



1 **Data inversion methods to determine sub-3 nm aerosol size**
2 **distributions using the Particle Size Magnifier**

3 Runlong Cai^{1,2,#}, Dongsen Yang^{3,#}, Lauri R. Ahonen², Linlin Shi³, Frans Korhonen², Yan Ma³, Jiming
4 Hao¹, Tuukka Petäjä², Jun Zheng^{3,*}, Juha Kangasluoma², and Jingkun Jiang^{1,*}

5 ¹ State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua
6 University, 100084 Beijing, China

7 ² Institute for Atmospheric and Earth System Research / Physics Faculty of Science, University of Helsinki, P.O. Box 64,
8 00014 Helsinki, Finland

9 ³ Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of
10 Information Science & Technology, 210044 Nanjing, China

11 #: Runlong Cai and Dongsen Yang contribute equally to this work

12 *: Correspondence to: J. Jiang (jiangjk@tsinghua.edu.cn) and J. Zheng (zheng.jun@nuist.edu.cn)

13 **Abstract.** Measuring particle size distribution accurately down to approximately 1 nm is needed for studying atmospheric
14 new particle formation. The scanning particle size magnifier (PSM) using diethylene glycol as the working fluid has been
15 used for measuring sub-3 nm atmospheric aerosol. A proper inversion method is required to recover the particle size
16 distribution from PSM raw data. Similar to other aerosol spectrometers and classifiers, PSM inversion can be deduced to
17 a problem described by the Fredholm integral equation of the first kind. We tested the performance of the step-wising
18 method, the kernel function method (Lehtipalo et al., 2014), the H&A linear inversion method (Hagen and Alofs, 1983),
19 and the expectation-maximization (EM) algorithm. The step-wising method and the kernel function method were used in
20 previous studies on PSM. The H&A method and the expectation-maximization algorithm were used in data inversion for
21 the electrical mobility spectrometers and the diffusion batteries (Maher and Laird., 1985), respectively. In addition, Monte
22 Carlo simulation and laboratory experiments were used to test the accuracy and precision of the particle size distributions
23 recovered using four inversion methods. When all of the detected particles are larger than 3 nm, the step-wising method
24 may report false sub-3 nm particle concentrations because of assuming an infinite resolution, while the kernel function
25 method and the H&A method occasionally reports false sub-3 nm particles because of using the unstable least square
26 method. The accuracy and precision of the recovered particle size distribution using the EM algorithm are the best among
27 the tested four inversion methods. Compared to the kernel function method, the H&A method reduces the uncertainty
28 while keeping a similar computational expense. The measuring uncertainties in the present scanning mode may contribute
29 to the uncertainties of the recovered particle size distributions. We suggest using the EM algorithm to retrieve the particle
30 size distributions using the particle number concentrations recorded by the PSM. Considering the relatively high
31 computation expenses of the EM algorithm, the H&A method is recommended to be used for preliminary data analysis.
32 We also gave practical suggestions on PSM operation based on the inversion analysis.



1 **1 Introduction**

2 The particle size magnifier (PSM) using diethylene glycol as the working fluid (Vanhanen et al., 2011) is widely used in
3 new particle formation studies (Kulmala et al., 2012; Kulmala et al., 2013; Kontkanen et al., 2017) and other industrial
4 applications (Nosko et al., 2016; Ahonen et al., 2017). A PSM can report particle size distributions in the 1-3 nm size
5 range, which is a key size region in the nucleation study. Particles in the PSM grow into larger sizes due to the
6 condensation of super saturated diethylene glycol, and these particles after the initial growth are detected using a
7 downstream condensation particle counter (CPC). The PSM detection efficiency (the CPC is included if not specially
8 mentioned) of particles with a certain diameter is a function of the super saturation ratio of diethylene glycol. Increasing
9 the flow rate passing through the chamber containing saturated diethylene glycol vapour, i.e., the saturator flow rate, can
10 enhance the super saturation ratio thus the particle detection efficiencies. The total particle number concentration detected
11 by the PSM varies with the varying saturator flow rate, and one can determine the particle size distribution according to
12 the observed relationship between the particle number concentration and the saturator flow rate.

13 A proper inversion method is required to recover the particle size distribution using the recorded relationship between the
14 particle number concentration and the saturator flow rate. The step-wising method and the kernel function method were
15 used in previous studies for PSM inversion (Lehtipalo et al., 2014). The step-wising method is a one-to-one linear
16 inversion method using the relationship between the 50% cut-off size and the saturator flow rate, which essentially
17 assumes infinite sizing resolutions, i.e., the particles of a specific size are activated at a certain saturator flow rate.
18 However, such an approximation may lead to non-negligible errors due to the relatively low resolution of the PSM. The
19 kernel function method accounts for the detection efficiency curves, and the particle size distribution is recovered using
20 the non-negative least square method.

21 Although the uncertainties of the particle size distribution determined using the PSM was discussed recently
22 (Kangasluoma and Kontkanen, 2017), the uncertainties introduced during the data inversion have not been systematically
23 addressed. There are always measuring uncertainties in practical conditions, thus one should account for the measuring
24 errors when evaluating the performance of a data inversion method. Because of the relatively low resolution of the PSM,
25 the matrix connecting the particle size distribution and the observed total number concentration is usually ill-conditioned.
26 The kernel function method may theoretically recover the observed particle size distribution when there are no random
27 errors. However, it sometimes leads to large uncertainties when there are small random errors because of the instability
28 of the least square method at a near collinear data set (Ellis, 1998).

29 The equation mapping the particle size distribution to the particle number concentration detected by the PSM is the
30 Fredholm integral equation of the first kind, which arises in many fields, e.g., when studying the molecular dynamics in



1 complex systems (Schäfer et al., 1996) and characterizing the transfer function of an ion drift tube (Buckley and Hogan,
2 2017). Various types of aerosol spectrometers or classifiers, e.g., cascade impactors, optical particle spectrometers,
3 electrical mobility spectrometers, and diffusional barriers, classify particles according to the signals recorded by a number
4 of channels. There is no strict one-to-one relationship between the particle number concentration in a certain size range
5 and the detected signal in a certain channel because of the finite sizing resolutions. The inversion methods used in the
6 previous aerosol spectrometers can possibly be applied to address the PSM inversion problem. The review of the inversion
7 methods for aerosol spectrometers can be found in Kandlikar and Ramachandran (1999), Knutson (1999), and
8 Ramachandran and Cooper (2011).

9 An inversion method with less prior information on the particle size distribution is preferable for the PSM inversion
10 problem. It is impossible to obtain a continuous particle size distribution using a finite number of the detected signals
11 without any constraints, e.g., a known analytical expression to describe the size distribution. Some inversion methods rely
12 on a presumed particle size distribution formula (Fuchs et al., 1962; Raabe, 1978; Ramachandran and Kandlikar, 1996)
13 or prior information on the detection efficiencies (e.g., Onischuk et al., 2017). However, approximating various shapes
14 of the observed sub-3 nm particle size distributions or the PSM detection efficiency curves using a specific formula may
15 lead to relatively large uncertainties. Some methods are feasible in certain conditions, however, sometimes they are not
16 convergent or may lead to high-frequency oscillations (Twomey, 1975; Ferri et al., 1989) due to practical random errors.
17 Some methods use smoothing criterions to deal with the oscillations (Markowski, 1987; Winklmayr et al., 1990), however,
18 they occasionally report an over-smoothed size distribution because of the relatively low resolution and limited size bins
19 of the PSM. The Tikhonov regularisation (Tikhonov, 1963) uses a regularisation parameter to determine the balance of
20 smoothing and the agreement with the PSM recorded data, thus the inverted result may be affected by the method to
21 determine the regularisation parameter (e.g., Wahba, 1977; Hansen, 1992).

22 Based on the reasons mentioned above, we chose the H&A linear inversion method (Hagen and Alofs, 1983) and the
23 expectation-maximization algorithm, and tested the feasibility to apply these methods in the PSM inversion problem. The
24 H&A method is a linear inversion method used in size distribution multi-charge correction, which has the relatively low
25 computational expense. The expectation-maximization algorithm is an iterative method based on probability theory
26 (Dempster et al., 1977), and it was used to reconstruct particle size distributions from diffusion battery data (Maher and
27 Laird, 1985; Wu et al., 1989).

28 In this study, we tested the performance of the step-wising method, the kernel function method, the H&A method, and
29 the expectation-maximization algorithm in PSM inversion. Experiments and Monte Carlo simulations accounting for
30 random errors were used to evaluate the sizing accuracies and the uncertainties of the particle size distributions recovered



1 using four inversion methods. The influence of particles larger than 3 nm on the reported sub-3 nm particle size
2 distributions was discussed. Based on the comparison, the methods with comparatively low uncertainties and high
3 stabilities were recommended to address the PSM inversion problem.

4 2 Theory

5 2.1 PSM measuring theory

6 A PSM measures the total particle number concentration of the activated particles. The sampled aerosol flow is mixed
7 with a high-temperature flow containing saturated diethylene glycol coming from the saturator, and then the mixed flow
8 passes through a low-temperature growth tube. The particles large than a specific diameter can overcome the Kelvin effect
9 and grow into larger sizes due to the condensation of super saturated diethylene glycol. The detection efficiency is mainly
10 determined by the particle diameter and the saturator flow rate. The chemical compositions and charging state may affect
11 the detection efficiencies (Kangasluoma et al., 2013; Kangasluoma et al., 2016a) and lead to errors in the reported particle
12 size distributions (Kangasluoma and Kontkanen, 2017), however, we mainly focus on the inversion method in this study
13 and assume the detection efficiency is only size dependent at a certain saturator flow rate. Since the temperatures in the
14 saturator and the growth tube are fixed, a higher saturator flow rate leads to a higher super saturation ratio of diethylene
15 glycol in the growth tube hence higher detection efficiencies (Fig. 1a). See Section 3.1 for the details on how to obtain
16 the detection efficiency curves. The detected total particle number concentration varies with the varying saturator flow
17 rate when the particle size distribution keeps unchanged. The relationship between the detected total particle
18 concentration, R , the saturator flow rate, s , and the particle size distribution function, n , can be expressed in the Fredholm
19 integral equation of the first kind:

$$20 \quad R_i = \int_0^{+\infty} \eta(s_i, d_p) \times n \times dd_p + \varepsilon_i, \quad (1)$$

21 where R_i is the number concentration recorded at the i^{th} saturator flow rate, s_i ; d_p is the electrical mobility diameter since
22 the calibrating particles are classified according to their electrical mobility; η is the overall detection efficiency determined
23 by s and d_p , including the detection efficiency and the sampling efficiency; n is the probability density of particle number
24 concentration (particle size distribution function), dN/dd_p and N is the accumulated number concentration of particles
25 smaller than d_p ; and ε_i is the error in the recorded particle concentration at s_i .

