Supplement of Atmos. Chem. Phys., 26, 3697–3722, 2026
https://doi.org/10.5194/acp-26-3697-2026-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Cloud condensation nuclei phenomenology: predictions based on aerosol chemical and optical properties

Inés Zabala et al.

Correspondence to: Inés Zabala (ineszabala@ugr.es) and Gloria Titos (gtitos@ugr.es)

Supplemental Materials

Shen methodology

The first step in the approach of S2019 demonstrates that a logarithmic function more accurately captures the dependence of N_{CCN} on SS than other commonly used fits (see Fig. 1 in S2019). Figure S7 shows the same result for the stations considered here. The second step explores the relationship between N_{CCN} and σ_{sp}, highlighting the role of BSF in modulating this dependence. S2019 introduce the ratio $R_{CCN/\sigma} = N_{CCN}/\sigma_{sp}$ and show that there is a linear relationship between $R_{CCN/\sigma}$ and BSF:

$$N_{\mathrm{CCN}} = (a_{\mathrm{SS}} \cdot \mathrm{BSF} + b_{\mathrm{SS}}) \cdot \sigma_{\mathrm{sp}} = ((a_1 \ln(\mathrm{SS}) + a_0) \cdot \mathrm{BSF} + b_1 \ln(\mathrm{SS}) + b_0) \cdot \sigma_{\mathrm{sp}} \tag{S1}$$

Equation S1 provides the starting point for the parameterization of CCN using aerosol optical properties. Fit coefficients at each SS for the sites analyzed in S2019 are listed in their Table 3, while those for the sites in this study are shown in Table S1. This relationship clearly differs among sites and for different SS. To eliminate the SS dependence, the slopes (a_{SS}) and offsets (b_{SS}) from the linear regressions are plotted against the SS, following the S2019 methodology. As shown in Figure S1, the data follow a logarithmic fit, leading to the reformulation of equation S1 as:

$$R_{\mathrm{CCN}/\sigma} = \frac{N_{\mathrm{CCN}}}{\sigma_{\mathrm{sp}}} = a \cdot \mathrm{BSF} + b \tag{S2}$$

Table S1. Slopes and offsets of linear regressions of $R_{CCN/\sigma}$ vs. BSF at different SS. SE: standard error of the respective coefficient obtained from the linear regressions. The coefficients are written as $[N_{CCN}]/[\sigma_{sp}] = a \cdot BSF + b$, in units of cm^{-3} Mm.

Site		**0.1%**	**0.2%**	**0.4%**	**0.6%**	**0.8%**	**1.0%**
ANX	$a \pm SE$	39 ± 9	81 ± 13	183 ± 19	237 ± 22	265 ± 25	326 ± 28
	$b \pm SE$	-2.1 ± 1.3	-3.3 ± 1.9	-12.9 ± 2.8	-18.8 ± 3.2	-21.2 ± 3.7	-29.3 ± 4.2
GUC	$a \pm SE$	85 ± 6	410 ± 7	648 ± 11	788 ± 16	917 ± 20	980 ± 24
	$b \pm SE$	-2.7 ± 1.0	-27.8 ± 1.2	-39.2 ± 2.0	-40.4 ± 2.8	-44.5 ± 3.6	-43.0 ± 4.2
COR	$a \pm SE$	239 ± 32	634 ± 43	998 ± 74	1430 ± 110	1724 ± 146	1919 ± 182
	$b \pm SE$	-21.5 ± 5.9	-66.2 ± 8.0	-84.0 ± 13.6	-120.7 ± 20.2	-133.8 ± 26.8	-137.9 ± 33.6
ENA	$a \pm SE$	88 ± 7	191 ± 19	312 ± 35	n/a	339 ± 23	377 ± 27
	$b \pm SE$	-8.2 ± 0.9	-19.7 ± 2.5	-33.1 ± 4.7	n/a	-34.7 ± 3.1	-39.5 ± 3.6
MAO	$a \pm SE$	-27 ± 121	162 ± 12	272 ± 22	363 ± 37	428 ± 48	455 ± 61
	$b \pm SE$	9.9 ± 14.8	-1.6 ± 1.7	-2.0 ± 3.2	-2.2 ± 5.2	-1.4 ± 6.9	2.4 ± 8.7
MOS	$a \pm SE$	405 ± 8	496 ± 10	674 ± 15	776 ± 16	862 ± 20	888 ± 22
	$b \pm SE$	-35.2 ± 1.0	-44.8 ± 1.2	-63.8 ± 1.8	-74.3 ± 1.9	-83.1 ± 2.4	-85.4 ± 2.7
SGP	$a \pm SE$	319 ± 7	624 ± 6	1175 ± 10	n/a	1899 ± 18	2042 ± 27
	$b \pm SE$	-24.9 ± 1.1	-56.2 ± 1.1	-115.5 ± 1.7	n/a	-193.4 ± 2.9	-214.1 ± 4.5

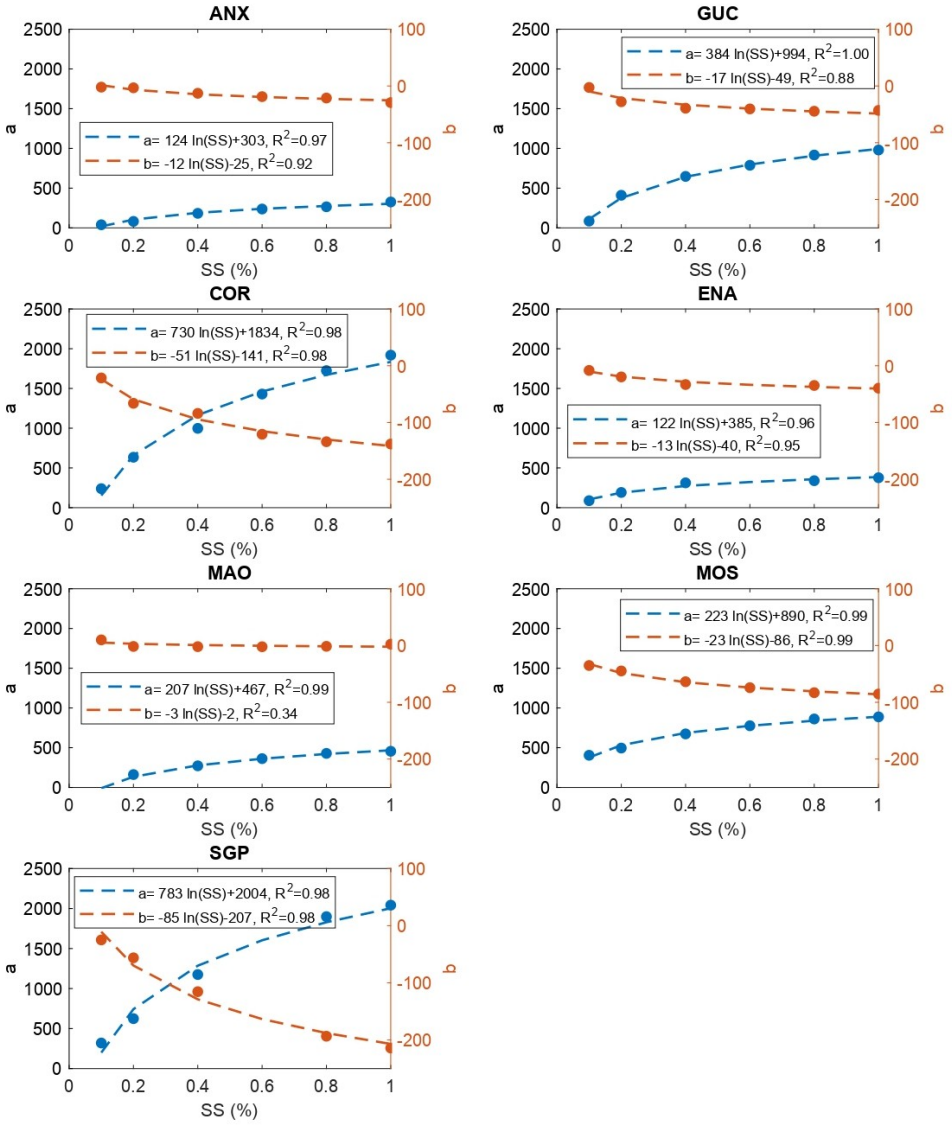

Figure S1. Slopes (a) and offsets (b) of the linear regressions $R_{CCN/\sigma} = a.BSF + b$ of each site (Table S1) as a function of SS. Logarithmic fitting applied to data.

