Supplementary information for

Machine Learning Assisted Inference of the Particle Charge Fraction and the Ion-induced Nucleation Rates during New Particle Formation Events

Pan Wang[1], Yue Zhao[1], Jiandong Wang[2,3], Veli-Matti Kerminen[4], Jingkun Jiang[5], Chenxi Li[1*]

[1]School of Environmental Science and Engineering, Shanghai Jiao Tong University, 200240, Shanghai, China

[2]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China

[3]China Meteorological Administration Aerosol-Cloud-Precipitation Key Laboratory, School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, China

[4]Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, 00014 Helsinki, Finland

[5]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, 100084 Beijing, China

Correspondence to: Chenxi Li (chenxi20@sjtu.edu.cn)

S1. Expressions for $P_{i,k}$ and $L_{i,k}$ in Eq. (9)

The expression for the mass production rate $P_{i,k}$ and loss rate $L_{i,k}$ are given as follows:

$$P_{i,k} = \sum_{a=1}^{NS} \sum_{b=1}^{NS} \sum_{x+y=k} \beta(a,b,x,y) N_{a,x} N_{b,y} f(i,a,b,x,y)(m_a + m_b) \quad \text{(S1)}$$

$$L_{i,k} = \sum_{a=1}^{NS} \sum_{x=-Q}^{Q} \beta(a,i,x,k) N_{a,x} \quad \text{(S2)}$$

The subscript *i*, *k* are the mass bin number and charge state, respectively, NS is the number of mass sections, $\beta(a,b,x,y)$ is the collision rate coefficient between collision rate coefficients between bin (a,x) and bin (b,y) (a,b are mass bin numbers and x, y are charge bin numbers), $N_{a,x}$ and $N_{b,y}$ are the particle number concentration in bin (a,x) and (b,y), m_a and m_b are the average particle mass in mass bin a and mass bin b, respectively, Q is maximum number of charges on the particle. The implementation of Eqs. (S1) and (S2) does conserve particle mass during coagulation (Matsui, 2017) . Therefore, a correction factor is multiplied to $P_{i,k}$, which is expressed as

$$f_{corr} = \frac{\sum_{i=1}^{NS} \sum_{k=-Q}^{Q} \frac{P_{i,k}}{1+dt \cdot L_{i,k}}}{\sum_{i=1}^{NS} \sum_{k=-Q}^{Q} \frac{M_{i,k} \cdot L_{i,k}}{1+dt \cdot L_{i,k}}} \quad \text{(S3)}$$

where $M_{i,k}$ is mass concentration in bin (i, k) and dt is the time step for coagulation (set to 2 s in the model).

S2. The evaluation of GR

The growth of particles mainly originates from the condensation of vapors, including SA and OOMs, when the particle number concentrations are low. In the CDMS-ion we adjusted the particle growth rate by controlling the concentration of condensing vapors in seven volatility bins by multiplying their default concentration by a scaling factor. The relative abundance of these vapors were fixed and was based on the vapor concentration observed during a typical event observed in urban Beijing, China (Fig. S1a). To extract particle growth rate, we simulated the growth of an individual particle exposed under a given vapor concentration for a few hours. As shown in Fig. S1b, the particle size increases almost linearly with time, and the particle growth rate by fitting the the particle size with a first order fitting curve.

