead method with only size-segregated chemical composition as the optimized parameter, and the red line represents the DREAM-MCMC optimization where both size-segregated composition and size distribution parameters are optimized. Panel (a) corresponds to 2019-05-11 05:00:00, and panel (b) corresponds to 2019-10-15 07:00:00.

The first example (Figure S7a and Figure S8) corresponds to a spring-time window, specifically 2019-05-11 05:00:00. There is a higher total CCNc, and the DREAM-MCMC improves the closure at all super-saturations. All 5 chains converge well with $\hat{R}$ -hat values very close to 1 for all parameters. The posterior distributions are relatively compact with the chain evolution indicating stable convergence, with broader distributions for $m_{\text{org,Ait}}$ (in this note denoted as M_{org1}) and N_{Ait} (in this note denoted as N_1). There is a significant shift in the median of the posterior distribution of M_{org1} compared to the initial guess, whereas the posteriors of the size distribution parameters are centred on values very close to the median of the observations, suggesting that an increased mass fraction of organics in the Aitken mode is the key to better CCN closure for this time-step. However, the posterior distribution for M_{org1} is relatively wide, indicating less sensitivity to the exact value compared to the very tight posterior distributions for D_1 , N_1 , and D_2 (the diameters and number concentrations corresponding to Aitken and accumulation modes), which suggest that the closure is very sensitive to those three parameters.

In the second example (Figure S7b and Figure S9) the situation is more complex. This corresponds to a fall window, specifically 2019-10-15 07:00:00, where the total CCNc is lower. While DREAM-MCMC significantly improves the closure, it struggles more compared to the previous case, over-predicting at 0.1% SS and underpredicting at 0.2% and 0.3%. The chains show slower mixing and broader posteriors for several parameters, particularly M_{org1} , N_1 and D_2 . In this case medians of the posterior distributions are significantly shifted from the initial guesses for all parameters. The $\hat{R}$ -hat values using all 5 chains are below 1.5 for all parameters but are larger for N_1 and D_1 (approx. 1.3 for both), indicating poorer mixing. Wider posteriors indicate possibly weak identifiability, meaning that multiple parameter combinations can yield similarly good agreement with the observed CCN spectra, pointing to stronger parameter correlations in the posterior. Here, M_{org1} is more poorly constrained than in the first case, but the number concentration parameters (N_1 and N_2) still play a more dominant role in reducing the CCN prediction.

