Manuscript No.: amt-2021-334

Title: Regularized inversion of aerosol hygroscopic growth factor probability density function: Application to humidity-controlled fast integrated mobility spectrometer measurements

We thank the anonymous referees for their valuable and constructive comments/suggestions on our manuscript. We have revised the manuscript accordingly and please find our point-to-point responses below.

Comments by Anonymous Referee #1:

General Comments:

The authors test of different inversion methods for aerosol hygroscopicity measurements with a specific focus on the humidity-controlled fast integrated mobility spectrometer. The manuscript is very well written, and concise. As inversion techniques other than least-square approaches have recently gained some attention in the community again (Petters, 2021; Sipkens et al., 2020a, 2020b; Stolzenburg et al., 2022), this manuscript certainly is another interesting approach focusing on an important problem, the retrieval of the GF-PDF. However, with respect to the above-mentioned literature, I have the feeling that the authors need to add additional work to their manuscript, such that it significantly adds to the body of knowledge and I can recommend publication in Atmos. Meas. Tech. I give my detailed suggestions below.

Responses: We thank the reviewer for the constructive comments and suggestions. Point-to-point responses to comments and questions are detailed below. Following the reviewer's suggestions, we examined the impact of additional Gaussian noises on the performance of inversion methods. The inversion methods were also tested with different forward and inverse models (i.e., assuming that the calibration of the FIMS/DMA is not perfect). We also applied the inversion algorithms to ambient measurement data and compared the performances. The new results and discussions are now included in the revised manuscript.

Detailed Comments:

1) The authors need to challenge their inversion technique with more difficult data. Using the same forward and inverse model is considered to be not sufficient in testing an inversion approach, leading to unrealistically good outputs (e.g. Colton and Kress, 2013). The authors need to run tests on their inversion where the forward and inverse model are different (i.e. assuming that the calibration of the FIMS is not perfect) and where statistics other than the counting error influence the result and their actual distribution can only be guessed (e.g. the standard deviation of an additional Gaussian error is different in forward and inverse model). Both cases are closer to real measurements and this would demonstrate the performance of the inversion algorithms under more challenging conditions.

Responses: The forward model and inverse model are not exactly identical. First, in the forward calculation, the HFIMS measurement R_i (i.e., the convolution of the GF-PDF, the DMA transfer function, and the WFIMS transfer function) is integrated using Eq. (4) with a much higher resolution of g (i.e., 120 bins over 0.8 - 2.0) while the inverse model casts the HFIMS kernel into a matrix with only 20 bins of g. In addition, the synthetic HFIMS measurements include noise due to counting statistics based on Poisson distribution.

Following the reviewer's suggestion, we also challenged our algorithm with different forward and inverse models to simulate the scenarios when DMA or WFIMS is not perfectly calibrated. The particle sizes measured by DMA and WFIMS are determined by the voltage and sheath flow, which can be calibrated straightforwardly. Therefore, the nonideality in DMA and WFIMS performance likely manifests in the deviation of instrument mobility resolution from the theoretical value. To test the performance of inversion algorithms for such scenarios when the transfer function of DMA or WFIMS is not fully calibrated, we generate the synthetic HFIMS measurements by perturbing DMA or WFIMS mobility resolution (i.e., R_{DMA} or R_{WFIMS}), while maintaining the theoretical R_{DMA} or R_{WFIMS} in the inverse model. The mobility resolution is perturbed by varying the ratio of sheath to aerosol flow for DMA or WFIMS (R_Q=$Q_{\mathrm{sh}}/Q_{\mathrm{a}}$) in the derivation of the transfer function. The default flow rate ratio for DMA and WFIMS are 10 and 50, respectively. Figures 1 and 2 show the inversion results when DMA R_Q in the forward model is varied from 8 to 12 while WFIMS R_Q is maintained at the actual value of 50 and when DMA R_Q is maintained at 10 while WFIMS R_Q is varied from 40 to 60. The results are based on inversions of 500 sets of synthetic HFIMS measurements (with the noise of counting statistics included) using Twomey's method. The average residual (χ^2), the GF-PDF error (γ^2), and the smoothness (ξ), all showed very minor variation with DMA or WFIMS R_Q used in the forward model, suggesting negligible impacts on Twomey inversion results due to imperfect calibration of DMA and WFIMS resolution. A possible explanation is that typical GF-PDFs of ambient aerosol particles are relatively broad such that the inverted GF-PDF is insensitive to DMA and WFIMS resolution. The impact of R_Q on other nonparametric methods was also investigated and found negligible.

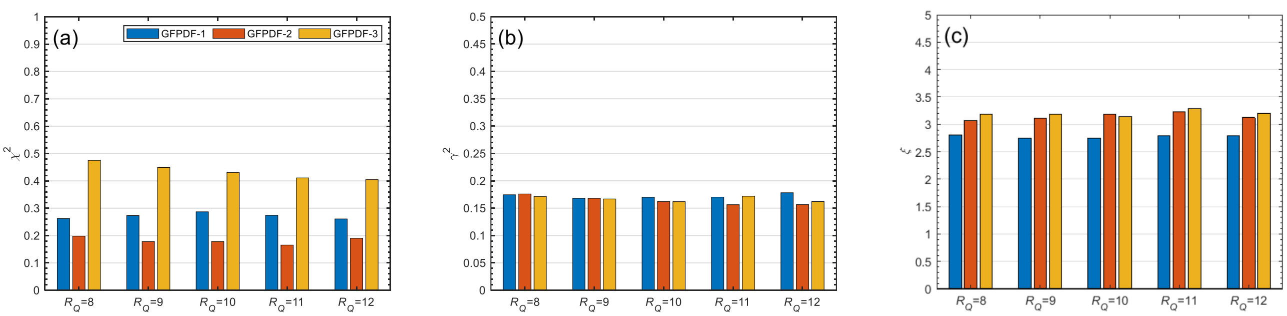


Figure 1. The reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of GF-PDF inverted using Twomey's method as a function of DMA R_Q used to calculate DMA transfer function in the forward model (WFIMS R_Q maintained at the actual value of 50). Actual DMA and WFIMS R_Q values of 10 and 50 are used to derive transfer functions in the inverse model (i.e., calculation of the inversion matrix). The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on the inversion of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

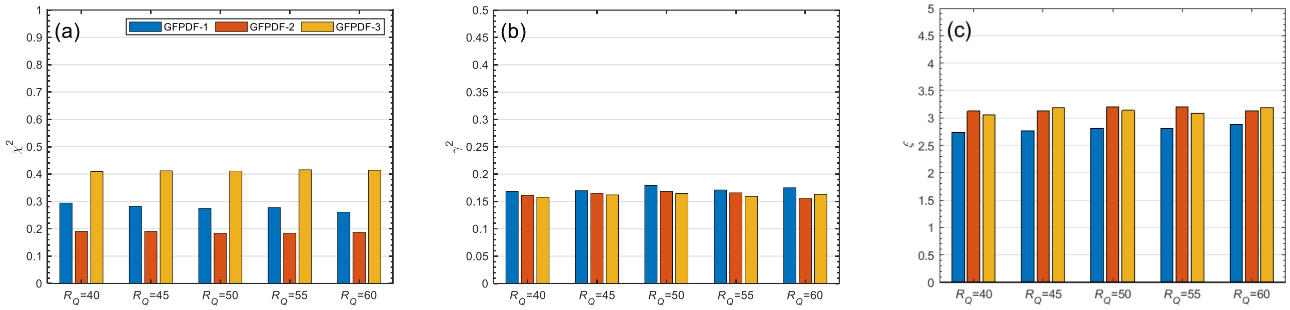


Figure 2. The reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of GF-PDF inverted using Twomey's method as a function of WFIMS R_Q used to calculate WFIMS transfer function in the forward model (DMA R_Q maintained at the actual value of 50). Actual DMA and WFIMS R_Q values of 10 and 50 are used to derive transfer functions in the inverse model (i.e., calculation of the inversion matrix). The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on the inversion of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

In addition to noise due to counting statistics, we also included additional Gaussian noise (e.g., due to variations of sample flow rate) ranging from 1% to 10% in generating the synthetic HFIMS data and examined the impact of the additional noise on inverted GF-PDFs. Figure 3 shows that Gaussian noises up to 10% have negligible impact on the reconstruction residual (χ^2), the error (γ^2), and the smoothness (ξ) of GF-PDFs inverted using Twomey's method. Similarly, the impact is also negligible for GF-PDF inverted using unweighted LSQ and 0th, 1st, and 2nd order Tikhonov regularizations (not shown). The negligible impacts indicate that the noise of typical HFIMS measurements is dominated by counting statistics. The above results and discussion are detailed in a new section (Section 3.2) titled "Effect of measurement uncertainties" in the revised manuscript (line 252-302).

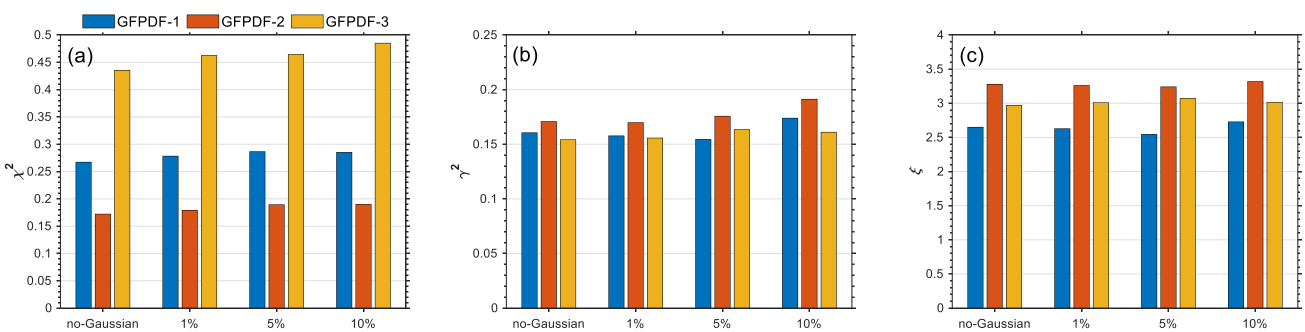


Figure 3. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the degree of smoothing, ξ (**c**) of GF-PDFs inverted using Twomey's methods from synthetic HFIMS data with additional Gaussian noises of different levels (i.e., none, 1%, 5%, and 10%). Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

2) The authors should at least present an application of their algorithms to real measurement data, such that the reader can judge how big differences we would expect in the retrieval of the GF-PDF for typical hygroscopicity measurements in the ambient air.

Responses: Following the reviewer's suggestion, we apply the nonparametric inversion methods to ambient HFIMS measurements, and the results are compared in Fig. 4. The HFIMS responses reconstructed from GF-PDF inverted using unregularized LSQ, Tikhonov, and Twomey's methods generally match the measurement (black circle) well. The GF-PDF at 85% RH for ambient 35 nm particles consist of a smaller less-hygroscopic mode and a larger more-hygroscopic mode. As expected, the HFIMS response reconstructed from LSQ inverted GF-PDF has the minimum deviation from the actual measurement whereas the GF-PDF exhibits more oscillations near the tail of the second mode. These oscillations create a small third mode that is absent from the smoother GF-PDFs inverted using regularized methods (i.e., Tikhonov and Twomey's methods). GF-PDF inverted using Twomey's method and 0th Tikhonov clearly distinguish the two growth factor modes. In comparison, the two modes become more overlapped in GF-PDF inverted using 1st and 2nd Tikhonov regularization, due to additional and possibly excessive regularization.

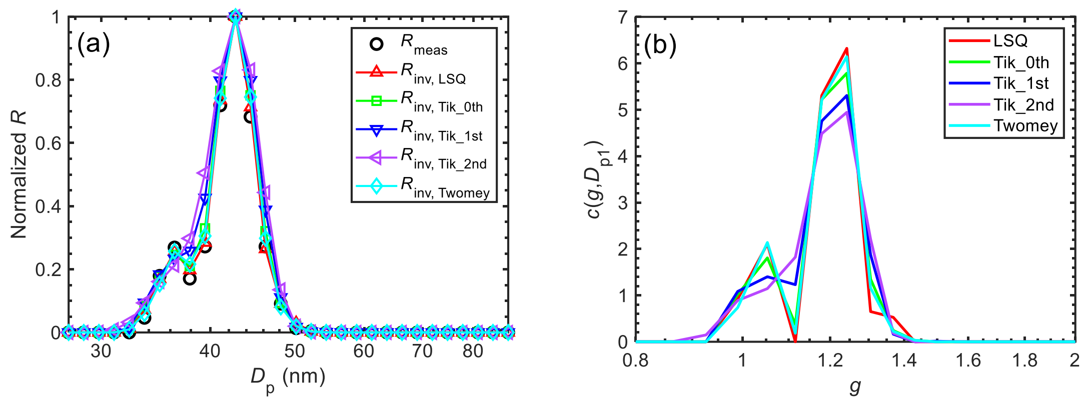


Figure 4. (**a**) Comparison between the HFIMS measured response (black circle) and the responses (marked lines) reconstructed from GF-PDF derived using different methods for 35 nm ambient aerosol at 85% RH. (**b**) Inverted GF-PDFs using different methods.

We also examined the statistics of the reconstruction residual and the smoothness of GF-PDF inverted from 3-day HFIMS measurements using the listed nonparametric methods. Among all nonparametric inversion methods, unregularized LSQ leads to the lowest reconstruction residual but the worst smoothness (Fig. 5). As regularizations are introduced in the Tikhonov algorithms, the inverted GF-PDFs become smoother at the expense of increased reconstruction residuals. The Tikhonov regularized solutions strongly depend on the regularization parameter λ. In this study, the value of λ has been derived using three approaches, including (1) the L-curve, (2) the Hanke-Raus rule, and (3) comparison of inverted GF-PDF with the true solution. Note the 3rd approach (i.e., comparison of inverted GF-PDF with the true solution) is not possible for ambient measurements. Inversions of synthetic data show that the L-curve approach generally underestimates the regularization parameter (Fig. 5 in the manuscript), resulting in insufficiently regularized solutions (Naseri et al., 2021). For the 3-day ambient measurements, when λ is derived using the L-curve approach, the reconstruction residuals for the GF-PDF inverted using Tikhonov algorithms are very close to those of the unregularized LSQ, consistent with underestimated λ values (Fig. 5a and d). In contrast, Tikhonov regularizations with λ value determined using the Hanke-Raus rule tend to over-smooth solutions due to overestimated λ values, resulting in significantly increased errors in reconstructed HFIMS measurements (Fig. 5b and e). The 3-day ambient measurements are also inverted using Tikhonov algorithms with an empirical λ value of 0.03 (Fig. 5c and f), which corresponds to the mean value of optimized λ values (i.e., derived using the 3rd approach) for the synthetic HFIMS data. The inverted GF-PDF shows improved smoothness compared to the solution from the LSQ method, without introducing excessive reconstruction errors. While the empirical λ value appears to work quite well for the 3-day measurements, using this fixed regularization parameter may not be appropriate for other ambient measurements. For Twomey's method, both the reconstruction residual and the smoothness are between those based on the 0th order and 1st order Tikhonov regularizations with the empirical regularization parameter (λ = 0.03), suggesting an appropriate trade-off between the GF-PDF smoothness and the fidelity in reproducing the HFIMS measurements. Note that the statistics of the GF-PDF error cannot be derived as the actual GF-PDF of ambient aerosols are unknown. As a result, it is difficult to draw a definite conclusion regarding which method has the best performance in retrieving the GF-PDF based on the ambient measurements. The above results and discussion have been added in Section S5 of the revised supplemental information.

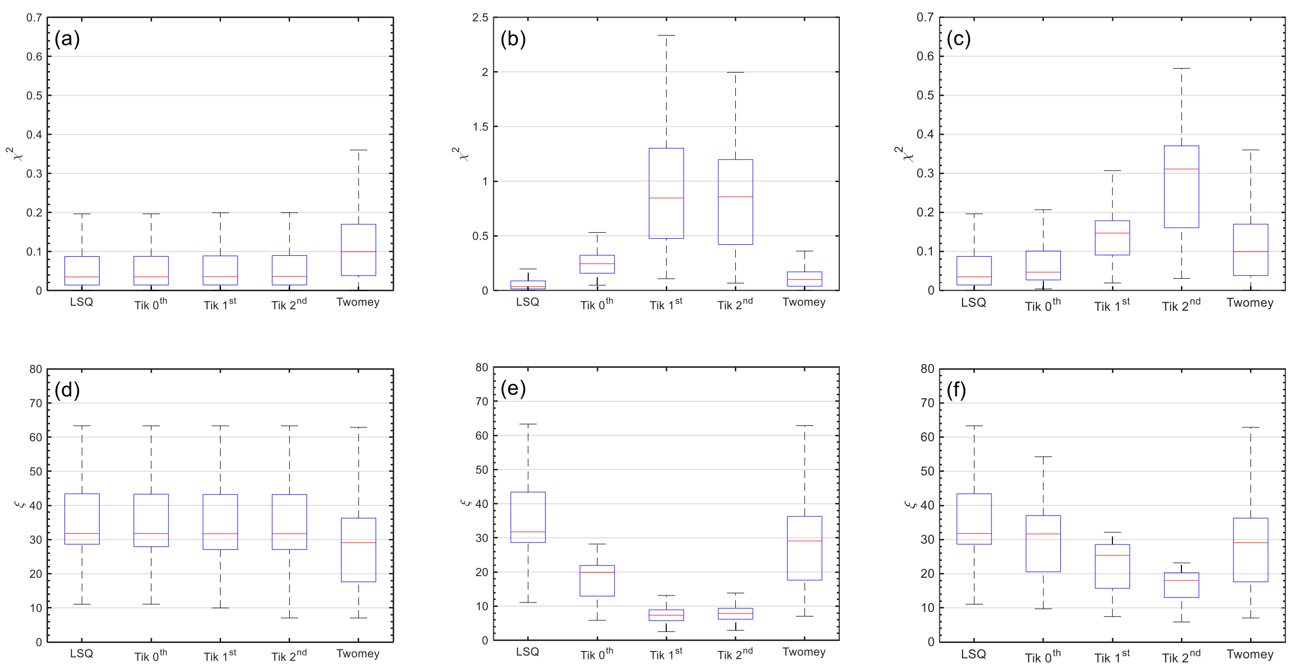


Figure 5. Comparison of reconstruction residual, χ^2 (**a, b, c**) and the smoothness, ξ (**d, e, f**) of GF-PDFs inverted using different methods, including LSQ, 0th, 1st, and 2nd order Tikhonov regularizations, and Twomey's method), based on 3-day HFIMS measurements of ambient

35 nm aerosols at five different RH levels (i.e., 20%, 40%, 60%, 75%, and 85%). The Tikhonov regularization parameters are derived using the L-curve approach (**a, d**), the Hanke-Raus rule (**b, e**), and an empirical value of 0.03 (**c, f**), respectively.