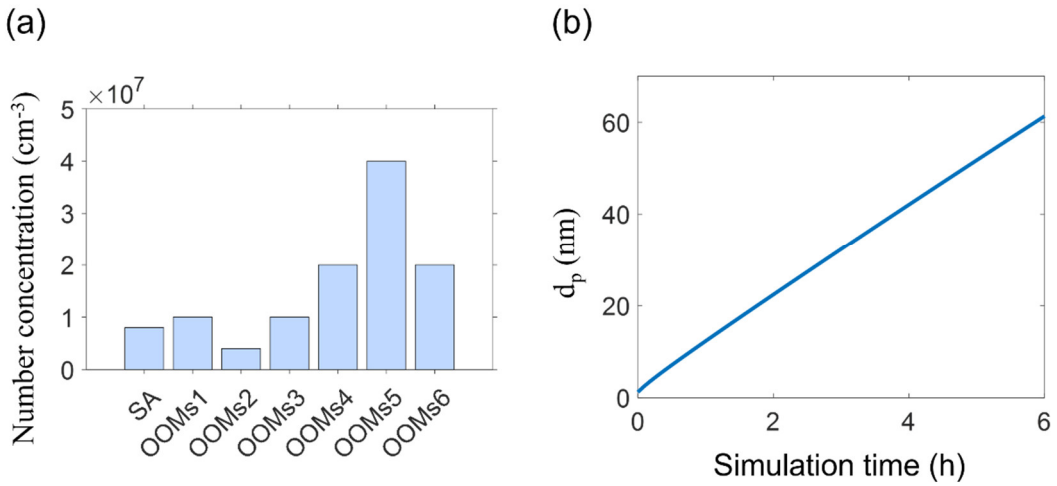

Figure S1. (a) The relative abundance of sulfuric acid and oxygenated organic molecules. The organic species are classified into 6 bins (OOMs1-6) with $\log_{10}C^*$= -9, -7, -5, -3, -1,0 (C^* is the saturation concentration in unit of μg m^{-3}). (b) Simulated particle size as a function of time.

S3. Analytical equations for τ_{ss}

At particle sizes relevant for atmospheric new particles (1-100 nm), most of particles are neutral or singly charged. Therefore, here we only consider singly charged and neutral particles. To be consistent with the main text, we also assume that the positive and negative ions have the same mobility and concentrations. For this simplified system, the following equations describe the dynamic charging process of a monodisperse particles,

$$\frac{dN_0}{dt} = -\beta_{0,+}n_+N_0 - \beta_{0,-}n_-N_0 + \beta_{+1,-}n_-N_{+1} + \beta_{-1,+}n_+N_{-1} \quad \text{(S4)}$$

$$\frac{dN_{+1}}{dt} = \beta_{0,+}n_+N_0 - \beta_{+1,-}n_-N_{+1} \quad \text{(S5)}$$

$$\frac{dN_{-1}}{dt} = \beta_{0,-}n_-N_0 - \beta_{-1,+}n_+N_{-1} \quad \text{(S6)}$$

where N_0, N_{+1}, N_{-1} represent the concentration of particles with zero, +1 and -1 charge, $n_{+/-}$ is the concentration of positive/negative ions, $\beta_{n,+/-}$ is the collision rates between particles with n charges and positive/negative ions. Because the system is symmetrical with respect to polarity, we have $N_{+1} = N_{-1} = N_1$, $n_+ = n_- = n_i$, $\beta_{0,+} = \beta_{0,-} = \beta_0$, $\beta_{+1,-} = \beta_{-1,+} = \beta_1$. This leads to

$$\frac{dN_0}{dt} = -2\beta_0 n_i N_0 + 2\beta_1 n_i N_1 \quad \text{(S7)}$$

$$\frac{dN_1}{dt} = \beta_0 n_i N_0 - \beta_1 n_i N_1 \quad \text{(S8)}$$

During the charging process, the total number of particles is conserved, i.e., $N_t = N_0 + 2N_1$. Substituting this relation to Eq. S8 leads to

$$\frac{dN_1}{dt} = \beta_0 n_i (N_t - 2N_1) - \beta_1 n_i N_1 \tag{S9}$$

The solution of Eq. S9 is

$$N_1(t) = \frac{\beta_0 N_t}{2\beta_0 + \beta_1}(1 - \exp(-(2\beta_0 + \beta_1)n_i t)) + N_1(0)\exp(-(2\beta_0 + \beta_1)n_i t)) \tag{S10}$$

According to Eq. S10, the steady-state N_1 value is $N_{1,ss} = \frac{\beta_0 N_t}{2\beta_0 + \beta_1}$. Rearranging Eq. S10 leads to

$$\begin{cases} 1 - \frac{N_1(t)}{N_{1,ss}} = \exp(-(2\beta_0 + \beta_1)n_i t), & \text{if } N_1(0) = 0 \\ \frac{N_1(t)}{N_{1,ss}} - 1 = \frac{\beta_1}{2\beta_0}\exp(-(2\beta_0 + \beta_1)n_i t), & if\ N_1(0) = \frac{N_t}{2} \end{cases} \tag{S11a, b}$$