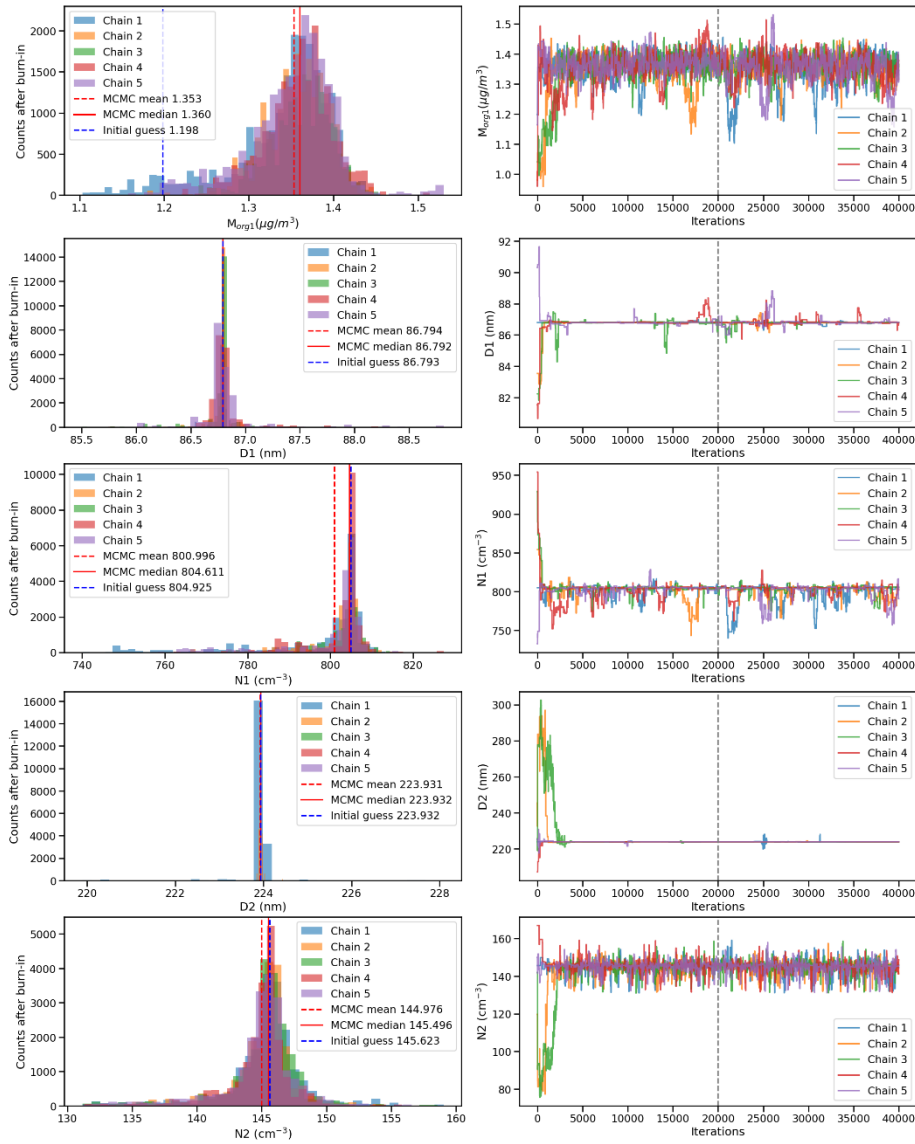


Figure S8. Chain evolution (right) and posterior distributions of the optimized parameters (left) from the DREAM-MCMC run. Each plot shows five parallel chains with 40,000 sampling steps. The vertical dashed line in the right panels marks the end of the burn-in period, and the posterior distributions on the left are taken from the final samples after burn-in which are used for posterior inference. Results correspond to observation 4200 (2019-05-11 05:00:00).

