

Supplemental Materials

Shen methodology

Table S1. Slopes and offsets of linear regressions of R_{CCN}/σ 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 $\text{cm}^{-3} \text{Mm}$.

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

Derivation of equation A3

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

Considering linear regressions obtained from Figure A1:

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

$$a_1 \ln(\mathbf{SS}) + a_0 \approx a_1 \ln(\mathbf{SS}) + (2.41 \pm 0.13)a_1 \approx a_1 (\ln(\mathbf{SS}) + (2.41 \pm 0.13))$$

$$b_1 \ln(\mathbf{SS}) + b_0 \approx b_1 \ln(\mathbf{SS}) + (2.42 \pm 0.12)b_1 \approx b_1 (\ln(\mathbf{SS}) + (2.42 \pm 0.12))$$

$$\approx ((-0.095 \pm 0.011)a_1 + (5.7 \pm 11.0)) (\ln(\mathbf{SS}) + (2.42 \pm 0.12))$$

$$R_{\text{CCN}} \approx (a_1 \ln(\mathbf{SS}) + a_0)\mathbf{BSF} + b_1 \ln(\mathbf{SS}) + b_0$$

$$\approx a_1 (\ln(\mathbf{SS}) + (2.41 \pm 0.13))\mathbf{BSF} + ((-0.095 \pm 0.011)a_1 + (5.7 \pm 11.0)) (\ln(\mathbf{SS}) + (2.42 \pm 0.12))$$

$$\text{Approximation, since } (2.42 \pm 0.12) \approx (2.41 \pm 0.13)$$

$$R_{\text{CCN}} \approx a_1 (\ln(\mathbf{SS}) + (2.41 \pm 0.12))\mathbf{BSF} - (0.095 \pm 0.011)a_1 (\ln(\mathbf{SS}) + (2.41 \pm 0.13)) + (5.7 \pm 11.0) (\ln(\mathbf{SS}) + (2.41 \pm 0.13))$$

$$\approx a_1 (\ln(\mathbf{SS}) + (2.41 \pm 0.13)) (\mathbf{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0) (\ln(\mathbf{SS}) + (2.41 \pm 0.13))$$

$$\approx \ln(\mathbf{SS}) + (2.41 \pm 0.13) (a_1 (\mathbf{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$

$$\approx (\ln(\mathbf{SS}) - \ln(\exp(-2.41) \pm \exp(-2.41)(0.13))) (a_1 (\mathbf{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$

$$\approx \ln\left(\frac{\mathbf{SS}}{0.089 \pm 0.011}\right) (a_1 (\mathbf{BSF} - (0.095 \pm 0.011)) + (5.7 \pm 11.0))$$

Figures and tables

Table S2. List of measurement sites considered in this study and measurements available at each site. ¹MOSAIC 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. ²CP=Christy Peak. ³SPL=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	ASI	COR	ENA	GUC
Station information	Location (Campaign/Site name)	Andenes, Norway (COMBLE)	Ascension Island (LASIC)	Cordoba, Argentina (CACTI)	Azores, Portugal (Eastern North Atlantic)	Colorado, USA (SAIL)
	Coordinates	69.14 N 15.68 E, 2 masl	7.97 S 14.35 W, 341 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
	Site type	Polar/ Marine	Marine	Continental	Marine	Mountain/ Continental
	Site reference	Geerts et al. (2022)	Zuidema et al. (2016)	Fast et al. (2024)	Wood et al. (2015)	Feldman et al. (2023)
CCN measurements	Time period	Dec 2019- May 2020	Jun 2016- Oct 2017	Oct 2018- Apr 2019	Nov 2022- Dec 2023	Nov 2021- Jun 2023
	CCNC Model	DMT2C	DMT2C	DMT2C	DMT2C	DMT2C
	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, 0.6, 0.8, 1	0.1, 0.2, 0.4, 0.8, 1	0.1, 0.2, 0.4, 0.6, 0.8, 1
Particle number size distribution measurements	Time period	Dec 2019- Apr 2020	May 2016- Oct 2017	Sep 2018- Apr 2019	Nov 2022- Jan 2024	Oct 2021- Jun 2023
	SMPS Model	TSI 3938	TSI 3936	TSI 3936	TSI 3938	TSI 3938
	Size range (nm)	10/11-496/514	11/22-461	11-461	11-496	13-496/514
Chemical composition measurements	Time period	n/a	Jun 2016- Nov 2017	Sep 2018- May 2019	Nov 2022- Apr 2023	Apr 2022- Apr 2023
	ACSM Model	n/a	Q-ACSM	Q-ACSM	Q-ACSM	Q-ACSM
Optical properties measurements	Time period	Dec 2019- Apr 2020	n/a	Sep 2018- Apr 2019	Nov 2022- Dec 2023	Oct 2021- Jun 2023
	Neph Model	TSI 3563	n/a	TSI 3563	TSI 3563	TSI 3563
	PSAP Model	RR 3W	n/a	RR 3W	RR 3W	RR 3W