3) Data availability. I think it is nowadays almost standard that the authors publish their code openly along with the corresponding manuscript (again referring to Petters, 2021; Sipkens et al., 2020b; Stolzenburg et al., 2022). If the authors want to ensure reuse of their methods this is highly encouraged.

Responses: We have organized our code and uploaded it with the datasets related to this manuscript to a GitHub repository (https://github.com/zjs023/Regularized_inversion_HFIMS).

4) I understand that the manuscript focuses on the inversion for the FIMS, but this instrument is not available to many researchers doing hygroscopicity studies. In my opinion, it would be highly beneficial if the authors could also test their inversion on classical TDMA data.

Responses: While we agree that it would be beneficial, a systematic investigation and comparison of the inversion methods for TDMA measurement are beyond the scope of this manuscript, and they will be a subject of a future study. Nevertheless, we have derived the equations for the inversion of HTDMA data and detailed them in Section S1 of supplementary information. Following the reviewer's suggestion, we have published the code, which could be conveniently modified to invert classical TDMA data (i.e., replace the subfunction that calculates the WFIMS transfer function with one for the transfer function of the 2nd DMA in a TDMA system.)

5) L.153 ff.: Please note that Tikhonov regularization (and LSQ) implicitly assume Gaussian statistics on the measurement noise, but that the underlying statistics used to generate the measurement data is of Poisson nature in this manuscript (see also comment above on imposing an additional Gaussian uncertainty to challenge the methods under more difficult conditions). You can refer to Stolzenburg et al. (2022) here who developed a Poisson approach for regularization.

Responses: We thank the reviewer for the comment. We agree that underlying statistics used to generate the HFIMS measurement are of Poisson nature whereas Tikhonov regularization (and LSQ) implicitly assumes Gaussian statistics on the measurement noise. For very low count numbers, the counting statistics may not be well approximated by a Gaussian distribution. We have clarified this and referenced Stolzenburg et al. (2022) in the revised manuscript (line 153-154).

6) L.159: Please be aware that the matrices described here impose specific boundary conditions (Donatelli and Reichel, 2014). I assume that Dirichlet boundary conditions have been used. Please specify this, as it is often neglected in inversion procedures that the boundary condition is quite important for the shape of the solution.

Responses: Rectangular regularization matrices were used for the first and second-order Tikhonov regularization inversions (Hansen, 1994). We also tried square regularization matrices incorporating the Dirichlet or Neumann boundary conditions. Figure 6 shows the reconstruction residual, the GF-PDF error, and the degree of smoothing based on the first and second-order Tikhonov inversions using a square regularization matrix incorporating the Dirichlet boundary condition. The results are very similar to those based on rectangular matrices (Fig. 4 in the manuscript).

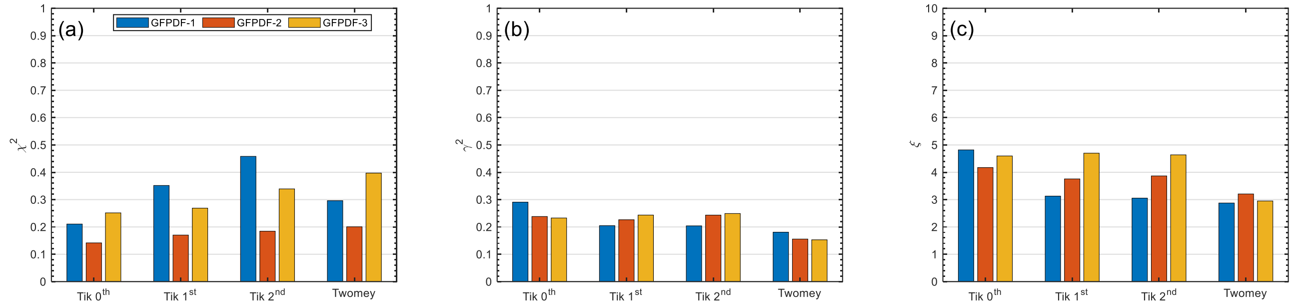


Figure 6. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), the degree of smoothing, ξ (**c**) inverted with square regularization matrices incorporating the Dirichlet boundary condition. Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow).

7) L.161 ff.: Which algorithm was used to locate the L-curve corner?

Responses: In this study, the regularization parameter λ is identified by using the "L_curve" algorithm (a routine of the regularization tool software package developed by Dr. Hansen, http://people.compute.dtu.dk/pcha/Regutools/index.html) to find the corner of the L-curve with the maximum curvature, as detailed in Hansen (1994). We clarified this in the revised manuscript (line 159-161).

8) L.242: The discussion about the computing time is difficult, as it crucially relies on the used device. I think it is better to make Figure 2d with relative computing times.

Responses: We agree the computing time strongly depends on the device. On the other hand, the implementing time of "lsqnonneg" function on our desktop (Intel's 8th generation processor Core i7-8700) is very close to 1 second, therefore, Figure 2d (Fig. 4d in the revised manuscript) essentially presents the relative computer time in reference to the LSQ method. We also think it is useful to provide the computing time required on a typical desktop computer. We have clarified this in the revised manuscript (line 327-331).

8) L.250 ff.: The measure smoothness is defined in Eq. (14) by the 2nd order derivative, so it is not surprising to me, that the 2nd order Tikhonov regularization achieves the best smoothness measure, as it optimizes for the same quantity.

Responses: We agree. The 2nd order Tikhonov regularization achieves the best smoothness measure, most likely because its regularization is based on the same quantity, as demonstrated by applications of the inversions from ambient HFIMS measurements (Section S5 in the SI).

References

Hansen, P. C.: REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems, Numerical Algorithms, 6, 1-35, 10.1007/BF02149761, 1994.

Naseri, A., Sipkens, T. A., Rogak, S. N., and Olfert, J. S.: An improved inversion method for determining two-dimensional mass distributions of non-refractory materials on refractory black carbon, Aerosol Science and Technology, 55, 104-118, 10.1080/02786826.2020.1825615, 2021.

Manuscript No.: amt-2021-334

Title: Regularized inversion of aerosol hygroscopic growth factor probability density function: Application to humidity-controlled fast integrated mobility spectrometer measurements

We thank the anonymous referees for their valuable and constructive comments/suggestions on our manuscript. We have revised the manuscript accordingly and please find our point-to-point responses below.

Comments by Anonymous Referee #2:

General Comments:

The authors apply different inversion methods to invert the aerosol GR-PDF from the measured signals from synthetic HFIMS signals. They found that for the few test cases, Markowski-Towmey's method generally outperforms other methods. By doing this, they convincingly improved the data inversion of HFIMS data and promisingly HTDMA data, which were mainly based on predefined size distributions or least square methods. This well-written manuscript is easy to follow. I recommend it to be published in Atmospheric Measurement Techniques, However, a major revision is necessary to convincingly demonstrate that the data inversion of HFIMS (and HTDMA) is improved. I feel that the authors are too optimistic about the representativity of their limited synthetic data on real laboratory experiments and atmospheric measurements. Further, this manuscript will have a broader impact on the community if its outcomes (e.g., inversion codes) can be readily used for HTDMA measurements. My detailed comments are given below.

Responses: We thank the reviewer for the constructive suggestions and comments. Point-to-point responses to comments and questions are detailed below. Following the reviewer's suggestions, we examined the impact of additional Gaussian noises on the performance of inversion methods. We also applied the inversion algorithms to ambient measurement data and compared the performances. Additional analyses were also carried out to elucidate why Twomey's method is statistically better than the Tikhonov regularization methods. The new results and discussions are now included in the revised manuscript.

Detailed Comments:

1) More tests and/or discussions are needed to provide supports for the argument that Towmey's method outperforms other tested inversion methods. The three test cases are perhaps sufficient to show that Towmey's method is better than least square methods because the least square methods are notorious for solving ill-conditioned problems. However, the reason why Towmey's method is better than the Tikhonov regularization methods needs more clarification and/or data to support.

Responses: We thank the reviewer for this comment. To elucidate why Twomey's method is better than the Tikhonov regularization methods, we compare the results of inversions based on Twomey's method and 1st order Tikhonov regularization with the regularization parameter derived using three different approaches. The first approach (i.e., the L-curve approach, Hansen and O'Leary, 1993), derives the λ by seeking a trade-off between minimizing the residual term and minimizing the regularization term (i.e., roughness of the solution). The 2nd approach is based on the Hanke-Raus rule, which selects the λ

value that minimizes the λ-dependent residual term $\frac{1}{\lambda}\left\|\mathbf{Mc}^{\boldsymbol{Tik}}(\lambda)-\mathbf{R}\right\|_{\mathbf{2}}$ (Hanke and Raus, 1996; Sipkens et al., 2020). In the 3rd approach, the value of λ is optimized by comparing the inverted GF-PDF with the correct solution, i.e., minimizing the Euclidean distance between the inverted and the pre-defined GF-PDF. Therefore, inversion based on λ derived using the 3rd approach represents the best possible performance of the Tikhonov regularization. It is worth noting that the 3rd approach is not possible for real measurements, as the true GF-PDF is unknown. As shown in Fig. 1, the Tikhonov regularized solution strongly depends on the regularization parameter λ. The Tikhonov regularization with the optimized λ (derived using approach #3) provides the most accurate solution (i.e., lowest GF-PDF error) as expected, and outperforms Twomey's method. However, when λ derived using the L-curve approach or the Hanke-Raus rule is used, GF-PDF inverted using 1st order Tikhonov regularization generally has a larger error (i.e., γ^2) than that inverted using Twomey's method. The above comparisons indicate that while Tikhonov regularization can outperform Twomey's method in theory, the optimal regularization parameter λ cannot be obtained reliably using existing methods in practice, leading to inferior performance than Twomey's method. For example, the L-curve approach does not work well if the curvature of the L-curve is negative everywhere, and in such scenario, the leftmost point (i.e., with smaller λ) on the L-curve is taken as the corner (Hansen, 1994), leading to insufficient regularizations of the solution (Naseri et al., 2021). On the other hand, the Hanke-Raus rule often chooses a much larger λ compared with the optimal value, which results in over-smoothed solutions potentially with even larger errors. We also carried out similar comparisons of Twomey's method with 0th and 2nd order Tikhonov regularizations with λ values derived using the three different approaches, and the results are consistent. We have included the above comparison and discussion in the revised manuscript (line 338-374).

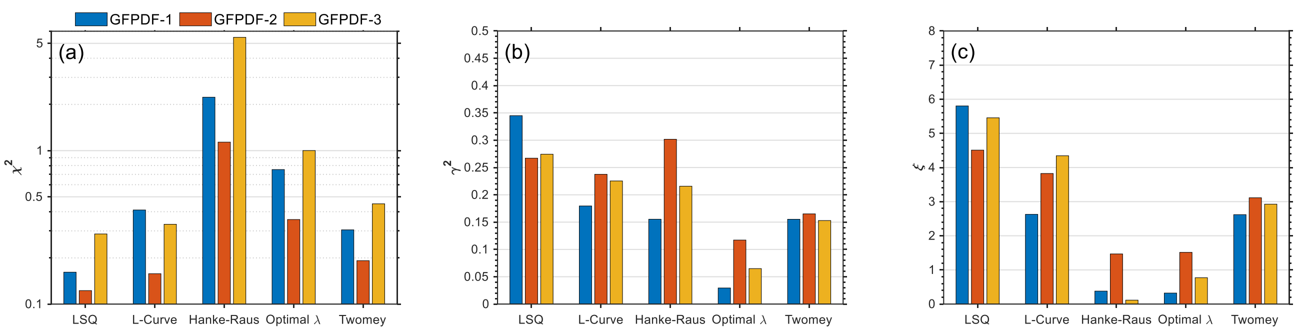


Figure 1. The reconstruction residual, $\boldsymbol{\chi}^{\mathbf{2}}$ (a), the GF-PDF error, $\boldsymbol{\gamma}^{\mathbf{2}}$ (b), and the smoothness, $\boldsymbol{\xi}$ (c) of GF-PDF inverted using LSQ, 1st order Tikhonov regularization with the regularization parameter derived from three different approaches (L-curve, Hanke-Raus rule, and optimized $\boldsymbol{\lambda}$), and Twomey's method. The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

2) The authors need to show the performance of Towmey's method with at least one dataset from either laboratory experiments or atmospheric measurements. Estimating the measurement uncertainties with only the counting uncertainties typically *underestimates the total uncertainties. Despite this, I am not concerned about the applicability of Towmey's to real datasets and its better performance of the than least square methods.*

Responses: Following the reviewer's suggestion, we apply the nonparametric inversion methods to ambient HFIMS measurements, and the results are compared in Fig. 2. The HFIMS responses reconstructed from GF-PDF inverted using unregularized LSQ, Tikhonov, and Twomey's methods generally match the measurement (black circle) well. The GF-PDF at 85% RH for ambient 35 nm particles consist of a smaller less-hygroscopic mode and a larger more-hygroscopic mode. As expected, the HFIMS response reconstructed from LSQ inverted GF-PDF has the minimum deviation from the actual measurement whereas the GF-PDF exhibits more oscillations near the tail of the second mode. These oscillations create a small third mode that is absent from the smoother GF-PDFs inverted using regularized methods (i.e., Tikhonov and Twomey's methods). GF-PDF inverted using Twomey's method and 0^{th} Tikhonov clearly distinguish the two growth factor modes. In comparison, the two modes become more overlapped in GF-PDF inverted using 1^{st} and 2^{nd} Tikhonov regularization, due to additional and possibly excessive regularization.

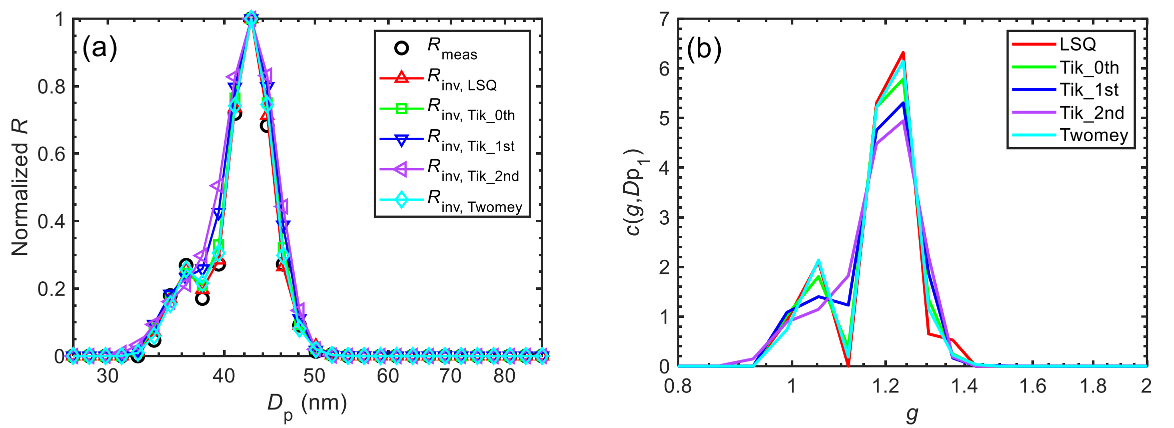


Figure 2. (**a**) Comparison between the HFIMS measured response (black circle) and the responses (marked lines) reconstructed from GF-PDF derived using different methods for 35 nm ambient aerosol at 85% RH. (**b**) Inverted GF-PDFs using different methods.