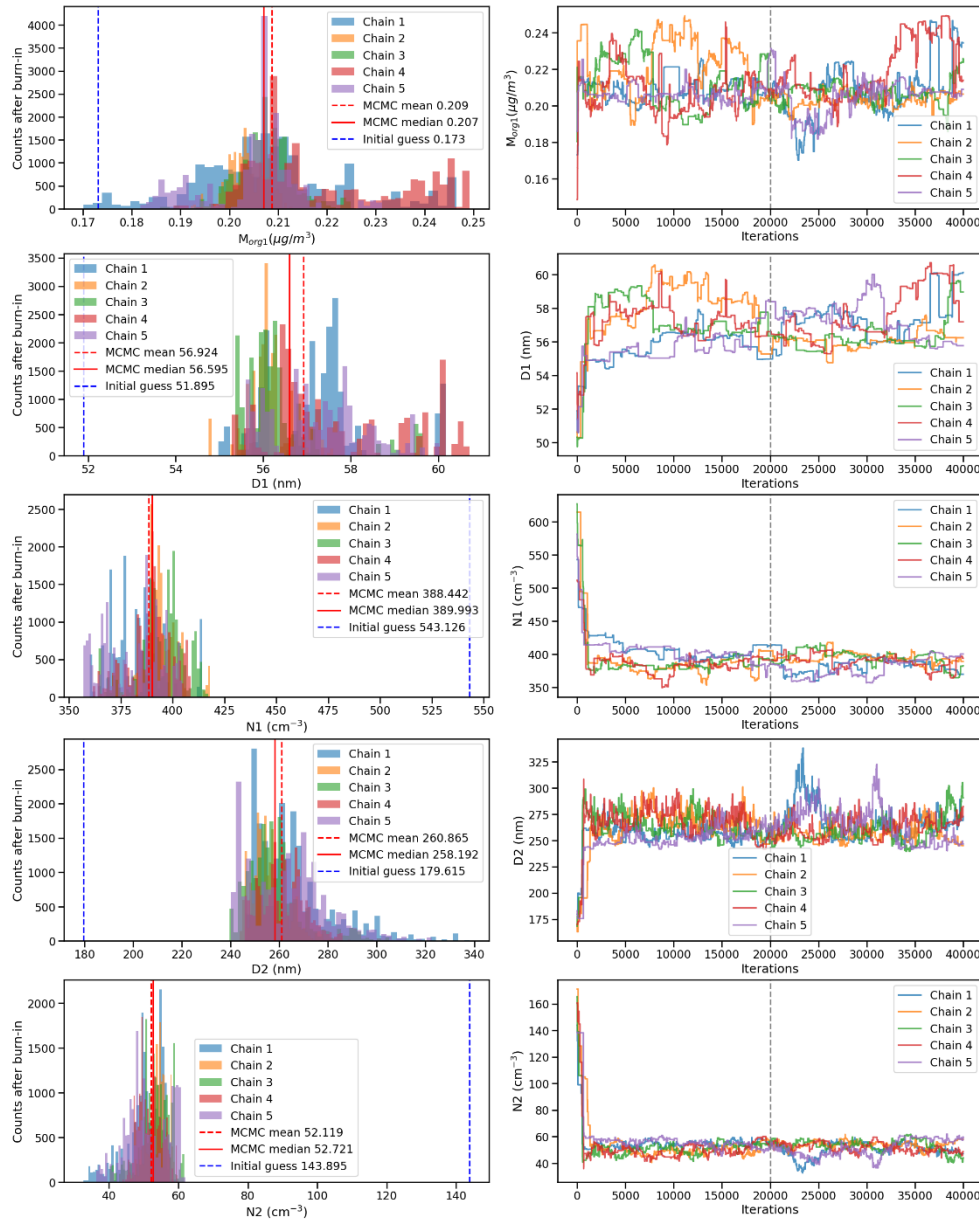


Figure S9. Chain evolution (right) and posterior distributions of the optimized parameters (left) from the DREAM-MCMC run. Each plot shows five parallel chains with 40,000 sampling steps. The vertical dashed line in the right panels marks the end of the burn-in period, and the posterior distributions on the left are based on the final samples after burn-in which are used for posterior inference. Results correspond to observation 5813 (2019-10-15 07:00:00).

S6. Distribution of median absolute deviation (MAD)

Figure S10 shows the probability density of median absolute deviation, MAD values fitted with chi-squared distributions for both the Aitken and accumulation modes. The results indicate that the variability is generally small, with distributions strongly centered close to zero and narrow tails. The fitted chi-squared parameters suggest that fluctuations in diameter and sigma are low, whereas number concentrations show comparatively larger spread. Overall, this analysis confirms that the derived parameters remain fairly stable within CCN cycle period, with occasional variability in particle number concentration.

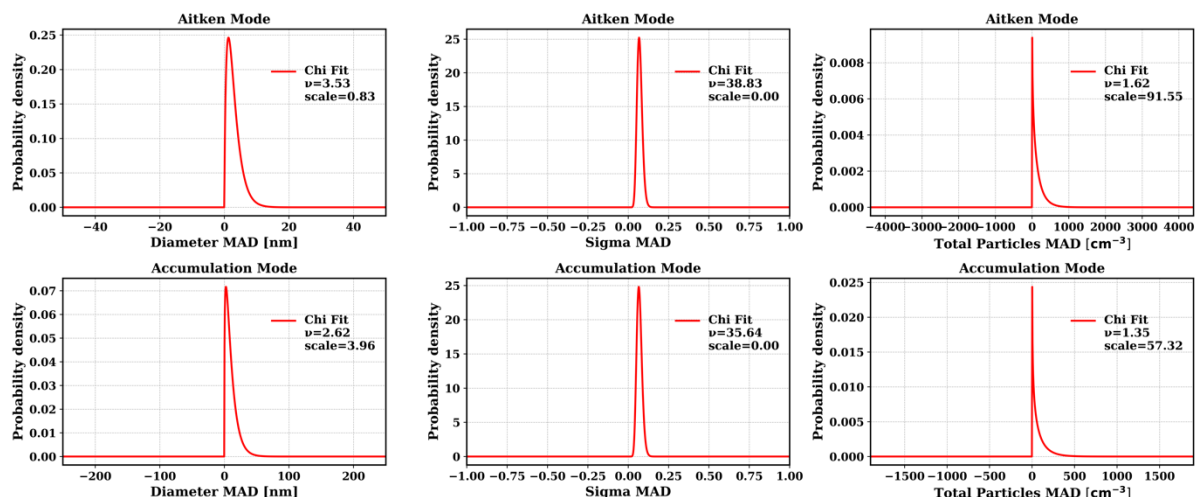


Figure S10. Chi-squared probability density functions (PDFs) fitted to the median absolute deviation (MAD) values of aerosol size distribution parameters calculated with respect to median of corresponding parameters during CCN cycle. The data represent entire 5 years period between 2016 to 2020. Panels show Aitken mode (top row) and accumulation mode (bottom row) MADs for mode diameter, geometric standard deviation (σ), and particle number concentration. The fitted parameters (degrees of freedom, ν , and scale) are reported in the legends. The fits are constrained to non-negative values to reflect the definition of MAD.

S7. Distribution of normalized root mean square error (NRMSE) over brute-force sampled mass fraction bins of organic in Aitken mode ($f_{\text{org, Aitken}}$)

To investigate how different combinations of mass fractions of different species in the Aitken and accumulation modes influence the model performance, we complemented the inverse modelling algorithms with 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 S11 illustrates the distribution of NRMSE values over this parameter space with respect to all samples of mass fraction of organics in Aitken mode ($f_{\text{org, Aitken}}$).