Campaign abbreviations: COMBLE = Cold-Air Outbreaks in the Marine Boundary Layer Experiment, LASIC = Layered Atlantic Smoke Interactions with Clouds, CACTI = Cloud Aerosol and Complex Terrain Interactions, SAIL = Surface Atmosphere Integrated field Laboratory.

Table S3. Continuation of Table S2

	Site ID	MAO	MOS	SBS-CP ²	SBS-SPL ³	SGP
Station information	Location (Campaign/ Site name)	Manacapuro, Brazil (GoAMAZON)	Arctic Ocean ¹ (MOSAIC)	Colorado, USA (STORMVEX)	Colorado, USA (STORMVEX)	Oklahoma, USA (Southern Great Plains)
	Coordinates	3.21 S 60.60 W, 50 masl	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	Continental/ Urban	Polar/ Marine	Mountain/ Continental	Mountain/ Continental	Continental
	Site reference	Martin et al. (2016)	Shupe et al. (2022)	Mace et al. (2010)	Mace et al. (2010)	Sheridan et al. (2001)
CCN measurements	Time period	Jan 2014- Nov 2015	Oct 2019- Oct 2020	Oct 2010- Apr 2011	Nov 2010- Apr 2011	Apr 2017- Aug 2023
	CCNC Model	DMTIC	DMT2C	DMTIC	DMTIC	DMT2C
	SS (%)	(1):0.2, 0.4, 0.6,0.8,1	0.1, 0.2, 0.4, 0.6, 0.8, 1	(1): 0.2, 0.4, 0.6, 0.8, 1	(1): 0.1,0.2, 0.4,0.6, 0.8,1	0.1, 0.2, 0.4, 0.8, 1
		(2): 0.1, 0.2, 0.4, 0.8		(2): 0.1, 0.2, 0.4, 0.8, 1	(2): 0.1, 0.2, 0.4,0.8,1	
Particle number size distribution measurements	Time period	Jan 2014- May 2015	Oct 2019- Oct 2020	Oct 2010- May 2011	Dec 2010- Mar 2011	Jan 2017- Dec 2023
	SMPS Model	TSI 3936	TSI 3938	TSI 3936	TSI 3936	TSI 3938
	Size range (nm)	10/13-461	10-496/514	9-334	9-334	10/11-461/514
Chemical composition measurements	Time period	n/a	n/a	n/a	n/a	Jan 2017- Apr 2023
	ACSM Model	n/a	n/a	n/a	n/a	Q-ACSM
Optical properties measurements	Time period	Jan 2014- May 2015	Oct 2019- Oct 2020	n/a	n/a	Jan 2017- Dec 2023
	Neph Model PSAP Model	n/a	TSI 3563 RR 3W	n/a	n/a	TSI 3563 RR 3W

Campaign abbreviations: GoAMAZON = Observations and Modeling of the Green Ocean Amazon, MOSAiC = Multidisciplinary drifting Observatory for the Study of Arctic Climate, STORMVEX = Storm Peak Lab Cloud Property Validation Experiment.

Table S4. 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 ($\kappa_{\text{BC}} = 0$) (Schmale et al., 2018; Deng et al., 2019). Values for NH_4NO_3 , $(\text{NH}_4)_2\text{SO}_4$ and $(\text{NH}_4)\text{HSO}_4$ 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	$(\text{NH}_4)_2\text{SO}_4$	$(\text{NH}_4)\text{HSO}_4$	H_2SO_4	NaCl	BC	<i>Organics</i>
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 S5. 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	ASI	MOS	ANX
Chemistry	Scheme 1	1199	15498	–	607	–	–	–	–
	Scheme 2	1237	17318	–	8318	618	8558	–	–
	Scheme 3	1467	17999	–	9626	800	10703	–	–
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