Equations S11a and S11b correspond to initially neutral and charged particles, respectively. The left hand side of Eq. S11 are the difference between $\frac{N_1(t)}{N_{1,ss}}$ and unity. According to our definition (see main text) of the characteristic time τ_{ss}, setting the left hand side of Eq. S11 to 1/e leads to the expression for τ_{ss}:

$$\tau_{ss} = \begin{cases} \frac{1}{(2\beta_0 + \beta_1)n_i}, & \text{if } N_1(0) = 0 \\ \frac{1 + \ln\left(\frac{\beta_1}{2\beta_0}\right)}{(2\beta_0 + \beta_1)n_i}, & if\ N_1(0) = \frac{N_t}{2} \end{cases} \tag{S12a, b}$$

Note that β_1 is considerably larger than $2\beta_0$ for small particles due to Coulombic attraction and $\ln\left(\frac{\beta_1}{2\beta_0}\right)$ is positive. Therefore, τ_{ss} for initially charged particles are larger than initially neutral particles, which explains the difference between Fig. 2a and 2b in the main text.

S4. Parametrization of d_b

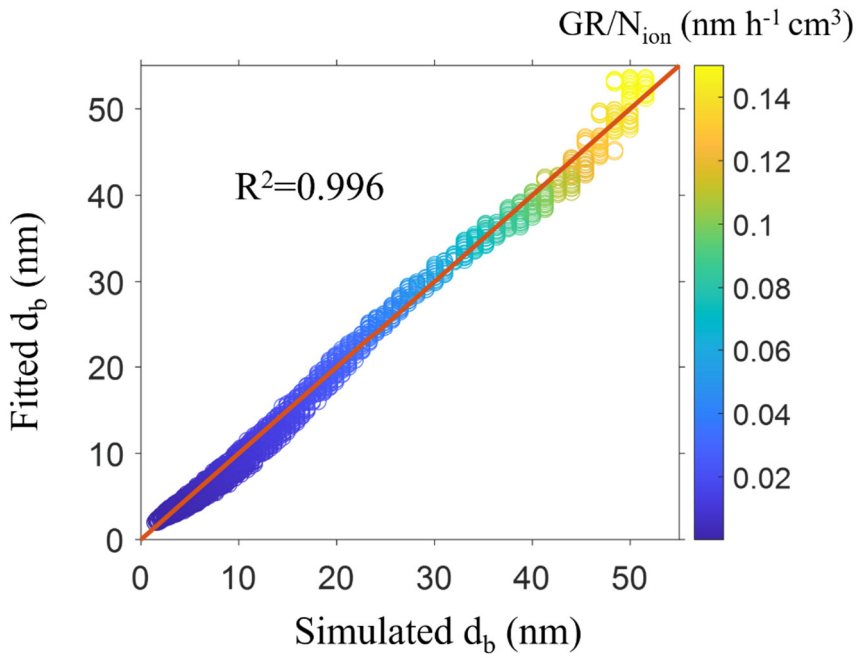

Figure S2. A comparison between the near-steady-state size (d_b) from the simulation and the correlation of Eq. S13.

As shown in section 3.1.2, during neutral NPF the particle charge fraction gradually approach the steady state value (i.e. r_c gradually increases to unity). To determine size boundary d_b above which the charge fraction of the new particle approximately reaches their steady state value , we obtained d_b from the simulations and parameterized its value as a function of particle growth rate (GR, in nm/h), ambient ion concentration (N_{ion}, in #/cm^3), the coagulation sink (CoagS, in s^{-1}) and the nucleation rate (J, in # cm^{-3} s^{-1}) . Here d_b is defined as the size at which the singly charged fraction of new particles reach 63%-136% (i.e., 1-1/e - 1+1/e) of the steady state value. In other words, we have treated a relative error of 1/e in particle charging fraction as acceptable in SMPS measurements.

