Atmos. Chem. Phys., 26, 16653–16666, 2026
https://doi.org/10.5194/acp-26-16653-2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Chemistry and Physics
Open Access
EGU

Technical note: Deep-GF-PRM – a deep learning framework for parameterizing aerosol hygroscopic growth factor probability density function

Huiping Lu[1,2], Jiaoshi Zhang[1], Jingyi Zhang[1,2], Shengqian Zhou[3], Yang Wang[4], Jianguo Liu[2], and Huaqiao Gui[2]

[1]Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, China
[2]Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
[3]Center for Aerosol Science and Engineering, Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA
[4]Department of Chemical, Environmental and Materials Engineering, University of Miami, Coral Gables, Florida 33146, USA

Correspondence: Jiaoshi Zhang (jiaoshi@ustc.edu.cn) and Jianguo Liu (jgliu@aiofm.ac.cn)

Received: 12 February 2026 – Discussion started: 10 March 2026
Revised: 14 August 2026 – Accepted: 16 October 2026 – Published: 23 November 2026

Abstract. The hygroscopic properties of atmospheric aerosols are crucial for quantifying their impact on radiation and cloud formation. They are often characterized by a growth factor probability density function (GF-PDF), which can be parameterized as a superposition of multiple Gaussian distributions. Conventionally, nonparametric inversion methods are developed to retrieve GF-PDF from the instrument responses, e.g., measurements of humidified tandem differential mobility analyzer (HTDMA). However, additional parametric fittings are required to extract modal parameters from the inverted GF-PDF, a process that is computationally intensive and susceptible to fitting errors.

In this study, we introduce Deep-GF-PRM, a deep learning framework that parameterizes the GF-PDF modal parameters directly from the instrument responses. The core of Deep-GF-PRM is a deep neural network that embeds the instrument's kernel function and physical constraints, creating end-to-end mapping from the instrument response to the GF-PDF modal parameters. Trained on a large dataset of synthetic instrument responses generated using a wide range of GF-PDFs and noise levels, Deep-GF-PRM accurately reproduces synthetic GF-PDFs and retrieves modal parameters with high fidelity. The model is applied to real-world measurements, and yields results highly consistent with nonparametric inversions. Deep-GF-PRM thus provides an efficient solution for parameterizing aerosol hygroscopic properties.

1 Introduction

The uptake of water by atmospheric aerosols alters their size and refractive index, thereby significantly affecting aerosol optical properties and radiative effects (Swietlicki et al., 2008; Pinterich et al., 2017; Pilinis et al., 1995). In addition, the hygroscopic growth of aerosols is also closely related to their cloud condensation nuclei (CCN) activity and is essential for the formation of cloud droplets (Wang et al., 2002; George and Abbatt, 2010). The hygroscopic growth of aerosols is governed primarily by their chemical composition and is a function of the ambient relative humidity (RH). It is commonly measured using a humidified tandem differential mobility analyzer (HTDMA), which exposes monodisperse dry particles to an elevated relative humidity (e.g., 90 %) and measures the size distribution of humidified particles (Liu et

al., 1978; Rader and McMurry, 1986; Tang et al., 2019). A recently developed humidity-controlled fast integrated mobility spectrometer (HFIMS) drastically accelerates aerosol hygroscopic growth measurements by employing a water-based fast integrated mobility spectrometer (WFIMS) that captures the humidified particle size distribution at a time resolution of 1 s (Pinterich et al., 2017). As a result, HFIMS extends the size-resolved hygroscopic growth measurements to multiple RH levels (Pinterich et al., 2017; Zhang et al., 2021).

The hygroscopic growth is characterized by a growth factor probability density function (GF-PDF), where the growth factor (GF) is defined as the ratio of the humidified particle size at a specified RH to its dry size. Due to the heterogeneity of aerosol chemical compositions, the growth factor is often not uniform and exhibits a wide range, leading to a broad and often multimodal distribution of GF-PDF (Swietlicki et al., 2008; Alonso-Blanco et al., 2019). While the first moment of the GF-PDF represents the ensemble mean growth factor, the shape and spread of the GF-PDF provide detailed insights into their chemical composition and mixing state. For better interpretation of the GF-PDF, it is characterized by a multimodal Gaussian or lognormal distribution (Stolzenburg and McMurry, 1988; Stolzenburg and McMurry, 2008) and parameterized by a series of modal parameters including number fraction (f), mean GF (G), and standard deviation (σ) of individual modes.