26 There are many potential sources of the error, ε . For instance, the uncertainties in the calibrated detection efficiencies, the
27 systematic errors caused by the non-ideal fitting formula of the detection efficiency curves, the CPC counting
28 uncertainties, the uncertainties in the super saturation ratio due to fluctuations in the flow rate and temperature, and the



1 unstable aerosol source will all contribute to the difference between the detected number concentration and the expected
2 particle concentration assuming there is no error.

3 As shown in Fig. 1b, the kernel function of the PSM, K , is defined as the derivative of the detection efficiency, η , with
4 respect to the saturator flow rate, s . The area of the kernel function is equal to the difference between the detection
5 efficiencies at the maximum and minimum saturator flow rates. Here we define r as the derivative of the detected number
6 concentration, R , with respect to s . According to Eq. 1, the relationship between r and s is also a Fredholm integral
7 equation of the first kind:

$$8 \quad r_m = \int_0^{+\infty} K(s_m, d_p) \times n \times dd_p + \varepsilon'_m, \quad (2)$$

9 where r_m is the r at the m^{th} saturator flow rate, s_m ; and ε'_m is the error in r_m . Although r is theoretically defined as the
10 derivative of R , practically one can only approximate r using the difference between two adjacent R_i over the increment
11 in s_i and approximate s_m with the mean value of the two corresponding s_i . These approximations also contribute to the
12 uncertainties, ε'_m in addition to the aforementioned sources for ε_i .

13 When using a PSM to determine particle size distributions, the PSM records the varying total particle concentration, R_i ,
14 and the corresponding saturator flow rate, s_i . The saturator flow rate may vary continuously in the scanning mode or fixed
15 at different flow rates in the stepping mode. The particle size distributions are recovered using the recorded relationship
16 between R_i and s_i or the relationship between the approximated r_m and s_m .

17 The size ability of the PSM can be described using the size resolution. Similar to the definition of the sizing resolution of
18 a differential mobility analyser (DMA, Flagan, 1999) to classify particles according to their electrical mobility, we define
19 the resolution of a PSM as:

$$20 \quad Res = \frac{s^*}{\Delta s} \quad (3)$$

21 where Res is the resolution at s^* ; s^* is the peak saturator flow rate of a kernel function; and Δs is the full width at half
22 maximum of the kernel function peak. A relationship between the saturator flow rate and the electrical mobility diameter
23 is defined to straightforwardly relate the resolution and the particle diameter. The peak saturator flow rate, s^* is defined
24 as the corresponding saturator flow rate of the particle diameter. This definition is similar to but different from the
25 definition using the saturator flow rate at the half maximum detection efficiency in Lehtipalo et al. (2014) and in the
26 commercialized PSM. The sizing resolution of a PSM can be estimated according to the relationship between s and d_p , as
27 shown in Fig. 2. However, one should especially keep in mind that the PSM does not measure particle diameter because
28 the relationship between s and d_p is only a definition rather than an intrinsic correlation. A PSM only record the varying



1 particle concentration against the varying saturator flow rate (as indicated in Eqs. 1 and 2). One can only obtain the
 2 particle diameters via proper data inversion.

3 2.2 The step-wising method

4 The resolution of the PSM is assumed infinite in the step-wising method. Thus, the integral equation relating n and r
 5 collapses into a one-to-one corresponding relationship (Lehtipalo et al., 2014),

$$6 \quad n_m = \frac{2(R_{i+1} - R_i)}{\eta(s_{\max}, d_i) + \eta(s_{\max}, d_{i+1})} \quad (4)$$

7 where n_m is the particle size distribution function (dN/dd_p) at d_m ; d_m , d_i , and d_{i+1} are the corresponding half-maximum cut-
 8 off diameters of s_m , s_i , and s_{i+1} , respectively; and s_m is the mean value of s_i and s_{i+1} . The relationship between particle
 9 diameter and the saturator flow rate is determined using the saturator flow rate at the half maximum detection efficiency
 10 (Lehtipalo et al., 2014). The step-wising method does not magnify the relative error in measurement since it is a one-to-
 11 one inversion method. However, the inverted results using the step-wising method are perhaps non-negligibly affected by
 12 the relatively low resolutions of the PSM.

13 2.3 The kernel function method

14 The kernel function method assumes that the particle size distribution can be approximated using several particle size
 15 bins and the detection efficiencies of particles in each size bin are the same. The mathematical description of this
 16 approximation is:

$$17 \quad r_m \approx \sum_{j=1}^J K(s_m, d_j) \times n_j \times \Delta d_j, \quad J \leq I - 1, \quad (5)$$

18 where d_j is the representing particle diameter of each size bin; J is the number of d_j ; n_j is the particle size distribution
 19 function (dN/dd_p) at d_j ; Δd_j is the length of each size bin; and I is the number of R_i . The symbol of \approx is to emphasize that
 20 Eq. 5 is an approximation even if there are no measuring errors because it approximates the integral with a finite discrete
 21 sum and estimates r_m using the recorded R_i . Using a matrix, Eq. 5 can be rewritten as:

$$22 \quad \overset{\cdot}{r}_{(I-1) \times 1} \approx \mathbf{G}_{(I-1) \times J} \cdot \overset{\cdot}{n}_{J \times 1}, \quad J \leq I - 1 \quad (6)$$

$$23 \quad \text{where } \mathbf{G}_{i,j} = K(s_i, d_j) \times \Delta d_j \quad (7)$$

24 The subscriptions in the uppercase of Eq. 6 indicate the dimensions of the matrix and the vectors, while the subscriptions
 25 in the lowercase of Eq. 7 represent the corresponding element. The particle size distribution is obtained via solving Eq. 6
 26 using the non-negative least square method.

27 2.4 The Hagen & Alofs method



1 The H&A method (Hagen and Alofs, 1983) was proposed to deal with the multi-charging correction problem when using
2 a DMA. It can also be used to solve the PSM inversion problem. Similar to the kernel function method, a discrete sum is
3 used to approximate the integral:

$$4 \quad R_i = \sum_{j=1}^J \eta(s_i, d_j) \times n_j \times \Delta d_j, \quad J \gg I \quad (8)$$

5 We use the symbol of = in Eq. 8 rather than \approx because the H&A method requires a J much larger than I . One should
6 increase J if the error in approximating the integral with the discrete sum is still large. Usually, J is determined as 30
7 times that of I considering the computational expenses. However, Eq. 8 itself is not solvable because there are more
8 unknown variables than the equations. Thus, additional constraints are required. The H&A method assumes that any n_j
9 can be approximated using n_i , i.e.,

$$10 \quad n_j \approx f(\vec{n}_i, d_j), \quad (9)$$

11 where f is the function relating n_j and n_i ; n_i is the particle size distribution function at d_i ; and the vector symbol indicates
12 that n_j is estimated using more than one single n_i . The determination of d_i is theoretically arbitrary as long as the number
13 of d_i is the same as the number of R_i . For the details to determine f , please refer to Hagen and Alofs (1983).

14 Similar to the kernel function method, the relationship between the particle size distribution and the number concentration
15 recorded by the PSM can be described in the vector form:

$$16 \quad \vec{R}_{I \times 1} \approx \mathbf{Q}_{I \times I} \cdot \vec{n}_{I \times 1} \quad (10)$$

17 where \mathbf{Q} is determined by η , f , and Δd_j . One can directly solve Eq. 10 (e.g., via Gaussian elimination) since \mathbf{Q} is usually
18 non-singular. However, different from the matrix obtained from a DMA, the matrix \mathbf{Q} in PSM inversion problem is
19 usually not a positive-definite matrix because the detected particle concentration sometimes decreases with the increasing
20 saturator flow rate due to random errors. Simply solving Eq (10) often obtains negative values in particle size distributions.
21 Thus, the non-negative least square method is suggested to determine the particle size distribution in the PSM inversion
22 problem. The H&A methods can also reconstruct the particle size distribution according to the relationship between r_m
23 and s_m . However, using the kernel functions instead of the detection efficiencies does not necessarily improve the accuracy
24 or precision of the results. On the contrast, we found that using the kernel functions usually lead to larger uncertainties
25 than using the detection efficiencies because of the errors caused by approximating r_m .

26 The H&A method is theoretically more stable than the kernel function method because of the more accurate assumption
27 of the true aerosol size distribution. However, the H&A method adapted for PSM inversion may still report size
28 distributions with large uncertainties because of using the least square method. The computational expense of the H&A



1 method is similar to that of the kernel function method because the rate-limiting step is to solve the least square question.
2 Their low computational expense is an advantage over other nonlinear inversion methods.

3 **2.5 The expectation-maximization algorithm**

4 The EM algorithm is a statistical method dealing with inversion problems with unobserved latent variables. An
5 explanation of the EM algorithm can be found in Do and Batzoglou (2008). In the PSM inversion problem, the latent
6 variable is $R_{i,j}$, defined as the contribution of particles with the diameter of d_j to the detected number concentration, R_i
7 (Maher and Laird, 1985). The algorithm obtains the recovered particle size distribution using two steps: the expectation
8 step and the maximization step. In the expectation step, the values of $R_{i,j}$ are estimated according to Bayesian theorem:

$$9 \quad R_{i,j} = \frac{n_j \times \eta(s_i, d_j) \times \Delta d_j}{\sum_{j=1}^J n_j \times \eta(s_i, d_j) \times \Delta d_j} \quad (11)$$

10 In the maximization step, the particle size distribution function is estimated according to the maximum likelihood:

$$11 \quad n_j = \frac{\sum_{i=1}^I R_{i,j}}{\sum_{i=1}^I \eta(s_i, d_j) \times \Delta d_j} \quad (12)$$

12 The EM algorithm obtains the recovered particle size distribution by repeating the expectation step and the maximization
13 step until convergence. The convergence can be measured by the likelihood function (Maher and Laird, 1985). The values
14 and the number of d_j are not limited when using the EM algorithm, and a larger J can reduce the errors in approximating
15 the integral using the discrete sum. Thus, the EM algorithm is able to report particle size distributions with more size bins
16 compared to the step-wising method, the kernel function method, and the H&A method.