The parametrized function is expressed as

$$d_b = \left(a_0 + a_1 \frac{GR}{N_{ion}} + a_2 \left(\frac{GR}{N_{ion}}\right)^2 + a_3 \left(\frac{GR}{N_{ion}}\right)^3\right) CoagS^b J^c \quad (S13)$$

where $a_0 = 1.41$, $a_1 = 889.53$, $a_2 = -10210.73$, $a_3 = 52421.03$, $b = -0.0025$, $c = 0.0077$. Figure S2 shows the comparison of Eq. S13 and d_b retrieved from the simulations. The simulated and the parametrized values are in good agreement, with an R^2 value of 0.996 and a maximum deviation of 40.1%. In Eq. S13, the dominant contributing variable is GR/N_{ion}, while the dependence of d_b on

CoagS and J is weak, in agreement the trend shown in Figure 3. At typical conditions in urban Bejing with GR= 3-4 nm h^{-1} and N_{ion}=122-224 cm^{-3}, d_b is 12.2-22.5 nm. This value is consistent with the observation by Li et al. (2022), who have shown that SMPS yields similar particle size distributions above the size of 13 nm, whether X-ray sources or atmospheric ions are used as a neutralizer for ambient particles.

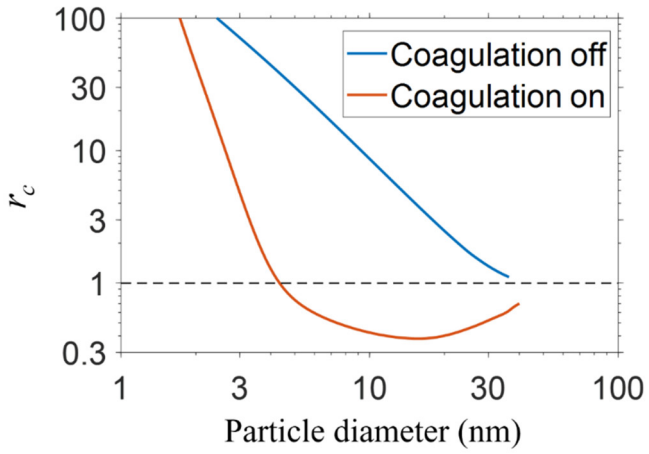

Figure S3. The simulated r_c with coagulation on and off in the simulations. The simulation conditions are as follows: F_{IIN} = 1, N_{ion} = 50 cm^{-3}, CoagS = 0.005 s^{-1}, GR = 4 nm h^{-1}, J = 100 cm^{-3} s^{-1},

S5. Particle number concentration and mode size

Figures S4a and S4b illustrate the effect of particle charging on N_{max} (the maximum particle number concentration during an NPF event) for neutral nucleation and IIN at CoagS = 0.005 s^{-1}, J = 5 cm^{-3} s^{-1}, respectively. The colors in the heatmap represent the ratio $r_N = \frac{N_{max,with\ charging}}{N_{max,without\ charging}}$. For neutral nucleation (F_{IIN}=0), r_N is less than one by only a few percent (Fig. S4a). The decrease of r_N to below 1 arises because charged particles interact with pre-existing particle through Coulombic force, resulting in an elevated coagulation sink (see Fig. S6). In the case of IIN with F_{IIN}=100%, the reduction in r_N is more pronounced (Fig. S4b). This is due to both the higher CoagS for charged particles and the enhanced coagulation between oppositely charged new particles. The maximum decrease in N_{max} is about 25%, which occurs at low N_{ion}: in this situation the newly formed particles are only slowly neutralized by the atmospheric ions and charge effect of higher CoagS and enhanced coagulation persists. As shown by Fig. S5b, when the NPF rate is higher (J = 50 cm^{-3} s^{-1}), r_N further decreases due to stronger coagulation between particles.