The GF-PDF can be retrieved from the instrument response (i.e., the HTDMA or HFIMS measurements) using both parametric and nonparametric methods. In the parametric method, the GF-PDF is represented with a specific functional form, and the function parameters are derived through parametric fitting that minimizes the residual between the measured and reconstructed size distributions of humidified particles. For example, the TDMAfit algorithm (Stolzenburg and McMurry, 1988) assumes the GF-PDF to be a superposition of multiple Gaussian functions and fits the modal parameters via least-squares minimization. However, the TDMAfit algorithm is often not robust in identifying closely overlapped modes, and the convergence of fitting strongly relies on initial guess of the modal parameters (Gysel et al., 2009).

In the nonparametric method (Cubison et al., 2005; Gysel et al., 2009; Petters, 2021; Zhang et al., 2022), a kernel matrix is designed to map the GF-PDF to the instrument response, such that the GF-PDF can be directly inverted without the knowledge of its functional form. It is unsupervised and more robust. The matrix-based inversion also facilitates the implementation of regularizations (e.g., Tikhonov regularization and Twomey's iterative regularization) to suppress noise amplification during data inversion (Zhang et al., 2022; Petters, 2021). However, the nonparametric inversion yields only a discrete GF-PDF spectrum without explicit modal information, and a post fitting is required to identify separate modes with distinct hygroscopic behavior and extract the parameters of each mode. This step is often performed by a least-squares fitting of multiple Gaussians to the inverted GF-PDF (Holmgren et al., 2014; Alonso-Blanco et al., 2019). In practice, this additional fitting could increase computational cost and introduce extra uncertainties, especially for noisy or overlapped modes.

Deep learning has emerged as a powerful tool and has been widely employed to address computational bottlenecks in atmospheric and climate models (Reichstein et al., 2019; Huang and Seinfeld, 2022). Pure data-driven deep learning networks rely on large training datasets for pattern recognition and often suffer from overfitting and limited generalization (Reichstein et al., 2019; Molina Catricheo et al., 2024). Incorporating physical knowledge into deep learning, e.g., by introducing physics-based loss terms, embedding known governing equations, or enforcing conservation constraints, can significantly improve model robustness and generalization, especially when training data are limited or the underlying problem is ill-posed (Karniadakis et al., 2021; Cai et al., 2021). This general paradigm encompasses a spectrum of approaches, from Physics-Informed Neural Networks (PINNs) that solve PDEs to physics-constrained models that regularize the learning process with physically meaningful consistency penalties, and has been successfully applied across various aspects of aerosol science. Recent examples include secondary organic aerosol formation in chemical transport models (Sharma et al., 2023), global dust deposition mapping with advection-diffusion constraints (Molina Catricheo et al., 2024), and hybrid aerosol retrieval from satellite data (Fu et al., 2024).

In this study, we propose a concept that nonparametric inversion and parameterization of GF-PDFs can be emulated by a deep learning model. We introduce Deep-GF-PRM (Lu and Zhang, 2026), a deep learning framework that parameterizes the GF-PDF modal parameters directly from the raw instrument responses. Deep-GF-PRM is designed to take instrument responses as input and directly outputs modal parameters (i.e., f, G, σ) of the GF-PDF. Trained on a large dataset of synthetic instrument responses generated using a wide range of GF-PDFs with various modal combinations and noise levels, Deep-GF-PRM achieves good predictive capability and strong generalizability across instruments with different transfer functions. The model performance is evaluated on synthetic datasets and compared with nonparametric inversion method and other machine learning algorithms. The model is also applied to parameterization of GF-PDFs from the real-world aerosol hygroscopic growth measurements by HFIMS and HTDMA.