We also examined the statistics of the reconstruction residual and the smoothness of GF-PDF inverted from 3-day HFIMS measurements using the listed nonparametric methods. Among all nonparametric inversion methods, unregularized LSQ leads to the lowest reconstruction residual but the worst smoothness (Fig. 3). As regularizations are introduced in the Tikhonov algorithms, the inverted GF-PDFs become smoother at the expense of increased reconstruction residuals. The Tikhonov regularized solutions strongly depend on the regularization parameter λ. In this study, the value of λ has been derived using three approaches, including (1) the L-curve, (2) the Hanke-Raus rule, and (3) comparison of inverted GF-PDF with the true solution. Note the 3^{rd} approach (i.e., comparison of inverted GF-PDF with the true solution) is not possible for ambient measurements. Inversions of synthetic data show that the L-curve approach generally underestimates the regularization parameter (Fig. 5 in the manuscript), resulting in insufficiently regularized solutions (Naseri et al., 2021). For the 3-day ambient measurements, when λ is derived using the L-curve approach, the reconstruction residuals for the GF-PDF inverted using Tikhonov algorithms are very close to those of the unregularized LSQ, consistent with underestimated λ values (Fig. 3a and d). In contrast, Tikhonov regularizations with λ value determined using the Hanke-Raus rule tend to over-smooth solutions due to overestimated λ values, resulting in significantly increased errors in reconstructed HFIMS measurements (Fig. 3b and e). The 3-day ambient measurements are also inverted using Tikhonov algorithms with an empirical λ value of 0.03 (Fig. 3c and f), which corresponds to the mean value of optimized λ values (i.e., derived using the 3rd approach) for the synthetic HFIMS data. The inverted GF-PDF shows improved smoothness compared to the solution from the LSQ method, without introducing excessive reconstruction errors. While the empirical λ value appears to work quite well for the 3-day measurements, using this fixed regularization parameter may not be appropriate for other ambient measurements. For Twomey's method, both the reconstruction residual and the smoothness are between those based on the 0th order and 1st order Tikhonov regularizations with the empirical regularization parameter (λ = 0.03), suggesting an appropriate trade-off between the GF-PDF smoothness and the fidelity in reproducing the HFIMS measurements. Note that the statistics of the GF-PDF error cannot be derived as the actual GF-PDF of ambient aerosols are unknown. As a result, it is difficult to draw a definite conclusion regarding which method has the best performance in retrieving the GF-PDF based on the ambient measurements. The above results and discussion have been added in Section S5 of the revised supplemental information.

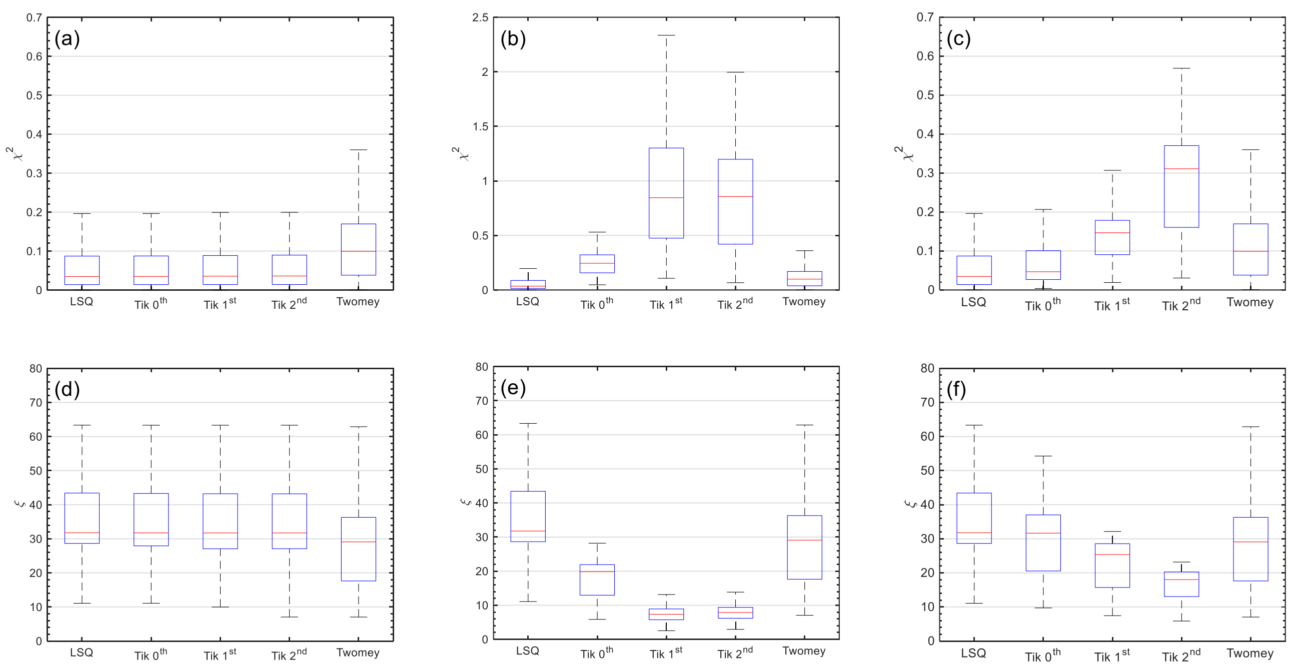


Figure 3. Comparison of reconstruction residual, χ^2 (**a, b, c**) and the degree of smoothing, ξ (**d, e, f**) of inverted GF-PDFs using different inversion methods (i.e., LSQ, Tikhonov of 0, 1, 2-th order, and Twomey's method), based on 3-day HFIMS measurements of ambient

aerosols of 35 nm at five different RH levels (20%, 40%, 60%, 75%, and 85%). The Tikhonov regularization parameters are derived using the L-curve approach (**a, d**), the Hanke-Raus rule (**b, e**), and an empirical value of 0.03 (**c, f**), respectively.

We agree that estimating the measurement uncertainties with only the counting uncertainties underestimates the total uncertainties, which also include the system noises (e.g., variations of the sample flow). In addition to noise due to counting statistics, we also included additional Gaussian noise (e.g., due to variations of sample flow rate) ranging from 1% to 10% in generating the synthetic HFIMS data and examined the impact of the additional noise on inverted GF-PDFs. Figure 4 shows that Gaussian noises up to 10% have negligible impact on the reconstruction residual (χ^2), the error (γ^2), and the smoothness (ξ) of GF-PDFs inverted using Twomey's method. Similarly, the impact is also negligible for GF-PDF inverted using unweighted LSQ and 0th, 1st, and 2nd order Tikhonov regularizations (not shown). The negligible impacts indicate that the noise of typical HFIMS measurements is dominated by counting statistics. The above results and discussion are detailed in a new section (Section 3.2) titled "Effect of measurement uncertainties" in the revised manuscript (line 252-302).

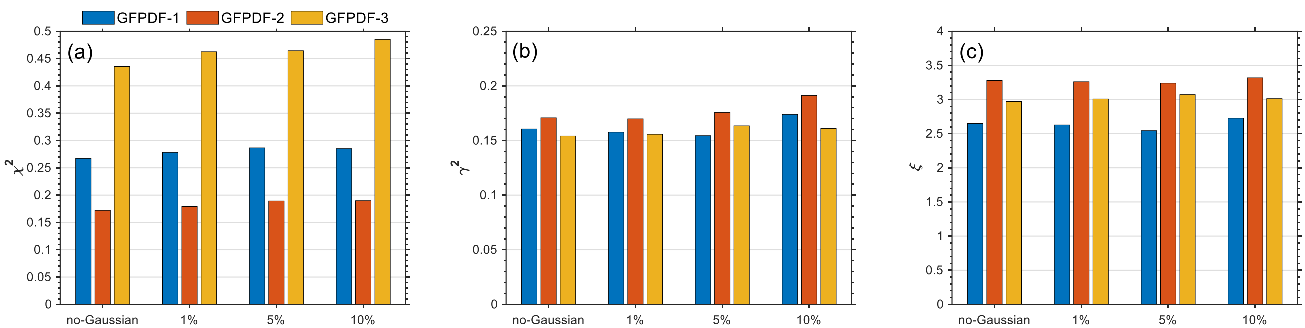


Figure 4. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the degree of smoothing, ξ (**c**) inverted GF-PDFs using Twomey's methods with additive Gaussian noises of different levels (i.e., none, 1%, 5%, and 10%). Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

3) Lines 30 - lines 115. The working principles of HFIMS are well summarized. However, they can also be removed or shortened to make space for more tests and discussion, as long as the inversion problem (e.g., Eq. 4) is clearly proposed.

Responses: We thank the reviewer for the comment. We did not include specific details of the HFIMS (i.e., the general principle and instrument setup), which has been presented in previous studies (Pinterich et al., 2017; Zhang et al., 2021). We summarized existing parametric inversion methods for both the HTDMA and HFIMS (i.e., ML and PL least-squares fitting). In Sect. 2.1, we present the mathematical derivation of both forward and inverse models, which are the key components of this study.

4) Line 116, Eq. 4. Please consider adding an error term (É) to Eq. 4 and other related equations to emphasize that the main challenge of data inversion is to deal will the uncertainties. The least-square methods are supposed to work pretty well if there is no error in the inversion problem as presented in Eq. 4.

Responses: We have clarified this in the revised manuscript (line 112-125).

5) Line 125. "The integration can be written as......". I recommend replacing "written as" with "approximated by". Discretizing a continuous distribution is also a step of inversion and there are inversion algorithms using improved discretization methods (e.g., Hagen and Alofs, doi.org/10.1080/02786828308958650).

Responses: We thank the reviewer for this constructive suggestion. “approximated by” is much more rigorous. We made the change in the revised manuscript (line 121).

6) As far as I am concerned, Towmey's method does not mathematically guarantee convergence of the inversion results. Optimized adjusting factor(s) are usually needed to guarantee that convergence without a great sacrifice of the computational expense. As a result, the convergence of Towmey's method for one dataset (e.g., synthetic data) does not guarantee its convergence for other datasets (e.g., laboratory experiments and atmospheric measurements). I recommend the authors address this very briefly in the main text. Considering broader applications of the inversion methods to HTDMA studies, I recommend the authors address this very briefly in the main text.

Responses: We thank the reviewer for this comment. To conduct a comprehensive test, we have carried out comprehensive tests of inversion algorithms using both synthetic HFIMS data based on representative GF-PDFs and ambient HFIMS measurements (i.e., as described in the revised SI), and the results indicate that Twomey’s method performs well for both datasets. While the convergence of Twomey’s method is achieved for the datasets used in this study, it is possible that Twomey’s method may not achieve a converged solution on rare occasions, especially for measurements with very poor counting statistics. We have clarified this in the revised manuscript (line 176-177).

References

Hanke, M., and Raus, T.: A General Heuristic for Choosing the Regularization Parameter in Ill-Posed Problems, SIAM Journal on Scientific Computing, 17, 956-972, 10.1137/0917062, 1996.

Hansen, P. C., and O’Leary, D. P.: The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, SIAM Journal on Scientific Computing, 14, 1487-1503, 10.1137/0914086, 1993.

Hansen, P. C.: REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems, Numerical Algorithms, 6, 1-35, 10.1007/BF02149761, 1994.

Naseri, A., Sipkens, T. A., Rogak, S. N., and Olfert, J. S.: An improved inversion method for determining two-dimensional mass distributions of non-refractory materials on refractory black carbon, Aerosol Science and Technology, 55, 104-118, 10.1080/02786826.2020.1825615, 2021.

Pinterich, T., Spielman, S. R., Wang, Y., Hering, S. V., and Wang, J.: A humidity-controlled fast integrated mobility spectrometer (HFIMS) for rapid measurements of particle hygroscopic growth, Atmos. Meas. Tech., 10, 4915-4925, 10.5194/amt-10-4915-2017, 2017.

Sipkens, T. A., Olfert, J. S., and Rogak, S. N.: Inversion methods to determine two-dimensional aerosol mass-mobility distributions: A critical comparison of established methods, Journal of Aerosol Science, 140, 105484, https://doi.org/10.1016/j.jaerosci.2019.105484, 2020.

Zhang, J., Spielman, S., Wang, Y., Zheng, G., Gong, X., Hering, S., and Wang, J.: Rapid measurement of RH-dependent aerosol hygroscopic growth using a humidity-controlled fast integrated mobility spectrometer (HFIMS), Atmos. Meas. Tech., 14, 5625-5635, 10.5194/amt-14-5625-2021, 2021.

Regularized inversion of aerosol hygroscopic growth factor probability density function: Application to humidity-controlled fast integrated mobility spectrometer measurements

Jiaoshi Zhang[1], Yang Wang[1,3], Steven Spielman[2], Susanne Hering[2], and Jian Wang[1]

[1]Center for Aerosol Science and Engineering, Washington University in St. Louis, St. Louis, Missouri, USA
[2]Aerosol Dynamics Inc, Berkeley, California, USA
[3]Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, Rolla, Missouri, USA

Correspondence to: Jian Wang (jian@wustl.edu)

Abstract. Aerosol hygroscopic growth plays an important role in atmospheric particle chemistry and the effects of aerosol on radiation and hence climate. The hygroscopic growth is often characterized by a growth factor probability density function (GF-PDF), where the growth factor is defined as the ratio of the particle size at a specified relative humidity to its dry size. Parametric, least-square methods are the most widely used algorithms for inverting the GF-PDF from measurements of humidified tandem differential mobility analyzers (HTDMA) and have been recently applied to the GF-PDF inversion from measurements of the humidity-controlled fast integrated mobility spectrometer (HFIMS). However, these least square methods suffer from noise amplification due to the lack of regularization in solving the ill-posed problem, resulting in significant fluctuations in the retrieved GF-PDF and even occasional failures of convergence. In this study, we introduce nonparametric, regularized methods to invert aerosol GF-PDF and apply them to HFIMS measurements. Based on the HFIMS kernel function, the forward convolution is transformed into a matrix-based form, which facilitates the application of the nonparametric inversion methods with regularizations, including Tikhonov regularization and Twomey's iterative regularization. Inversions of the GF-PDF using the nonparameteric methods with regularization are demonstrated using HFIMS measurements simulated from representative GF-PDFs of ambient aerosols. The characteristics of reconstructed GF-PDFs resulting from different inversion methods, including previously developed least-square methods, are quantitively compared. The result shows that Twomey's method generally outperforms other inversion methods. The capabilities of Twomey's method in reconstructing the pre-defined GF-PDFs and recovering the mode parameters are validated.

1 Introduction

The hygroscopic growth of aerosol particles influences heterogeneous reactions, light extinction, and visibility, whereby aerosol water is most relevant for the direct radiative forcing of Earth's climate (Tang and Munkelwitz, 1994; Pilinis et al., 1995; Swietlicki et al., 2008). The ability of aerosols to absorb water depends mainly on their compositions, hence the hygroscopic properties reflect the variability of the key chemical components (Gysel et al., 2007; Zheng et al., 2020). Therefore, the variation of aerosol hygroscopic growth can be used to infer the potential chemical composition, especially for small aerosols that are beyond the size range of the aerosol mass spectrometer. Aerosol hygroscopic growth under atmospheric relative humidity (RH) is commonly measured by a humidified tandem differential mobility analyzer system (HTDMA) (Liu et al., 1978; Rader and McMurry, 1986; Swietlicki et al., 2008). In an HTDMA system, monodisperse particles classified by the 1st DMA are exposed to an elevated RH in a humidity conditioner, and the size distribution of humidified particles is then measured by a 2nd DMA and a particle detector using scanning mobility technique. The particle hygroscopic growth is then derived from the size distribution of the humidified particles. Recently, a humidity-controlled fast integrated mobility spectrometer (HFIMS) was developed. The HFIMS replaces the 2nd DMA and particle detector within the HTDMA system with a water-based fast integrated mobility spectrometer (WFIMS), which captures the size distribution of humidified particles instantly (Pinterich et al., 2017a). As a result, the HFIMS drastically accelerates aerosol hygroscopic growth measurements (Pinterich et al., 2017b; Wang et al., 2019; Zhang et al., 2021), making it feasible to characterize ambient aerosol hygroscopic growth at a wide range of sizes and RH levels under ~ 25 min.

The HTDMA measurement, i.e., the mobility-concentration distribution of humidified particles, is a convolution of the aerosol hygroscopic growth factor probability density function (GF-PDF) and the transfer functions of both DMAs. Similarly, the HFIMS measurement represents a convolution of the aerosol GF-PDF together with the transfer functions of the DMA and the WFIMS (Wang et al., 2019). Two inversion algorithms, TDMAfit (Stolzenburg and McMurry (1988)) and TDMAinv (Gysel et al. (2009)) were developed and widely used to retrieve the GF-PDF from HTDMA measurements. In both algorithms, the GF-PDF is represented with a specific functional form, and the function parameters were derived by least-squares fitting. For example, the TDMAfit algorithm assumes the GF-PDF as a superposition of multiple Gaussian distribution functions (Stolzenburg and McMurry, 1988) or a summation of multiple lognormal (ML) distribution functions (Stolzenburg and McMurry, 2008). Likewise, TDMAinv describes the GF-PDF as a piecewise linear (PL) function at predefined growth factor values (Gysel et al., 2009). The function parameters are derived using least-squares fitting that minimizes the residual between the measured and reconstructed size distributions of humidified particles. Similar methods have been applied to invert GF-PDFs from HFIMS measurements by Wang et al. (2019).