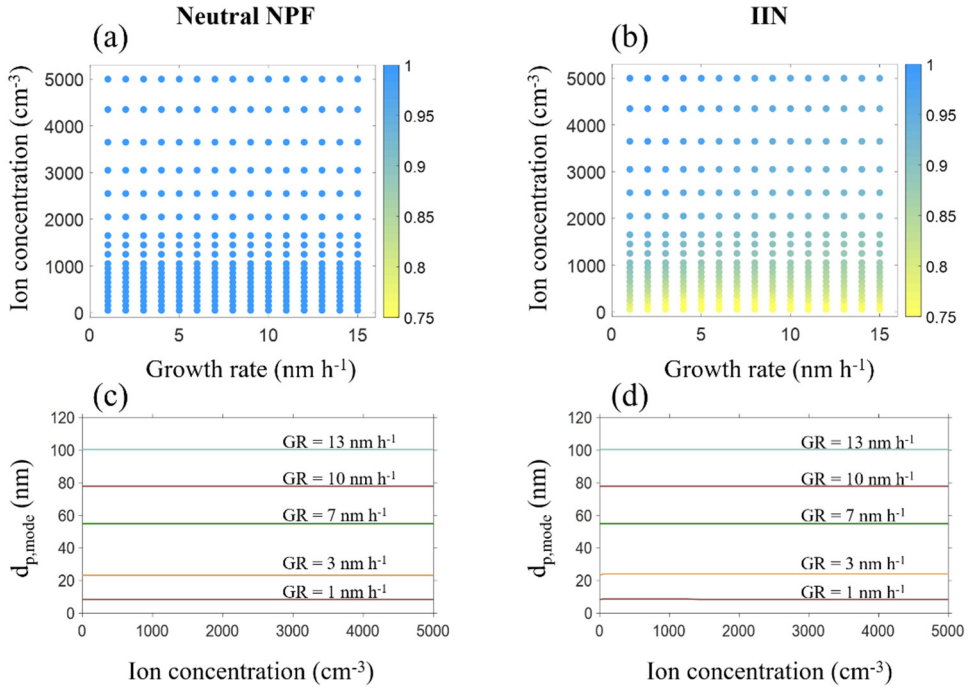

Figure S4. Comparison of the maximum particle concentration and particle mode diameter between simulations with and without considering the particle charging. (a)-(b) the ratio of the maximum particle number concentration N_{max}. The magnitude of this ratio is represented by the colormap for different GR and N_{ion} values. (c)-(d) the mode diameter $d_{p,mode}$ as a function of N_{ion} for different GR

values. N_{ion} =0 corresponds to the case in which charging is not included in the simulation. The J and CoagS are set to 5 cm^{-3} s^{-1} and 0.005 s^{-1} in these simulations.

Figures S4c and S4d compare the simulated mode diameter $d_{p,mode}$ at a time of 8 hours after NPF onset across several GR values. Notably, including charging in the simulation does not significantly alter the particle mode diameter: the change of $d_{p,mode}$ is smaller than size bin resolution used in the model. This observation still stands when the NPF rate J is increased from 5 cm^{-3} s^{-1} to 50 cm^{-3} s^{-1} (Fig. S5c-d).