2 Method

2.1 The GF-PDF inversion problem

Figure 1 depicts the typical data processing procedure for aerosol hygroscopic growth measurements, which can be briefed as a two-step process: First, the nonparametric

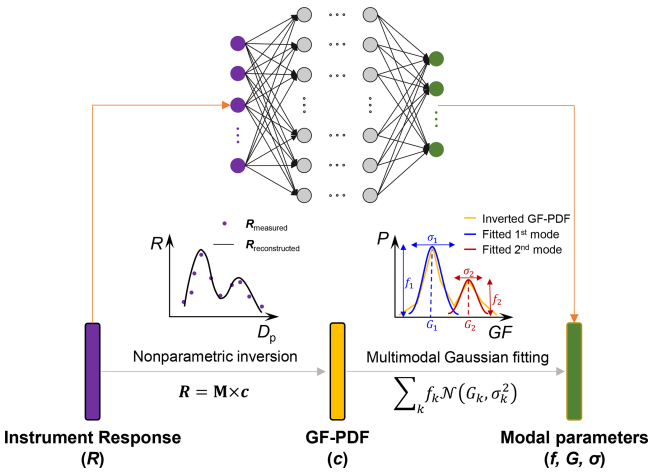


Figure 1. Scheme of parameterization of aerosol hygroscopic growth factor probability density function (GF-PDF). The conventional approach, i.e., nonparametric inversion of the GF-PDF spectrum from the instrument response followed by a multimodal Gaussian fitting to retrieve the modal parameters, can be substituted by a deep learning model which directly maps the instrumental responses to the GF-PDF modal parameters.

method is employed to retrieve GF-PDF from the raw instrument response. Second, to parameterize the inverted GF-PDF, an additional parametric fitting is required to extract the modal parameters from the inverted GF-PDF spectrum. The rest of this section elaborates on derivation of nonparametric inversion algorithm and parameterization of the inverted GF-PDF using a sum of multimodal Gaussian distributions.

The instrument response of aerosol hygroscopic growth measurements is determined by the aerosol size distribution, the GF-PDF, and the instrument transfer function (e.g., HTDMA or HFIMS). According to Stolzenburg and McMurry (2008), the theoretical HTDMA response of the ith D_{p}^{*} bin can be derived as:

$$\begin{aligned} R_i = & \frac{Q_{\mathrm{a2}}}{Q_{\mathrm{s2}}}\eta_{\mathrm{CPC}}\left(D_{\mathrm{p2},i}^{*}\right)\eta_{\mathrm{p,cond}}\left(D_{\mathrm{p2},i}^{*}\right) \\ & \times \frac{Q_{\mathrm{a1}}}{Q_{\mathrm{s1}}}\eta_{\mathrm{chg}}\left(D_{\mathrm{p1}}^{*}\right)n\left(D_{\mathrm{p1}}^{*}\right) \\ & \times \iint c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)\Omega_1\left[V_1,\tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right] \\ & \times \Omega_2\left[V_{2,i},\tilde{Z}_{\mathrm{p}}(D_{\mathrm{p2}})\right]\mathrm{d}D_{\mathrm{p2}}\mathrm{d}D_{\mathrm{p1}}+\epsilon_i \end{aligned} \quad (1)$$

where Q_{a1} and Q_{s1}, Q_{a2} and Q_{s2} are the aerosol inlet and outlet flows of the 1st DMA and 2nd DMA, respectively. $\eta_{\mathrm{p,cond}}\left(D_{\mathrm{p2},i}^{*}\right)$ and $\eta_{\mathrm{CPC}}\left(D_{\mathrm{p2},i}^{*}\right)$ are the penetration efficiency of particles going through the RH conditioner and the CPC detection efficiency, and are assumed to be 1. $n\left(D_{\mathrm{p1}}^{*}\right)$ is the aerosol number size distribution and $\eta_{\mathrm{chg}}\left(D_{\mathrm{p1}}^{*}\right)$ is the aerosol charging efficiency. D_{p1}^{*} is the centroid diameter of the 1st classifying DMA, and $D_{\mathrm{p2},i}^{*}$ is the geometric mean diameter of the ith D_{p}^{*} bin of the 2nd scanning DMA. $\Omega_1\left[V_1,\tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right]$ is the transfer function of the 1st DMA with a classifying voltage of V_1, and $\tilde{Z}_{\mathrm{p}}\left(D_{\mathrm{p1}}\right)$ is the particle mobility normalized by the centroid mobility of the 1st DMA corresponding to V_1. $\Omega_2\left[V_{2,i},\tilde{Z}_{\mathrm{p}}\left(D_{\mathrm{p2}}\right)\right]$ is the transfer function of the 2nd DMA corresponding to a scanning voltage $V_{2,i}$, and $\tilde{Z}_{\mathrm{p}}\left(D_{\mathrm{p2}}\right)$ is the mobility of humidified particles normalized by the centroid mobility of the 2nd DMA. In other words, Ω_1 is the transmission probability of the 1st DMA at the selected dry size, and Ω_2 is the transmission probability of the 2nd DMA at each voltage scan step, which resolves the size (or GF) distribution of humidified particles. ϵ_i is the random error. $c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)$ represents the GF-PDF for particles with a dry diameter of D_{p1} growing to a diameter of D_{p2} during the humidity conditioning process and integrates to unity, i.e., $\int_{D_{\mathrm{p2}}=0}^{D_{\mathrm{p2}}=\infty}c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)\mathrm{d}D_{\mathrm{p2}}=1$.