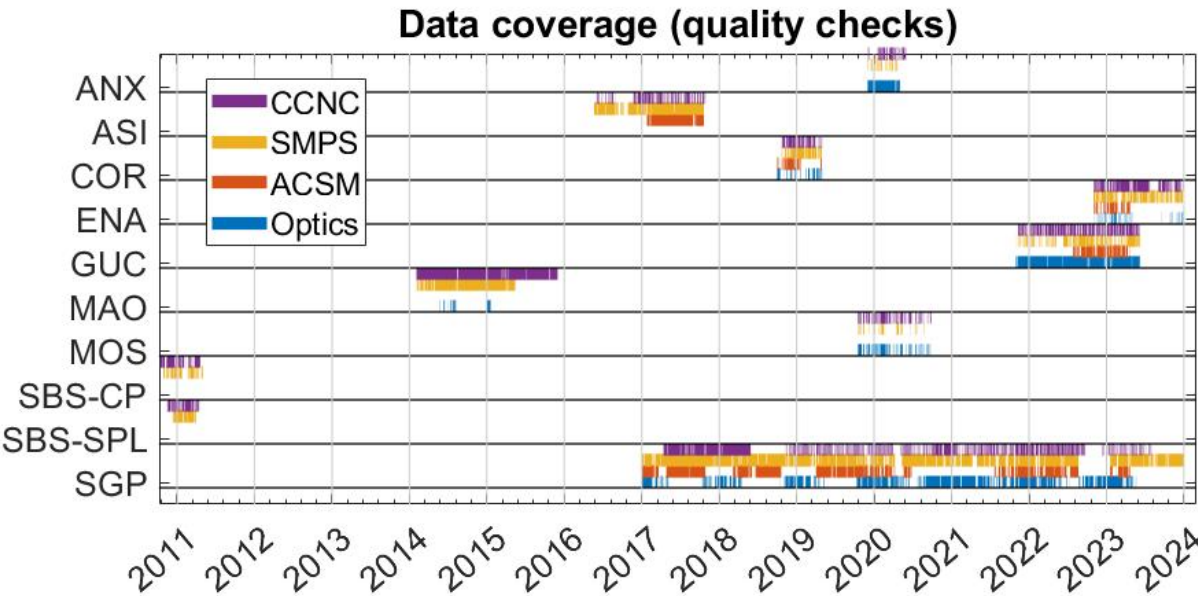


Figure S1. Data coverage of the instruments used in this study for all the sites. Total particle concentration comparison and CCN counter column comparison quality checks applied.

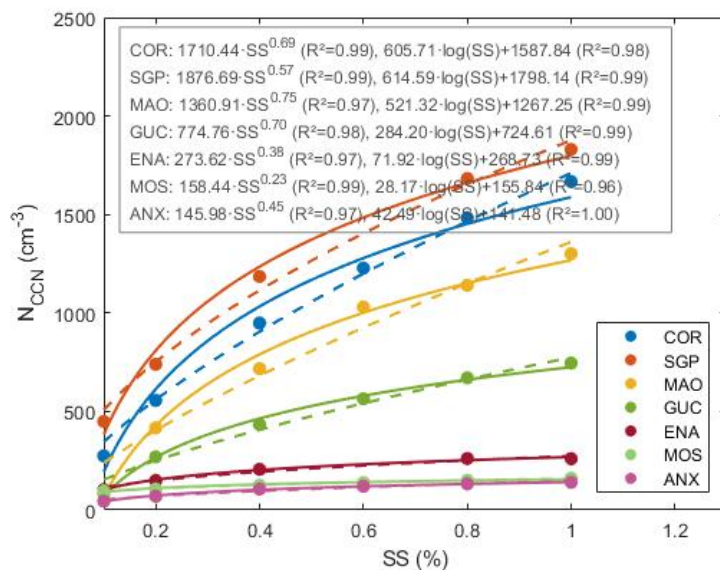


Figure S2. N_{CCN} measured average values at different SS. The Twomey equation power function is represented by the solid lines for each site. A logarithmic fit to the same data is represented by the dashed lines. The fit equations for both types of functions are provided in the legend for each site.

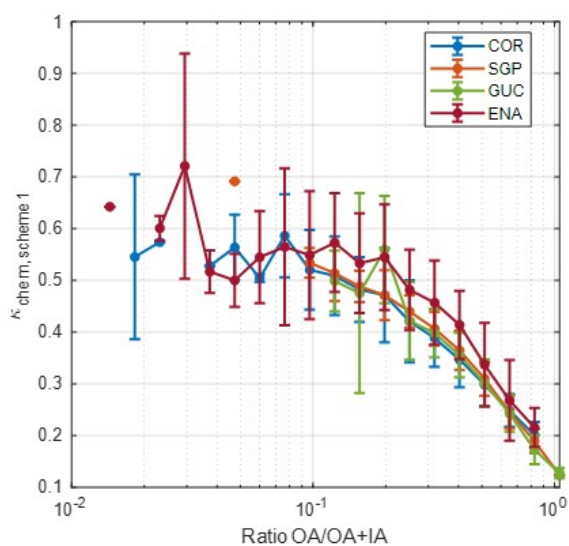


Figure S3. Relationship of the composition-derived κ_{chem} from Scheme 1 to the binned and averaged ratio of organic (OA) to total (OA+IA) aerosol components. The vertical bars denote the standard deviation.