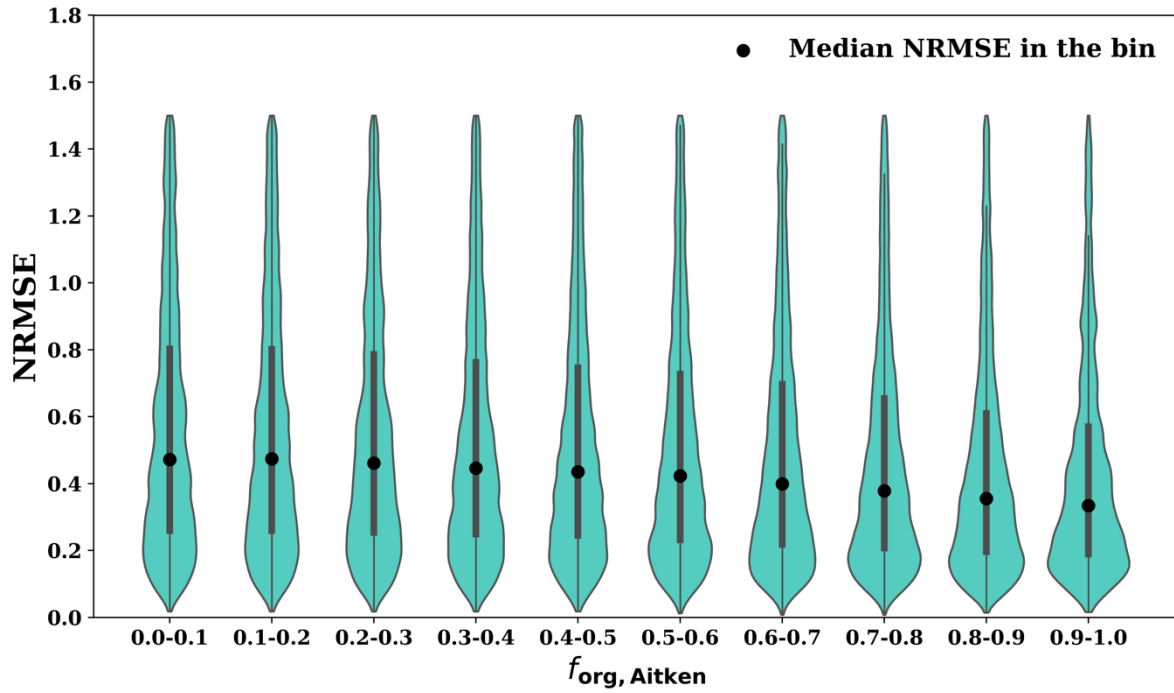


Figure S11 Violin plots of the normalized root mean square error (NRMSE) between measured and modelled CCN spectra during optimization, grouped by the mass fraction of organics in the Aitken mode ($f_{\text{org, Aitken}}$).

S8. Calculation of activation diameter

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:

$$D_{\text{act}} = D_0 + \frac{N_{\text{CCN}} - N_0}{N_1 - N_0} \times (D_1 - D_0) \quad (\text{S15})$$

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).

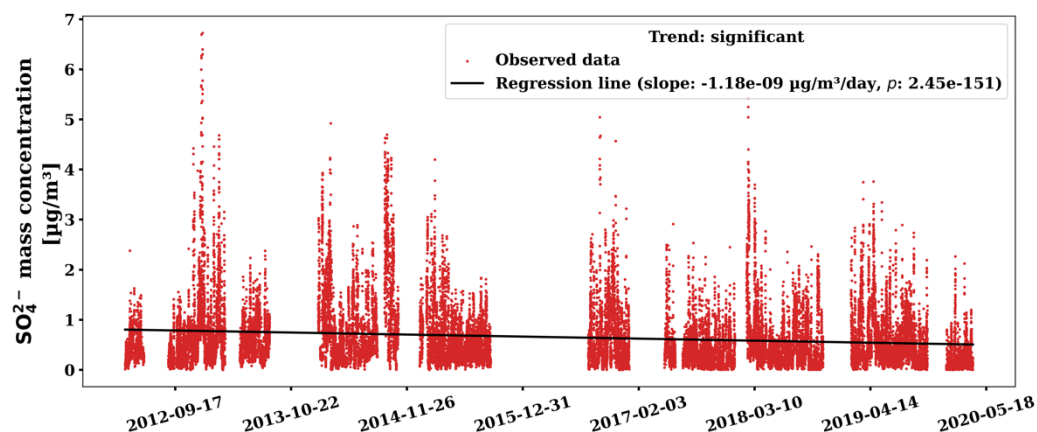


Figure S12 Time series of sulfate ion mass concentration measured by the ACSM. Scatter points represent individual sulfate ion concentration values, while the black solid line indicates the linear regression fit. The slope and p -value of the regression are provided in the legend. The statistical significance of the trend is assessed at the 95% confidence level ($p < 0.05$).