While these efficiencies and size distribution are relatively constant within the narrow non-zero mobility range of the DMA transfer function, the first term in the right-hand side of Eq. (1) is regarded as constant and defined as E, and Eq. (1) can be simplified as:

$$\begin{aligned} R_i = & E\iint c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)\Omega_1\left[V_1,\tilde{Z}_{\mathrm{p}}(D_{\mathrm{p1}})\right] \\ & \times \Omega_2\left[V_{2,i},\tilde{Z}_{\mathrm{p}}(D_{\mathrm{p2}})\right]\mathrm{d}D_{\mathrm{p2}}\mathrm{d}D_{\mathrm{p1}}+\epsilon_i \end{aligned} \quad (2)$$

Note that the GF-PDF $c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)$ can be normalized as a function of growth factor g (i.e., $D_{\mathrm{p2}}/D_{\mathrm{p1}}$), which satisfies $c_{\mathrm{cond}}\left(g,D_{\mathrm{p1}}\right)\mathrm{d}g=c_{\mathrm{cond}}\left(D_{\mathrm{p2}},D_{\mathrm{p1}}\right)\mathrm{d}D_{\mathrm{p2}}$. Given the narrow particle size range classified by the 1st DMA, it is assumed that the GF-PDF is the same for all particles classified by the 1st DMA, and therefore $c_{\mathrm{cond}}\left(g,D_{\mathrm{p1}}\right)$ is independent of D_{p1} for the integration in Eq. (2) and can be simplified as $c_{\mathrm{cond}}(g)$ or $c(g)$. Rewriting the GF-PDF as $c_{\mathrm{cond}}(g)$ and replacing D_{p2} with gD_{p1} in Eq. (2) gives:

$$\begin{aligned} R_i = & E\int_0^{+\infty}\mathrm{d}g\;c_{\mathrm{cond}}(g)\int_0^{+\infty}\mathrm{d}D_{\mathrm{p1}} \\ & \times \Omega_1\left[V_1,\tilde{Z}_{\mathrm{p}}\left(D_{\mathrm{p1}}\right)\right]\Omega_2\left[V_{2,i},\tilde{Z}_{\mathrm{p}}\left(gD_{\mathrm{p1}}\right)\right]+\epsilon_i \\ \approx & E\sum_{j=1}^{J}c_{\mathrm{cond}}\left(g_j\right)\int_{g_{j-\frac{1}{2}}}^{g_{j+\frac{1}{2}}}\mathrm{d}g\int_0^{+\infty}\mathrm{d}D_{\mathrm{p1}} \\ & \times \Omega_1\left[V_1,\tilde{Z}_{\mathrm{p}}\left(D_{\mathrm{p1}}\right)\right]\Omega_2\left[V_{2,i},\tilde{Z}_{\mathrm{p}}\left(gD_{\mathrm{p1}}\right)\right]+\epsilon_i \end{aligned} \quad (3)$$

where the integration can be approximated by a sum over J GF bins, with the assumption that $c_{\mathrm{cond}}(g)$ is a constant value within each GF bin. This forward response is then cast into a matrix equation:

$$\boldsymbol{R}=\mathbf{M}\times\boldsymbol{c}+\boldsymbol{\epsilon} \quad (4)$$