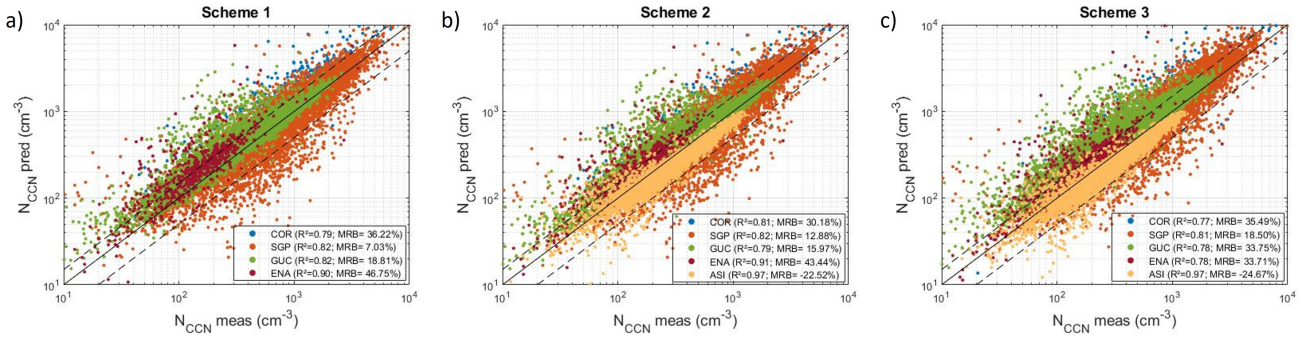


Figure S4. Log-log scatter plot of predicted CCN concentrations ($N_{CCN, pred}$) with respect to the observed CCN concentrations ($N_{CCN, meas}$) at all SS colored for different sites using the three prediction schemes. (a) Scheme 1 ($\kappa_{chem, Sch1}$), (b) Scheme 2 ($\kappa_{chem, Sch2}$) and (c) Scheme 3 (fixed κ). The solid black line represents the 1:1 line and the dashed lines are the $\pm 50\%$. Coefficient of determination (R^2) and median relative bias (MRB) shown for each site. ASI not included in Scheme 1 due to the lack of BC.

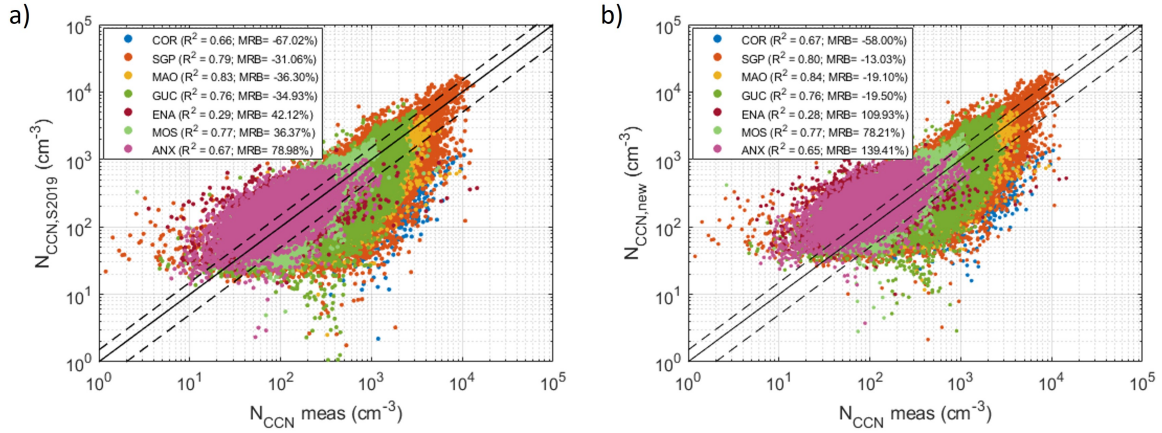


Figure S5. Log-log scatter plot of predicted CCN concentrations with respect to the observed CCN concentrations ($N_{CCN, meas}$) colored for different sites considering (a) equation in S2019 ($N_{CCN, S2019}$; based on 6 sites) and (b) new equation ($N_{CCN, new}$; based on 13 sites). Coefficient of 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 $\pm 50\%$.