17 The EM algorithm is more stable compared to the algorithms based on the least square methods (Maher and Laird, 1985).
18 The convergence of the EM algorithm has been proved (Dempster et al., 1977), however, the convergence speed is not
19 mathematically guaranteed. Compared to the kernel function method and the H&A method, the computational expense
20 of the EM algorithm is much higher. In addition, the EM algorithm is a greedy algorithm such that the iteration is easily
21 trapped in a local optimum. To start the first expectation step, an initial guess of the particle size distribution is required.
22 We suggest the initial guess to be a vector of all ones. Note that the EM algorithm is sensitive to the initial guess and
23 using a recovered particle size distribution obtained from another method, e.g., the step-wising method does not
24 necessarily improve the iteration results.



1 3 Methods

2 3.1 Experiments

3 Laboratory experiments using particles with known peak size or size distribution were conducted to test the inversion
4 methods (Fig. 3). Sub-10 nm tungsten oxide particles were generated using a wire generator (Peineke et al., 2006;
5 Kangasluoma et al., 2015). In the narrow peak measurement, the negatively charged particles were classified using a
6 high-resolution Herrmann DMA. The sizing resolutions of the Herrmann DMA in the experimental conditions were no
7 smaller than 25 (Kangasluoma et al., 2016b). Thus, the classified aerosols out of the Herrmann DMA can be
8 approximately regarded as monodisperse. The relationship between the Herrmann DMA voltage and the classified particle
9 size was calibrated using standard molecular ions (Ude and de la Mora, 2005). A TSI 3068B aerosol electrometer using
10 the same aerosol flow rate with the PSM (2.5 liters per minute, lpm) was used as the reference.

11 In the wide peak measurement, the particle size distributions classified using a TSI nanoDMA have wider peaks than
12 those generated in the narrow peak measurement. The aerosol and sheath flow rates of the nanoDMA were 2 and 10 lpm,
13 respectively. It should be clarified that the particle size distribution classified using the nanoDMA in the wide peak
14 measurement were still narrow due to the limitation of the nanoDMA. A lower sizing resolution either achieved by a
15 higher aerosol-to-sheath flow ratio will cause the nanoDMA out of work due to significant turbulence. A half-mini DMA
16 (Fernández de la Mora and Kozłowski, 2013) with calibrated penetration efficiency and a downstream Faraday cage
17 electrometer (FCE) were used to measure the classified particle size distributions in parallel.

18 The PSM (Airmodus A11) was calibrated using negatively charged tungsten oxide particles before the test. The
19 experimental setup for the calibration was the same with that used in the narrow peak measurement. The influence of the
20 finite resolution of the Herrmann DMA on the calibrated efficiency curves was negligible. The saturator flow rate of the
21 tested PSM varied from 0.05 to 1.3 lpm. This saturator flow rate range is wider than that of a typical PSM to obtain a
22 complete kernel function curve of 3 nm particles. The maximum background noise of the PSM was approximately 1
23 No./cm³, which was negligible compared to the usually detected particle concentrations. The detection efficiency is
24 determined as the ratio of the particle number concentrations reported by PSM over the number concentration reported
25 by the electrometer. The detection efficiency curves of the PSM were fitted using a function (Eq. 13) modified from the
26 Chapman-Richards growth curve (Richards, 1959) which fitted better than other tested functions for the tested PSM,

$$27 \quad \eta = a \times \left[1 + |b| \times (s - s_{\max}) \right] \times \left[1 - \exp(-c \times s) \right]^d, \quad (13)$$

28 where s_{\max} is the maximum saturator flow rate (1.3 lpm); a , b , c , and d are the fitting parameters. If not specially
29 mentioned, the PSM was fixed at 18 different saturator flow rates when measuring the particle size distributions in this
30 study. This operation in the stepping mode was to avoid the potential uncertainties introduced in the scanning mode. The



1 stability of the particle size distribution was monitored using the reference FCE during the relatively long measuring
2 period.

3 **3.2 Simulation**

4 The performance of the four inversion methods was also studied using Monte Carlo simulations. The detection efficiencies
5 used in the simulations were determined according to the calibrated efficiencies but slightly adjusted towards smoother
6 curves. The uncertainties in practical calibration were neglected in the simulation.

7 The particle number concentrations detected at different saturator flow rates were simulated using a certain initial particle
8 size distribution. The random error, ε_i , was inserted into the simulated particle concentration, R_i . The random errors were
9 determined experimentally. The relative random errors were larger than the statistical relative errors predicted using
10 Poisson distribution (Iida, 2008; Kuang et al., 2012; Kangasluoma and Kontkanen, 2017) and independent of the particle
11 concentrations at a certain instrumental configuration, indicating that random errors were governed by the fluctuations of
12 the source and/or the instrumental parameters (e.g., flow rate). We used the mean relative random standard deviation
13 observed in the experimental tests, 3.7%, as the representative value. Totally 10 data points were assumed to be collected
14 at each saturator flow rate. Thus, the random errors inserted into the simulated particle concentrations, i.e., the relative
15 standard deviations of the mean particles concentrations, were assumed to be 1.2% ($= 3.7\% / \sqrt{10}$). A relatively large
16 random error of 10% obtained from the ambient measurements was also tested. The Monte Carlo simulation was
17 conducted for 10000 times using each inversion method to estimate the accuracy and precision of the recovered particle
18 size distribution indicated by the mean values and the standard deviations of inverted results.

19 **4 Results and discussion**

20 **4.1 Sizing accuracy**

21 The inversion methods tested in this study, i.e., the step-wising method, the kernel function method, the H&A method,
22 and the EM algorithm are able to estimate the classified particle diameters when the PSM was measuring nearly
23 monodisperse sub-3 nm particles. When the classified particle diameters were 1.51 nm and 2.41 nm, respectively, all of
24 the four inversion methods can recover single peaks around the classified diameter (Figs 4a, 4b). The size distribution
25 reported by the step-wising method was the widest because the step-wising method does not account for the resolution of
26 the PSM. Note that the peak diameters reported by the kernel function method and the H&A method were also affected
27 by the selection of the particle size bins. The total particle concentrations obtained via inversion were similar to the
28 number concentration detected by the reference FCE, except for the number concentration of 1.51 nm particles reported
29 by the kernel function method.



1 None of the four inversion methods could size particles larger than 3 nm with relatively good sizing accuracies. When
2 the classified particle diameter was 3.93 nm, the four inversion methods failed to report narrow peaks with peak diameters
3 approximating 3.93 nm (Fig. 4c). This is because the PSM resolution for particles larger than 3 nm is low, i.e., the
4 resolution was ~ 1.0 when measuring the classified 3.93 nm particles (Fig. 2). When focusing on the sub-3 nm particle
5 size range, the kernel function method, the H&A method, and the EM algorithm reported nearly no sub-3 nm particles.
6 However, the step-wising method reported a non-negligible amount of sub-3 nm particles with a total number
7 concentration of 1591 No./cm³ due to the low sizing resolution.

8 We further tested the sizing ability of the four inversion methods using the sum of the recorded particle concentrations
9 when the PSM was measuring 1.51, 2.41, and 3.93 nm particles (Fig. 4d). The kernel function method, the H&A method,
10 and the EM algorithm distinguished the particles with different sizes, and the reconstructed peaks were similar to the
11 corresponding peaks when the PSM was measuring monodisperse particles. The inverted results using the step-wising
12 method was also unaffected by the summation, however, it was difficult to distinguish the isolated peaks from the
13 recovered particle size distribution due to the broadened size distribution.

14 The size distributions of particles larger than 3 nm could not be successfully retrieved via data inversion because of the
15 low resolution of PSM for these particles, however, it helped to recover sub-3 nm particle size distributions. Most of the
16 reported particle sizes using the kernel function method, the H&A method, and the EM algorithm were larger than 3 nm
17 when the PSM was measuring 3.93 nm particles (Fig. 4c). This estimation of particles larger than 3 nm assured a relatively
18 accurate sizing of sub-3 nm particle size distribution (Fig. 4d). Thus, we recovered the particle size distribution up to 5
19 nm using different inversion methods but focus only on the sub-3 nm size range.

20 **4.2 Uncertainties using different inversion methods**

21 The step-wising method, the kernel function method, and the H&A method may report false sub-3 nm particles when
22 there are only particles are larger than 3 nm in the input aerosol. A particle size distribution with a peak diameter of 5 nm
23 and nearly no sub-3 nm particles was simulated (Fig. 5a). The detected particle concentrations were assumed to fluctuate
24 with a 1.2% relative standard deviation due to measuring uncertainties (Fig. 5b). The EM algorithm reported nearly no
25 sub-3 nm particles except for the smallest size bin at 1.16 nm (Fig. 5c). The expected values of particle concentrations in
26 the bins smaller than 3 nm recovered using the H&A method were near zero, however, false sub-3 nm particle
27 concentrations were occasionally reported (Fig. 5d). Compared to the H&A method, the size distribution recovered using
28 the kernel function method was more unstable, especially in the sub-2 nm size range (Fig. 5e). Different from the H&A
29 method and the kernel function method that reported false results due to their instability, the step-wising method reported
30 false particle size distributions with nonzero mean values (Fig. 5f). This is because the step-wising method assumes a



1 simple one-to-one relationship between the saturator flow rate and the recovered particle diameter instead of accounting
2 for the wide kernel function peaks.

3 The false sub-3 nm particle concentrations due to improper inversion methods were tested experimentally. Particles larger
4 than 5 nm were classified using the nanoDMA (Fig. 6a). No sub-3 nm particles were reported using the EM algorithm
5 and the H&A method. On the contrast, the kernel function method and the step-wising method reported approximately
6 3×10^3 particles when the total particle concentration measured using the DMA-FCE system was approximately 2.4×10^4 .
7 Based on both the simulating and experimental results, we conclude that the PSM may report false sub-3 nm particle size
8 distributions when there are actually no sub-3 nm particles because of using non-ideal data inversion methods, especially
9 the step-wising method. Note that large particles whose detection efficiencies do not vary with the saturator flow rate do
10 not lead to a bias in the recovered sub-3 nm particle concentrations. We examined this theoretical deduction
11 experimentally using a PSM to measure ambient particles existing in the room air and the recorded particle concentration
12 did not significantly vary with the saturator flow rate.