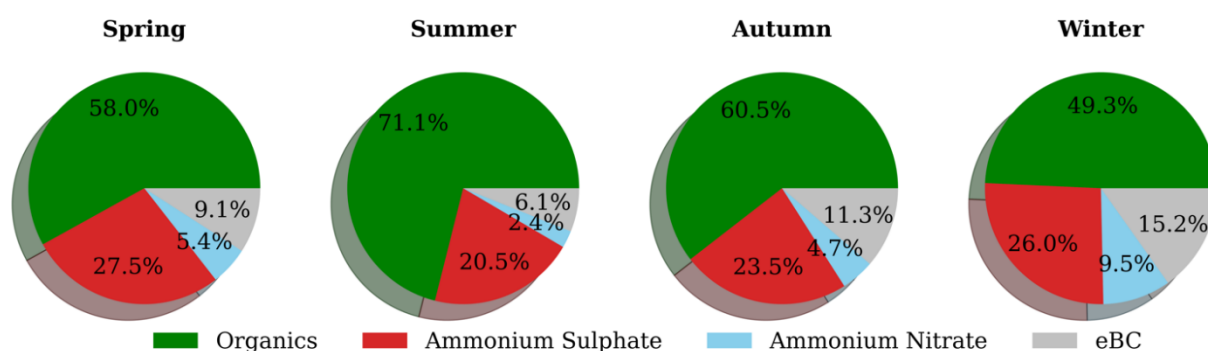


Figure S13 Mass fractions of various chemical species during different seasons.

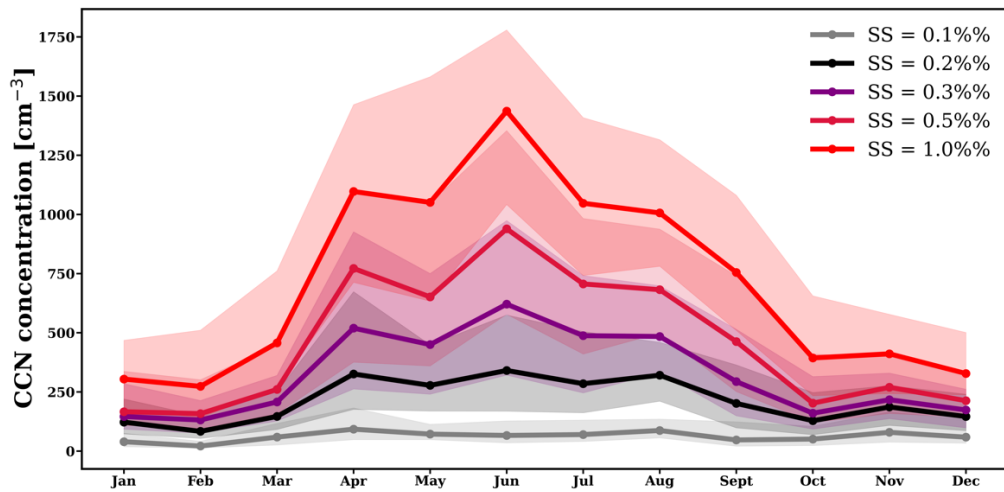


Figure S14 Monthly median CCN concentrations at different supersaturations with shaded areas indicating the interquartile range.

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

Methods	GMB (R) $SS = 0.1\%$	GMB (R) $SS = 0.2\%$	GMB (R) $SS = 0.3\%$	GMB (R) $SS = 0.5\%$	GMB (R) $SS = 1.0\%$
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)
$\kappa_{\text{org}} = 0$	0.92 (0.83)	0.87 (0.87)	0.94 (0.87)	1.086 (0.89)	1.938 (0.87)
κ_{MCMC}	1.32 (0.85)	0.95 (0.97)	0.96 (0.99)	1.05 (0.99)	0.99 (0.99)