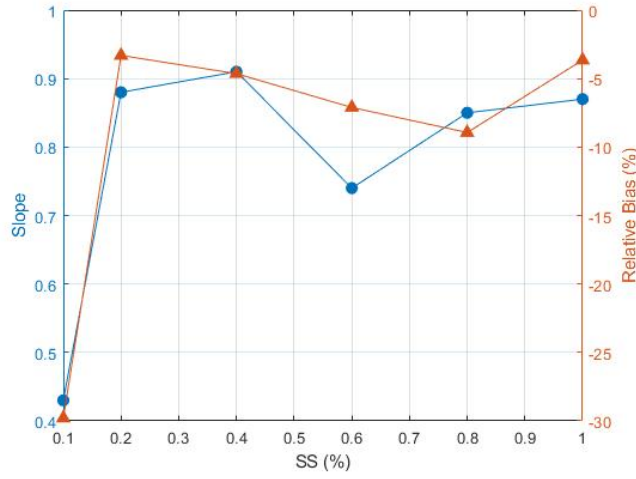


Figure S6. Slope and median relative bias (%) for each supersaturation (SS) level between the predicted and measured N_{CCN} , based on the new equation proposed in this study.

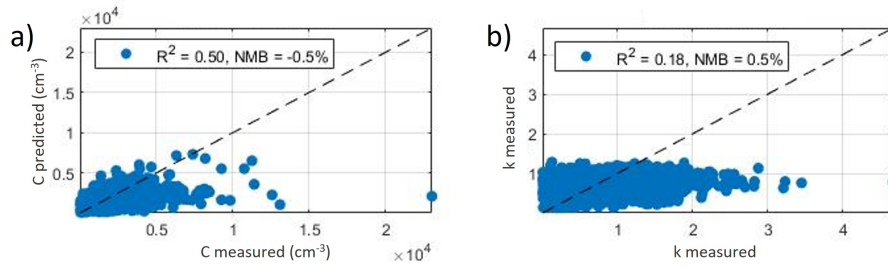


Figure S7. Comparison of observed versus Random Forest–predicted Twomey parameters: (a) parameter C and (b) parameter k . The model was trained using the AOPs used in S2019 (σ_{sp} , BSF, and SAE). For each parameter, the coefficient of determination (R^2) and normalized mean bias (NMB) are reported. The dashed line represents the 1:1 line.

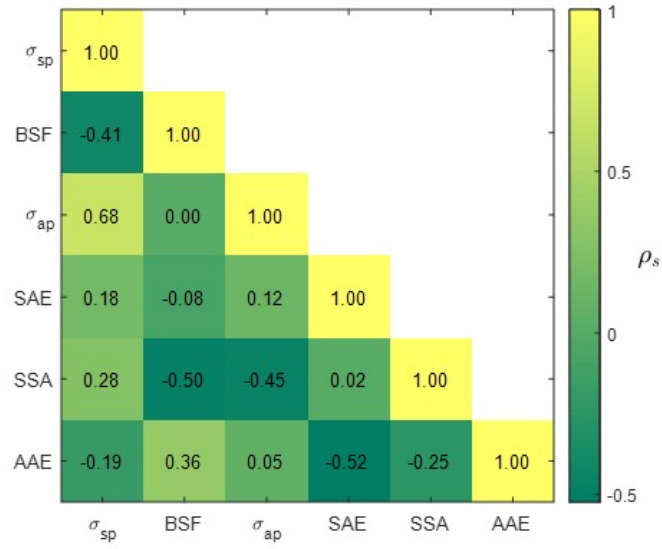


Figure S8. Spearman correlation coefficients (ρ_s) between all the AOPs.

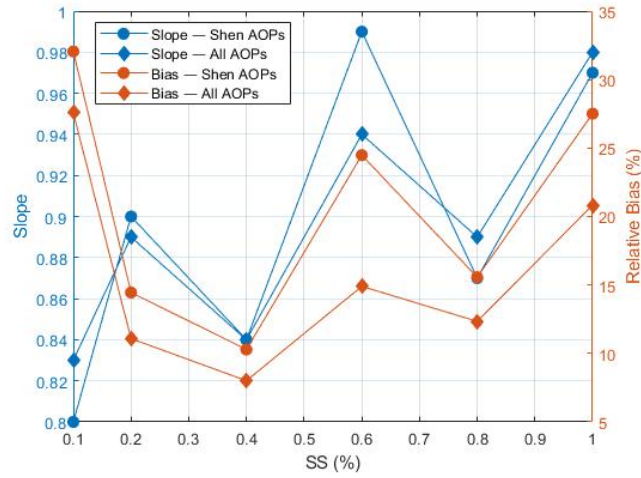


Figure S9. Slope and median relative bias (%) for each supersaturation (SS) level between the RF-predicted and measured N_{CCN} , based on the AOPs presented in S2019 and all available AOPs.