13 The performance of the four inversion methods in the sub-3 nm size range under the influences of larger particles was
14 tested using a bimodal distribution (Fig. 7a). Similar particle size distributions are usually observed in the atmospheric
15 new particle formation events (Jiang et al., 2011) and in flame (Tang et al., 2017). As shown in Fig. 7, the particle size
16 distribution recovered using the EM algorithm had the highest accuracy and the smallest uncertainties among the four
17 methods. The recovered particle size distribution using the EM algorithm had a slightly different shape compared to the
18 initial distribution because the results were trapped in the local optimum. However, the differences between the recovered
19 and the initial size distributions were the smallest. The standard deviations of the size distribution recovered using the
20 H&A method and the kernel function method were relatively large due to the unstable least square method. Because of a
21 better assumption of the initial particle size distribution, the H&A method resulted in smaller uncertainties compared to
22 the kernel function method, especially in the sub-2 nm size range. The size distribution recovered using the EM algorithm
23 has higher accuracy and stability compared to both the H&A method and the kernel method because the one-to-one
24 inversion method does not magnify relative errors.

25 The experimental tests using bimodal distributions agreed with the simulation results. The particles with a peak diameter
26 at approximately 2.3 nm were classified using the nanoDMA. We added the observed number concentration to those
27 detected in Fig. 6a (particles larger than 5 nm) to account for the influence of large particles. Unfiltered room air served
28 as the makeup flow to provide background particles. As shown in Fig. 8, all the four inversion methods recovered the
29 peak around 2.3 nm, while the results reported by the H&A method and the kernel function method were less smooth
30 compared to the EM algorithm and the step-wising method.



1 Smoothing the size distribution recovered using the H&A method and the kernel function method into fewer size bins
2 can reduce the uncertainties. We determined the number of the size bins of the recovered distributions according to the
3 number of the fixed saturator flow rates. Too many size bins will lead to relatively large uncertainties, however, the
4 uncertainties can be reduced by sacrificing the resolution, i.e., reporting the size distribution in fewer bins. The size
5 distributions recovered using the kernel function method were reported in typical 4-6 bins (Lehtipalo et al., 2014). This
6 was achieved by assuming fewer discrete particle diameters in Eq. 5. Another option is to merge bins into fewer numbers
7 after inversion rather than assume fewer bins at the beginning. Note that the H&A method cannot assume fewer discrete
8 size bins at the beginning. Instead, the H&A method assumes an adequate number of size bins to guarantee a relatively
9 smooth distribution (Eq. 8). As shown in Fig. 9, the standard deviations of the reported size distribution with fewer size
10 bins were comparatively smaller than the corresponding standard deviations with more size bins shown in Fig. 7. The
11 H&A method reported size distributions with smaller standard deviations than the kernel function method, and the kernel
12 function reported in merged size bins had smaller standard deviations than the kernel function method using fewer size
13 bins at the beginning. This is because approximating the true particle size distribution, which is usually a smooth curve,
14 with fewer discrete size bins will lead to larger uncertainties. Thus, we suggest merging the recovered particle size
15 distribution into a few size bins to reduce the uncertainties when using the H&A method and the kernel function method.
16 Relatively large uncertainties were found when recovering sub-1.3 nm particle size distributions. A particle size
17 distribution with an increasing dN/dd_p as a function of the decreasing particle diameter, which is a typical particle size
18 distribution observed in the atmospheric new particle formation events (Jiang et al., 2011), was used to test the four
19 inversion methods (Fig. 10). None of the inversion methods reported a particle size distribution with relatively small
20 uncertainties comparable to the inverted results shown in Fig. 7c, especially in the sub-1.3 nm size range. Similar to the
21 results for particles larger than 3 nm, the low resolution of particles smaller than 1.3 nm (Fig. 2) is possibly the cause of
22 the large uncertainties. In addition, incomplete kernel function peaks and the relatively low detection efficiencies of sub-
23 1.3 nm particles may also contribute to the uncertainties (Fig. 1).

24 The performance of the inversion methods under relatively large random errors was also tested. The relative standard
25 deviation used in the above simulations, 3.7%, was estimated according to laboratory experiments. The relative standard
26 deviations of the recorded particle number concentration obtained from the atmospheric measurement were usually
27 similar to the value obtained in the laboratory, indicating the random errors were governed by instrumental factors.
28 However, relatively large uncertainties in the recorded particle number concentrations were sometimes observed due to
29 the unstable atmospheric aerosol source. Thus, we simulated the performance of the four inversion methods using a
30 relative standard deviation of 10%. It should be clarified that the value 10% only characterizes the random errors of the



1 CPC since it was estimated using the data when the recorded particle number concentration did not vary with the saturator
2 flow rate. Compared to the results in Fig.7 simulated using the same aerosol size distribution, the uncertainties in the
3 recovered particle size distributions using the larger relative standard deviation of 10% was larger (Fig. 10). The EM
4 algorithm still reported smaller uncertainties compared to the H&A method and the kernel function method. Note the
5 expected value of sub-2 nm particle size distribution recovered using the kernel method was close to the input size
6 distribution when the uncertainty was 3.7% (Fig. 7); however, the recovered size distribution in the sub-2 nm size range
7 was non-negligibly overestimated when the uncertainty was 10% (Fig. 10).

8 **4.3 Uncertainties in the scanning mode**

9 The PSM instrumental factors limiting the accuracy of the inversion were also tested. Although using the EM algorithm
10 and the H&A method can reduce the errors of the recovered size distributions compared to the kernel function method
11 and the step-wise method, relatively small measuring uncertainties are still vital to retrieve a particle size distribution
12 with relatively high accuracies. The uncertainties in the scanning mode, for example, is one of the potential sources of
13 the measuring uncertainties. The saturator flow rate of a scanning PSM increases linearly with time in previous studies.
14 However, the relationship between the particle diameters and the saturator flow rates at the kernel function peaks is
15 nonlinear (Fig. 2). The detection efficiencies of particles larger than 1.6 nm vary mainly in the flow rate range from 0.05
16 to 0.3 lpm while the corresponding scanning time is only 20% of the whole scanning cycle. This nonlinear relationship
17 may result in non-negligible uncertainties in the recovered particle size distributions (Fig. 12). The EM algorithm
18 recovered the single peak when using the particle concentrations recorded in the stepping mode. However, the recovered
19 particle size distribution using the EM algorithm was not a single smooth peak when using data recorded in the scanning
20 mode (Fig. 12). This difference can be illustrated using the raw data. The curves of the particle number concentration
21 recorded in the stepping mode and the scanning mode are similar to each other and they both appear to be smooth (Fig.
22 13a). When presenting in the derivate of the particle number concentration with the respect to saturator flow rate, however,
23 the curve corresponding to the stepping mode appeared to be a single peak while the other curve corresponding to the
24 scanning mode seemed to be composed of multiple single peaks (Fig. 13b). Since none of the four inversion methods
25 tested in this study add smoothing constraints when solving the Fredholm integral equation of the first kind, this roughness
26 in the raw data will lead to split peaks in the recovered particle size distribution unless one report the size distribution
27 using only a few size bins.



1 **4.4 Implications on using the PSM**

2 According to the discussion above, we provide the following suggestions on using a PSM to determine particle size
3 distributions:

4 (a) Particle size range and saturator flow rate range. Complete efficiency curves are preferable to determine the particle
5 size distribution in a certain size range. For example, to reduce the uncertainties in the recovered size distribution of ~3
6 nm particles, the saturator flow rate in this study was extended from the commonly used 0.1 lpm to 0.05 lpm where the
7 detection efficiency of 3.11 nm particles was almost zero. The detection efficiency curves of particles larger than the
8 maximum concerned diameter should also be calibrated to reduce the influence of large particles on the recovered particle
9 size distribution and total concentration. The PSM can theoretically estimate particle size distributions larger than 3 nm
10 or smaller than 1.3 nm, however, the uncertainties are usually large due to the low resolution and the incomplete detection
11 efficiency curves. The particles whose detection efficiency are constant values in the measuring saturator flow rate range
12 cannot be determined using a PSM and they do not influence the recovered particle size distributions if their
13 concentrations are stable during each scanning cycle.

14 (b) Scanning scheme. The scanning scheme of the saturator flow rate is suggested to be improved to reduce the measuring
15 uncertainties. The scanning scheme is preferably determined to ensure that the particle diameter corresponding to the
16 saturator flow rate increases linearly with time so that the numbers of the recorded particle number concentration at each
17 saturator flow rate are the same when the recovered particle size increases linearly. A convex function between the
18 saturator flow rate and the scanning time, e.g., an exponentially increasing saturator flow rate, is also better than the linear
19 scanning scheme. Such improvement may require updating both the hardware and the software.

20 (c) Inversion method. We suggest using the EM algorithm to address the PSM inversion problem because the particle size
21 distributions recovered using the EM algorithm have the best accuracy and stability among the four tested methods.
22 However, considering the relatively high computational expense of the EM algorithm, the H&A method reporting in
23 merged size bins is recommended to be used for preliminary data analysis and for meeting the need of fast inversion, e.g.,
24 real-time display on the instrumental screen. The accuracy of the recovered size distribution is also determined by the
25 uncertainties in the recorded number concentration rather than the inversion method alone. The inversion methods
26 suggested in this study does not necessarily assure an accurate inverted result without properly determined detection
27 efficiencies and an improved scanning scheme.

28 (d) Uncertainties in atmospheric measurement. One should be always aware of the potential uncertainties in the recovered
29 particle size distribution, especially when conducting atmospheric measurement. The reported sub-3 nm particle
30 concentrations may be false results due to systematic and random error, especially when using the step-wise method.



1 The number of the reported size bins should also be carefully limited. For example, the EM algorithm can theoretically
2 provide infinite size bins; however, we suggest reducing the reported size bins to avoid false fluctuations.