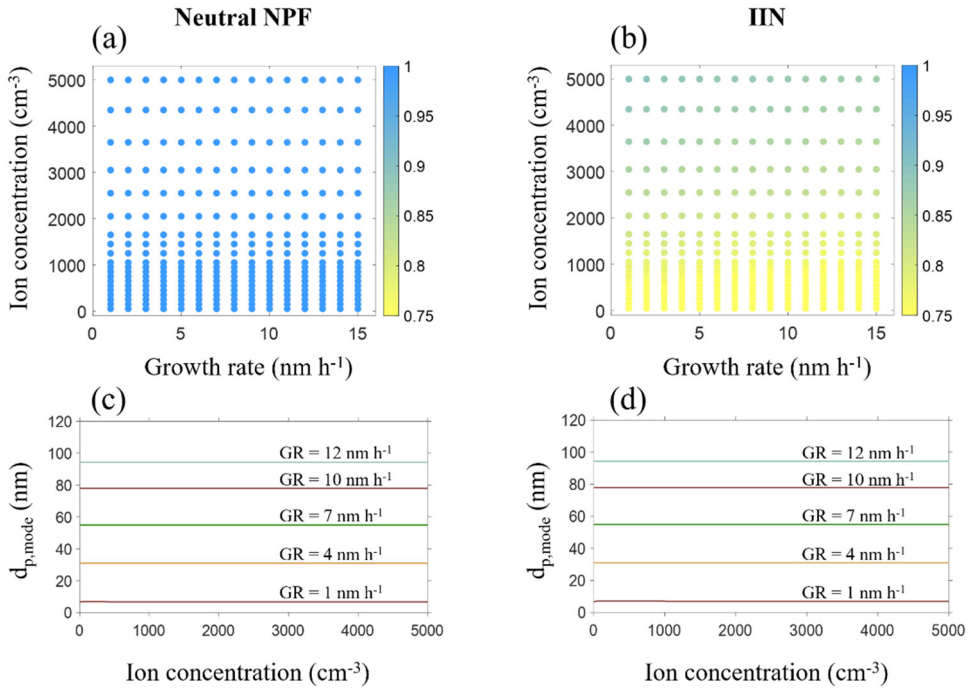

Figure S5. Comparison of the maximum particle concentration and particle mode diameter between simulations with and without considering the particle charging. The J and CoagS are respectively set to 50 cm^{-3} s^{-1} and 0.005 s^{-1} in these simulations. (a)-(b) the ratio of the maximum particle number concentration N_{max}. The magnitude of this ratio is represented by the colormap at different GR and N_{ion} values. (c)-(d) the mode diameter $d_{p,mode}$ at t = 8 h as a function of N_{ion} for different GR values. N_{ion} =0 corresponds to the case in which particle charging is not included in the simulation.

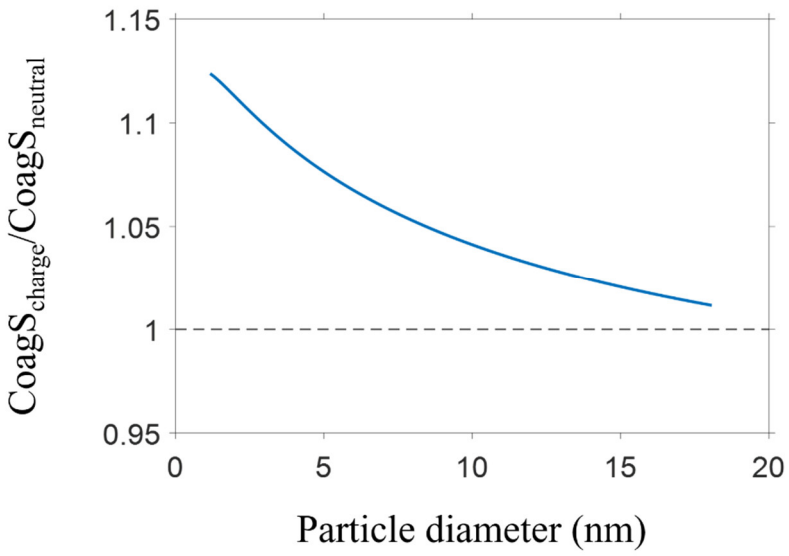

Figure S6. The ratio of the CoagS for singly charged particles and neutral particles as a function of particle size. In this calculation the pre-existing particles (which serve as the coagulation sink for newly formed particles) are lognormally distributed with a geometric mean diameter of 100 nm and a geometric standard deviation of 1.4. We also assume that the charge state of the pre-existing particles have reached steady state. Coulomb interactions between particles are considered in the calculation (Chahl and Gopalakrishnan, 2019; Gopalakrishnan and Hogan, 2011), but charge-dipole interactions are not.

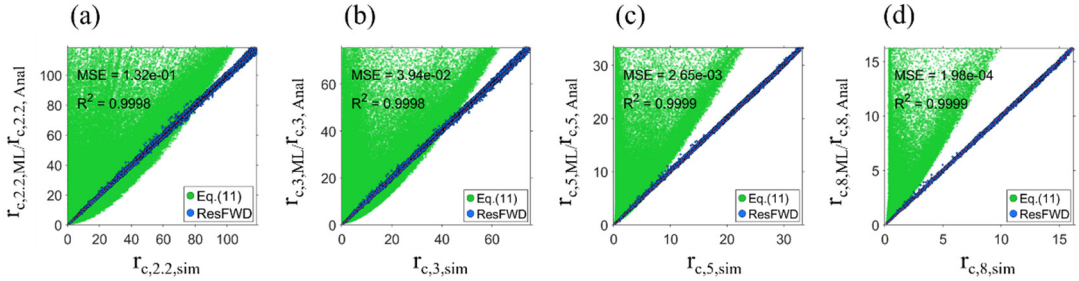

Figure S7. Comparison between the simulated r_c ($r_{c,sim}$) , the ResFWD-predicted r_c ($r_{c,ML}$) and the r_c calculated with Eq. (11) ($r_{c,Anal}$) at particle diameters of 2.2 nm, 3nm, 5 nm and 8 nm. The numbers in the subscript of r_c denote the particle size. The R^2 and MSE obtained from testing the ResFWD model against $r_{c,sim}$ are shown in the panels.