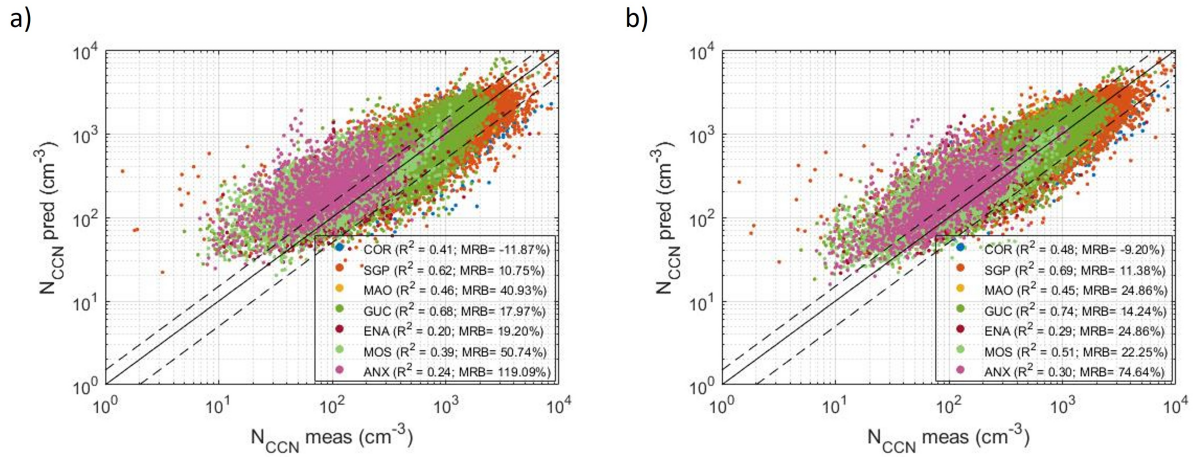


Figure S10. Log-log scatter plot of predicted CCN concentrations (N_{CCN} pred) versus observed concentrations (N_{CCN} meas) colored for different sites using a random forest model to estimate the parameters of the Twomey equation considering (a) AOPs used in S2019 (σ_{sp} , BSF and SAE) and (b) all 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 $\pm 50\%$.

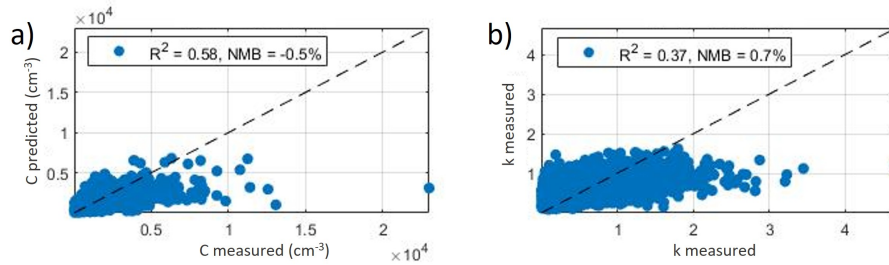


Figure S11. Comparison of observed versus Random Forest-predicted Twomey parameters: (a) parameter C and (b) parameter k . The model was trained using all AOPs (σ_{sp} , BSF, SAE, σ_{ap} , AAE, SSA). For each parameter, the coefficient of determination (R^2) and normalized mean bias (NMB) are reported. The dashed line represents the 1:1 line.