Inversion of the GF-PDF from the HTDMA or HFIMS measurements is an ill-posed problem (Gysel et al., 2009). Least-squares methods such as TDMAfit and TDMAinv provide simple and effective ways to solve this ill-posed problem by representing the GF-PDF in a specific functional form (Kandlikar and Ramachandran, 1999). However, the GF-PDF inverted by the TDMAfit algorithm often relies on the initial guess of the parameters, resulting in occasional failures of convergence (Gysel et al., 2009). For example, it was reported that the TDMAfit algorithm may not be robust in cases of closely multiple overlapped modes and the successful convergence depends on the initial guess (Swietlicki et al., 2008). Moreover, it is well-known that the unregularized least-squares method amplifies the measurement noise (Kandlikar and Ramachandran, 1999; Sipkens et al., 2020), resulting in significant fluctuations in the retrieved GF-PDF. It has been shown that the derived GF-PDF using the TDMAinv algorithm may oscillate strongly when a higher bin resolution is chosen, while a too low resolution may not be adequate to reproduce complex shapes of true GF-PDF (Gysel et al., 2009). This may lead to incorrect interpretation of the aerosol mixing state (Wang et al., 2019). The approach to overcoming noise amplification is to regularize the problem by including additional information, such as smoothness (Kandlikar and Ramachandran, 1999). Tikhonov regularization is among the most common regularization methods and has been applied to inversions of the aerosol size distribution (Talukdar and Swihart, 2003) and mass-mobility distribution (Sipkens et al., 2020). Recently, a software package was developed to invert HTDMA data using Tikhonov regularization (Petters, 2021). Twomey's method (Twomey, 1975), one of the most common iterative regularization methods, has been widely used to invert aerosol size distributions (Collins et al., 2002; Olfert et al., 2008; Wang et al., 2018) and two-dimensional mass-mobility distributions (Rawat et al., 2016; Sipkens et al., 2020). However, to our best knowledge, Twomey's method has not been applied to invert GF-PDF from HTDMA or HFIMS measurements.

In this study, we present nonparametric, regularized inversions of the GF-PDF from HFIMS measurements. These inversion methods can be adapted to HTDMA measurements straightforwardly. The forward model (i.e., the convolution of the GF-PDF, the transfer function of DMA, and the transfer function of WFIMS) is derived analytically and cast into a matrix form such that nonparametric inversion methods can be conveniently applied. The nonparametric inversions are demonstrated by retrieving GF-PDF from HFIMS measurements of ambient aerosols. The dependence of retrieved GF-PDF on GF bin resolutions is investigated, and an optimal GF bin resolution is identified. Synthetic data are generated using representative GF-PDFs of ambient aerosols and are applied to evaluate different inversion methods, including (1) parametric, least-squares fittings, (2) nonparametric, unregularized least-squares, (3) Twomey's method, and (4) Tikhonov regularization. The performances of the different inversion methods including reconstruction accuracy, GF-PDF fidelity, smoothness, and computation time are presented and discussed.

2 Methods

This section presents the GF-PDF inversion routine from the HFIMS measurement, which includes the mathematical derivation of the matrix-based inverse problem, the description of different inversion algorithms, and the generation of synthetic data for evaluating the inversion algorithms.

2.1 A matrix form for the forward model

The integrated response of HFIMS is determined by the aerosol size distribution, the DMA transfer function, the GF-PDF, and the WFIMS transfer function (Wang et al., 2019). The number concentration of particles with diameters between D_{p1} and $D_{p1} + \mathrm{d}D_{p1}$ downstream of the DMA inside the HFIMS is given by:

$$\mathrm{d}N_{\mathrm{DMA}} = \frac{Q_{\mathrm{a,DMA}}}{Q_{\mathrm{s,DMA}}}\eta_{\mathrm{chg}}\left(D_{\mathrm{p1}}\right)\eta_{\mathrm{p,DMA}}\left(D_{\mathrm{p1}}\right)\Omega\left(V_{\mathrm{DMA}},\tilde{Z}_{\mathrm{p1}}\right)\mathrm{d}N \tag{1}$$

where $Q_{\mathrm{a,DMA}}$ and $Q_{\mathrm{s,DMA}}$ are the DMA aerosol and sample (i.e., monodispersed) flow rates, respectively, $\eta_{\mathrm{chg}}\left(D_{\mathrm{p1}}\right)$ is the aerosol charging efficiency, $\eta_{\mathrm{p,DMA}}\left(D_{\mathrm{p1}}\right)$ is the particle penetration efficiency through the DMA, and $\Omega\left(V_{\mathrm{DMA}},\tilde{Z}_{\mathrm{p1}}\right)$ is the transfer function of the DMA operated with the classifying voltage of V_{DMA}, $\tilde{Z}_{\mathrm{p1}}$ is the particle mobility (Z_{p1}) normalized by the DMA centroid mobility corresponding to V_{DMA}. $\mathrm{d}N = n(D_{\mathrm{p1}})\mathrm{d}D_{\mathrm{p1}}$ represents the number concentration of particles with diameters between D_{p1} and $D_{\mathrm{p1}} + \mathrm{d}D_{\mathrm{p1}}$. The number concentration of particles with diameters between D_{p2} and $D_{\mathrm{p2}} + \mathrm{d}D_{\mathrm{p2}}$ at the outlet of the conditioner is

$$\mathrm{d}N_{\mathrm{cond}} = \mathrm{d}D_{\mathrm{p2}} \int_{D_{\mathrm{p1}}=0}^{D_{\mathrm{p1}}=\infty} \eta_{\mathrm{p,cond}}(D_{\mathrm{p2}}) c_{\mathrm{cond}}(D_{\mathrm{p2}}, D_{\mathrm{p1}}) \mathrm{d}N_{\mathrm{DMA}} \tag{2}$$

where the integration considers all possible values of D_{p1}. $\eta_{\mathrm{p,cond}}(D_{\mathrm{p2}})$ is the penetration efficiency of the conditioned particles, assuming the particle growth from D_{p1} to D_{p2} is instantaneous. $c_{\mathrm{cond}}(D_{\mathrm{p2}}, D_{\mathrm{p1}})$ is the growth factor probability density function (GF-PDF) for particles with a dry diameter of D_{p1} growing to a diameter of D_{p2} during the humidity conditioning process. The GF-PDF satisfies $\int_{D_{\mathrm{p2}}=0}^{D_{\mathrm{p2}}=\infty} c_{\mathrm{cond}}(D_{\mathrm{p2}}, D_{\mathrm{p1}}) \mathrm{d}D_{\mathrm{p2}} = 1$.

The WFIMS response to particles with diameters between D_{p2} and $D_{\mathrm{p2}} + \mathrm{d}D_{\mathrm{p2}}$ in the i^{th} D_{p}^* bin during any time interval (t) is calculated by

$$\mathrm{d}R_i = \frac{Q_{\mathrm{a,WFIMS}} N_F}{\dot{N}_F} \eta_{\mathrm{p,WFIMS}}(D_{\mathrm{p2}}) \eta_{\mathrm{det}}(D_{\mathrm{p2}}) \Omega_{\mathrm{WFIMS},i}(Z_{\mathrm{p2}}) \mathrm{d}N_{\mathrm{cond}} \tag{3}$$

$Q_{\mathrm{a,WFIMS}}$ is the inlet flow rate through the WFIMS, N_F is the number of frames being used to count $\mathrm{d}R_i$, $\dot{N}_F$ is the frame rate. $N_F/\dot{N}_F$ represents the time interval (t) of counting, $\eta_{\mathrm{p,WFIMS}}$ is the penetration efficiency of particles going through the WFIMS separator, $\Omega_{\mathrm{WFIMS},i}(Z_{\mathrm{p2}})$ is the transfer function of the i^{th} bin of the instrument response diameter (D_{p}^*) of the WFIMS. Note that the detection efficiency for particles above 8 nm has been shown to be 1 (i.e., $\eta_{\mathrm{det}}(D_{\mathrm{p2}}) = 1$, Pinterich et al., 2017a). The theoretical response of the i^{th} D_{p}^* bin of the HFIMS, R_i, can be derived by combining the above equations as detailed in Wang et al. (2019):

$$R_i = E \iint \frac{1}{D_{\mathrm{p1}}} c_{\mathrm{cond}}(g, D_{\mathrm{p1}}) \Omega(V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p1}}) \Omega_{\mathrm{WFIMS},i}(Z_{\mathrm{p2}}) \mathrm{d}D_{\mathrm{p2}} \mathrm{d}D_{\mathrm{p1}} + \epsilon_i \tag{4}$$

Where $E = R_{\mathrm{tot}} \frac{b}{b_{\mathrm{view}}} \frac{Q_{sh,\mathrm{DMA}}}{Q_{a,\mathrm{DMA}}} \frac{d\tilde{Z}_{\mathrm{p1}}}{dD_{\mathrm{p1}}}\Big|_{D_{\mathrm{p1}}^*}$. R_{tot} is the total counts of particles detected within the WFIMS viewing window, i.e., $R_{\mathrm{tot}} = \sum_i R_{\mathrm{i}}$, where R_{i} is the response of the i^{th} D_{p}^* bin of the WFIMS. b_{view} and b are the length of the viewing area of the CCD-captured image and the length of the WFIMS mobility separator. ϵ_i is the error in the measured response. In Eq. (4), the GF-PDF is written as a function of growth factor g (i.e., $D_{\mathrm{p2}}/D_{\mathrm{p1}}$), and it satisfies $c_{\mathrm{cond,n}}(g, D_{\mathrm{p1}}) dg = c_{\mathrm{cond}}(D_{\mathrm{p2}}, D_{\mathrm{p1}}) dD_{\mathrm{p2}}$. Given the narrow particle size range classified by the DMA, we assume the GF-PDF is the same for all particles classified by the DMA at a given voltage, i.e., $c_{\mathrm{cond}}(g, D_{\mathrm{p1}})$ is independent of D_{p1} for the integration in Eq. (4). Rewriting the GF-PDF as $c_{\mathrm{cond}}(g)$ and replacing D_{p2} with gD_{p1} in Eq. (4) gives:

$$R_i = E \int_0^{+\infty} dg\, c_{\mathrm{cond}}(g) \int_0^{+\infty} dD_{\mathrm{p1}} \Omega[V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p1}}(D_{\mathrm{p1}})] \Omega_{\mathrm{WFIMS},i}[Z_{\mathrm{p}}(gD_{\mathrm{p1}})] + \epsilon_i \tag{5}$$

The integration can be approximated by~~written as~~ a sum over J GF bins, with the assumption that $c_{\mathrm{cond}}(g)$ is a constant value within each GF bin:

$$R_{i,\mathrm{theo}} = E\sum_{j=1}^{J} c_{\mathrm{cond}}(g_j)\int_{g_{j-\frac{1}{2}}}^{g_{j+\frac{1}{2}}} dg \int_0^{+\infty} dD_{\mathrm{p1}}\Omega\left[V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right]\Omega_{\mathrm{WFIMS},i}\left[Z_{\mathrm{p}}(gD_{\mathrm{p1}})\right] + \epsilon_i \quad (6)$$

where $g_{j-1/2}$ and $g_{j+1/2}$ (j = 1, 2, 3, …, J) are the lower and upper bounds of the jth GF bin. Eq. (6) can be further arranged into a matrix form (neglecting the error term) as

$$\mathbf{R} = \mathbf{M} \times \mathbf{c} \quad (7)$$

where the HFIMS response $\mathbf{R}$ is an $I \times 1$ array composed of R_i (i = 1, 2, 3, …, I). I is the selected size bins of the WFIMS that covers the size range of ($0.8D_{\mathrm{p1}}^*$, $2.0D_{\mathrm{p1}}^*$) according to the settings of the DMA centroid diameter D_{p1}^*. The unknown GF-PDF $\mathbf{c}$, an $J \times 1$ array composed of c_j (j = 1, 2, 3, …, J), can be found by solving the Fredholm integral equation (7). The element of the HFIMS kernel matrix, $\mathbf{M}$, is calculated by

$$M_{ij} = E_i\int_{g_{j-\frac{1}{2}}}^{g_{j+\frac{1}{2}}} dg \int_0^{+\infty} dD_{\mathrm{p1}}\Omega\left[V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right]\Omega_{\mathrm{WFIMS},i}\left[Z_{\mathrm{p}}(gD_{\mathrm{p1}})\right] \quad (8)$$

The HFIMS kernel describes the probability of particles with GF between $g_{j-1/2}$ and $g_{j+1/2}$ that is measured between the channel limits between $Z_{\mathrm{p},i-1/2}^*$ and $Z_{\mathrm{p},i-1/2}^*$. As described above, the inversion of the GF-PDF ($\mathbf{c}$) becomes an ill-posed problem due to overlapping of the HFIMS kernel function, like that of the aerosol size spectrometers (Kandlikar and Ramachandran, 1999; Collins et al., 2002; Talukdar and Swihart, 2003). It is worth noting that the derivation of the HFIMS kernel function can be easily applied to HTDMA measurement by replacing the WFIMS transfer function ~~$\Omega_{\mathrm{WFIMS},i}\left[Z_p(gD_{\mathrm{p1}})\right]$~~ with the transfer function of the 2nd DMA ~~$\Omega\left[V_t, Z_p(gD_{\mathrm{p1}})\right]$~~ in Eq. (8), ~~where V_t is the classifying voltage of the 2nd DMA~~as detailed in the supplementary information (SI).

2.2 Inversion methods

A number of techniques have been developed to solve the Fredholm integration (Kandlikar and Ramachandran, 1999). With Eqs. (7) and (8), nonparametric algorithms can be straightforwardly applied to invert GF-PDF, hence no prior knowledge of the functional form of GF-PDF is needed.

Unregularized least-squares

The simplest route is the ordinary least-squares (LSQ) which seeks to minimize the square of the residual:

$$\mathbf{c}^{\mathbf{LSQ}} = arg\min_{\mathbf{c}}\left\{\|\mathbf{Mc} - \mathbf{R}\|_2^2\right\} \quad (9)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Here, the least-squares solution is solved by using the *lsqnonneg* function from MATLAB. As the uncertainty in measurements can vary substantially for different D_{p}^* bins, the residue is often weighted by measurement uncertainty. A weighted LSQ (WLSQ) seeks to minimize the weighted sum of squares (Sipkens et al., 2020):

$$\mathbf{c}^{\mathbf{WLSQ}} = arg\min_{\mathbf{c}}\left\{\|\mathbf{W}(\mathbf{Mc} - \mathbf{R})\|_2^2\right\} \quad (10)$$

where $\mathbf{W}$ denotes a diagonal weight matrix, whose i^{th} diagonal element is the reciprocal of the standard deviation for data point i.

Tikhonov regularization

Tikhonov regularization is a common regularization method that overcomes noise amplification, and it has been used to invert aerosol size distribution and 2-D aerosol mass-mobility distributions (Talukdar and Swihart, 2003; Petters, 2021; Stolzenburg et al., 2022). In Tikhonov regularization, an additional regularization term is included in the least-squares approach:

$$\mathbf{c}^{\mathbf{Tik}} = arg \min_{\mathbf{c}} \{ \|\mathbf{Mc} - \mathbf{R}\|_2^2 + \lambda^2 \|\mathbf{Lc}\|_2^2 \} \tag{11}$$

where $\lambda^2 \|\mathbf{Lc}\|_2^2$ represents the regularization term designed to minimize the derivative of a specific order and λ is the regularization parameter that controls the degree of regularization. The penalization matrix **L** is often set as the identity matrix **I**, the bidiagonal (-1, 1) matrix, and the upper tridiagonal (1, -2, 1) matrix for the 0th, 1st, and 2nd order regularization, respectively (Hansen and O'Leary, 1993; Hansen, 1994). The parametric L-curve of $\|\mathbf{Mc}_\lambda - \mathbf{R}\|_2$ vs $\|\mathbf{Lc}_\lambda\|_2$ is plotted and the corner of the L-curve with the maximum curvature is identified using the "L-curve" routine from the regularization tools package developed by Hansen (1994). This ~~as the~~ optimal regularization parameter λ ~~which~~ corresponds to a good balance between minimization of the residual and reduction of the noise in the inverted **c** (Hansen, 1992; Hansen and O'Leary, 1993). Similarly, a weighted Tikhonov regularization (WTik) can be applied by (Sipkens et al., 2020):

$$\mathbf{c}^{\mathbf{WTik}} = arg \min_{\mathbf{c}} \{ \|\mathbf{W}(\mathbf{Mc} - \mathbf{R})\|_2^2 + \lambda^2 \|\mathbf{Lc}\|_2^2 \} \tag{12}$$

The effect of introducing the weight in the LSQ inversion and Tikhonov regularization is examined in Section 3.2.

Twomey's method

Twomey's method is commonly used to find solutions for ill-posed problems and has been proved to be effective in inversions of the aerosol size distribution (Collins et al., 2002; Olfert et al., 2008) and aerosol mass-mobility distribution (Rawat et al., 2016; Sipkens et al., 2020). It is a non-linear optimization method and provides iterative regularizations. An initial guess solution is iteratively multiplied by small multiples of the HFIMS kernel function which are proportional to the ratio of the measured to calculated measurements as follows:

$$c_j^{k+1} = \left[1 + \left(\frac{R_i}{\mathbf{m}_i\mathbf{c}^k} - 1\right) M_{ij}\right] \cdot c_j^k \tag{13}$$

where $\mathbf{m}_i$ is the ith row of the HFIMS kernel function $\mathbf{M}$, and $R_i/\mathbf{m}_i\mathbf{c}^k$ denotes the relative divergence between actual and reconstructed HFIMS measurements. The positively constrained, least-squares solution is set as the initial guess (Olfert et al., 2008). Then, the initial guess is smoothed using a three-term moving average (Markowski, 1987) and input into the iterative Twomey's routine which is then repeated until a Chi-squared criterion is satisfied. It is worth noting that Twomey's method may require sufficient counting statistics to ensure converged solutions.