3 **5 Conclusions**

4 We tested the performance of four inversion methods to recover particle size distributions from the particle size magnifier
5 data using Monte Carlo simulation and experiments. The four inversion methods are the step-wising method, the kernel
6 function method, the H&A method, and the EM algorithm, respectively. The step-wising method may report false sub-3
7 nm particle concentrations when there are no sub-3 nm particles in the input aerosol because it does not account for the
8 influence of particles large than 3 nm. The kernel function method and the H&A method may lead to relatively large
9 uncertainties in the recovered particle size distribution because of using the unstable least square method, and they
10 occasionally report false sub-3 nm concentrations due to the large uncertainties. Compared to the kernel function method,
11 the H&A lead to smaller uncertainties while having a similar computation expense. This is because that the H&A method
12 assumes a near continuous size distribution rather than a discrete distribution with limited size bins. One can reduce the
13 uncertainties via merging the particle size distribution reported by the H&A method into fewer size bins. Among the
14 tested inversion methods, the EM algorithm has the highest accuracy and stability. Another advantage of the EM algorithm
15 over the other three methods is that it does not limit the number of the particle size bins. The instrumental factors also
16 limit the accuracy and precision of the recovered particle size distribution. The uncertainties of the recovered size
17 distributions of particle smaller than 1.3 nm or larger than 3 nm may be significant due to the incomplete kernel function
18 curves, the low resolution and/or the low detection efficiency. The measuring uncertainties in the scanning mode may
19 also increase the uncertainties of the recovered size distribution.

20 Based on this study, we suggest that a) the EM algorithm is used to recover the particle size distribution measured by the
21 PSM and the H&A method can be used for preliminary data analysis and for fast inversion purposes; b) the hardware and
22 software of the PSM should be improved to reduce the measuring uncertainties, e.g., via changing the scanning scheme
23 of the saturator flow rate; c) one should carefully distinguish the false inverted results from the true sub-3 nm particles,
24 especially in the sub-2 nm size range and/or when using the step-wising method.

25 **Data availability**

26 The characterizations of the tested PSM are shown in the figures. The Matlab scripts for the inversion methods are
27 available upon request.



1 Competing interests

2 The authors declare that they have no conflict of interest.

3 Acknowledgement

4 Financial supports from the National Key R&D Program of China (2017YFC0209503), the National Natural Science
5 Foundation of China (21521064 & 41730106), ACTRIS-2 (grant agreement No. 654109), the Academy of Finland
6 (project No. 307331), and Faculty of Science, University of Helsinki, are acknowledged. R. Cai appreciates the support
7 from China Scholarship Council (CSC) for his visit to University of Helsinki.

8 References

- 9 Ahonen, L. R., Kangasluoma, J., Lammi, J., Lehtipalo, K., Hämeri, K., Petäjä, T., and Kulmala, M.: First
10 measurements of the number size distribution of 1–2 nm aerosol particles released from manufacturing
11 processes in a cleanroom environment, *Aerosol Science and Technology*, 51, 685–693, 2017.
- 12 Buckley, D. T. and Hogan, C. J.: Determination of the transfer function of an atmospheric pressure drift tube ion
13 mobility spectrometer for nanoparticle measurements, *Analyst*, 142, 1800–1812, 2017.
- 14 Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm,
15 *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38, 1977.
- 16 Do, C. B. and Batzoglou, S.: What is the expectation maximization algorithm?, *Nature Biotechnology*, 26, 897–899,
17 2008.
- 18 Ellis, S. P.: Instability of least square, least absolute deviation and least median of squares linear regression,
19 *Statistical Science*, 13, 337–350, 1998.
- 20 Fernández de la Mora, J. and Kozlowski, J.: Hand-held differential mobility analyzers of high resolution for 1–
21 30nm particles: Design and fabrication considerations, *Journal of Aerosol Science*, 57, 45–53, 2013.
- 22 Ferri, F., Giglio, M., and Perini, U.: Inversion of light scattering data from fractals by the Chahine iterative algorithm,
23 *Applied Optics*, 28, 3074–3082, 1989.
- 24 Flagan, R. C.: On Differential Mobility Analyzer Resolution, *Aerosol Science and Technology*, 30, 556–570, 1999.
- 25 Fuchs, N. A., Stechkina, I. B., and Starosselskii, V. I.: On the determination of particle size distribution in
26 polydisperse aerosols by the diffusion method, *British Journal of Applied Physics*, 13, 280–281, 1962.
- 27 Hagen, D. E. and Alofs, D. J.: Linear inversion method to obtain aerosol size distributions from measurements with
28 a differential mobility analyzer, *Aerosol Science and Technology*, 2, 465–475, 1983.
- 29 Hansen, P. C.: Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Review*, 34, 561–580, 1992.
- 30 Iida, K.: Atmospheric nucleation: Development and application of nanoparticle measurements to assess the roles of
31 ion-induced and neutral processes, PhD, University of Minnesota, Minneapolis, 2008.
- 32 Jiang, J., Zhao, J., Chen, M., Eisele, L. F., Scheckman, J., Williams, J. B., Kuang, C., and McMurry, H. P.: First
33 Measurements of Neutral Atmospheric Cluster and 1–2 nm Particle Number Size Distributions During
34 Nucleation Events, *Aerosol Research Letter*, 45, ii–v, 2011.
- 35 Kandlikar, M. and Ramachandran, G.: Inverse methods for analysing aerosol spectrometer measurements: a critical
36 review, *Journal of Aerosol Science*, 30, 413–437, 1999.
- 37 Kangasluoma, J., Junninen, H., Lehtipalo, K., Mikkilä, J., Vanhanen, J., Attoui, M., Sipilä, M., Worsnop, D.,
38 Kulmala, M., and Petäjä, T.: Remarks on Ion Generation for CPC Detection Efficiency Studies in Sub-3-nm
39 Size Range, *Aerosol Science and Technology*, 47, 556–563, 2013.
- 40 Kangasluoma, J., Attoui, M., Junninen, H., Lehtipalo, K., Samodurov, A., Korhonen, F., Sarnela, N., Schmidt-Ott,
41 A., Worsnop, D., Kulmala, M., and Petäjä, T.: Sizing of neutral sub 3nm tungsten oxide clusters using
42 Airmodus Particle Size Magnifier, *Journal of Aerosol Science*, 87, 53–62, 2015.
- 43 Kangasluoma, J., Samodurov, A., Attoui, M., Franchin, A., Junninen, H., Korhonen, F., Kurtén, T., Vehkamäki, H.,
44 Sipilä, M., Lehtipalo, K., Worsnop, D. R., Petäjä, T., and Kulmala, M.: Heterogeneous Nucleation onto Ions
45 and Neutralized Ions: Insights into Sign-Preference, *The Journal of Physical Chemistry C*, 120, 7444–7450,
46 2016a.
- 47 Kangasluoma, J., Attoui, M., Korhonen, F., Ahonen, L., Siivola, E., and Petäjä, T.: Characterization of a Herrmann-
48 type high-resolution differential mobility analyzer, *Aerosol Science and Technology*, 50, 222–229, 2016b.



- 1 Kangasluoma, J. and Kontkanen, J.: On the sources of uncertainty in the sub-3 nm particle concentration
2 measurement, *Journal of Aerosol Science*, 112, 34-51, 2017.
- 3 Knutson, E. O.: History of Diffusion Batteries in Aerosol Measurements, *Aerosol Science and Technology*, 31, 83-
4 128, 1999.
- 5 Kontkanen, J., Lehtipalo, K., Ahonen, L., Kangasluoma, J., Manninen, H. E., Hakala, J., Rose, C., Sellegri, K., Xiao,
6 S., Wang, L., Qi, X., Nie, W., Ding, A., Yu, H., Lee, S., Kerminen, V.-M., Petäjä, T., and Kulmala, M.:
7 Measurements of sub-3 nm particles using a particle size magnifier in different environments: from clean
8 mountain top to polluted megacities, *Atmospheric Chemistry and Physics*, 17, 2163-2187, 2017.
- 9 Kuang, C., Chen, M., Zhao, J., Smith, J., McMurry, P. H., and Wang, J.: Size and time-resolved growth rate
10 measurements of 1 to 5 nm freshly formed atmospheric nuclei, *Atmospheric Chemistry and Physics*, 12, 3573-
11 3589, 2012.
- 12 Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P.,
13 Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E., Laaksonen, A., and Kerminen, V. M.: Measurement
14 of the nucleation of atmospheric aerosol particles, *Nature protocols*, 7, 1651-1667, 2012.
- 15 Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H. E., Nieminen, T., Petaja, T., Sipilä, M.,
16 Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Jarvinen, E., Aijala, M., Kangasluoma, J., Hakala, J.,
17 Aalto, P. P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T.,
18 Mauldin, R. L., 3rd, Duplissy, J., Vehkamäki, H., Back, J., Kortelainen, A., Riipinen, I., Kurten, T., Johnston,
19 M. V., Smith, J. N., Ehn, M., Mentel, T. F., Lehtinen, K. E., Laaksonen, A., Kerminen, V. M., and Worsnop,
20 D. R.: Direct observations of atmospheric aerosol nucleation, *Science*, 339, 943-946, 2013.
- 21 Lehtipalo, K., Leppä, J., Kontkanen, J., Kangasluoma, J., Franchin, A., Wimmer, D., Schobesberger, S., Junninen,
22 H., Petäjä, T., Sipilä, M., Mikkilä, J., Vanhanen, J., Worsnop, D. R., and Kulmala, M.: Methods for determining
23 particle size distribution and growth rates between 1 and 3 nm using the Particle Size Magnifier, *Boreal
24 Environment Research*, 19, 215-236, 2014.
- 25 Maher, E. F. and Laird, N. M.: EM algorithm reconstruction of particle size distributions from diffusion battery
26 data, *Journal of Aerosol Science*, 16, 557-570, 1985.
- 27 Markowski, G. R.: Improving Twomey's Algorithm for Inversion of Aerosol Measurement Data, *Aerosol Science
28 and Technology*, 7, 127-141, 1987.
- 29 Nosko, O., Vanhanen, J., and Olofsson, U.: Emission of 1.3–10 nm airborne particles from brake materials, *Aerosol
30 Science and Technology*, 51, 91-96, 2016.
- 31 Onischuk, A. A., Baklanov, A. M., Valiulin, S. V., Moiseenko, P. P., and Mitrochenko, V. G.: Aerosol diffusion
32 battery: The retrieval of particle size distribution with the help of analytical formulas, *Aerosol Science and
33 Technology*, doi: 10.1080/02786826.2017.1387642, 2017. 1-17, 2017.
- 34 Peineke, C., Attoui, M. B., and Schmidt-Ott, A.: Using a glowing wire generator for production of charged,
35 uniformly sized nanoparticles at high concentrations, *Journal of Aerosol Science*, 37, 1651-1661, 2006.
- 36 Raabe, O. G.: A general method for fitting size distributions to multicomponent aerosol data using weighted least-
37 squares, *Environmental Science and Technology*, 12, 1162-1167, 1978.
- 38 Ramachandran, G. and Kandlikar, M.: Bayesian analysis for inversion of aerosol size distribution data, *Journal of
39 Aerosol Science*, 27, 1099-1112, 1996.
- 40 Ramachandran, G. and Cooper, D. W. (Eds.): *Size distribution data analysis and presentation*, John Wiley & Sons,
41 New York, 2011.
- 42 Richards, F. J.: A flexible growth function for empirical use, *Journal of Experimental Botany*, 10, 290-301, 1959.
- 43 Schäfer, H., Sternin, E., Stannarius, R., Arndt, M., and Kremer, F.: Novel Approach to the Analysis of Broadband
44 Dielectric Spectra, *Physical Review Letter*, 76, 2177-2180, 1996.
- 45 Tang, Q., Cai, R., You, X., and Jiang, J.: Nascent soot particle size distributions down to 1 nm from a laminar
46 premixed burner-stabilized stagnation ethylene flame, *Proceedings of the Combustion Institute*, 36, 993-1000,
47 2017.
- 48 Tikhonov, A. N.: On the solution of ill-posed problems and the method of regularization, *Doklady Akademii Nauk
49 SSSR*, 151, 501-504, 1963.
- 50 Twomey, S.: Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect
51 estimation of particle size distributions, *Journal of Computational Physics*, 18, 188-200, 1975.
- 52 Ude, S. and de la Mora, J. F.: Molecular monodisperse mobility and mass standards from electrosprays of tetra-
53 alkyl ammonium halides, *Journal of Aerosol Science*, 36, 1224-1237, 2005.
- 54 Vanhanen, J., Mikkilä, J., Lehtipalo, K., Sipilä, M., Manninen, H. E., Siivola, E., Petäjä, T., and Kulmala, M.:
55 Particle Size Magnifier for Nano-CN Detection, *Aerosol Science and Technology*, 45, 533-542, 2011.
- 56 Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy, *SIAM Journal on
57 Numerical Analysis*, 14, 651-667, 1977.