The coefficients a_1, a_0, b_1 and b_0 with their respective errors from both this study and Shen et al. (2019) are shown in Table S2. Next, to obtain a site-independent parametrization, the different coefficients from all sites are combined. Figure S2 shows the relationships of the coefficients a_0 vs. a_1, b_0 vs. b_1, a_1 vs. b_1, and a_0 vs. b_0. Linear regressions yield a_0 = (2.41 $\pm$ 0.13)a_1, b_0 = (2.42 $\pm$ 0.12)b_1 and b_1 = (-0.095 $\pm$ 0.011)a_1+(5.7$\pm$ 11.0). Considering these relationships and, after the development

shown in "Derivation of equation S3 ", equation S1 can be expressed as:

$$N_{\mathrm{CCN}} \approx \ln\left(\frac{\mathrm{SS}}{0.089 \pm 0.011}\right)\left[a_1\left(\mathrm{BSF} - (0.095 \pm 0.011)\right) + (5.7 \pm 11.0)\right]\sigma_{\mathrm{sp}} \tag{S3}$$

Table S2. Coefficients a_1, a_0, b_1 and b_0 from the logarithmic fitting of coefficients in Table S1 to SS. Shen et al. values are given in Table 4 of Shen et al. (2019). SE: standard error of the respective coefficient obtained from the linear regressions.

	Site	$a_1 \pm$ SE	$a_0 \pm$ SE	$b_1 \pm$ SE	$b_0 \pm$ SE
Shen et al.	SMEAR II	464 ± 11	1170 ± 16	-49 ± 1.5	-118 ± 0.67
	SORPES	331 ± 12	817 ± 18	-26 ± 0.9	-62 ± 1.4
	PGH	205 ± 30	385 ± 41	-6.3 ± 1.5	-9.1 ± 2.0
	PVC	810 ± 17	1933 ± 21	-70 ± 1.7	-160 ± 2.1
	MAO	393 ± 45	858 ± 40	-25 ± 6.6	-60 ± 5.8
	ASI	52 ± 17	164 ± 26	-2.9 ± 1.6	-6.3 ± 2.3
This work	ANX	124 ± 18	303 ± 14	-11 ± 2.9	-25 ± 2
	GUC	384 ± 20	994 ± 17	-17 ± 5	-49 ± 4
	COR	730 ± 96	1834 ± 77	-51 ± 6	-141 ± 5
	ENA	122 ± 30	385 ± 23	-13 ± 4	-40 ± 3
	MAO	207 ± 16	467 ± 13	-3 ± 4	-2 ± 3
	MOS	222 ± 23	889 ± 18	-23 ± 2	-86 ± 2
	SGP	783 ± 140	2003 ± 106	-85 ± 16	-206 ± 12

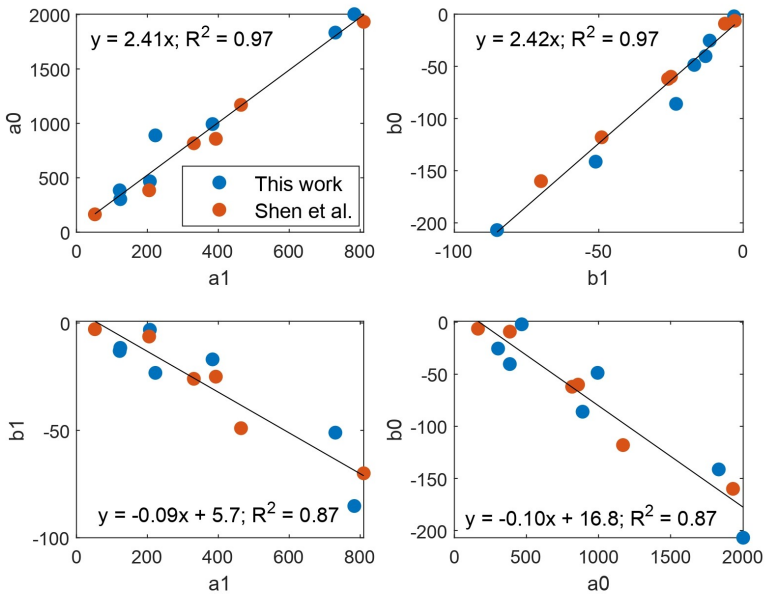

Figure S2. Relationship between the coefficients a_0, a_1, b_0, and b_1 of Eq. S1 for each site shown in Table S2. The coefficients units are cm^{-3} Mm.

Derivation of equation S3

$$N_{\text{CCN}} = (a_{\text{SS}} \cdot \text{BSF} + b_{\text{SS}}) \cdot \sigma_{\text{sp}}$$
$$= [(a_1 \ln(\text{SS}) + a_0) \cdot \text{BSF} + b_1 \ln(\text{SS}) + b_0] \cdot \sigma_{\text{sp}}$$

Considering linear regressions obtained from Figure S1:
$a_0 \approx (2.41 \pm 0.13)a_1$, $b_0 \approx (2.42 \pm 0.12)b_1$ and $b_1 \approx$ (-0.095 $\pm$ 0.011)a_1+(5.7$\pm$11.0)

$$a_1 \ln(\text{SS}) + a_0 \approx a_1 \ln(\text{SS}) + (2.41 \pm 0.13)a_1 \approx a_1 (\ln(\text{SS}) + (2.41 \pm 0.13))$$
$$b_1 \ln(\text{SS}) + b_0 \approx b_1 \ln(\text{SS}) + (2.42 \pm 0.12)b_1 \approx b_1 (\ln(\text{SS}) + (2.42 \pm 0.12))$$
$$\approx ((-0.095 \pm 0.011)a_1 + (5.7 \pm 11.0))(\ln(\text{SS}) + (2.42 \pm 0.12))$$

$$R_{\text{CCN}} \approx (a_1 \ln(\text{SS}) + a_0)\text{BSF} + b_1 \ln(\text{SS}) + b_0$$
$$\approx a_1 (\ln(\text{SS}) + (2.41 \pm 0.13))\,\text{BSF} + ((-0.095 \pm 0.011)a_1 + (5.7 \pm 11.0))(\ln(\text{SS}) + (2.42 \pm 0.12))$$
Approximation, since $(2.42 \pm 0.12) \approx (2.41 \pm 0.13)$

$$R_{\text{CCN}} \approx a_1 (\ln(\text{SS}) + (2.41 \pm 0.12))\,\text{BSF} - (0.095 \pm 0.011)a_1 (\ln(\text{SS}) + (2.41 \pm 0.13)) + (5.7 \pm 11.0)(\ln(\text{SS}) + (2.41 \pm 0.13))$$
$$\approx a_1 (\ln(\text{SS}) + (2.41 \pm 0.13))(\text{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0)(\ln(\text{SS}) + (2.41 \pm 0.13))$$
$$\approx \ln(\text{SS}) + (2.41 \pm 0.13)(a_1 (\text{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$
$$\approx (\ln(\text{SS}) - \ln(exp(-2.41) \pm exp(-2.41)(0.13)))(a_1 (\text{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$
$$\approx \ln\left(\frac{\text{SS}}{0.089 \pm 0.011}\right)(a_1 (\text{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$