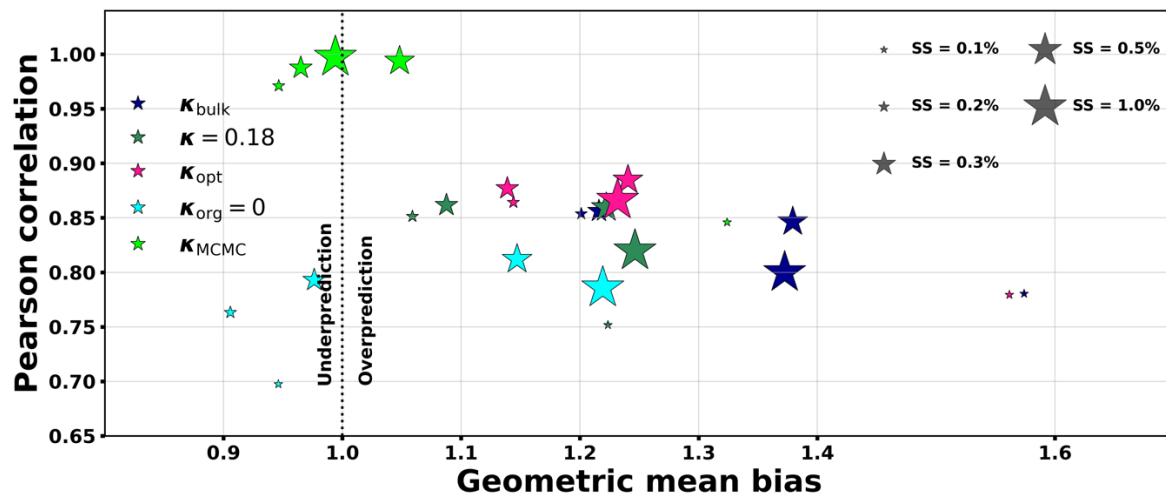


Figure S15 Scatter plot illustrating the geometric mean bias (GMB) and Pearson correlation for different supersaturation (SS) levels, comparing four methods: κ_{bulk} , $\kappa_{0.18}$, and κ_{opt} , $\kappa_{\text{org}} = 0$. Each marker represents a different SS level, with size proportional to SS 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).

S9. Performance Metrics of Various CCN Prediction Methods

In figure S15, we looked at GMB but importantly, another quantity, the standard deviation (SD) of GMB provides insight into the robustness of each method: a lower SD indicates that the bias is consistent across cases, while a higher SD suggests larger variability in performance (Fig. S16). This makes it a key metric to consider when evaluating the stability of CCN prediction methods.

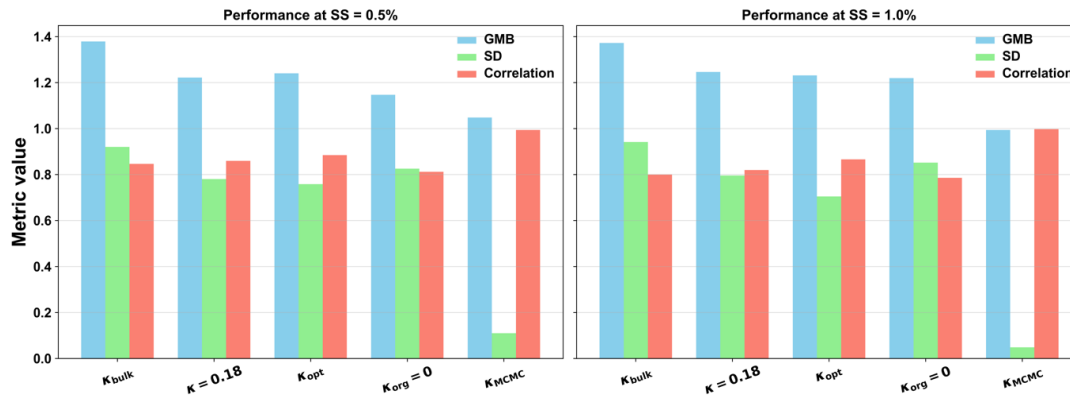


Figure S16 Comparison of CCN prediction performance at supersaturations of 0.5 % and 1.0 % using different methods. Metrics shown are geometric mean bias (GMB), its standard deviation (SD), and correlation between predicted and observed CCN.

The results show that DREAM-MCMC (κ_{MCMC}) performs best overall, with a geometric mean bias very close to unity, extremely high correlation with observations, and the lowest spread. This indicates that it provides the most accurate and stable CCN predictions. The κ_{opt} method, which optimizes only size-segregated composition, ranks second: it achieves high correlation and relatively low variability, though the bias remains slightly above one, suggesting moderate overprediction. $\kappa_{\text{org}} = 0$ falls in the middle, with moderate bias, correlation, and variability, making it balanced but not particularly strong in any metric. The fixed $\kappa = 0.18$ approach performs similarly to $\kappa_{\text{org}} = 0$ but is slightly weaker, with somewhat lower correlation. Finally, κ_{bulk} performs the worst, showing the highest bias and systematic overprediction of CCN, along with only moderate correlation and spread.