- 1 Winklmayr, W., Wang, H.-C., and John, W.: Adaptation of the Twomey Algorithm to the Inversion of Cascade
- 2 Impactor Data, *Aerosol Science and Technology*, 13, 322-331, 1990.
- 3 Wu, J. J., Cooper, D. W., and Miller, R. J.: Evaluation of aerosol deconvolution algorithms for determining
- 4 submicron particle size distribution with diffusion battery and condensation nucleus counter, *Journal of*
- 5 *Aerosol Science*, 20, 477-482, 1989.
- 6
- 7

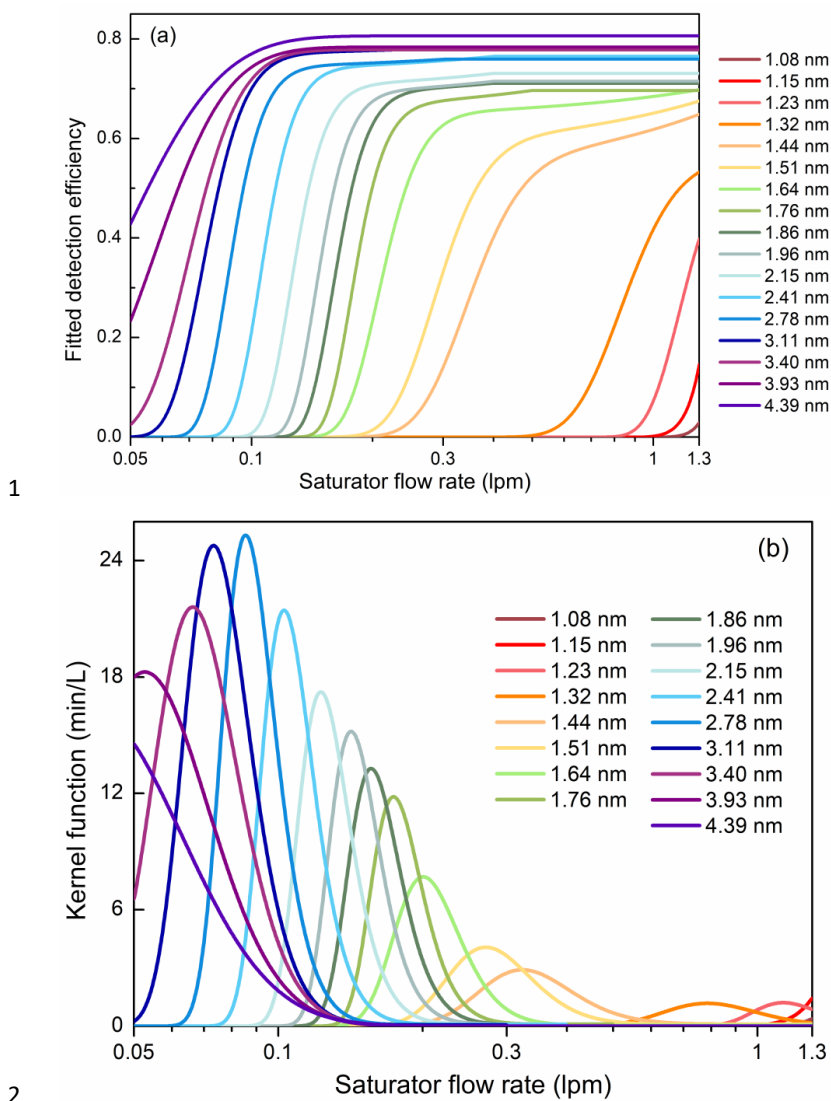
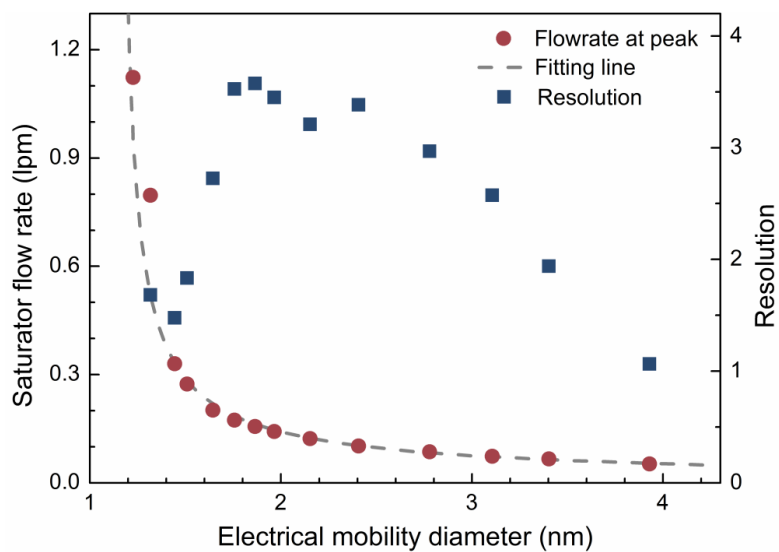
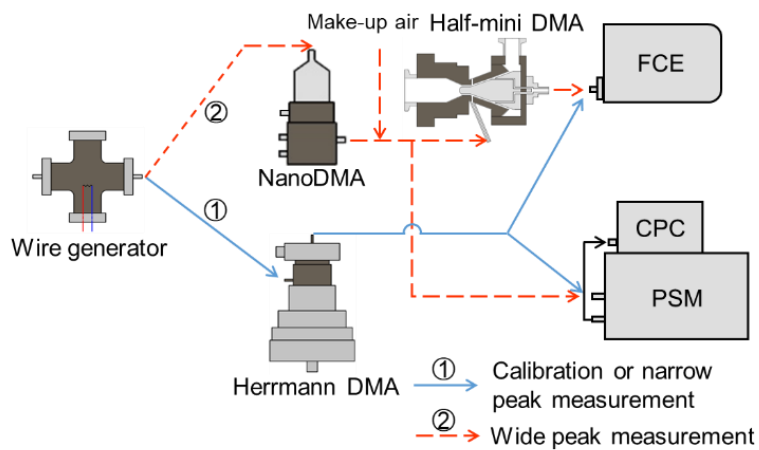


Figure 1 (a) The fitted detection efficiency curves according to calibration data. (b) The estimated kernel function curves according to the fitted detection efficiencies. The kernel function is equal to the derivative of the detection efficiency with the respect to the saturator flow rate.



1
2 **Figure 2** The saturator flow rate at kernel function peak and the resolution as functions of the particle diameter. Note that the resolution
3 is defined using the saturator flow rate, however, the horizontal axis is shown in the particle diameter corresponding to the peak
4 saturator flow rate for more straightforward understanding.
5



1
2
3
4

Figure 3 The experimental setup to calibration the PSM and test the inversion methods.