It was shown in Shen et al. (2019) that when the number of hourly samples exceeds approximately 1000 — a condition also met at all our sites — the uncertainty in the minimum BSF (BSF_{min}) becomes sufficiently low. Therefore, instead of subtracting a fixed offset of (0.095 ± 0.025) from the BSF, we use the observed minimum BSF value (BSF_{min}; 1st percentile of BSF). In addition, as shown in the derivation presented in Supplementary Section S4 of Shen et al. (2019), the final term (5.7$\pm$11.0) is treated as a constant C, which depends on R_{min}, defined as the minimum (first percentile) of N_{CCN}/σ_{sp}. Taking all this into account, Eq. S3 can be reformulated by incorporating these terms, and is written as follows.

$$N_{\text{CCN}} \approx \left(a_1 \ln\left(\frac{\text{SS}}{0.089 \pm 0.011}\right)(\text{BSF} - \text{BSF}_{\text{min}}) + R_{\text{min}}\right) \cdot \sigma_{\text{sp}}. \quad \text{(S4)}$$

The final step consists of relating the coefficient a_1 in Eq. S4 to the scattering Ångström exponent (SAE), which is the only parameter among optical properties found to be positively correlated with a_1. Based on the median values from Shen et al. (2019) and from this study, linear regression yields $a_1 \approx (320 \pm 78) \cdot \mathrm{SAE}\ \mathrm{cm}^3$ Mm (Fig. S3). Additionally, the minimum value of R in Eq. S4, R_{min}, was estimated as the 1st percentile of $R_{CCN/\sigma}$ at each site and supersaturation, resulting in an average value of $R_{\mathrm{min}} = 8.7 \pm 9.3\ \mathrm{cm}^{-3}$ Mm. Consequently, the parameterization becomes

$$N_{\mathrm{CCN}} \approx \left[(320 \pm 78)\,\mathrm{SAE} \cdot \ln\left(\frac{\mathrm{SS}}{0.089 \pm 0.011} \right) (\mathrm{BSF} - \mathrm{BSF}_{\mathrm{min}}) + (8.7 \pm 9.3) \right] \cdot \sigma_{\mathrm{sp}}. \tag{S5}$$

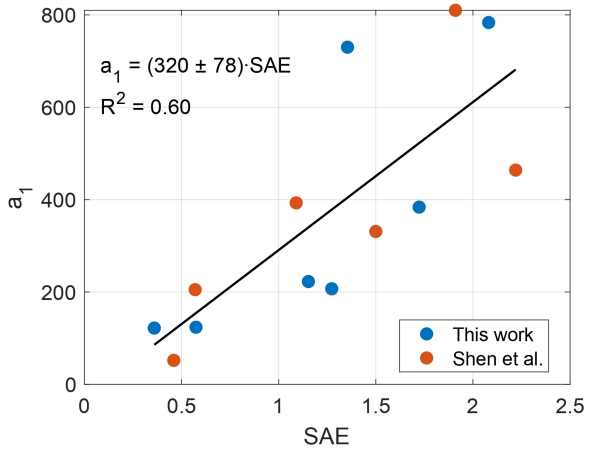

Figure S3. Relationship of the a_1 coefficient in Eq. S3 with the average PM_{10} scattering Ångström exponent (SAE).

Random forest performance

To assess the robustness of the random forest models, detailed training and test performance metrics for the two parameters of the Twomey equation (C and k) are reported. The models were trained using two sets of predictor configurations: the S2019 AOPs (σ_{sp}, BSF, SAE) and all AOPs (σ_{sp}, BSF, SAE, σ_{ap}, AAE, SSA). Table S3 summarizes RF performance metrics for the (a) S2019 AOPs and (b) all AOPs predictor sets. Metrics include the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) for both training and test datasets.

Table S3. Random forest performance metrics for the Twomey parameters C and k for both predictor sets, with training and test datasets shown: (a) S2019 AOPs and (b) All AOPs.

a)

Parameter	Dataset	R^2	RMSE	MAE
C	train	0.50	778	478 cm^{-3}
C	test	0.48	802	480 cm^{-3}
k	train	0.18	0.35	0.26
k	test	0.18	0.35	0.26

b)

Parameter	Dataset	R^2	RMSE	MAE
C	train	0.59	709	412 cm^{-3}
C	test	0.58	720	412 cm^{-3}
k	train	0.36	0.31	0.22
k	test	0.36	0.30	0.22

Overall, the models show similar performance on training and test datasets, indicating consistent behavior and no signs of under- or overfitting. Parameter C is moderately well predicted (R^2 = 0.50 - 0.59), while predictive skill for k is lower (R^2 = 0.18 - 0.36), likely due to weaker relationships between the predictors and k rather than model deficiencies.

All models were implemented in MATLAB using the `TreeBagger` function with 500 regression trees. At each split, all predictors were considered, and no explicit maximum tree depth was imposed. Model performance during training was evaluated using out-of-bag (OOB) predictions, which serve as an intrinsic cross-validation estimate. Convergence analysis of OOB RMSE and R^2 (Figure S4) confirms stable performance beyond $\sim$ 400-500 trees, justifying the choice of 500 trees for all models.

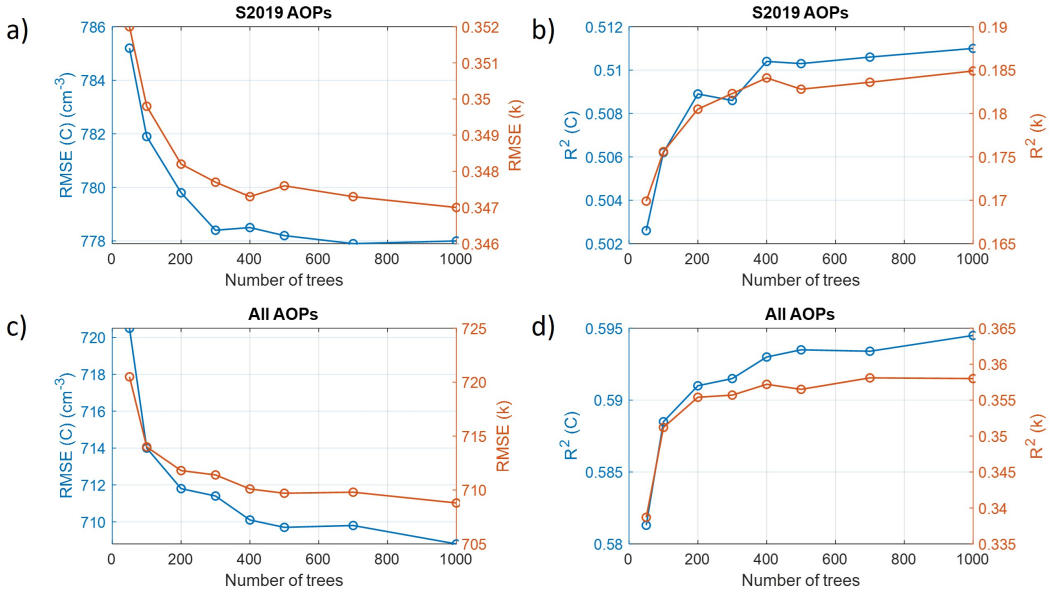

Figure S4. Convergence analysis of RF model performance as a function of the number of trees. Panels (a) and (c) show RMSE for C (left axis) and k (right axis) for the S2019 AOPs and All AOPs, respectively. Panels (b) and (d) show corresponding R^2. Metrics stabilize for $\sim$400–500 trees.