where the HTDMA response $\boldsymbol{R}$ is a $I\times 1$ array composed of R_i ($i=1,2,3,\ldots,I$), the unknown GF-PDF $\boldsymbol{c}$ is a $J\times 1$

array composed of c_j ($j = 1, 2, 3, \ldots, J$), and the HTDMA kernel function $\mathbf{M}$ is a $I \times J$ matrix that 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 $D^*_{\mathrm{p},i-1/2}$ and $D^*_{\mathrm{p},i+1/2}$. $\boldsymbol{\epsilon}$ is a vector that denotes the random error superimposed as a result of measurement uncertainties. Note that Eq. (4) is also valid for describing the HFIMS response as detailed in Zhang et al. (2022).

Due to the overlapping of the instrument kernel function $\mathbf{M}$, inversion of the GF-PDF from the HTDMA or HFIMS measurements is an ill-posed problem (Gysel et al., 2009). Regularizations are often applied to nonparametric inversions, which can greatly reduce the error in inverted GF-PDF by eliminating noise amplification. Zhang et al. (2022) compared Tikhonov regularization and Twomey's iterative regularization for GF-PDF retrieval from HFIMS measurements and demonstrated that Twomey's method better balances the smoothness of the inverted GF-PDF with fidelity in reproducing the instrument response. As a result, we adopt Twomey's method as the representative nonparametric inversion method throughout this work.

Due to the unknown functional form, the inverted GF-PDF using nonparametric methods is a discrete curve with respect to the predefined GF values (e.g., between 0.8–2.0), with an integral area of unity (Gysel et al., 2009). For the GF-PDF with a narrow, single-mode distribution, e.g., produced by aerosols with a single/uniform composition, the first moment and the standard deviation of the GF-PDF represent the mean GF and the spread factor of GF-PDF. However, the hygroscopic growth of ambient aerosols often exhibits multimodal patterns (i.e., with two or three modes), as a result of the existence of particles with distinctly different hygroscopic properties (Swietlicki et al., 2008; Alonso-Blanco et al., 2019), and hence the first moment and standard deviation of the GF-PDF is inadequate to characterize different hygroscopic modes. According to Stolzenburg and McMurry (1988), the GF-PDF can be represented by a sum of multimodal Gaussian distributions:

$$c_{\mathrm{cond}}(g) = \sum_{k=1}^{K} \frac{f_k}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(g-G_k)^2}{2\sigma_k^2}\right] \tag{5}$$

where K is the mode number, f_k, G_k, and σ_k are the number fraction, mean GF, and standard deviation of each individual mode. The modal parameters can be obtained by fitting the inverted GF-PDF with least-squares minimizations (Holmgren et al., 2014; Zhang et al., 2022). A common practice is to fit the GF-PDF with different mode numbers (i.e., K ranging from 1 to 3), and then select the best fit with the smallest mean square error between the inverted GF-PDF and the reconstructed from the fitted modal parameters. This additional fitting is not only computationally inefficient but also introduces extra uncertainties (see Sect. 3.1), as the convergence of parametric fittings relies on the proper selection of initial guess. As a result, although the nonparametric inversion of GF-PDF is unsupervised and can be fully automated, the parametric fitting may encounter occasional convergence failures, making the automated analysis of large datasets difficult.

2.2 Deep-GF-PRM framework

The conventional GF-PDF parameterization involves a two-step process, i.e., nonparametric inversion of the GF-PDF spectrum from the noise-blurred instrument response followed by a multimodal Gaussian fitting. To simplify this procedure, we propose Deep-GF-PRM, an end-to-end framework that directly maps the raw instrument responses of HTDMA or HFIMS (R) to the GF-PDF modal parameters (f, G, σ), as shown in Fig. 1. The Deep-GF-PRM framework, as schematized in Fig. 2, consists of an input layer, a series of hidden layers, and an output layer. The input layer receives the synthetic instrument response R (64×1). Since the HTDMA measurements are usually conducted at multiple sizes (e.g., 5 sizes ranging from 50 to 250 nm) and hence the measured size ranges are not identical, R is first scaled to a uniform GF range of 0.8–2.0 to eliminate the size information. Here we use R generated using a fixed HTDMA kernel function (i.e., at a fixed size of 50 nm) for training the model. The output layer consists of 9 neurons representing the modal parameters of up to three modes and are grouped into three categories, i.e., number fractions (f_1, f_2, f_3), mean growth factors (G_1, G_2, G_3), and standard deviations (σ_1, σ_2, σ_3). For samples with fewer than three modes, the unused parameter slots in the label vector are set to -1, to keep consistent vector dimensions. During evaluation, only modes with $f > 0.01$ and $\sigma > 0.01$ are retained as valid and accounted for reconstructing the GF-PDF.