Parametric LSQ fittings

The parametric fitting methods assume a prior known distribution of the GF-PDF and calculate the forward model problem (Eq. 4) to reconstruct the HFIMS measurements. A nonlinear least-squares fitting with boundary constraints is performed to search for the least-squares solution within the bounds. The ML and PL fitting routines for the GF-PDF inversion from HFIMS measurements have been developed by Wang et al. (2019). The influence of counting statistics and GF-PDF parameters (i.e., the number of modes of ML GF-PDF and the number of sections of PL GF-PDF) has been statistically studied. In this work, the GF-PDF inverted using ML and PL fitting routines with the optimized parameters are compared with those retrieved using nonparametric inversion methods described above.

2.3 Generation of synthetic data to evaluate inversion algorithms

HFIMS measurements are synthesized to evaluate the performance of different inversion methods. The synthetic data are based on three representative GF-PDFs that consist of one, two, and three lognormal modes, respectively. The mode parameters of the pre-defined GF-PDFs are listed in Table 1, similar to those listed in Wang et al. (2019). The parameters of f, G, σ are the fractional weight, mean diameter growth factor, and geometric standard deviation of each mode. The theoretical HFIMS response (i.e., R_i) is derived using Eq. (4) based on each of the three GF-PDFs, and Gaussian and Poisson noise are then added to the response using the following approach. First, a zero-mean Gaussian noise component is added to the theoretical HFIMS response to simulate the system noise such as fluctuation of the sample flow rate:

$$R_{i,\mathrm{G}} = R_i\left(1 + \alpha n_i^{\mathrm{G}}\right) \tag{14}$$

where R_i is the derived theoretical response of the i^{th} D_{p}^* bin, n_i^{G} is the i^{th} element of a standard normally-distributed random vector, n^G, with zero mean and variance of 1. The magnitude of the Gaussian noise is varied using a factor, α. The HFIMS measurement is then simulated using the following Poisson distribution to reflect the discrete nature of the particle counting process:

$$P(x) = \frac{R_{i,\mathrm{G}}^x}{x!}\exp\left(-R_{i,\mathrm{G}}\right) \tag{15}$$

where $P(x)$ is the probability that x number of particles are detected by HFIMS in the i^{th} D_{p}^* bin (i.e., actual measurements). The impact of the Gaussian noise on the performance of the inversion methods is examined for different noise levels in Section 3.2. Five hundred sets of HFIMS measurements are generated using Monte Carlo methods with constant counting statistics (i.e., R_{tot} of 100). These synthetic HFIMS measurements are then used to evaluate the inversion methods described above. Note that in the forward model for deriving the theoretical HFIMS response (i.e., Eq. 4), a higher resolution of g (i.e., 120 bins over 0.8 - 2.0) is used than that of the HFIMS kernel matrix (i.e., 20 bins of g, Eq. 8). The difference between the forward and inverse models, together with the inclusion of Gaussian and Poisson noises, minimizes the effect of inverse crime (Colton et al., 1998).

Table 1. Mode parameters of representative GF-PDFs for generating synthetic HFIMS measurements.

Predefined GF-PDF	Mode 1			Mode 2			Mode 3		
	f	G	σ	f	G	σ	f	G	σ
1	1.0	1.40	1.15		NA			NA	

2	0.45	1.10	1.05	0.55	1.30	1.05		NA	
3	0.39	1.05	1.10	0.32	1.40	1.05	0.29	1.70	1.10

3 Results and discussion

3.1 Optimal numbers of Growth factor bins and HFIMS size bins (D_p^*)

The numbers of GF bins (J) and D_p^* ~~-~~bins (I) determine the dimensions of HFIMS kernel function, which affects the inversion of GF-PDF. The optimal number of D_p^*~~-~~ bin is a trade-off between sizing resolution and counting statistics. Wang et al. (2019) examined the influence of WFIMS D_p^* bin number (I) on the inverted GF-PDF and found an optimal range of 23-32 for total particle counts of 100. For representative remote continental and urban aerosols, the number of particles measured by the HFIMS often exceed 100 in 20 seconds (Pinterich et al., 2017b; Zhang et al., 2021), ensuring sufficient counting statistics for ambient measurements. The dynamic range of WFIMS is roughly a factor of 10 in mobility, corresponding to a factor of ~3 in the size range (Zhang et al., 2021). In this study, 30 size bins (i.e., I = 30) that are evenly spaced on a logarithmic scale over the WFIMS size range are used in the inversions.

The influence of growth factor bin number (J) on the inverted GF-PDF is examined using the synthetic HFIMS measurements described above. The GF-PDF was inverted from each set of the simulated HFIMS measurements using different GF bin numbers ranging from 10 to 50 (i.e., corresponding to a GF resolution range of 0.024 – 0.12). To facilitate the comparison of GF-PDFs inverted with different GF bin numbers, we interpolate the inverted GF-PDFs to 120 fixed growth factors that are evenly distributed from 0.8 to 2.0. The average error of the inverted GF-PDF γ is defined as:

$$\gamma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(c_{i,inv} - c_{i,sim}\right)^2$$ (~~13~~16)

where $c_{i,inv}$ and $c_{i,sim}$ are the interpolated GF-PDF and pre-defined GF-PDF (i.e., true values) at the 120 fixed growth factors, respectively. N is the number of points of fixed growth factors (i.e., 120). The smoothness of the inverted GF-PDF is evaluated using the absolute second-order derivative:

$$\xi = \sum_{i=2}^{N-1}\left|2c_{inv}(g_i) - c_{inv}(g_{i+1}) - c_{inv}(g_{i-1})\right|$$ (~~14~~17)

To evaluate how well the inverted GF-PDF reproduces the HFIMS measurement, we define the residual of the reconstructed HFIMS measurement (i.e., reconstruction error) as:

$$\chi^2 = \sum_{i=1}^{L}\left(\tilde{R}_{i,inv} - \tilde{R}_i\right)^2$$ (~~15~~18)

where $\tilde{R}_{i,inv}$ is the normalized HFIMS measurement that is reconstructed using Eq. (7) (i.e., forward calculation).~~-~~ $\tilde{R}_i$ is the normalized synthetic HFIMS measurement (i.e., true values).

Figure 1 shows the smoothness of the inverted GF-PDF ~~inverted using Twomey methods~~ (ξ) versus the residual of reconstructed HFIMS measurement (χ^2) for different GF bin numbers (J). The variation of ξ with χ^2 exhibits an L-shaped curve for all three representative PF-PDF. The initial increase of J from 10 to 20 substantially improves the agreement between the reconstructed and simulated HFIMS measurements, as indicated by a much reduced χ^2 value. At the same time, ξ remains relatively small, indicating a high smoothness of the inverted GF-PDF. In contrast, an increase of J above 20 leads to a minor reduction of χ^2 value but a drastic increase of ξ, suggesting strong noise in the inverted GF-PDF. The optimal solution lies near the corner of the "L-curve" (Hansen and O'Leary, 1993) that strikes a balance between the smoothness and the fidelity to the HFIMS measurements. For all three pre-defined GF-PDFs, the corner of the L-curve corresponds to a J value of 20. GF-PDF inverted with 20 growth factor bins generally shows the smallest error (γ^2), indicating best agreements between the inverted and the true GF-PDFs. Note that the above results are based on inversions using Twomey's method. The same type of L-curves for GF-PDFs inverted using unregularized LSQ and Tikhonov regularizations are shown in SI (Section S2), and they also reveal a corner that corresponds to a J value of 20. These results suggest an optimal J value of 20 for a range of representative GF-PDFs and different inversion methods.

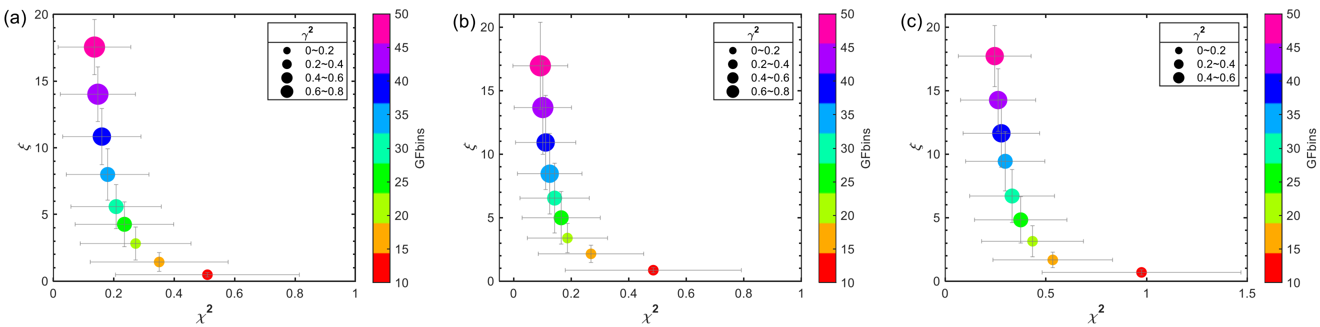


Figure 1. "L-curve" showing the dependence of reconstruction residual, χ^2, and the smoothness, ξ, on the number of GF bins of pre-defined GF-PDFs with (**a**) one mode, (**b**) two modes, and (**c**) three modes, respectively. The symbol size represents the error in inverted GF-PDF, γ^2. Whiskers represent standard deviation. The inversion is conducted using Twomey's method.

3.2 Effect of measurement uncertainties

The uncertainty in HFIMS measurements consists of mainly normal distributed random instrumental noise (e.g., sample flow fluctuation) and Poisson noise due to counting statistics. As the uncertainty varies among different HFIMS D_p^* bins, we first compare the performance of weighted and unweighted inversion methods, including LSQ and Tikhonov regularizations. For this comparison, inversion methods are applied to HFIMS data synthesized with α=0.05, a typical value used in previous studies (Gysel et al., 2009). A total of 500 sets of synthetic data are generated for each of the three pre-defined GF-PDFs. The values of synthesized HFIMS response ($R_{i,s}$) are integers, which reflect the discrete nature of particle counting. For weighted LSQ and Tikhonov regularizations, the weight for D_p^* bins (i.e., diagonal elements in **W**) is derived as $1/\sqrt{R_{i,s}}$. However, this approach leads to a weight of infinite when $R_{i,s}$ has a value of zero (i.e., no particle detected within the D_p^* bin). To overcome this issue, we replace zero $R_{i,s}$ values with a fixed number $R_{i,\min}$ when deriving the weight. Figure 2 compares the reconstruction residual, the GF-PDF error, and the smoothness of GF-PDF inverted using unweighted LSQ and weighted LSQ with $R_{i,\min}$ values of 1, 0.1, 0.01, respectively. Whereas statistically no substantial difference is found among the smoothness of GF-PDFs inverted using unweighted and weighted LSQ, unweighted LSQ leads to lower reconstruction residual and the error in inverted GF-PDF compared to the weighted LSQ. For the weighted LSQ inversions, both the reconstruction residue and the error in inverted GF-PDF increase with increasing weight for $R_{i,s}$ of zeros values (i.e., $1/\sqrt{R_{i,\min}}$). The measurement uncertainty is larger and therefore the weight is lower for channels with higher $R_{i,s}$, which corresponds to higher probability densities (i.e., higher $c(g)$ values). As a result, the GF-PDF inverted using weighted LSQ may have relatively larger errors for high $c(g)$ values, and consequently the average GF-PDF error (γ^2). The same comparisons are also carried out for weighted and unweighted Tikhonov algorithms, and again the weighted algorithms do not provide better performances (i.e., lower error in inverted GF-PDFs) than the unweighted ones. Therefore, subsequent analyses of this study are focused on unweighted algorithms for LSQ and Tikhonov regularizations. It is worth noting that derivation of the weight as $1/\sqrt{R_{i,s}}$ implicitly assumes that the noise in HFIMS measurements is due to counting statistics only, whereas the synthetic HFIMS data are generated with 5% Gaussian noise. As shown next, the noise in the synthetic HFISM data is dominated by the counting statistics. In addition, for real measurements, the level of Gaussian noise (i.e., α) is often not accurately known. We also repeated the above comparisons by deriving the weight as $1/\sqrt{(R_{i,s} + \alpha^2 R_{i,s}^2)}$, which accounts for both Poisson and Gaussian noises. The results are essentially the same.

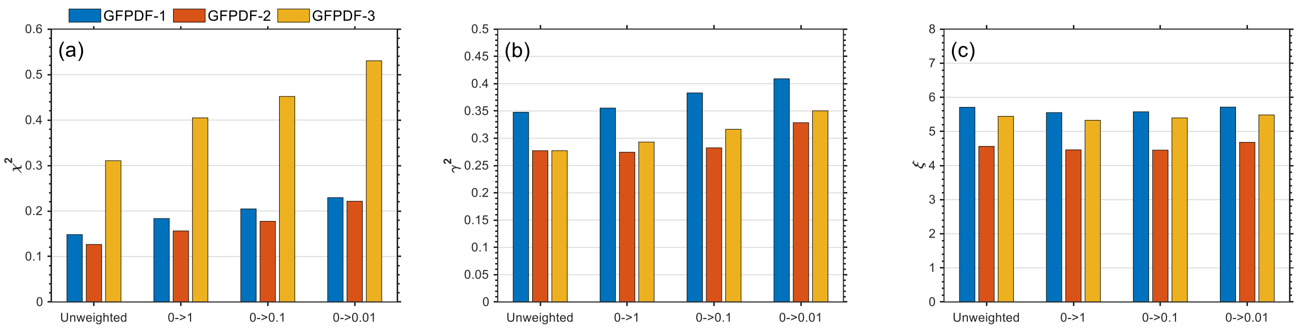


Figure 2. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of GF-PDFs inverted using the unweighted and weighted LSQ methods with different weighting schemes for zero value D_p^* bins (i.e., replacing zero values by 1, 0.1, and 0.01, respectively). Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

The effect of the level of Gaussian noise on the inverted GF-PDF is examined. Synthetic HFIMS measurements are generated following the approach described above (Eq. 13 and 14) at four Gaussian noise levels (i.e., α= 0%, 1%, 5%, and 10%). At each α level, 500 sets of synthetic data are generated and inverted using Twomey's method for each of the three pre-defined GF-PDF. All retrieved inversion parameters, including the reconstruction residual, the GF-PDF error, and the smoothness, are statistically the same for all four Gaussian noise levels (Fig. 3), indicating that HFIMS measurements noise is dominated by counting statistics, and the inclusion of the Gaussian noise has negligible impact on the GF-PDF inverted by Twomey's method.

Similarly, the impact of Gaussian noise is also negligible for GF-PDF inverted using unweighted LSQ and 0^{th}, 1^{st}, and 2^{nd} order Tikhonov regularizations (not shown).

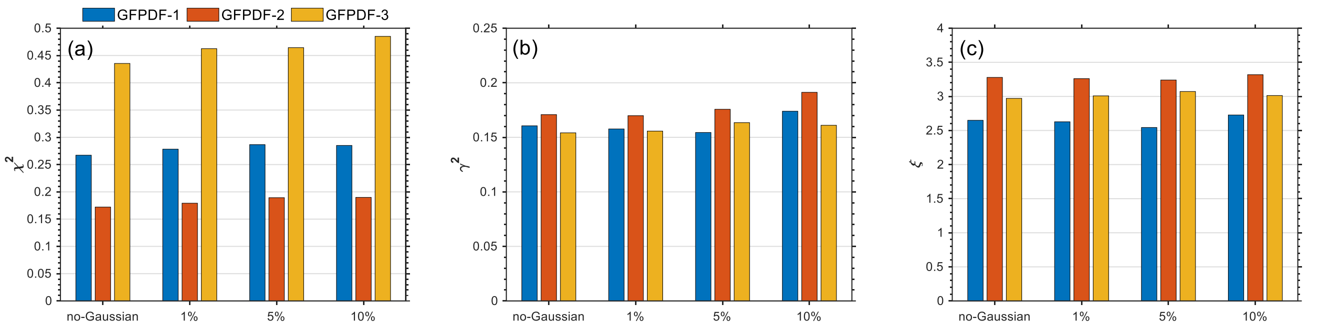


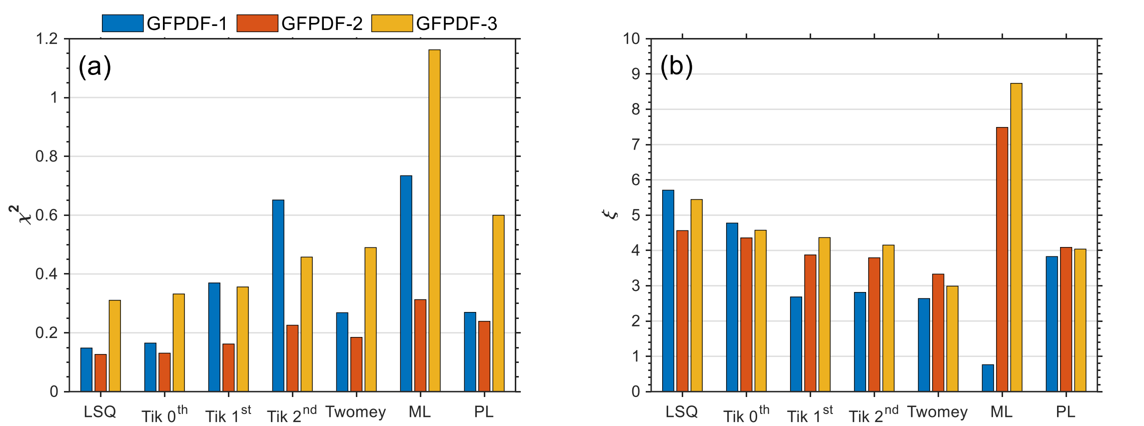
Figure 3. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the degree of smoothing, ξ (**c**) of GF-PDFs inverted using Twomey's methods from synthetic HFIMS data with additional Gaussian noises of different levels (i.e., none, 1%, 5%, and 10%). Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

We also challenged the inversion algorithms with different forward and inverse models to simulate the scenarios when DMA or WFIMS is not perfectly calibrated. A different DMA or WFIMS transfer function width is used to generate the synthetic HFIMS measurements than that used to calculate the inversion matrix. We found that up to ±20% variation of the DMA or WFIMS transfer function width has negligible impacts on the inverted GF-PDF. The results and discussion are detailed in Section S3 of the SI.