Figures and tables

Table S4. List of measurement sites considered in this study and measurements available at each site. [1]MOSAiC moved over the course of the year as the ship drifted with the ice - the latitude and longitude provided represent approximately the midpoint of the campaign. [2]CP=Christy Peak. [3]SPL=Storm Peak Laboratory. Sites where DMT2C is available, colB measures at 0.4% fixed SS. (1) or (2) in SS indicate periods with different SS values. If the lower or upper size bin range of the PNSD measurements changed over the course of the measurement process, this is indicated with a slash.

	Site ID	**ANX**	**COR**	**ENA**	**GUC**	**MAO**
Station information	Location (Campaign/ Site name)	Andenes, Norway (COMBLE)	Cordoba, Argentina (CACTI)	Azores, Portugal (Eastern North Atlantic)	Colorado, USA (SAIL)	Manacapuro, Brazil (GoAMAZON)
	Coordinates	69.14 N 15.68 E, 2 masl	32.13 S 64.73 W, 1141 masl	39.09 N 28.02 W, 30 masl	38.90 N 106.94 W, 3137 masl	3.21 S 60.60 W, 50 masl
	Site type	Polar/ Marine	Continental	Marine	Mountain/ Continental	Continental/ Urban
	Site reference	Geerts et al. (2022)	Fast et al. (2024)	Wood et al. (2015)	Feldman et al. (2023)	Martin et al. (2016)
CCN measurements	Time period	Dec 2019- May 2020	Oct 2018- Apr 2019	Nov 2022- Dec 2023	Nov 2021- Jun 2023	Jan 2014- Nov 2015
	CCNC Model	DMT2C	DMT2C	DMT2C	DMT2C	DMT1C
	SS (%)	0.1, 0.2, 0.4, 0.6, 0.8, 1	0.1, 0.2, 0.4, 0.6, 0.8, 1	0.1, 0.2, 0.4, 0.8, 1	0.1, 0.2, 0.4, 0.6, 0.8, 1	(1):0.2,0.4, 0.6,0.8,1 (2):0.1,0.2, 0.4,0.8
Particle number size distribution measurements	Time period	Dec 2019- Apr 2020	Sep 2018- Apr 2019	Nov 2022- Jan 2024	Oct 2021- Jun 2023	Jan 2014- May 2015
	SMPS Model	TSI 3938	TSI 3936	TSI 3938	TSI 3938	TSI 3936
	Size range (nm)	10/11-496/514	11-461	11-496	13-496/514	10/13-461
Chemical composition measurements	Time period	n/a	Sep 2018- May 2019	Nov 2022- Apr 2023	Apr 2022- Apr 2023	n/a
	ACSM Model	n/a	Q-ACSM	Q-ACSM	Q-ACSM	n/a
Optical properties measurements	Time period	Dec 2019- Apr 2020	Sep 2018- Apr 2019	Nov 2022- Dec 2023	Oct 2021- Jun 2023	Jan 2014- May 2015
	Neph Model PSAP Model	TSI 3563 RR 3W	TSI 3563 RR 3W	TSI 3563 RR 3W	TSI 3563 RR 3W	n/a

Campaign abbreviations: COMBLE = Cold-Air Outbreaks in the Marine Boundary Layer Experiment, CACTI = Cloud Aerosol and Complex Terrain Interactions, SAIL = Surface Atmosphere Integrated field Laboratory, GoAMAZON = Observations and Modeling of the Green Ocean Amazon.

Table S5. Continuation of Table S4

	Site ID	MOS	SBS-CP[2]	SBS-SPL[3]	SGP
Station information	Location (Campaign/ Site name)	Arctic Ocean[1] (MOSAiC)	Colorado, USA (STORMVEX)	Colorado, USA (STORMVEX)	Oklahoma, USA (Southern Great Plains)
	Coordinates	85 N 15 E, 12 masl	40.45 N 106.79 W, 2438 masl	40.45N 106.73 W, 3220 masl	36.61 S 97.49 W, 318 masl
	Site type	Polar/ Marine	Mountain/ Continental	Mountain/ Continental	Continental
	Site reference	Shupe et al. (2022)	Mace et al. (2010)	Mace et al. (2010)	Sheridan et al. (2001)
CCN measurements	Time period	Oct 2019- Oct 2020	Oct 2010- Apr 2011	Nov 2010- Apr 2011	Apr 2017- Aug 2023
	CCNC Model	DMT2C	DMT1C	DMT1C	DMT2C
	SS (%)	0.1, 0.2, 0.4, 0.6, 0.8, 1	(1): 0.2, 0.4, 0.6, 0.8, 1 (2): 0.1, 0.2, 0.4, 0.8, 1	(1): 0.1,0.2, 0.4,0.6, 0.8,1 (2): 0.1, 0.2, 0.4,0.8,1	0.1, 0.2, 0.4, 0.8, 1
Particle number size distribution measurements	Time period	Oct 2019- Oct 2020	Oct 2010- May 2011	Dec 2010- Mar 2011	Jan 2017- Dec 2023
	SMPS Model	TSI 3938	TSI 3936	TSI 3936	TSI 3938
	Size range (nm)	10-496/514	9-334	9-334	10/11-461/514
Chemical composition measurements	Time period	n/a	n/a	n/a	Jan 2017- Apr 2023
	ACSM Model	n/a	n/a	n/a	Q-ACSM
Optical properties measurements	Time period	Oct 2019- Oct 2020	n/a	n/a	Jan 2017- Dec 2023
	Neph Model PSAP Model	TSI 3563 RR 3W	n/a	n/a	TSI 3563 RR 3W

Campaign abbreviations: MOSAiC = Multidisciplinary drifting Observatory for the Study of ArctIc Climate, STORMVEX = Storm Peak Lab Cloud Property Validation Experiment.

Table S6. Densities (ρ) and hygroscopicity parameters (κ) for the aerosol species considered in the κ_{chem} calculation. Chloride is assumed to be present as NaCl, and black carbon (BC) particles are considered completely hydrophobic ($\kappa_{BC} = 0$) (Schmale et al., 2018; Deng et al., 2019). Values for NH_4NO_3, $(NH_4)_2SO_4$ and $(NH_4)HSO_4$ were taken from Wu et al. (2016) and organic aerosol values from Jiang et al. (2019) and Dusek et al. (2010). Values for H_2SO_4 are based on Gysel et al. (2007) and Petters and Kreidenweis (2007), while NaCl properties are from Zieger et al. (2017).

	NH_4NO_3	$(NH_4)_2SO_4$	$(NH_4)HSO_4$	H_2SO_4	$NaCl$	BC	$Organics$
ρ (g/cm^3)	1.72	1.77	1.78	1.83	2.16	1.7	1.4
κ	0.58	0.48	0.56	0.9	1.5	0	0.1

Table S7. Number of points considered for each site and each methodology (chemistry- or optics-derived methodologies). Number of points shown for random forest methodologies correspond to the 30% data subset used for testing.