The hidden layers are composed by a series of cascaded residual blocks, with each block consists of two fully connected layers. Each hidden layer contains 128 neurons with a Rectified Linear Unit (ReLU) activation function, as shown in Fig. 2. Given that the mapping between GF-PDF modal parameters and the instrument response is highly nonlinear, a neural network with greatly increased depth is required to capture the high-dimensional features. However, simply stacking fully connected layers often subjects to gradient vanishing issues and leads to degradation of training accuracy with increasing depth (He et al., 2016). Here we adopted the deep residual learning architecture proposed by He et al. (2016) and add residual connections between each two stacked layers, allowing the gradients propagating directly through the shortcut path without exploding gradients. The output of each residual block is expressed as:

$$x_{\mathrm{out}} = \mathrm{ReLU}(W_2 \cdot \mathrm{ReLU}(W_1 \cdot x_{\mathrm{in}} + b_1) + b_2) + x_{\mathrm{in}} \tag{6}$$

where x_{in} and x_{out} are the input and output features, and W_1, W_2 and b_1, b_2 are the weights and biases of each fully connected layer. Note that the depth of residual blocks (i.e.,

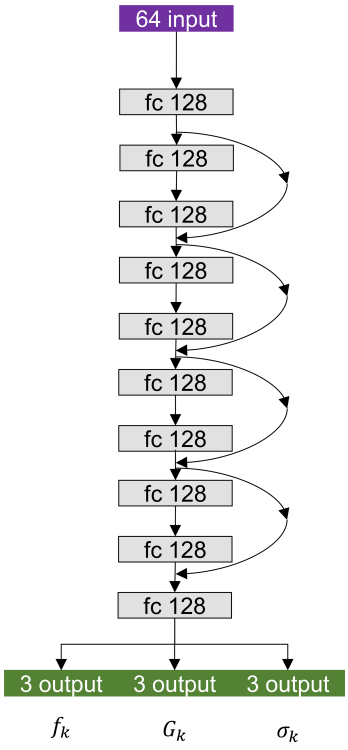


Figure 2. The network architecture of Deep-GF-PRM. The synthetic instrument response R (64 × 1) are regarded as input features, and the 3 sets of modal parameters (f, G, σ) are output targets (9 × 1). The outputs of absent modes are set to −1. The hidden layers consist of 4 fully connected blocks with residual connections, which allows gradients propagating directly back to previous blocks.

4) and neurons number of each layer (i.e., 128) were selected via hyperparameter tuning, which strikes the balance between model prediction accuracy and overfitting risks.

The inversion of GF-PDF from the instrument response is inherently an ill-posed problem, and regularization is required to stabilize the solution (Gysel et al., 2009; Zhang et al., 2022). Here we proposed a physics-constrained approach that integrates physics-based regularization into the learning process; specifically, the loss function combines parameter loss and physics loss terms:

$$L = \lambda_{\text{param}} \cdot \mathcal{L}_{\text{param}} + \lambda_{\text{phys}} \cdot \mathcal{L}_{\text{phys}} + \lambda_{\text{sp}} \cdot \mathcal{L}_{\text{sp}} \quad (7)$$

The parameter loss $\mathcal{L}_{\text{param}}$ is defined as Smooth L_1 loss on the raw 9-parameter vector. $\mathcal{L}_{\text{phys}}$ is the Smooth L_1 loss between the GF-PDF curves (i.e., $c_{\text{cond}}(g)$) reconstructed from the predicted and true parameters (i.e., Eq. 5), which constrains that the predicted parameters faithfully reproduce the true GF-PDF shape. $\mathcal{L}_{\text{sp}}$ is the L_1 norm of the predicted mode fractions f_k. It resolves the non-uniqueness in the GF-PDF parameterization, i.e., a single broad mode can be approximated by two narrower modes, by driving superfluous mode fractions to zero. Among parameter combinations that produce the same GF-PDF shape, this penalty steers the model toward the representation with the fewest modes. The weight of each loss ($\lambda_{\text{param}} = 0.3$, $\lambda_{\text{phys}} = 1.0$, and $\lambda_{\text{sp}} = 0.01$), were determined through Bayesian hyperparameter optimization with cross-validation.