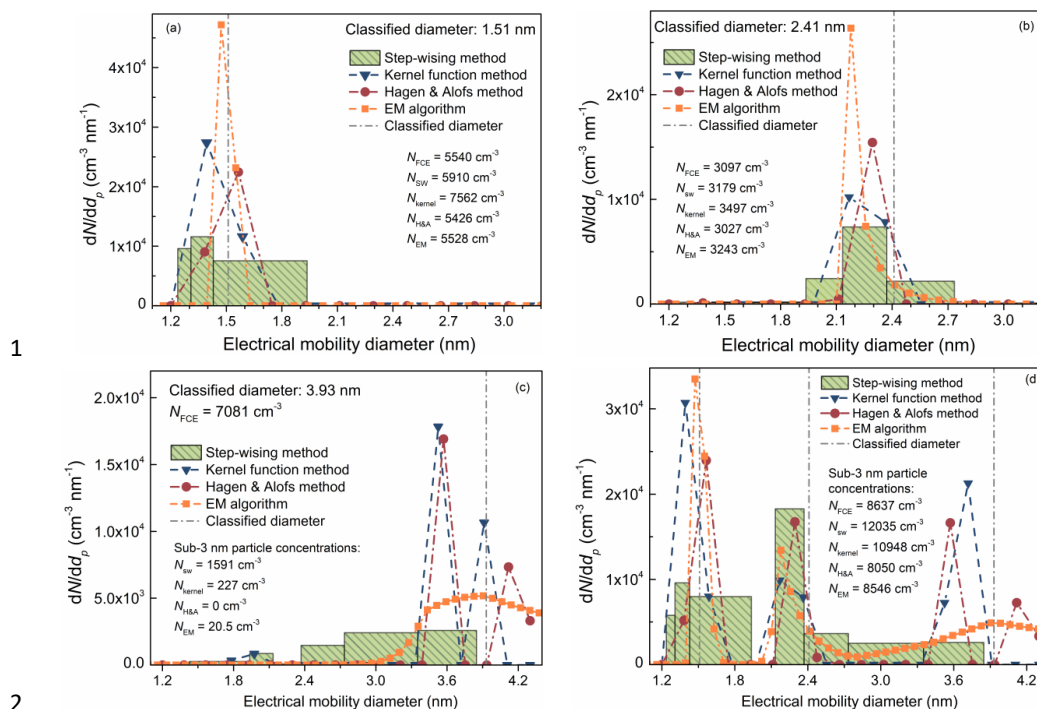
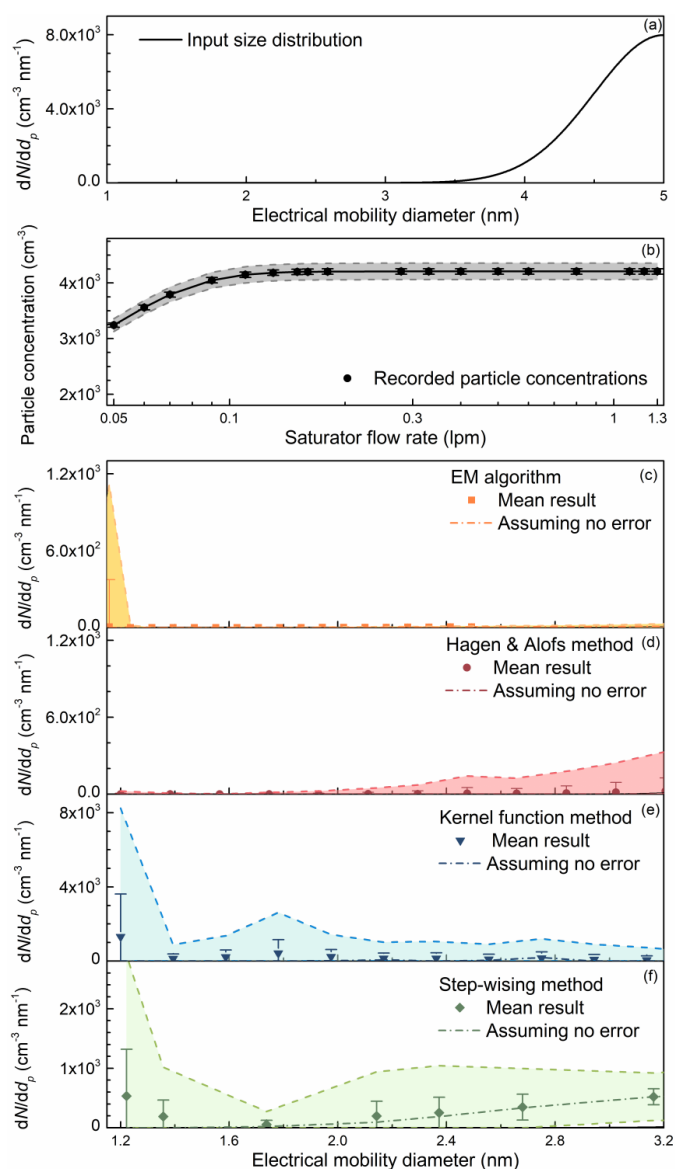
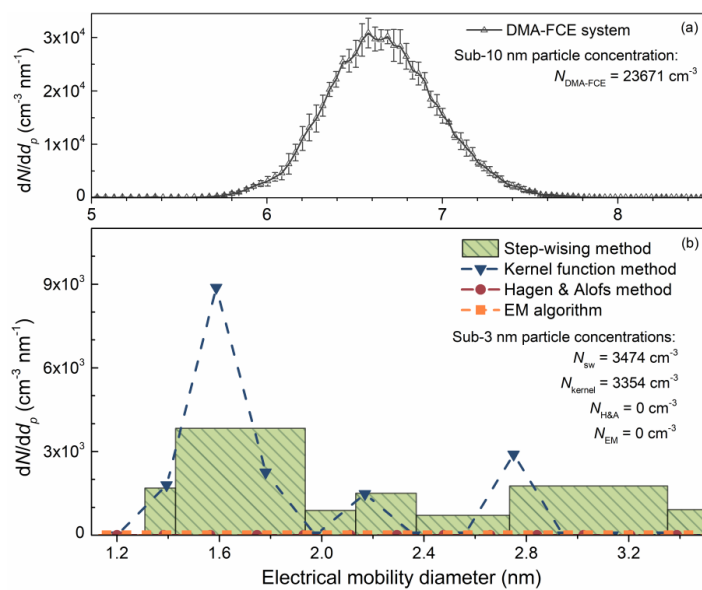


Figure 4 The recovered particle size distributions using different inversion methods when measuring monodisperse particles. FCE, SW, kernel, H&A, and EM are short for the Faraday cage electrometer, the step-wising method, the kernel function method, the H&A method, and the expectation-maximization algorithm, respectively. The number concentration detected by the reference FCE and the sum of recovered sub-3 nm particle concentration in each size bin are shown in the text. The size distributions in (d) were recovered using the sum of the recorded number concentrations in (a), (b), and (c), i.e., assuming the PSM was measuring 1.51, 2.41, and 3.93 nm particles simultaneously.

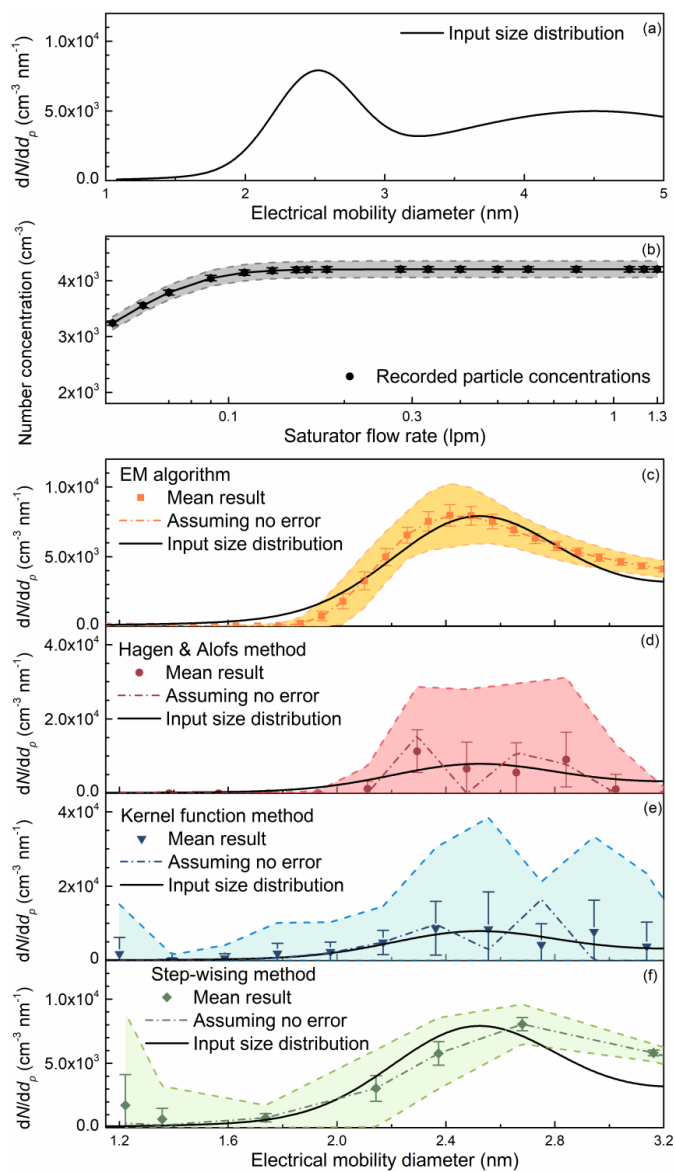


1
 2 **Figure 5** The recovered sub-3 nm particle size distributions simulated using the Monte Carlo method when the detected particles were
 3 larger than 3 nm. (a) The assumed true particle size distribution. (b) The simulated particle concentrations recorded by the PSM. The
 4 concentrations were assumed to fluctuate due to random errors. The particle size distributions were recovered using (c) the EM
 5 algorithm, (d) the H&A method, (e) the kernel function method, and (f) the step-wise method. The error bar represents the standard
 6 deviation of the recorded particle concentration or the recovered size distribution, and the shaded area indicates the range determined
 7 by three times the standard deviation. The dashed lines represent the inverted results assuming there were no random errors in the
 8 recorded particle number concentrations. Note that the scale of the vertical axis in (c-f) is different and the appearing possibility of
 9 recorded counts or the recovered size distribution is not uniform in the shaded area.

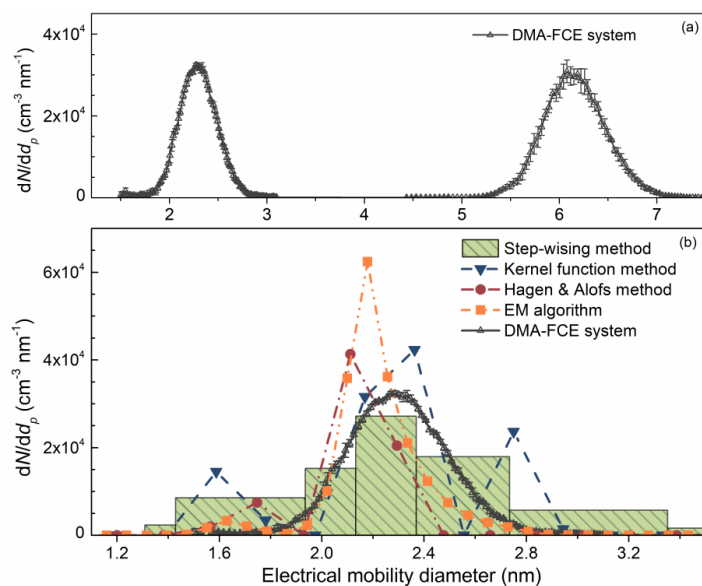
10



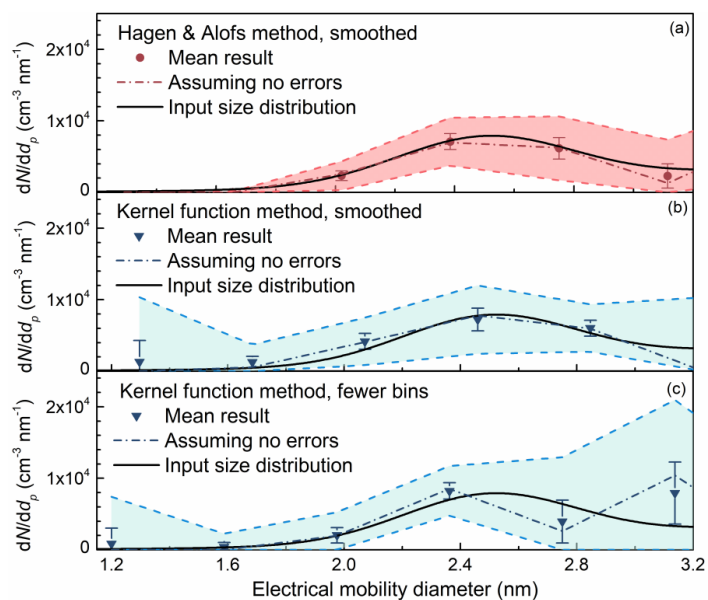
1
2 **Figure 6** The experimental testing results of the four inversion methods when the PSM was measuring particles larger than 3 nm. (a)
3 The particle size distribution detected by the reference halfmini DMA-FCE system. (b) The recovered particle distributions using
4 different inversion method.
5