3.~~2~~ 3 Comparisons of different inversion methods

The performances of different inversion methods described in Section 2.2 are systematically compared. A total of 500 sets of synthetic HFIMS data are generated and inverted for each of three pre-defined GF-PDFs. For all nonparametric methods, the inversions were carried out using the optimal numbers of GF bins (J) and D_{p}^{*} bins (I), 20 and 30, respectively. Figure ~~2~~ 4 shows the residual of reconstructed HFIMS measurements (χ^2), the smoothness (ξ), the error of inverted GF-PDF ~~residual~~ (γ^2), and the computing time for different inversion methods. ~~The results are the averages for the 500 sets of synthetic HFIMS data are shown for each of three pre-defined GF-PDFs.~~ Compared with parametric counterparts (i.e., ML and PL least-squares fitting), the nonparametric methods generally retrieve more accurate GF-PDFs. Note that the ML least-squares fitting fails to converge to a valid solution occasionally, resulting in the abnormally large error in the inverted GF-PDFs, particularly for the pre-defined GF-PDFs with two and three modes. It may be due to the assumed spectral shape of GF-PDFs or the finite range of the boundary constraints that lead to a failure of searching for a least-squares solution in the presence of random noise. Among all nonparametric inversion methods, the unregularized LSQ provides the solution with the lowest reconstruction residual but largest noise and error in the inverted GF-PDFs, consistent with the noise amplification in unregularized methods. In comparison, regularized inversion methods generally produce smoother solutions at the expense of increased reconstruction residuals. Among different Tikhonov regularization methods, higher-order regularizations (i.e., 1st and 2nd) tend to produce smoother solutions, although the errors in inverted GF-PDF are very similar statistically. The ξ value of the GF-PDF inverted using 1st and 2nd order Tikhonov regularizations increase with the mode number of GF-PDF, consistent with the increasingly more complex spectral shape of GF-PDF. Overall, Twomey's method outperforms the ~~other regularized inversion~~Tikhonov regularization methods regardless of the shapes of the pre-defined GF-PDFs. On average, the GF-PDF inverted using Twomey's method has the smallest error (γ^2) and lowest ξ value, indicative of the best performance. Note that the results are based on synthetic data generated with relatively low counting statistics (i.e., R_{tot} of 100). We also synthesized HFIMS data with R_{tot} of 500 and compared the performance of different inversion methods for measurements with the improved counting statistics, and the results are consistent with those shown in Fig. 4 (Fig. S7). We, therefore, expect the results reflect the general performances of different inversion methods for a typical range of counting statistics of HFIMS measurements.

Figure~~.~~ ~~2~~4(d) shows that once the matrix is generated, the implementation of the nonparametric methods requires a much shorter computing time than the parametric fitting methods. Here, the computing time is recorded on a desktop with Intel's 8th generation processor Core i7-8700. On average, a single-time implementation of the unregularized LSQ (i.e., the "lsqnonneg" function in MATLAB) requires ~1s for all three pre-defined GF-PDFs, and the computing times for all other ~~any~~ nonparametric methods are similar (with the ~~requires only ~1s for all three pre-defined GF-PDFs.~~ largest difference of only ~4%), indicative of equally good computing efficiencies. In contrast, both ML and PL least-squares fitting routines require more than one order of magnitude longer time.



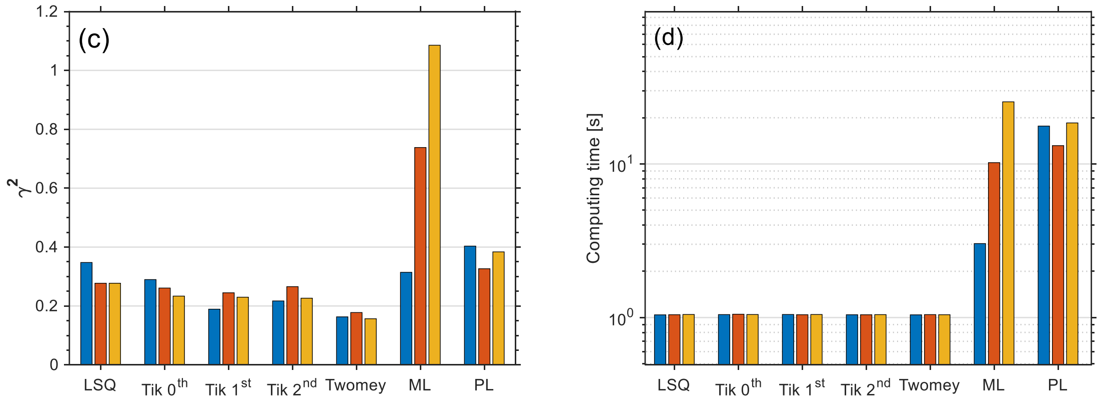


Figure ~~2~~4. Comparison of reconstruction ~~error~~residual, χ^2 (**a**), the smoothness, ξ (**b**), the GF-PDF ~~residual~~error, γ^2 (**c**), and the computing time (**d**) of GF-PDFs inverted using different inversion methods. Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

3.4 Comparison of Tikhonov regularization and Twomey's method

In this section, we investigate why Twomey's method performs better than Tikhonov regularizations. The Tikhonov regularized solutions depend on the regularization parameter, λ. The value of λ is often determined by heuristic methods, including the L-curve approach (Hansen and O'Leary, 1993) and the Hanke-Raus rule (Hanke and Raus, 1996). The L-curve approach determines λ by seeking a trade-off between minimizing the residual term and minimizing the regularization term (i.e., roughness of the solution), and the Hanke-Raus rule selects a computable λ that minimizes the λ-dependent residual term $\frac{1}{\lambda}\left\|\mathbf{Mc}^{\boldsymbol{Tik}}(\lambda)-\mathbf{R}\right\|_2$ (Hanke and Raus, 1996; Sipkens et al., 2020). As the pre-defined GF-PDFs are known for the synthetic HFIMS data, the value of λ can be optimized by comparing the inverted GF-PDF with the true solution, i.e., minimizing the error in inverted GF-PDF (γ^2). Figure 5 shows the comparison of the statistics of inversions using LSQ, Twomey's method, and 1st order Tikhonov. The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs. Here, the 1st order Tikhonov regularization is chosen as it shows better performance (i.e., lower GF-PDF error) than 0th and 2nd Tikhonov regularizations (Fig. 4). The Tikhonov regularization parameter is identified by all three methods: (1) the L-curve, (2) the Hanke-Raus rule, and (3) optimization through minimizing the error in inverted GF-PDFs. It is worth noting that the 3rd method is not feasible for real measurements, as the true GF-PDF is unknown. Figure 5b shows that the Tikhonov regularization with the optimized λ (i.e., the 3rd method) provides the most accurate solution (i.e., lowest GF-PDF error), and outperforms Twomey's method. However, when λ derived using the L-curve approach or Hanke-Raus rule is used, GF-PDF inverted using 1st order Tikhonov regularization generally has a larger error (i.e., γ^2) than that inverted using Twomey's method. The above comparisons indicate that while the Tikhonov regularization can outperform Twomey's method in theory, the optimal regularization parameter λ cannot be obtained reliably using existing methods in practice, leading to inferior performance than Twomey's method. For example, the L-curve approach does not work well if the curvature of the L-curve is negative everywhere, and in such scenario, the leftmost point (i.e., with smaller λ) on the L-curve is taken as the corner (Hansen, 1994), leading to insufficient regularizations of the solution (Naseri et al., 2021). On the other hand, the Hanke-Raus rule often chooses a much larger λ compared with the optimal value, which results in over-smoothed solutions with even larger errors. We also carried out similar comparisons of Twomey's method with 0th and 2nd order Tikhonov regularizations using λ values derived from the three different methods, and the results are consistent with those shown in Fig. 5.

The nonparametric inversion methods are also applied to HFIMS measurements of ambient particles with a dry diameter of 35 nm (Zhang et al., 2021), as detailed in the SI (Section S5). As the true GF-PDF of ambient aerosols is unavailable, the performance of the inversion methods can not be directly compared. Nevertheless, the comparison of the reconstruction residual and the smoothness of inverted GF-PDF paints a similar picture that Twomey's method strikes a good balance between the smoothness of the inverted GF-PDF and the fidelity in reproducing the HFIMS measurements, and it likely outperforms Tikhonov regularizations in practice.

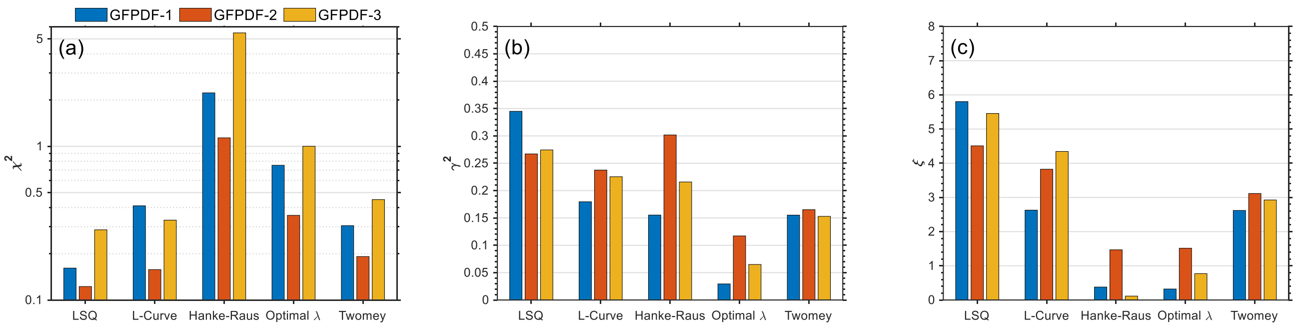


Figure 5. The reconstruction residual, χ^2 (a), the GF-PDF error, γ^2 (b), and the smoothness, ξ (c) of GF-PDF inverted using LSQ, 1st order Tikhonov regularization with the regularization parameter derived from three different approaches (L-curve, Hanke-Raus rule, and optimized λ), and Twomey's method. The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow).

3.~~3~~ 5 Inversion by Twomey's method

As Twomey's method is shown to be the best among all inversion methods examined, we characterize the accuracy of the GF-PDFs inverted using Twomey's method and the recovered mode parameters. Figure ~~3~~ 6 compares the GF-PDFs inverted with the optimized GF and D_p^* bin numbers and with the pre-defined GF-PDFs. The reconstructed and the simulated HFIMS measurements are also presented in the top panel. Both the inverted GF-PDF and reconstructed HFIMS measurements are averaged over the inversions of 500 sets of synthetic data. The results demonstrate excellent agreement of the reconstructed HFIMS measurements with the synthetic data (i.e., simulated HFIMS measurements) for all three pre-defined GF-PDFs. Both the spectral shapes and peak locations of the inverted GF-PDFs agreed well with that of the pre-defined GF-PDFs. In addition, the inverted GF-PDFs are also in better agreement as compared with those inverted from parametric least-squares approaches (i.e., ML and PL GF-PDFs, Wang et al., 2019).

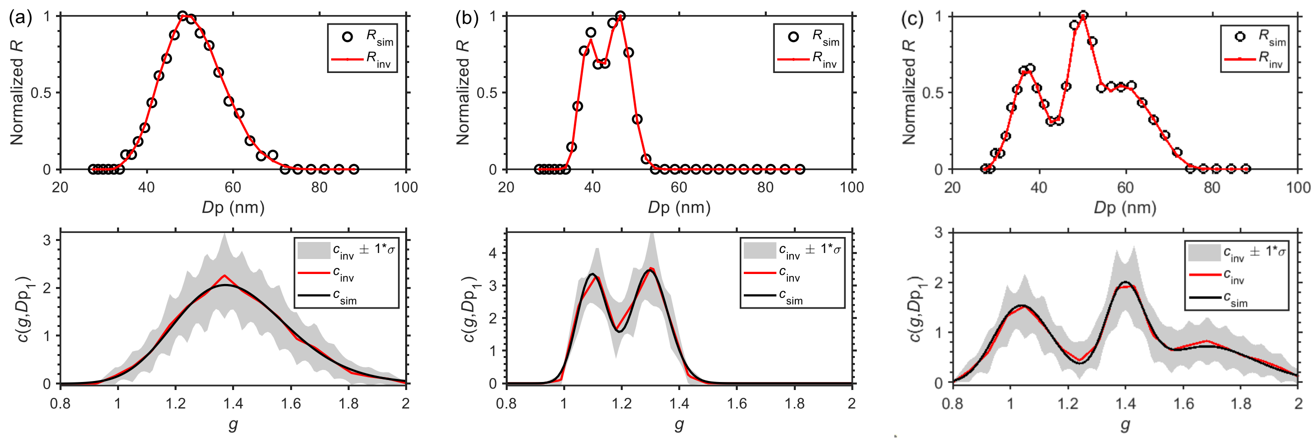


Figure ~~3~~6. (*Top panels*) Comparisons between the averaged reconstructed HFIMS measurements and the simulated HFIMS measurements corrupted with Poisson noises for pre-defined GF-PDFs of one mode (**a**), two modes (**b**), and three modes (**c**), respectively. (*Bottom panels*) Comparisons between the pre-defined GF-PDFs and the GF-PDFs inverted using Twomey's method with the optimized value of GF bins. The shaded area represents GF-PDF solution spaces within one standard deviation.

To quantify the accuracy of the inverted GF-PDFs, we fitted the inverted GF-PDFs to recover the mode parameters as shown in Table ~~1~~2. The pre-set mode parameters of the pre-defined GF-PDFs are shown in Table 1. The results show that both the mode geometric means and the multimodal number fractions can be recovered accurately with minor uncertainties.

Table 2. Recovered mode parameters of pre-defined GF-PDFs from inverted GF-PDFs.

Predefined	**Mode 1**			**Mode 2**			**Mode 3**		
GF-PDF	f	G	σ	f	G	σ	f	G	σ
1	1.00 ± 0	1.39 ± 0.03	1.09 ± 0.01		NA			NA	
2	0.46 ± 0.10	1.10 ± 0.02	1.02 ± 0.01	0.54 ± 0.10	1.30 ± 0.02	1.03 ± 0.01		NA	
3	0.37 ± 0.09	1.05 ± 0.03	1.05 ± 0.02	0.34 ± 0.13	1.40 ± 0.03	1.03 ± 0.02	0.28 ± 0.12	1.69 ± 0.08	1.06 ± 0.03

4 Conclusion

In this study, we develop and evaluate nonparametric regularized methods for inverting GF-PDF from HFIMS measurements. The integrated response of HFIMS, which is a convolution of the aerosol hygroscopic GF-PDF, the transfer function of the DMA, and the transfer function of the WFIMS, is first cast into a matrix form. With the matrix form, nonparametric regularized methods can be applied straightforwardly to invert the GF-PDF. Synthetic HFIMS measurements are generated using Monte-

Carlo simulations for representative aerosol GF-PDFs, and the synthetic data are used to investigate the dependence of inverted GF-PDF on the number of GF bins (i.e., GF resolutions) and the performances of different inversion methods. We show an optimal GF bin number of 20 for all nonparametric methods and representative GF-PDFs. The performances of unregularized least-squares, Twomey's algorithm, Tikhonov regularizations, and commonly used parametric inversion methods (i.e., ML and PL least-squares fitting) are compared. Nonparametric methods based on the matrix form have substantial advantages in the inversion of GF-PDF over the parametric fitting methods as (1) no prior assumption of GF-PDF distributions is required; (2) the matrix-based form facilitates the application of different regularizations (e.g., Tikhonov regularization and Twomey's iterative regularization), which reduce the error in inverted GF-PDF by eliminating noise amplification; (3) they are much more computationally efficient once the matrix is generated. The Tikhonov regularized solutions depend on the regularization parameter, λ. While the Tikhonov regularization can outperform Twomey's method in theory, the optimal λ value cannot be obtained reliably using existing methods in practice, leading to inferior performances than Twomey's method. On average, the GF-PDF inverted using Twomey's method has the smallest error compared to solutions using the other inversion methods regardless of the shapes of the pre-defined GF-PDFs, and it accurately reproduces the true GF-PDF, including the mode parameters and other key statistics.