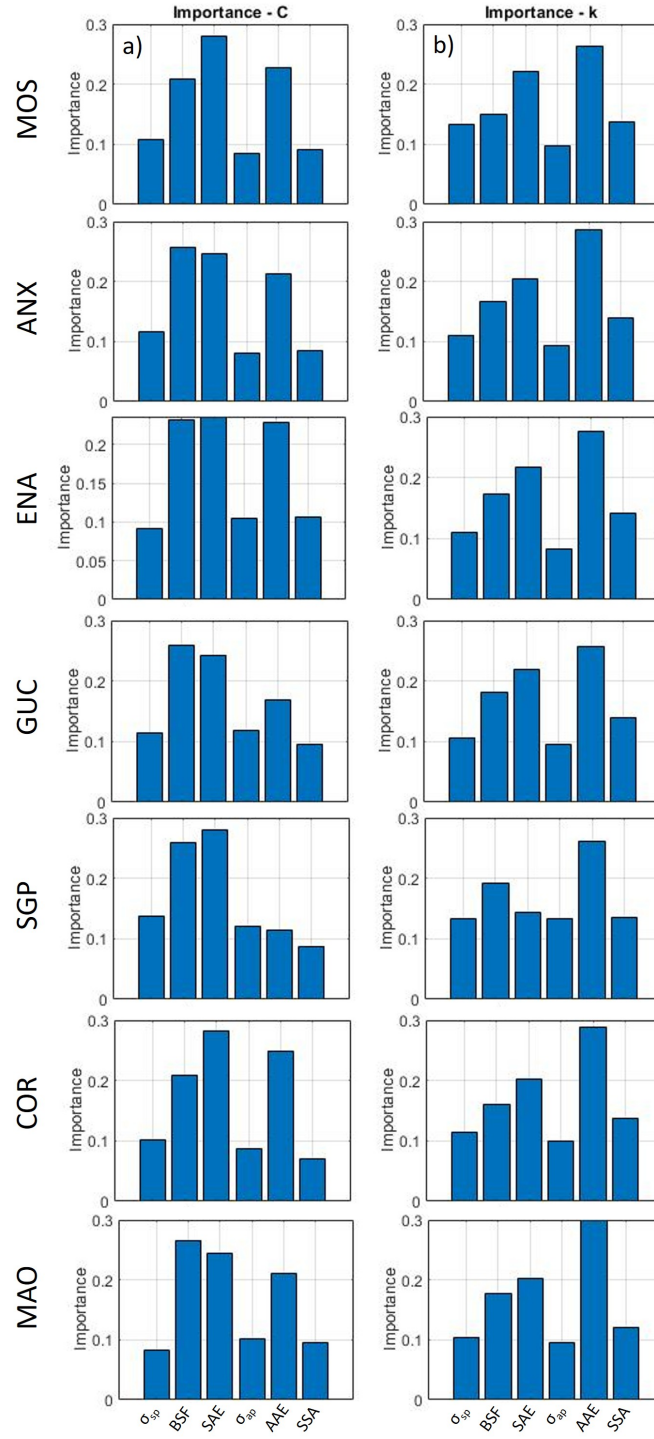


Figure S12. Importance of input variables in the Random Forest model considering all AOPs (σ_{sp} , BSF, SAE, σ_{ap} , AAE and SSA) for (a) C and (b) k parameters in LO SO cross-validation. Site name indicates the site excluded in RF model.