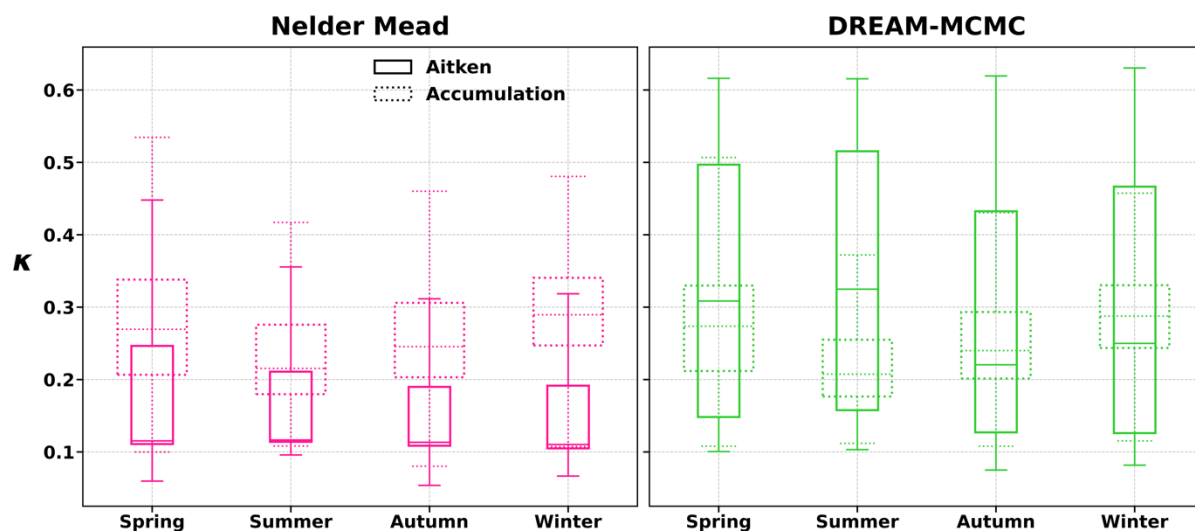


Figure S17. Seasonal variability of κ for Aitken and accumulation mode particles derived using Nelder–Mead optimization (left) and DREAM–MCMC inversion (right). Boxes represent interquartile ranges, whiskers the 5th–95th percentiles, and horizontal lines the medians.

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	0.0017	0.0001	0.0001	0.0001
eBC	0.10	0.075	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

Table S5. Mean of optimized mass fractions of various chemical species in Aitken mode in different seasons

	Spring	Summer	Autumn	Winter
Organics	0.70	0.75	0.72	0.69
Inorganics	0.20	0.19	0.16	0.16
eBC	0.1	0.06	0.11	0.15

Table S6. Mean of optimized mass fractions of various chemical species in Accumulation mode in different seasons

	Spring	Summer	Autumn	Winter
Organics	0.54	0.65	0.55	0.44

Inorganics	0.37	0.28	0.33	0.41
eBC	0.09	0.06	0.11	0.15

Table S7. Median mass fractions by group and optimization method.

Groups	Organics (Aitken)	Inorganics (Aitken)	BC (Aitken)	Organics (Acc)	Inorganics (Acc)	BC (Acc)
$\kappa_{\text{Aitken}} > \kappa_{\text{accumulation}}$ (MCMC)	0.11	0.78	0.11	0.69	0.23	0.08
$\kappa_{\text{Aitken}} > \kappa_{\text{accumulation}}$ (Nelder-Mead)	0.0004	0.87	0.12	0.68	0.23	0.087
$\kappa_{\text{accumulation}} > \kappa_{\text{Aitken}}$ (MCMC)	0.87	0.04	0.09	0.54	0.37	0.088
$\kappa_{\text{accumulation}} > \kappa_{\text{Aitken}}$ (Nelder-Mead)	0.91	0.0001	0.09	0.09	0.58	0.087

Table S8. Median κ and fraction of data points by group and optimization method.

Groups	Median κ_{Aitken}	Median $\kappa_{\text{accumulation}}$	Fraction of data
$\kappa_{\text{Aitken}} > \kappa_{\text{accumulation}}$ (MCMC)	0.47	0.21	0.54
$\kappa_{\text{Aitken}} > \kappa_{\text{accumulation}}$ (Nelder-Mead)	0.50	0.20	0.23
$\kappa_{\text{accumulation}} > \kappa_{\text{Aitken}}$ (MCMC)	0.13	0.27	0.46
$\kappa_{\text{accumulation}} > \kappa_{\text{Aitken}}$ (Nelder-Mead)	0.11	0.26	0.77