Data availability. Datasets and code packages related to this paper ~~will be provided by the corresponding author (Jian Wang, jian@wustl.edu) upon request~~is provided in a GitHub repository (https://github.com/zjs023/Regularized_inversion_HFIMS). More information is provided with the README in the repository.

Author contributions. JZ and JW designed the study. JZ developed the code. JZ and JW prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the funding support from the U. S. Department of Energy's Small Business Innovation Research (SBIR) program under contract DE-SC0013103 and Small Business Technology Transfer (STTR) program under contract DESC0006312.

References

Collins, D. R., Flagan, R. C., and Seinfeld, J. H.: Improved Inversion of Scanning DMA Data, Aerosol Science and Technology, 36, 1-9, 10.1080/027868202753339032, 2002.

Colton, D. L., Kress, R., and Kress, R.: Inverse acoustic and electromagnetic scattering theory, Springer, 1998.

Gysel, M., Crosier, J., Topping, D. O., Whitehead, J. D., Bower, K. N., Cubison, M. J., Williams, P. I., Flynn, M. J., McFiggans, G. B., and Coe, H.: Closure study between chemical composition and hygroscopic growth of aerosol particles during TORCH2, Atmos. Chem. Phys., 7, 6131-6144, 10.5194/acp-7-6131-2007, 2007.

Gysel, M., McFiggans, G. B., and Coe, H.: Inversion of tandem differential mobility analyser (TDMA) measurements, Journal of Aerosol Science, 40, 134-151, https://doi.org/10.1016/j.jaerosci.2008.07.013, 2009.

Hanke, M., and Raus, T.: A General Heuristic for Choosing the Regularization Parameter in Ill-Posed Problems, SIAM Journal on Scientific Computing, 17, 956-972, 10.1137/0917062, 1996.

Hansen, P. C.: Analysis of Discrete Ill-Posed Problems by Means of the L-Curve, SIAM Review, 34, 561-580, 10.1137/1034115, 1992.

Hansen, P. C., and O'Leary, D. P.: The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems, SIAM Journal on Scientific Computing, 14, 1487-1503, 10.1137/0914086, 1993.

Hansen, P. C.: REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems, Numerical Algorithms, 6, 1-35, 10.1007/BF02149761, 1994.

Kandlikar, M., and Ramachandran, G.: INVERSE METHODS FOR ANALYSING AEROSOL SPECTROMETER MEASUREMENTS: A CRITICAL REVIEW, Journal of Aerosol Science, 30, 413-437, https://doi.org/10.1016/S0021-8502(98)00066-4, 1999.

Liu, B. Y. H., Pui, D. Y. H., Whitby, K. T., Kittelson, D. B., Kousaka, Y., and McKenzie, R. L.: The aerosol mobility chromatograph: A new detector for sulfuric acid aerosols, Atmospheric Environment (1967), 12, 99-104, https://doi.org/10.1016/0004-6981(78)90192-0, 1978.

Markowski, G. R.: Improving Twomey's Algorithm for Inversion of Aerosol Measurement Data, Aerosol Science and Technology, 7, 127-141, 10.1080/02786828708959153, 1987.

Naseri, A., Sipkens, T. A., Rogak, S. N., and Olfert, J. S.: An improved inversion method for determining two-dimensional mass distributions of non-refractory materials on refractory black carbon, Aerosol Science and Technology, 55, 104-118, 10.1080/02786826.2020.1825615, 2021.

Olfert, J. S., Kulkarni, P., and Wang, J.: Measuring aerosol size distributions with the fast integrated mobility spectrometer, Journal of Aerosol Science, 39, 940-956, https://doi.org/10.1016/j.jaerosci.2008.06.005, 2008.

Petters, M. D.: A Software Package to Simplify Tikhonov Regularization with Examples for Matrix-Based Inversion of SMPS and HTDMA Data, Atmos. Meas. Tech. Discuss., 2021, 1-27, 10.5194/amt-2021-51, 2021.

Pilinis, C., Pandis, S. N., and Seinfeld, J. H.: Sensitivity of direct climate forcing by atmospheric aerosols to aerosol size and composition, Journal of Geophysical Research: Atmospheres, 100, 18739-18754, https://doi.org/10.1029/95JD02119, 1995.

Pinterich, T., Spielman, S. R., Hering, S., and Wang, J.: A water-based fast integrated mobility spectrometer (WFIMS) with enhanced dynamic size range, Aerosol Science and Technology, 51, 1212-1222, 10.1080/02786826.2017.1338664, 2017a.

Pinterich, T., Spielman, S. R., Wang, Y., Hering, S. V., and Wang, J.: A humidity-controlled fast integrated mobility spectrometer (HFIMS) for rapid measurements of particle hygroscopic growth, Atmos. Meas. Tech., 10, 4915-4925, 10.5194/amt-10-4915-2017, 2017b.

Rader, D. J., and McMurry, P. H.: Application of the tandem differential mobility analyzer to studies of droplet growth or evaporation, Journal of Aerosol Science, 17, 771-787, https://doi.org/10.1016/0021-8502(86)90031-5, 1986.

Rawat, V. K., Buckley, D. T., Kimoto, S., Lee, M.-H., Fukushima, N., and Hogan, C. J.: Two dimensional size–mass distribution function inversion from differential mobility analyzer–aerosol particle mass analyzer (DMA–APM) measurements, Journal of Aerosol Science, 92, 70-82, https://doi.org/10.1016/j.jaerosci.2015.11.001, 2016.

Sipkens, T. A., Olfert, J. S., and Rogak, S. N.: Inversion methods to determine two-dimensional aerosol mass-mobility distributions: A critical comparison of established methods, Journal of Aerosol Science, 140, 105484, https://doi.org/10.1016/j.jaerosci.2019.105484, 2020.

Stolzenburg, D., Ozon, M., Kulmala, M., Lehtinen, K. E. J., Lehtipalo, K., and Kangasluoma, J.: Combining instrument inversions for sub-10 nm aerosol number size-distribution measurements, Journal of Aerosol Science, 159, 105862, https://doi.org/10.1016/j.jaerosci.2021.105862, 2022.

Stolzenburg, M., and McMurry, P.: TDMAFIT user's manual, University of Minnesota, Department of Mechanical Engineering, Particle Technology Laboratory, Minneapolis, 1-61, 1988.

Stolzenburg, M. R., and McMurry, P. H.: Equations Governing Single and Tandem DMA Configurations and a New Lognormal Approximation to the Transfer Function, Aerosol Science and Technology, 42, 421-432, 10.1080/02786820802157823, 2008.

Swietlicki, E., Hansson, H.-C., Hämeri, K., Svenningsson, B., Massling, A., McFiggans, G., McMurry, P. H., Petäjä, T., Tunved, P., and Gysel, M.: Hygroscopic properties of submicrometer atmospheric aerosol particles measured with H-TDMA instruments in various environments—a review, Tellus B: Chemical and Physical Meteorology, 60, 432-469, 2008.

Talukdar, S. S., and Swihart, M. T.: An Improved Data Inversion Program for Obtaining Aerosol Size Distributions from Scanning Differential Mobility Analyzer Data, Aerosol Science and Technology, 37, 145-161, 10.1080/02786820300952, 2003.

Tang, I., and Munkelwitz, H.: Aerosol phase transformation and growth in the atmosphere, Journal of Applied Meteorology, 33, 791-796, 1994.

Twomey, S.: Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions, Journal of Computational Physics, 18, 188-200, https://doi.org/10.1016/0021-9991(75)90028-5, 1975.

Wang, Y., Pinterich, T., and Wang, J.: Rapid measurement of sub-micrometer aerosol size distribution using a fast integrated mobility spectrometer, Journal of Aerosol Science, 121, 12-20, https://doi.org/10.1016/j.jaerosci.2018.03.006, 2018.

Wang, Y., Zheng, G., Spielman, S. R., Pinterich, T., Hering, S. V., and Wang, J.: Retrieval of high time resolution growth factor probability density function from a humidity-controlled fast integrated mobility spectrometer, Aerosol Science and Technology, 53, 1092-1106, 2019.

Zhang, J., Spielman, S., Wang, Y., Zheng, G., Gong, X., Hering, S., and Wang, J.: Rapid measurement of RH-dependent aerosol hygroscopic growth using a humidity-controlled fast integrated mobility spectrometer (HFIMS), Atmos. Meas. Tech., 14, 5625-5635, 10.5194/amt-14-5625-2021, 2021.

Zheng, G., Kuang, C., Uin, J., Watson, T., and Wang, J.: Large contribution of organics to condensational growth and formation of cloud condensation nuclei (CCN) in the remote marine boundary layer, Atmospheric Chemistry and Physics, 20, 12515-12525, 2020.

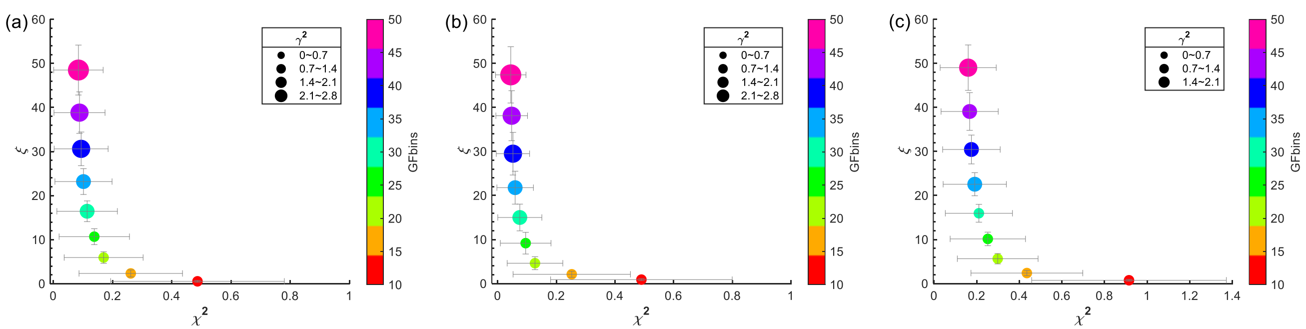
S1. Mathematical derivation of the integrated response of humidified tandem differential mobility analyzer (HTDMA)

The theoretical response of the i^{th} D_{p}^{*} bin of the HTDMA is

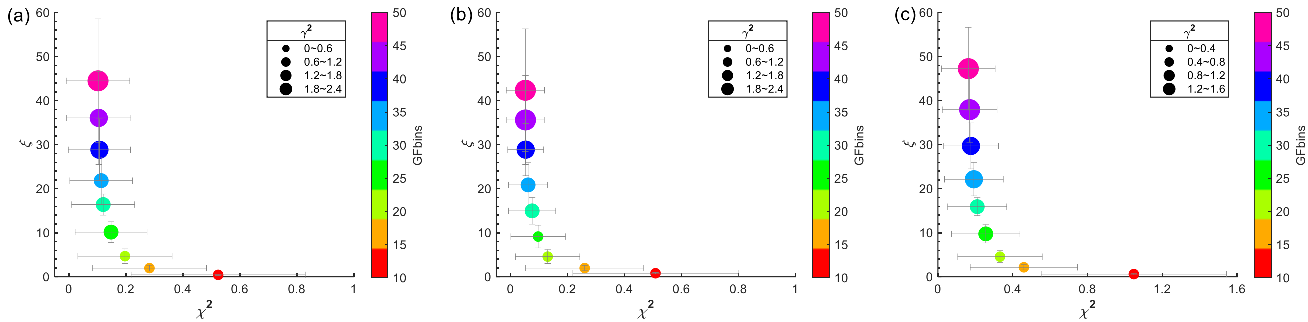
$$R_{i,\mathrm{theo}} = E \sum_{j=1}^{J} c_{\mathrm{cond}}(g_j) \int_{g_{j-\frac{1}{2}}}^{g_{j+\frac{1}{2}}} dg \int_{0}^{+\infty} dD_{\mathrm{p1}} \Omega_1\left[V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right] \Omega_2\left[V_i, \tilde{Z}_p(gD_{\mathrm{p1}})\right] + \epsilon_i$$

where $E = R_{\mathrm{tot}} \frac{Q_{sh,1}}{Q_{a,1}} \frac{d\tilde{Z}_{\mathrm{p1}}}{dD_{\mathrm{p1}}}\Big|_{D_{\mathrm{p1}}^{*}}$. R_{tot} is the total counts of particles detected by the CPC downstream the 2nd DMA with a detection efficiency of 1 for particles above 30 nm (i.e., $\eta_{\mathrm{det}}(D_{\mathrm{p2}}) = 1$). $Q_{\mathrm{a,1}}$ and $Q_{\mathrm{sh,1}}$ are the sample and sheath flow rates of the 1st DMA. $\Omega_1(V_{\mathrm{DMA}}, \tilde{Z}_{\mathrm{p1}})$ is the transfer function of the 1st DMA operated with the classifying voltage of V_{DMA}. $\tilde{Z}_{\mathrm{p1}}$ is the particle mobility (Z_{p1}) normalized by the 1st DMA centroid mobility corresponding to V_{DMA}. $\Omega_2\left[V_i, \tilde{Z}_p(gD_{\mathrm{p1}})\right]$ is the transfer function of the 2nd DMA with a scanning voltage V_i, and $\tilde{Z}_p(gD_{\mathrm{p1}})$ is the mobility of humidified particles normalized the centroid mobility of 2nd DMA. $c_{\mathrm{cond,n}}(g, D_{\mathrm{p1}})$ represents the growth factor probability density function (GF-PDF) for particles with respect to growth factor g, and it is assumed that the GF-PDF is the same for all particles classified by the 1st DMA at a given voltage, i.e., $c_{\mathrm{cond}}(g, D_{\mathrm{p1}})$ is independent of D_{p1}. ϵ_i is the error in the measured response. Note that equal aerosol flows are assumed for both the 1st and 2nd DMAs (i.e., $Q_{\mathrm{a,1}} = Q_{\mathrm{s,1}}$, $Q_{\mathrm{a,2}} = Q_{\mathrm{s,2}}$).

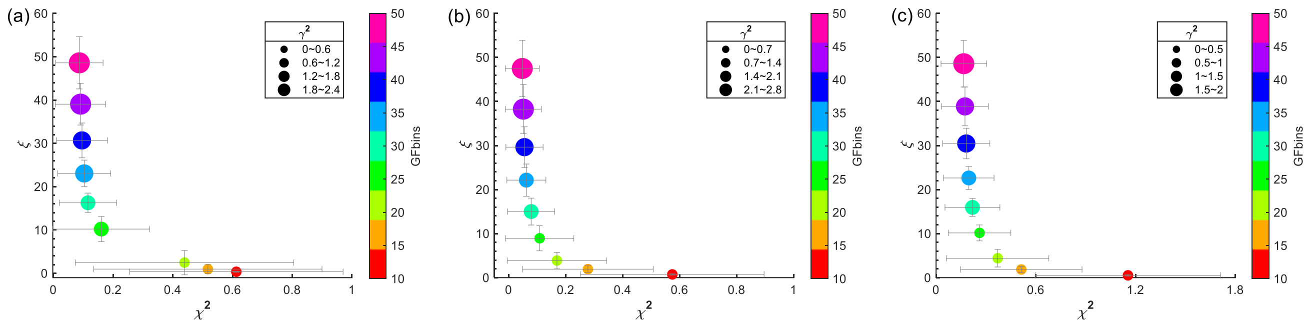
S2. Dependences of reconstruction residual and smoothness on the number of GF bins for pre-defined GF-PDFs inverted using unregularized LSQ and Tikhonov regularization



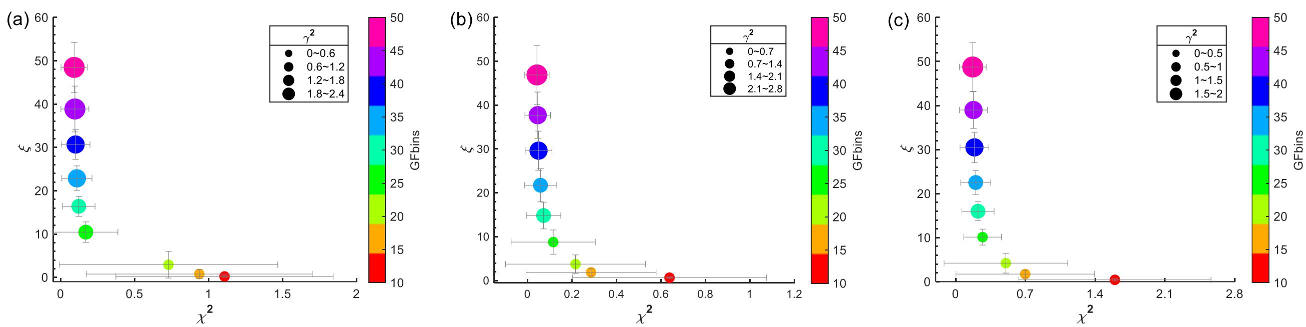
~~Fig~~Figure~~.~~ **S1.** “L-curve” showing the dependence of reconstruction residual, χ^2, and the smoothness, ξ, on the number of GF bins of pre-defined GF-PDFs with (**a**) one mode, (**b**) two modes, and (**c**) three modes, respectively. The symbol size represents the error in inverted GF-PDF, γ^2. Whiskers represent standard deviation. The inversion is conducted using unregularized LSQ method.