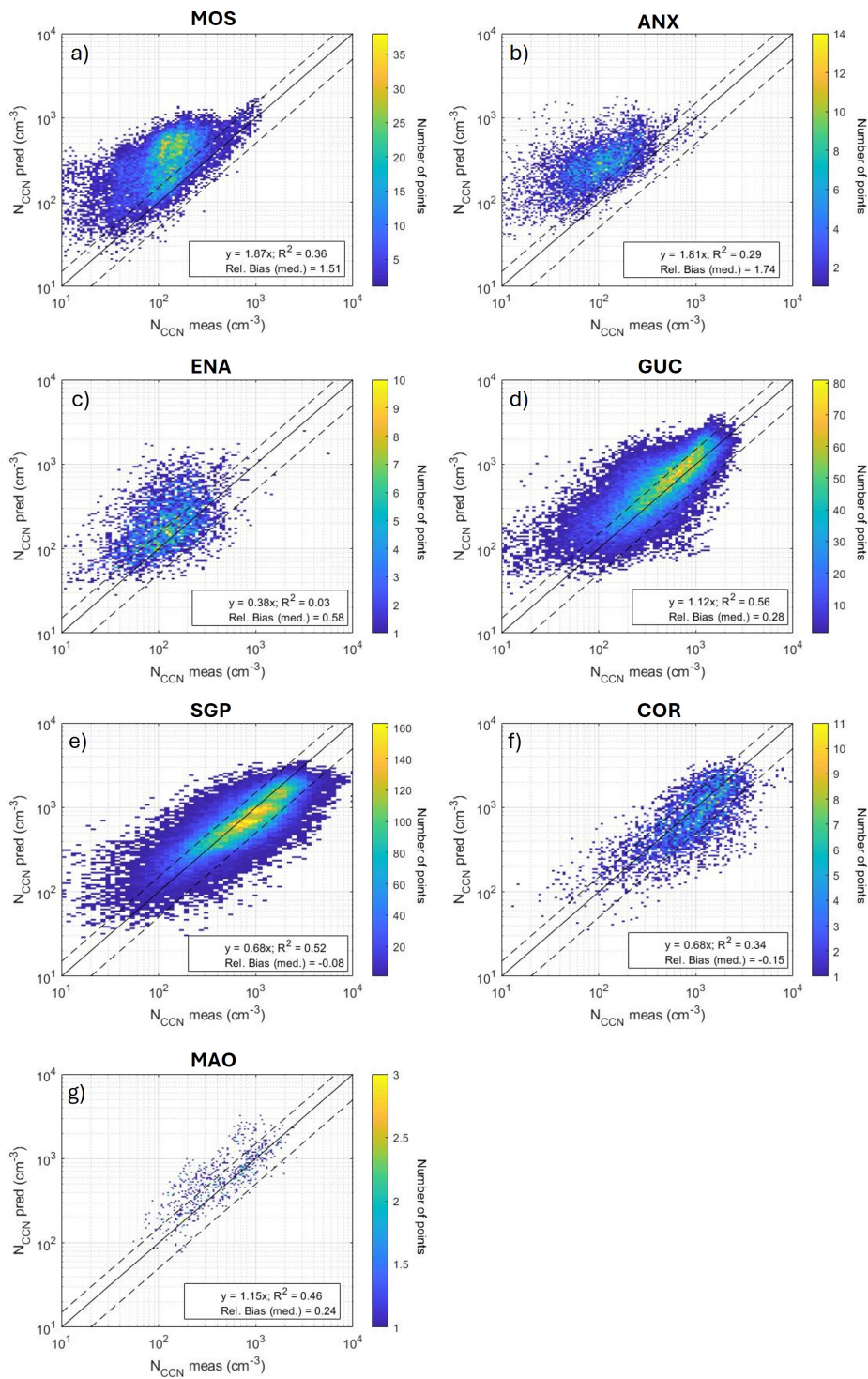


Figure S13. Log-log scatter plot of predicted CCN concentrations (N_{CCN} pred) versus observed concentrations (N_{CCN} pred) using a RF model to estimate the parameters of the Twomey equation. The solid black line represents the 1:1 line and the dashed lines are the $\pm 50\%$. Site name indicates the site excluded in RF model and used for testing.

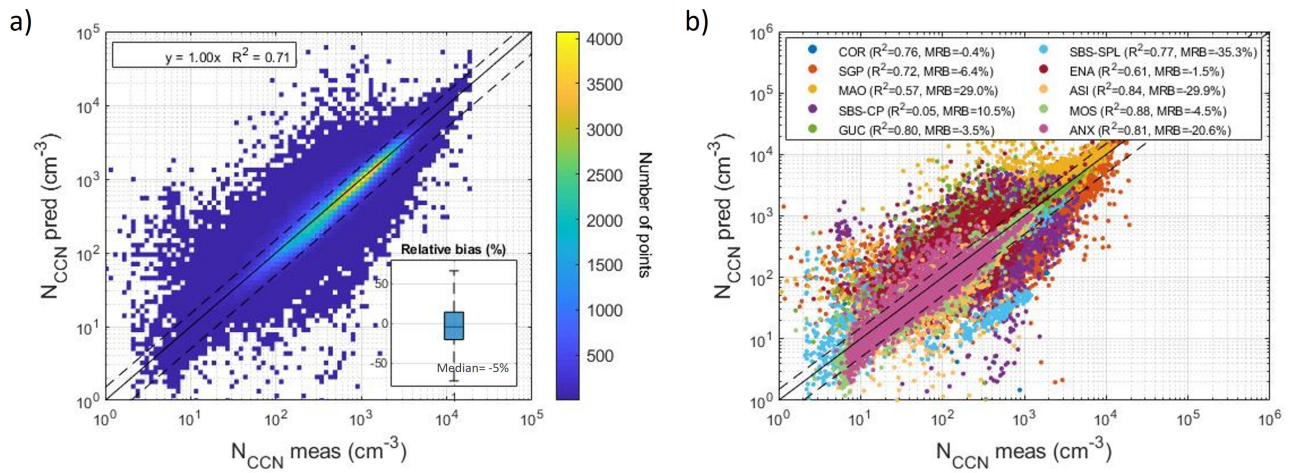


Figure S14. (a) Log-log scatter plot of predicted vs. observed CCN concentrations ($N_{CCN, meas}$), assuming a common D_{crit} for each SS. (b) Same as (a), but color-coded by site. The coefficient of determination (R^2) and median relative bias (MRB) are reported for all sites combined and individually. The solid black line represents the 1:1 line and the dashed lines are the $\pm 50\%$.