	Method	COR	SGP	MAO	GUC	ENA	MOS	ANX
Chemistry	Scheme 1	1199	15498	–	607	–	–	–
	Scheme 2	1237	17318	–	8318	618	–	–
	Scheme 3	1467	17999	–	9626	800	–	–
Optics Shen	S2019	4364	98254	9299	34988	7528	14323	7183
	New	4364	98254	9299	34988	7528	14323	7183
Optics RF	S2019 AOPs	1144	17222	244	13105	1187	5831	1895
	All AOPs	1135	17104	244	13002	1126	5726	1841

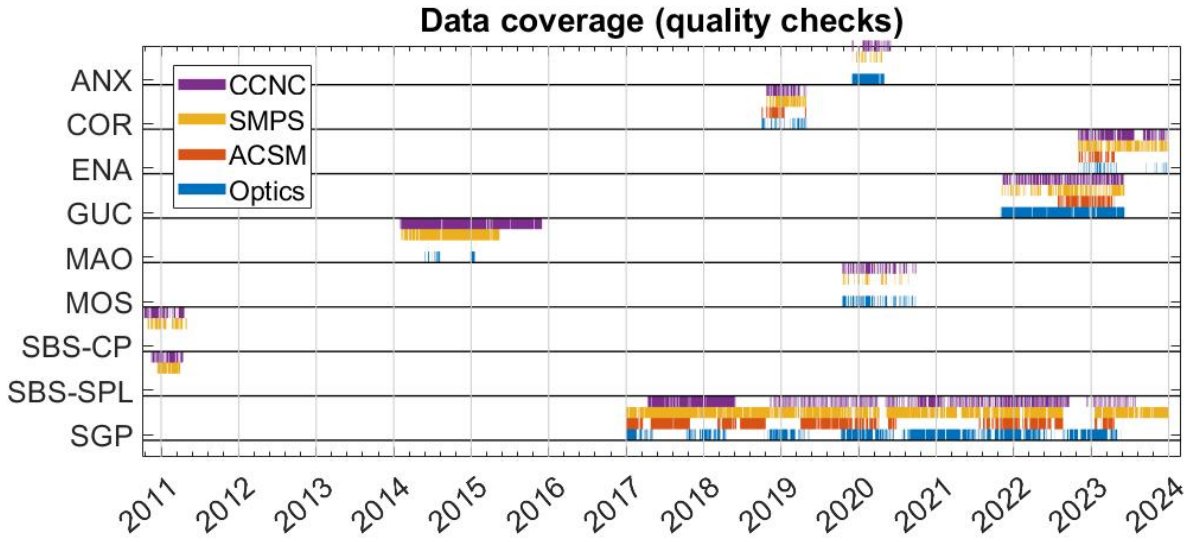

Figure S5. Data coverage of the instruments used in this study for all the sites. Total particle concentration comparison and CCN counter column comparison quality checks applied.

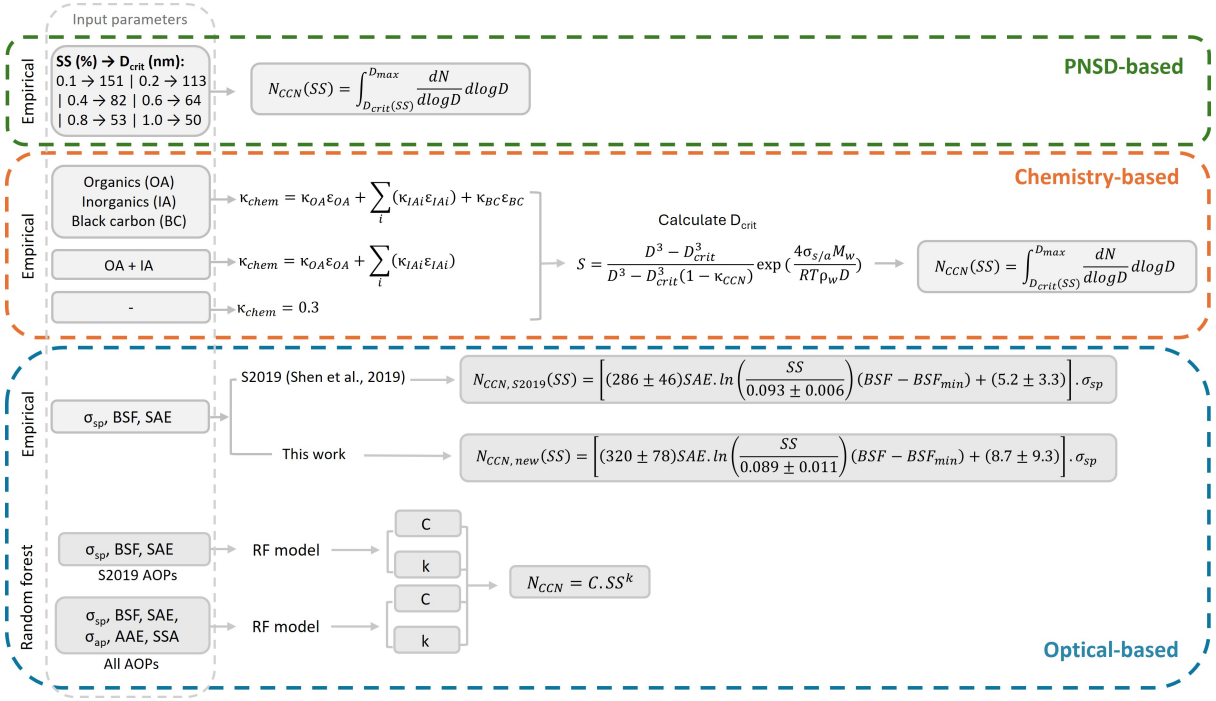

Figure S6. Flowchart of the models used to predict CCN concentrations.

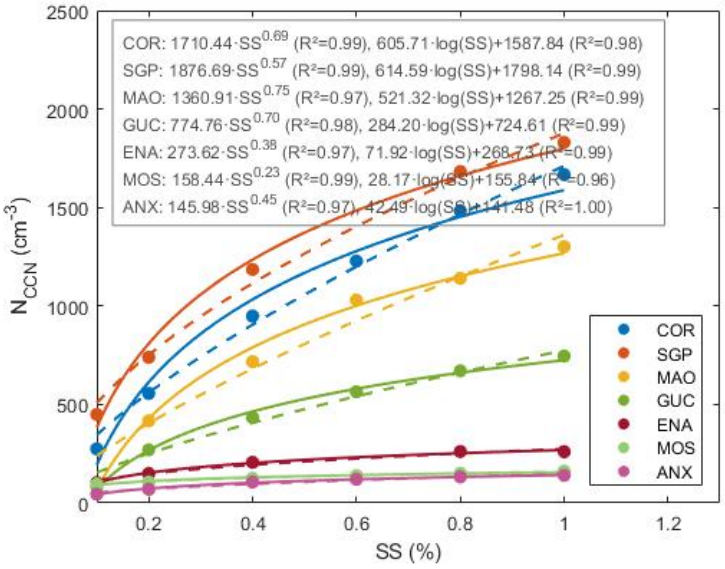

Figure S7. N_{CCN} measured average values at different SS. The Twomey equation power function is represented by the solid lines for each site. A logarithmic fit to the same data is represented by the dashed lines. The fit equations for both types of functions are provided in the legend for each site.

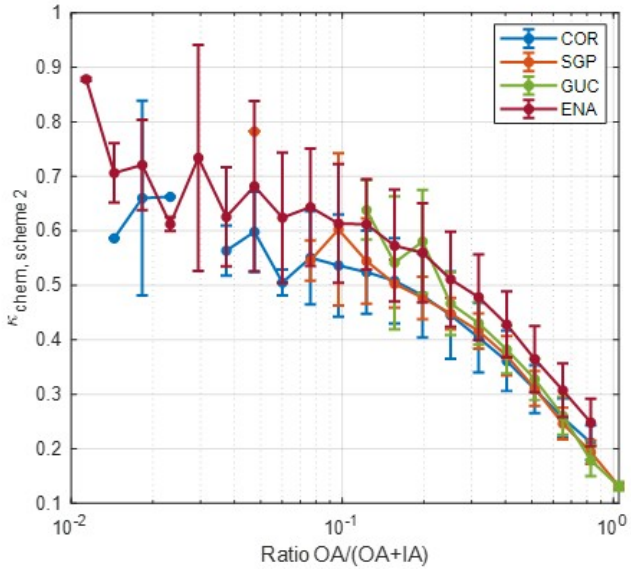

Figure S8. Relationship of the composition-derived κ_{chem} from Scheme 2 to the binned and averaged ratio of organic (OA) to total (OA+IA) aerosol components. The vertical bars denote the standard deviation.