S10. Insights on effect of the assumption of fixed eBC mass in both modes

The CCN spectra obtained using bulk chemical composition generally depicts overprediction from the observations, with the largest bias at both lowest supersaturation (0.1%) and the highest supersaturation (1.0%). A way to reduce this overprediction is to have a size-segregated composition that makes the Aitken mode as low in hygroscopicity as possible. As a conservative approach, when eBC is fixed only in accumulation mode (leaving no eBC mass in Aitken mode), the 5 year averaged (median) NRMSE remains almost similar at 0.28 (a little on higher side), while Aitken kappa is a bit higher as a non-hygroscopic element i.e. eBC is now completely in accumulation mode. In another approach, we allowed eBC to vary as an additional optimization parameter along with organics in the Aitken mode instead of keeping its mass fraction fixed in both modes. As expected, this adjustment lowers the NRMSE from 0.28 to 0.23, while the optimized κ remains nearly unchanged, but a little on lower side than the setup we use (see Table S10). Interesting to note that, this setup suggest that there should be 70% eBC in Aitken mode which seems unrealistic. Complementary results from DREAM-MCMC optimization (optimizing modal BC mass fraction while considering variability of lognormal parameters of size distribution during CCN cycle, see Table S11) of the same problem suggests similar results with around on an average 41% BC in Aitken mode – again a very high number considering the expected aerosol sources at the cite.

Table S9. Median of optimized mass fractions of various chemical species in Aitken and accumulation mode when all eBC are kept in accumulation mode – results from Nelder-Mead

Organics (Aitken)	Inorganics (Aitken)	eBC (Aitken)	Organics (accumulation)	Inorganics (accumulation)	eBC (accumulation)
0.99	0.01	0.00	0.59	0.43	0.10

Table S10. Median of optimized κ in different seasons when all eBC are kept in accumulation mode– results from Nelder-Mead

	Spring	Summer	Autumn	Winter
κ_{Aitken}	0.12	0.12	0.12	0.12
$\kappa_{\text{accumulation}}$	0.27	0.21	0.24	0.29

Table S11. Median of optimized mass fractions of various chemical species in Aitken and accumulation mode when eBC is also an optimized parameter

	Organics (Aitken)	Inorganics (Aitken)	eBC (Aitken)	Organics (accumulation)	Inorganics (accumulation)	eBC (accumulation)
Nelder-Mead	0.26	0.046	0.70	0.62	0.31	0.07
MCMC	0.32	0.27	0.41	0.63	0.30	0.07

S11. Optimizing organic properties

We conducted additional inverse-closure sensitivity studies to assess how the assumed organic aerosol properties influence the optimized parameters. Three types of inverse-closure tests were performed:

- Using bulk-composition but optimizing only the organic density, ρ_{org} and organic hygroscopicity parameter, κ_{org}
- Optimizing ρ_{org} and κ_{org} while also accounting for variability of size distribution lognormal parameters during CCN cycles
- Optimizing $m_{\text{org,Ait}}$, ρ_{org} and κ_{org} but not accounting for variability of size distribution lognormal parameters during CCN cycles

In all tests, κ_{org} was varied between 0.05 and 0.15, while ρ_{org} ranged from 1000 to 3000 kg m⁻³. Method (a) resulted in a lower NRMSE but yielded a median optimized organic density of 1000 kg m⁻³, which appears unrealistic. Method (b) produced a slightly higher NRMSE (0.085 compared to 0.079 from κ_{MCMC}) but provided a more realistic optimized organic density of 2179 kg m⁻³, exhibiting clear seasonal variability, with a minimum of approximately 1750 kg m⁻³ in summer. This value is notably higher than the 1500 kg m⁻³ assumed in the original inverse-closure approach. Both methods produced optimized κ_{org} values between 0.05 and 0.07, depending on the season. Overall, we obtained a reasonable estimate of optimized κ_{org} however, due to certain

unknown factors, the optimized ρ_{org} values remain physically implausible. In contrast, method (c) produced highly consistent and realistic results, as summarized in the table below.

Table S12: Median of optimized quantities in different seasons when $m_{\text{org,Ait}}$, ρ_{org} and κ_{org} are considered the for optimization while not accounting for variability of size distribution lognormal parameters

	κ_{Aitken}	$\kappa_{\text{accumulation}}$	$\kappa_{\text{org}}^{\text{Aitken}}$	$\kappa_{\text{org}}^{\text{accumulation}}$	$\rho_{\text{org}} \text{ (kg m}^{-3}\text{)}$
Spring	0.16	0.24	0.07	0.07	1228.56
Summer	0.15	0.18	0.07	0.07	1304.12
Autumn	0.13	0.23	0.07	0.08	1307.69
Winter	0.11	0.26	0.06	0.07	1224.46