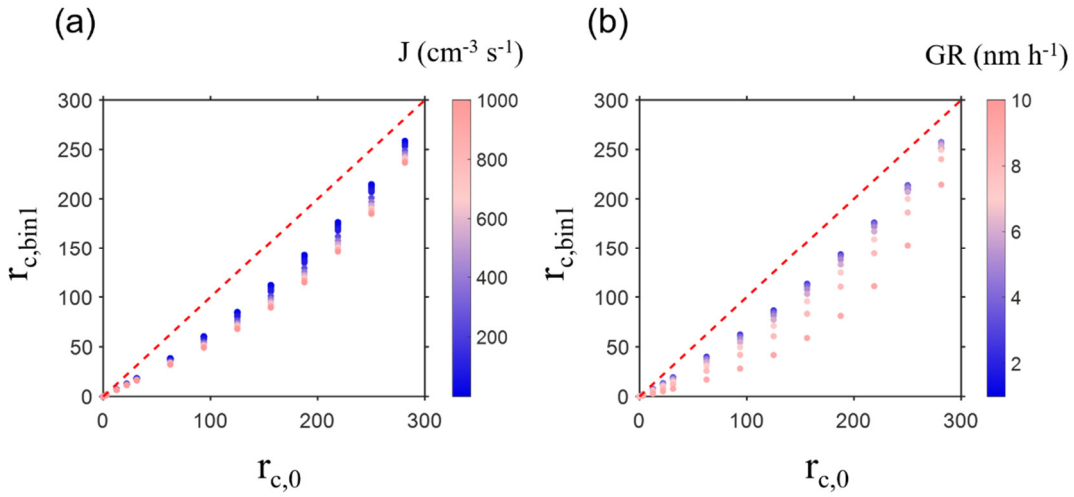

Figure S8. A comparison of the charge state of the particles in the smallest size bin ($r_{c,bin1}$) and $r_{c,0}$ (the ratio of F_{IIN} and the theoretical charge ratio in the first bin). $r_{c,bin1}$ should be used in Eq. (11)(Kerminen et al., 2007), but must be found out through simulation with CDMS-ion. In contrast, $r_{c,0}$ can be calculated with F_{IIN}, which is a model input as we develop ResFWD. The conditions for (a) are N_{ion} = 350 cm^{-3}, CoagS = 0.003 s^{-1}, GR = 4 nm/h, and the colors represent different values of J. The increase of J causes stronger coagulation between oppositely charged particles and decreases the value of $r_{c,bin1}$. The conditions for (b) are N_{ion} = 350 cm^{-3}, CoagS = 0.003 s^{-1}, J = 10 cm^{-3} s^{-1}, and the colors represent different values of GR. The increase of GR, which is an indicator of condensing vapor concentration in this study, causes the charged particles to move faster out of the first bin than neutral particles since vapor condensation on the charged particles is slightly faster due to the charge-dipole interactions. Both (a) and (b) show that $r_{c,bin1}$ is less than $r_{c,0}$, which partially causes Eq.(11) to overestimate the charge fraction of larger particles (see Fig. S7).

References:

Chahl, H. S. and Gopalakrishnan, R.: High potential, near free molecular regime Coulombic collisions in aerosols and dusty plasmas, Aerosol Sci. Technol., 53, 933-957, 10.1080/02786826.2019.1614522, 2019.

Gopalakrishnan, R. and Hogan, C. J.: Determination of the Transition Regime Collision Kernel from Mean First Passage Times, Aerosol Sci. Technol., 45, 1499-1509, 10.1080/02786826.2011.601775, 2011.

Kerminen, V.-M., Anttila, T., Petäjä, T., Laakso, L., Gagné, S., Lehtinen, K. E. J., and Kulmala, M.: Charging state of the atmospheric nucleation mode: Implications for separating neutral and ion-induced nucleation, J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2007JD008649, 2007.

Li, Y., Chen, X., and Jiang, J.: Measuring size distributions of atmospheric aerosols using natural air ions, Aerosol Sci. Technol., 56, 655-664, 10.1080/02786826.2022.2060795, 2022.

Matsui, H.: Development of a global aerosol model using a two-dimensional sectional method: 1. Model design, Journal of Advances in Modeling Earth Systems, 9, 1921-1947, https://doi.org/10.1002/2017MS000936, 2017.