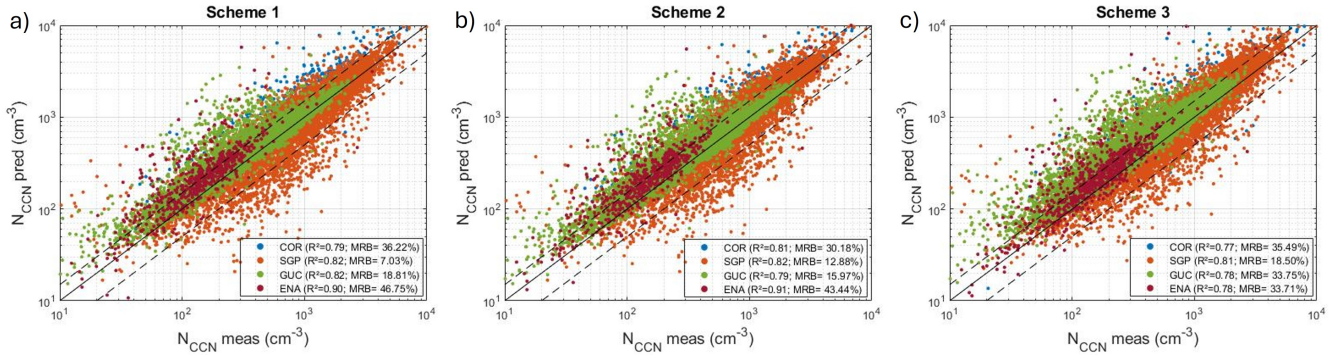

Figure S9. Log-log scatter plot of predicted CCN concentrations (N_{CCN} pred) with respect to the observed CCN concentrations (N_{CCN} pred) at all SS colored for different sites using the three prediction schemes. (a) Scheme 1 ($\kappa_{\text{chem,Sch1}}$), (b) Scheme 2 ($\kappa_{\text{chem,Sch2}}$) and (c) Scheme 3 (fixed κ). The solid black line represents the 1:1 line and the dashed lines are the +/-50%. Coefficient of determination (R^2) and median relative bias (MRB) shown for each site.

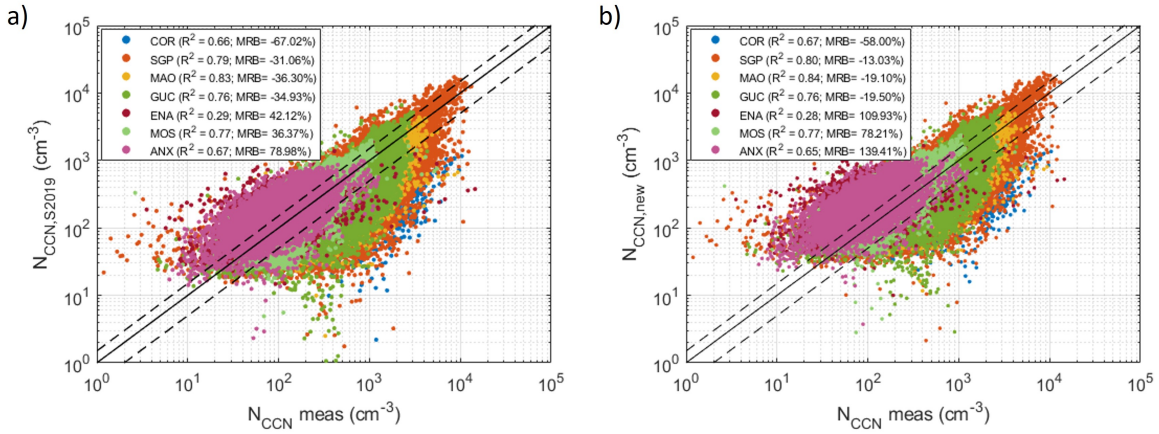

Figure S10. Log-log scatter plot of predicted CCN concentrations with respect to the observed CCN concentrations (N_{CCN} meas) colored for different sites considering (a) equation in S2019 ($N_{CCN,S2019}$; based on 6 sites) and (b) new equation ($N_{CCN,new}$; based on 13 sites). Coefficient of determination (R^2) and median relative bias (MRB) shown for each site. The solid black line represents the 1:1 line and the dashed lines are the +/-50%.

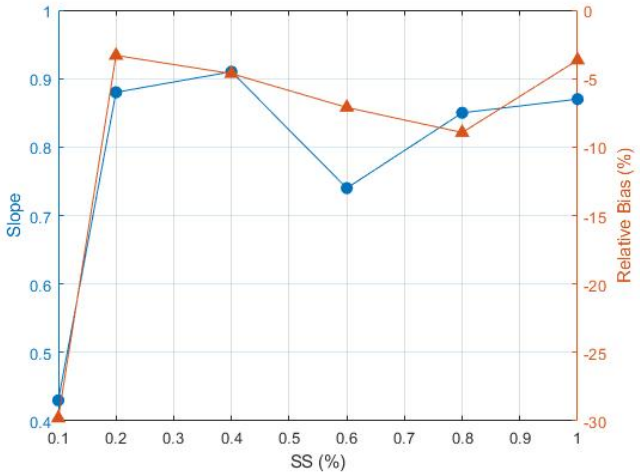

Figure S11. Slope and median relative bias (%) for each supersaturation (SS) level between the predicted and measured N_{CCN}, based on the new equation proposed in this study.

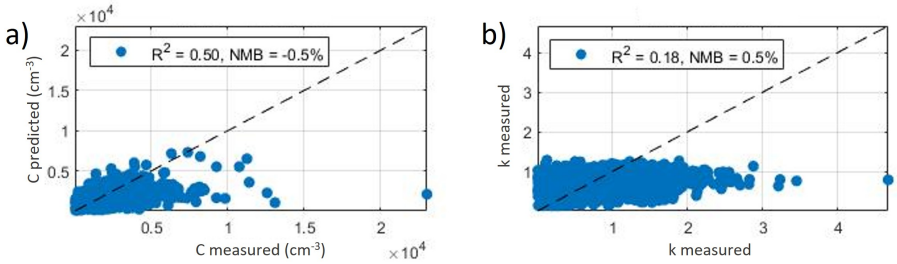

Figure S12. Comparison of observed versus Random Forest–predicted Twomey parameters: (a) parameter C and (b) parameter k. The model was trained using the AOPs used in S2019 (σ_{sp}, BSF, and SAE). For each parameter, the coefficient of determination (R^2) and normalized mean bias (NMB) are reported. The dashed line represents the 1:1 line.

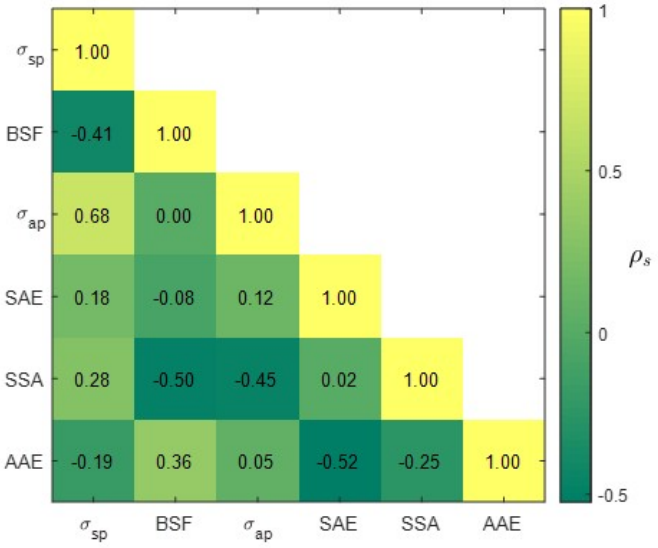

Figure S13. Spearman correlation coefficients (ρ_s) between all the AOPs.