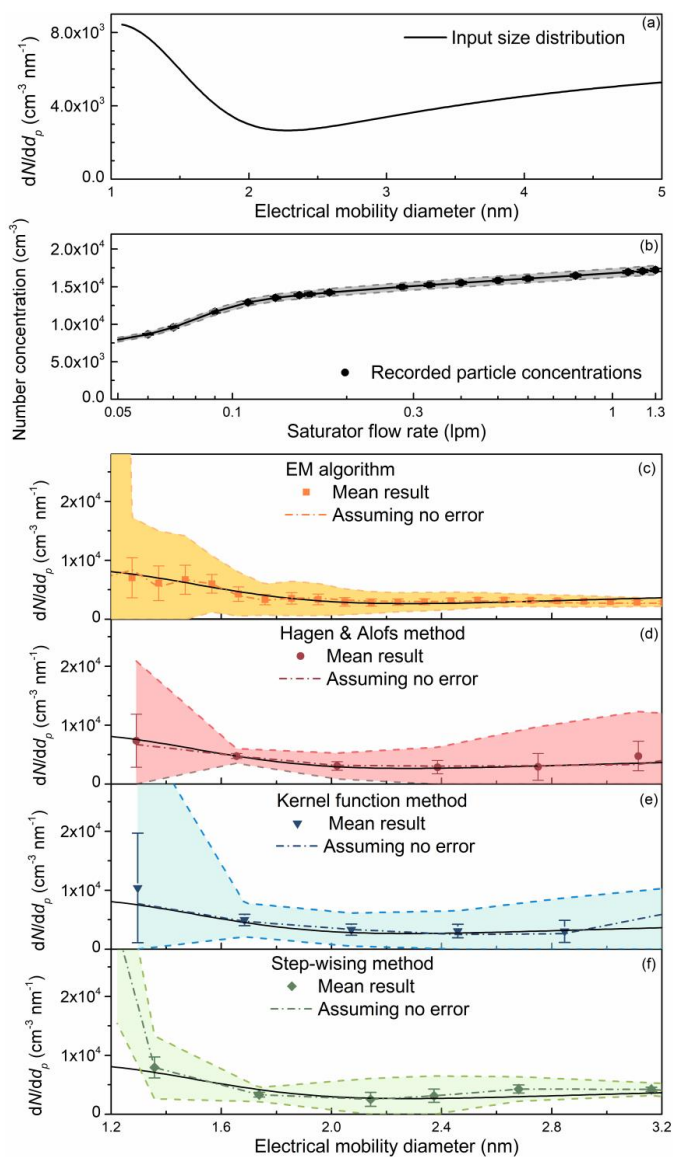
1
 2 **Figure 7** The input and recovered sub-3 nm particle size distributions simulated using the Monte Carlo method. Note the vertical axes
 3 in panel (c-f) are not the same.
 4



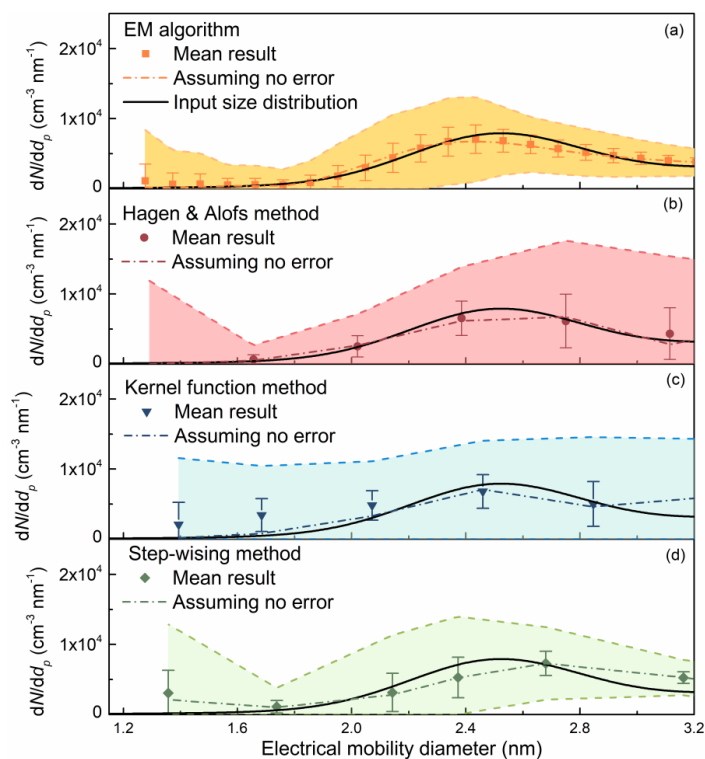
1
 2 **Figure 8** The experimental testing results of the four inversion methods when the PSM was measuring sub-3 nm particles with the
 3 influence of larger particles. The particle number concentrations for inversion and the particle size distribution detected using the
 4 DMA-FCE system were the sums of two separate experiments rather than real data obtained in a single experiment.
 5



1
2 **Figure 9** Comparisons of the inverted results using (a) the H&A method smoothing the particle size distribution via merging size bins;
3 (b) the kernel function method smoothing the particle size distribution via merging size bins; and (c) the kernel function method
4 assuming fewer discrete particle sizes in Eq. 5.
5



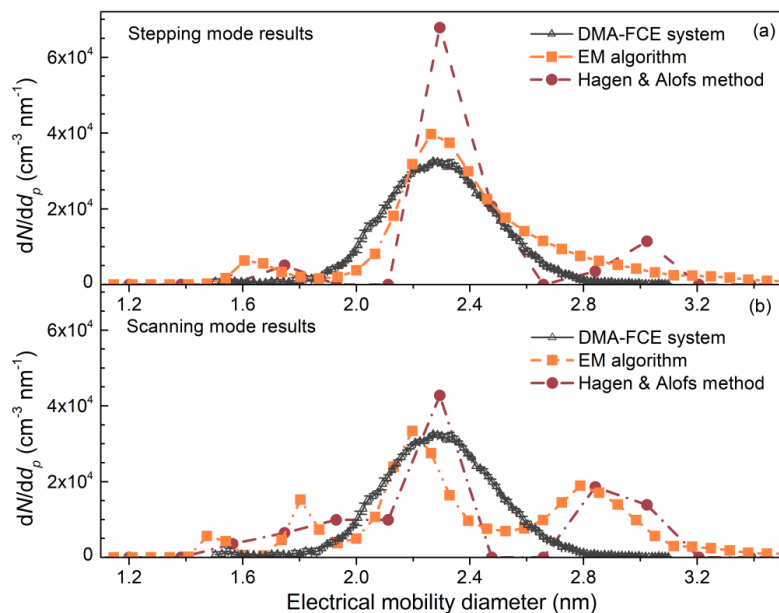
1
 2 **Figure 10** The input and recovered sub-3 nm particle size distributions simulated using the Monte Carlo method when the particle size
 3 distribution increases with the decreasing particle diameter.
 4



1
 2 **Figure 11** The recovered particle size distributions simulated using the Monte Carlo method when assuming the relative standard
 3 deviation of the recorded particle number concentration is 10%. The reported size bins smaller than 1.3 nm recovered using the kernel
 4 function method and the step-wise method are not shown because of the large uncertainties.
 5



1



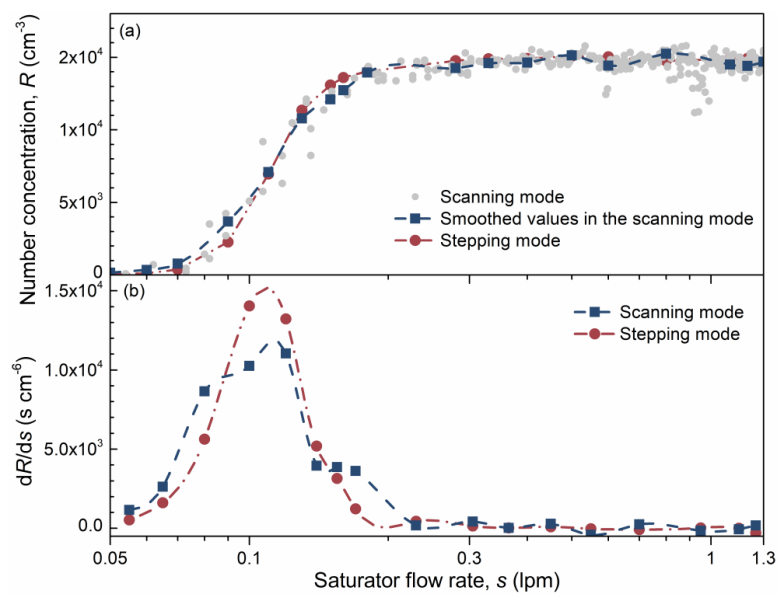
2

3

Figure 12 The recovered particle size distributions using the particle number concentration recorded in (a) the stepping mode and (b) the scanning mode.

4

5



1
2 **Figure 13** (a) The relationship between the recorded particle number concentration and the saturator flow rate in the scanning mode
3 and the stepping mode. (b) The derivative of number concentration with the respect to the saturator flow rate.
4