Figure~~Fig.~~ **S2.** "L-curve" showing the dependence of reconstruction residual, χ^2, and the smoothness, ξ, on the number of GF bins of pre-defined GF-PDFs with (**a**) one mode, (**b**) two modes, and (**c**) three modes, respectively. The symbol size represents the error in inverted GF-PDF, γ^2. Whiskers represent standard deviation. The inversion is conducted using 0th order Tikhonov regularization method.



Figure~~Fig.~~ **S3.** "L-curve" showing the dependence of reconstruction residual, χ^2, and the smoothness, ξ, on the number of GF bins of pre-defined GF-PDFs with (**a**) one mode, (**b**) two modes, and (**c**) three modes, respectively. The symbol size represents the error in inverted GF-PDF, γ^2. Whiskers represent standard deviation. The inversion is conducted using 1st order Tikhonov regularization method.



Figure~~Fig.~~ **S4.** "L-curve" showing the dependence of reconstruction residual, χ^2, and the smoothness, ξ, on the number of GF bins of pre-defined GF-PDFs with (**a**) one mode, (**b**) two modes, and (**c**) three modes, respectively. The symbol size represents the error in inverted GF-PDF, γ^2. Whiskers represent standard deviation. The inversion is conducted using 2nd order Tikhonov regularization method.

S3. Effect of instrument uncertainties due to calibration nonideality

We also challenged our algorithm with different forward and inverse models to simulate the scenarios when DMA or WFIMS is not perfectly calibrated. The particle sizes measured by DMA and WFIMS are determined by the voltage and sheath flow, which can be calibrated straightforwardly. Therefore, the nonideality in DMA and WFIMS performance likely manifests in the deviation of instrument mobility resolution from the theoretical value. To test the performance of inversion algorithms for such scenarios when the transfer function of DMA or WFIMS is not fully calibrated, we generate the synthetic HFIMS measurements by perturbing DMA or WFIMS mobility resolution (i.e., R_{DMA} or R_{WFIMS}), while maintaining the theoretical R_{DMA} or R_{WFIMS} in the inverse model. The mobility resolution is perturbed by varying the ratio of sheath to aerosol flow for DMA or WFIMS (R_Q=Q_{sh}/Q_{a}) in the derivation of the transfer function. The default flow rate ratio for DMA and WFIMS are 10 and 50, respectively. Figures S5 and S6 show the inversion results when DMA R_Q in the forward model is varied from 8 to 12 while WFIMS R_Q is maintained at the actual value of 50 and when DMA R_Q is maintained at 10 while WFIMS R_Q is varied from 40 to 60. The results are based on inversions of 500 sets of synthetic HFIMS measurements (with the noise of counting statistics included) using Twomey's method. The average residual (χ^2), the GF-PDF error (γ^2), and the smoothness (ξ), all showed very minor variation with DMA or WFIMS R_Q used in the forward model, suggesting negligible impacts on Twomey inversion results due to imperfect calibration of DMA and WFIMS resolution. A possible explanation is that typical GF-PDFs of ambient aerosol particles are relatively broad such that the inverted GF-PDF is insensitive to DMA and WFIMS resolutions. The impact of R_Q on other nonparametric methods was also investigated and found negligible.

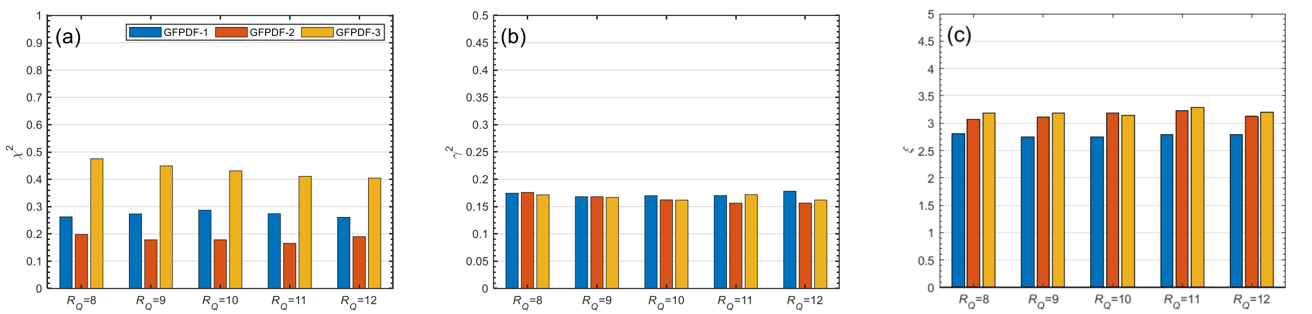


Figure S5. The reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of GF-PDF inverted using Twomey's method as a function of DMA R_Q used to calculate DMA transfer function in the forward model (WFIMS R_Q maintained at the actual value of 50). Actual DMA and WFIMS R_Q values of 10 and 50 are used to derive transfer functions in the inverse model (i.e., calculation of the inversion matrix). The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on the inversion of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

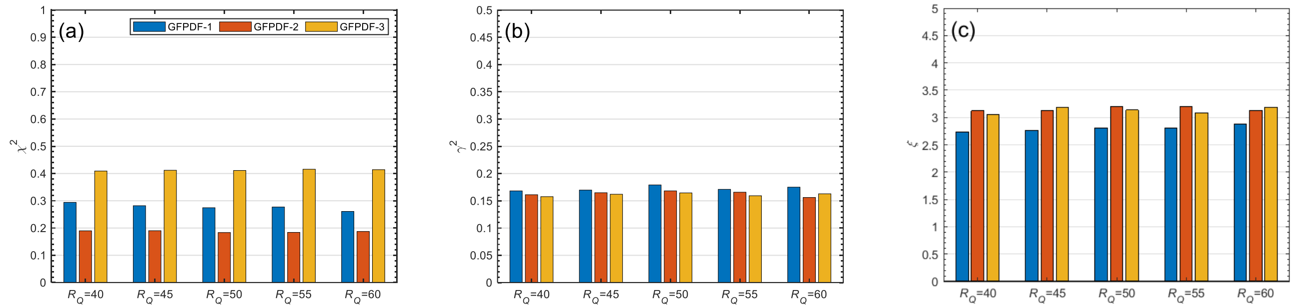


Figure S6. The reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of GF-PDF inverted using Twomey's method as a function of WFIMS R_Q used to calculate WFIMS transfer function in the forward model (DMA R_Q maintained at the actual value of 10). Actual DMA and WFIMS R_Q values of 10 and 50 are used to derive transfer functions in the inverse model (i.e., calculation of the inversion matrix). The colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on the inversion of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs.

S4. Effect of different level of counting noises

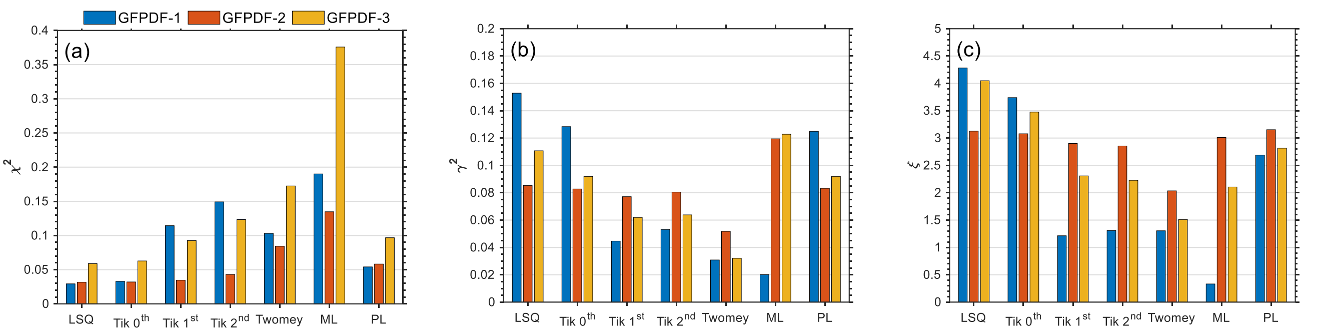


Figure S7. Comparison of reconstruction residual, χ^2 (**a**), the GF-PDF error, γ^2 (**b**), and the smoothness, ξ (**c**) of inverted GF-PDFs using different inversion methods. Colors correspond to the pre-defined GF-PDFs with one mode (blue), two modes (orange), and three modes (yellow). The results are averages based on inversions of 500 sets of synthetic HFIMS data for each of three pre-defined GF-PDFs with total particle counts of 500.

S5. Applications of nonparametric inversion methods to ambient HFIMS measurements

We apply the nonparametric inversion methods to ambient HFIMS measurements, and the results are compared in Fig. S8. The HFIMS responses reconstructed from GF-PDF inverted using unregularized LSQ, Tikhonov, and Twomey's methods generally match the measurement (black circle) well. The GF-PDF at 85% RH for ambient 35 nm particles consist of a smaller less-hygroscopic mode and a larger more-hygroscopic mode. As expected, the HFIMS response reconstructed from LSQ inverted GF-PDF has the minimum deviation from the actual measurement whereas the GF-PDF exhibits oscillations near the tail of the second mode. These oscillations create a small third mode that is absent from the smoother GF-PDFs inverted using regularized methods (i.e., Tikhonov and Twomey's methods). GF-PDF inverted using Twomey's method and 0th Tikhonov clearly distinguish the two growth factor modes. In comparison, the two modes become more overlapped in GF-PDF inverted using 1st and 2nd Tikhonov regularization, due to additional and possibly excessive regularization.

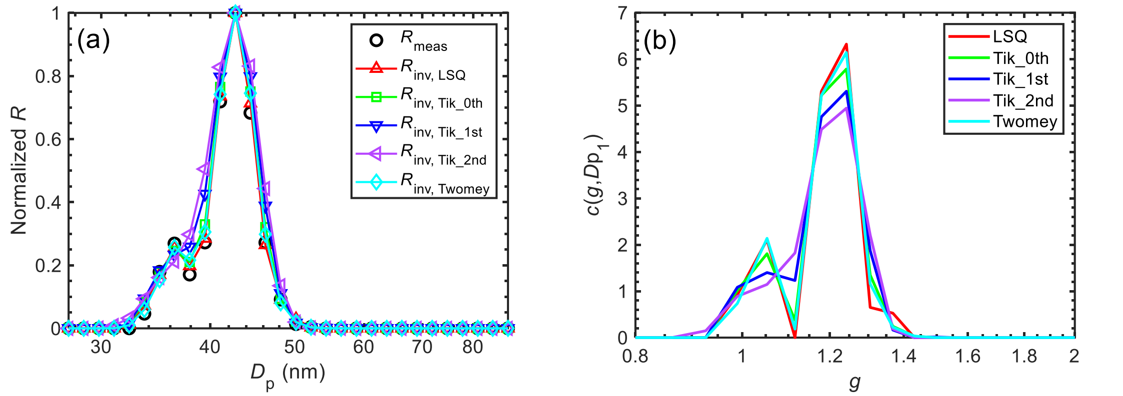


Figure S8. (**a**) Comparison between the HFIMS measured response (black circle) and the responses (marked lines) reconstructed from GF-PDF derived using different methods for 35 nm ambient aerosol at 85% RH. (**b**) Inverted GF-PDFs using different methods.

We also examined the statistics of the reconstruction residual and the smoothness of GF-PDF inverted from 3-day HFIMS measurements using the listed nonparametric methods. Among all nonparametric inversion methods, unregularized LSQ leads to the lowest reconstruction residual but the worst smoothness (Fig. S9). As regularizations are introduced in the Tikhonov algorithms, the inverted GF-PDFs become smoother at the expense of increased reconstruction residuals. The Tikhonov regularized solutions strongly depend on the regularization parameter λ. In this study, the value of λ has been derived using three approaches, including (1) the L-curve, (2) the Hanke-Raus rule, and (3) comparison of inverted GF-PDF with the true solution. Note the 3rd approach (i.e., comparison of inverted GF-PDF with the true solution) is not possible for ambient measurements. Inversions of synthetic data show that the L-curve approach generally underestimates the regularization parameter (Fig. 5 in the manuscript), resulting in insufficiently regularized solutions. For the 3-day ambient measurements, when λ is derived using the L-curve approach, the reconstruction residuals for the GF-PDF inverted using Tikhonov algorithms are very close to those of the unregularized LSQ, consistent with underestimated λ values (Fig. S9a and d). In contrast, Tikhonov regularizations with λ value determined using the Hanke-Raus rule tend to over-smooth solutions due to overestimated λ values, resulting in significantly increased errors in reconstructed HFIMS measurements (Fig. S9b and e). The 3-day ambient measurements are also inverted using Tikhonov algorithms with an empirical λ value of 0.03 (Fig. S9c and f), which corresponds to the mean value of optimized λ values (i.e., derived using the 3rd approach) for the synthetic HFIMS data. The inverted GF-PDF shows improved smoothness compared to the solution from the LSQ method, without introducing excessive reconstruction errors. While the empirical λ value appears to work quite well for the 3-day measurements, using this fixed regularization parameter may not be appropriate for other ambient measurements. For Twomey's method, both the reconstruction residual and the smoothness are between those based on the 0th order and 1st order Tikhonov regularizations with the empirical regularization parameter (λ = 0.03), suggesting an appropriate trade-off between the GF-PDF smoothness and the fidelity in reproducing the HFIMS measurements. Note that the statistics of the GF-PDF error cannot be derived as the actual GF-PDF of ambient aerosols are unknown. As a result, it is difficult to draw a definite conclusion regarding which method has the best performance in retrieving the GF-PDF based on the ambient measurements.

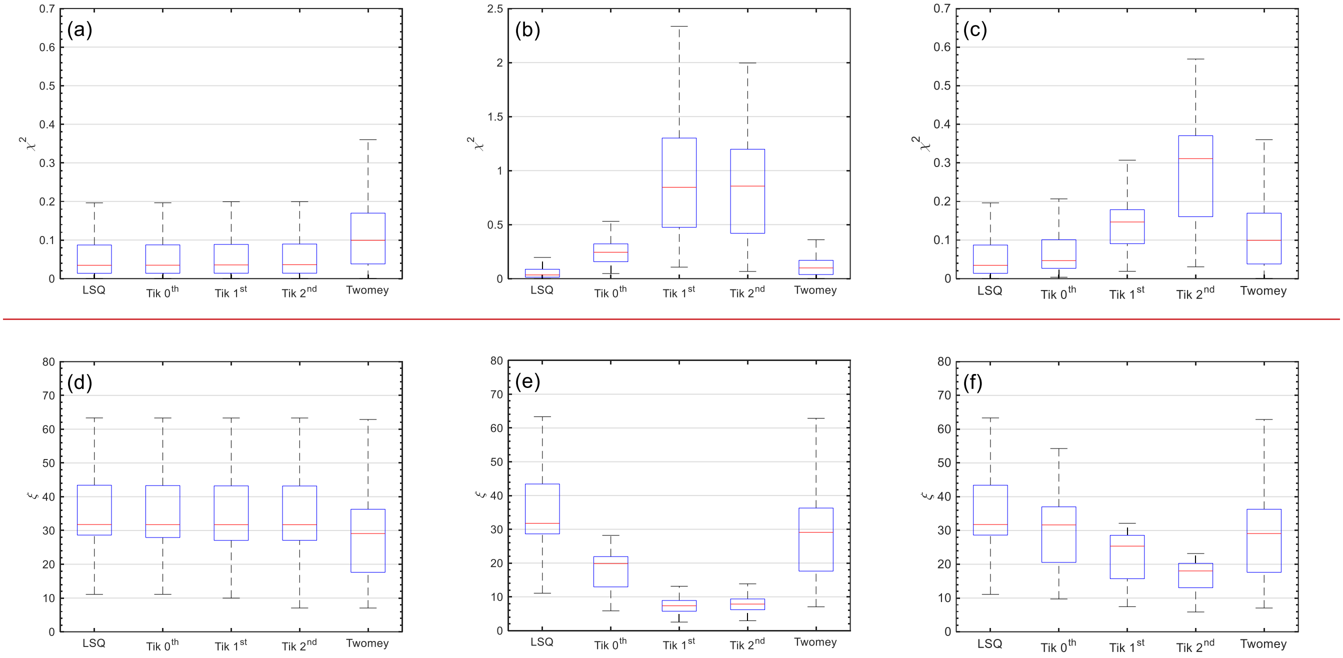


Figure S9. Comparison of reconstruction error, χ^2 (**a, b, c**) and the smoothness, ξ (**d, e, f**) of inverted GF-PDFs using different inversion methods (i.e., LSQ, Tikhonov of 0, 1, 2-th order, and Twomey's method), based on 3-day HFIMS measurements of ambient aerosols of 35 nm at five different RH levels (20%, 40%, 60%, 75%, and 85%). The Tikhonov regularization parameters are derived using the L-curve approach (**a, d**), the Hanke-Raus rule (**b, e**), and an empirical value of 0.03 (**c, f**), respectively.