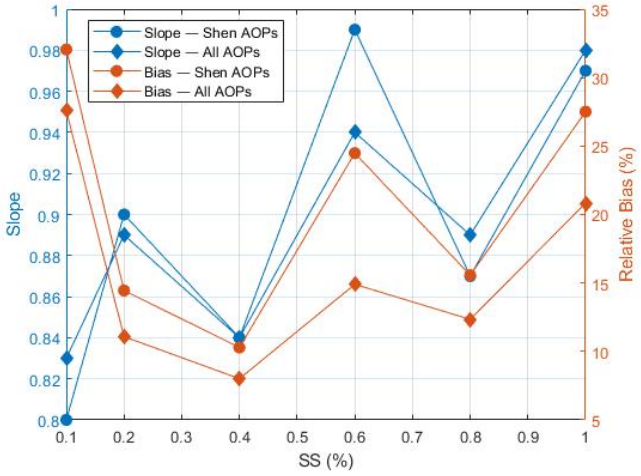

Figure S14. Slope and median relative bias (%) for each supersaturation (SS) level between the RF-predicted and measured N_{CCN}, based on the AOPs presented in S2019 and all available AOPs.

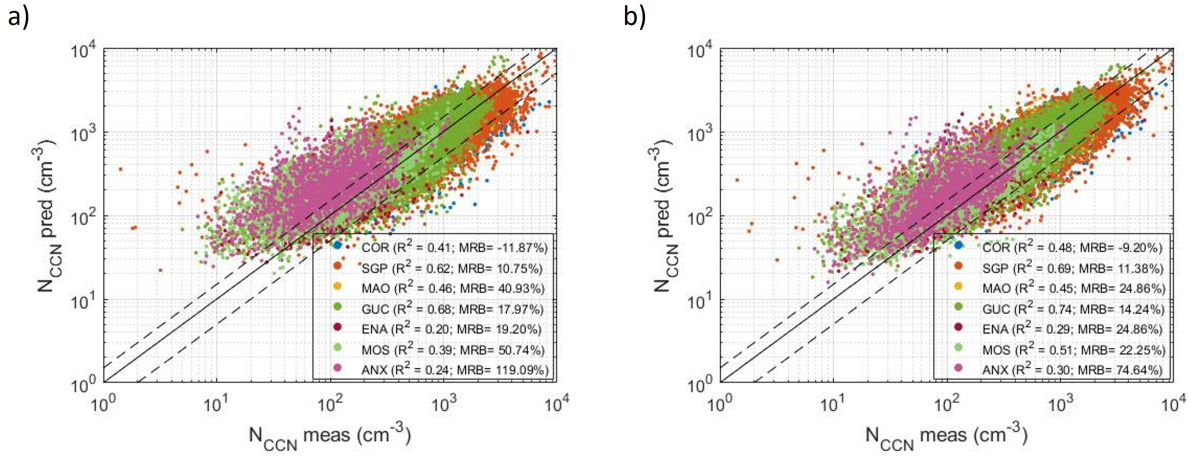

Figure S15. Log-log scatter plot of predicted CCN concentrations (N_{CCN} pred) versus observed concentrations (N_{CCN} meas) colored for different sites using a random forest model to estimate the parameters of the Twomey equation considering (a) AOPs used in S2019 (σ_{sp}, BSF and SAE) and (b) all AOPs (σ_{sp}, BSF, SAE, σ_{ap}, AAE, SSA). Coefficient of determination (R^2) and median relative bias (MRB) shown for each site. The solid black line represents the 1:1 line and the dashed lines are the +/-50%.

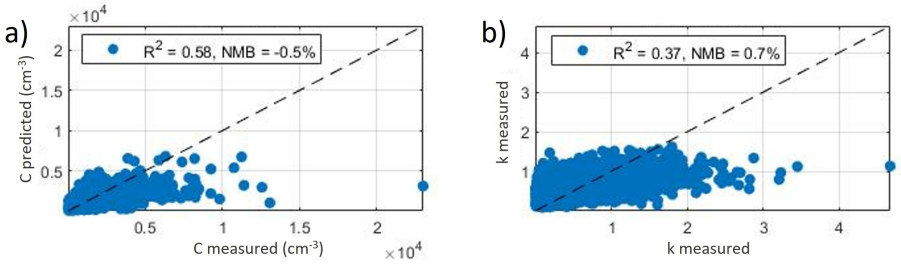

Figure S16. Comparison of observed versus Random Forest–predicted Twomey parameters: (a) parameter C and (b) parameter k. The model was trained using all AOPs (σ_{sp}, BSF, SAE, σ_{ap}, AAE, SSA). For each parameter, the coefficient of determination (R^2) and normalized mean bias (NMB) are reported. The dashed line represents the 1:1 line.

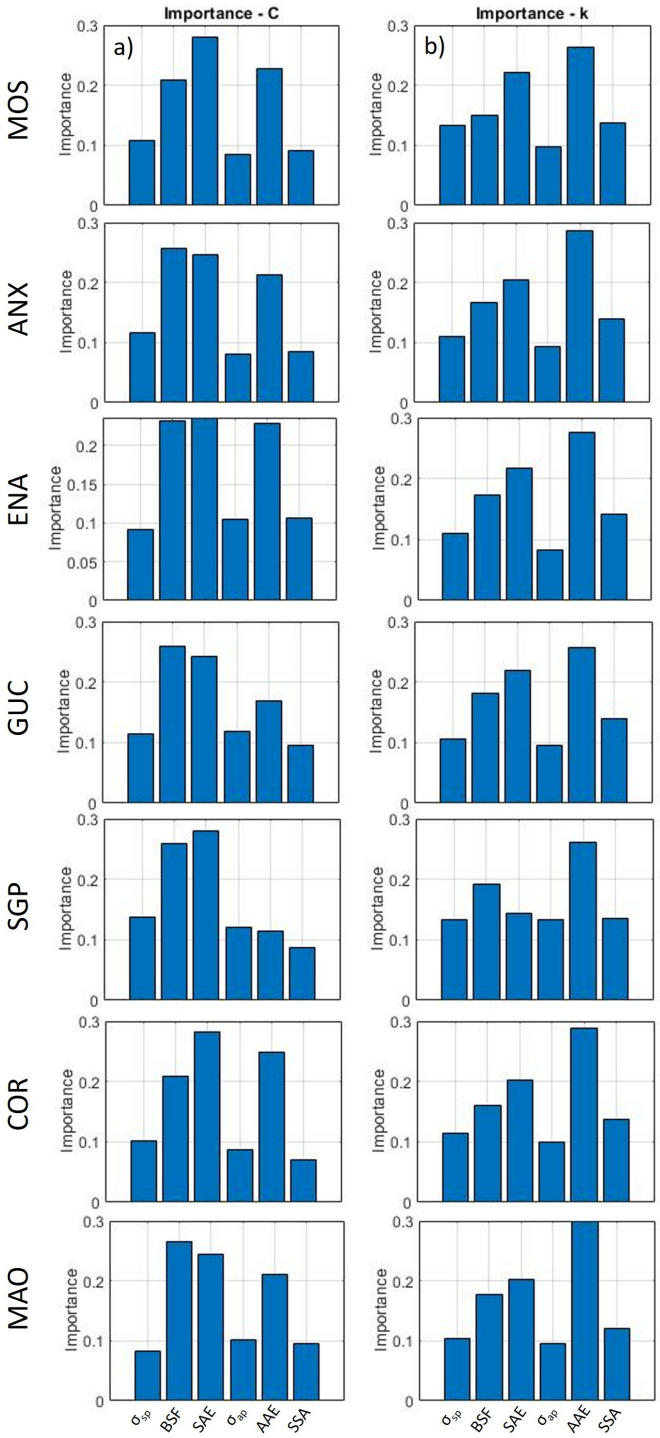

Figure S17. Importance of input variables in the Random Forest model considering all AOPs (σ_{sp}, BSF, SAE,σ_{ap}, AAE and SSA) for (a) C and (b) k parameters in LOSO cross-validation. Site name indicates the site excluded in RF model.

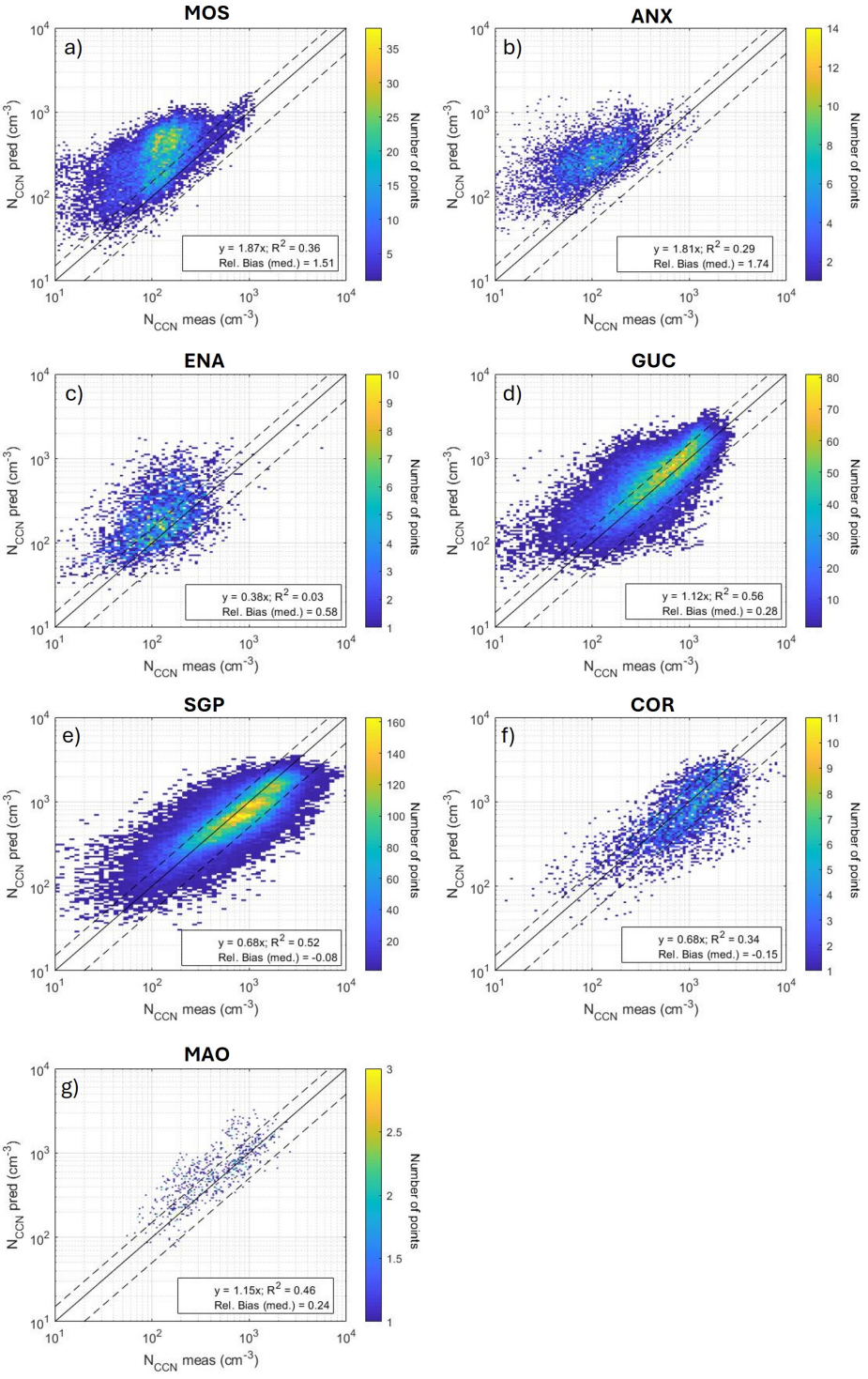

Figure S18. Log-log scatter plot of predicted (N_{CCN} pred) versus measured (N_{CCN} meas) CCN concentrations using a RF model to estimate the parameters of the Twomey equation. The solid black line represents the 1:1 line and the dashed lines are the +/-50%. Colored areas indicate the density of paired measurements, with color intensity representing the number of points within each log-spaced 2D bin (105×105 bins). Site name indicates the site excluded in RF model and used for testing.

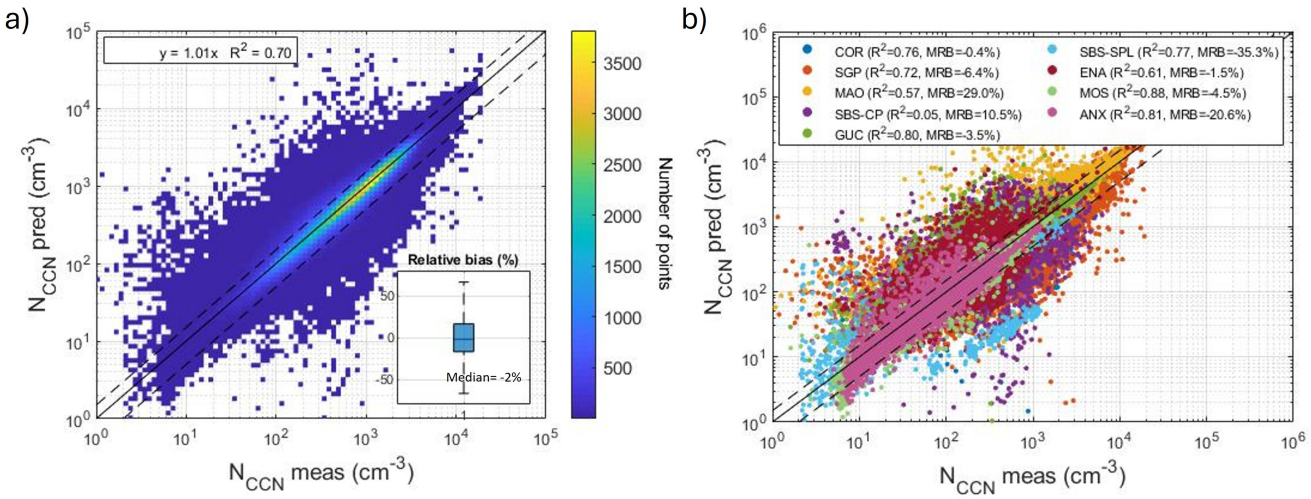

Figure S19. (a) Log–log scatter plot of predicted vs. observed CCN concentrations ($N_{CCN,meas}$) assuming a common D_{crit} for each SS. Colored areas indicate the density of paired measurements, with color intensity representing the number of points within each log-spaced 2D bin (105×105 bins). (b) Same as (a), but color-coded by site. The coefficient of determination (R^2) and median relative bias (MRB) are reported for all sites combined and individually. The solid black line represents the 1:1 line and the dashed lines are the +/-50%.