2.3 Generation of synthetic data for model training and evaluation

The training and evaluation of Deep-GF-PRM require a large dataset with ground-truth labels. Since the true GF-PDF is unknown from real-world measurements, synthetic data corrupted with measurement noise were generated following the procedure detailed in Wang et al. (2019) and Zhang et al. (2022), which is briefly described here. For a representative GF-PDF with known modal parameters (i.e., f, G, σ), the GF-PDF spectrum (i.e., $c_{\text{cond}}(g)$) is first reconstructed and the theoretical HTDMA response (i.e., R_i) is derived, following Eqs. (5) and (2), respectively. Gaussian and Poisson noise are then added to the response to represent the random instrumental noise and account for counting statistics.

In Zhang et al. (2022), only three representative GF-PDFs were compiled in the synthesized HTDMA measurements (i.e., three sets of GF-PDF modal parameters). Here, for supervised learning of the nonlinear mapping between instrument response and GF-PDF modal parameters, we generalize the synthesized data by expanding the parameter space to represent the diversified distribution of GF-PDFs in real-world measurements. To ensure the synthetic GF-PDFs are physically valid without any unrealistic overlaps, the sampled modal parameters are constrained as follows:

1. Each mode should satisfy $0.05 \le f \le 1$, $0.95 \le G \le 1.8$, and $0.015 \le \sigma \le 0.15$;
2. For unimodal cases, $0.95 \le G \le 1.6$, and $\sigma \in [0.015, 0.15]$;
3. For bimodal cases, $0.95 \le G_1 \le 1.3$, $G_2 \le 1.6$, and $G_2 - G_1 > 0.05$, and $\sigma \in [0.025, 0.15]$ for both modes;
4. For trimodal cases, $0.95 \le G_1 \le 1.3$, $G_2 \le 1.6$, $G_3 \le 1.8$, $G_2 - G_1 > 0.1$, and $G_3 - G_2 > 0.2$, and $\sigma \in [0.025, 0.15]$ for all three modes.

These ranges were randomly sampled to generate combinations of GF-PDF modal parameters with up to three modes. Each parameter set was sorted in the G-ascending order, and the absent modes were filled with −1 to maintain a fixed data dimension. Moreover, the synthesized data were further screened by matching the number of peaks in the GF-PDF spectrum and noise-free HTDMA responses with the preset mode counts. For multimodal cases, synthesized GF-PDFs or noise-free HTDMA responses with closely overlapped modes appearing as a single peak were discarded to eliminate ambiguities of one-to-many mapping in the training dataset.

The measurement uncertainty can be represented by random Gaussian noise (e.g., due to flow or voltage fluctuations) and Poisson noise (i.e., due to counting processes) superimposed onto the theoretical noise-free HTDMA response (Zhang et al., 2022). Depending on the size distribution of ambient aerosols, the size-resolved measurement of aerosol hygroscopic growth often demonstrates different counting statistics (Fig. 2 in Zhang et al., 2021). Besides, the level of Gaussian noise is not accurately known. Here, for each synthesized noise-free HTDMA response, we replicated it with total particle counts ranging from 100 to 1500 and corrupted with Gaussian noise of various magnitudes (i.e., 3 %–8 % relative to total response), to mimic the strong heterogeneity of signal-to-noise ratios in real-world measurements. Since GF-PDF is an intensive quantity and is independent of concentration, the synthetic noise-added instrument responses are then normalized to a unified range of 0 to 1.

The resulting synthetic dataset comprises approximately 3.8 million training samples and 27 000 independent test samples, covering a broad range of Gaussian noise (3 %–8 %) and counting statistics (100–1500). The ratio of unimodal, bimodal, and trimodal samples is approximately 1 : 4 : 5, to reflect the prevalence of complex mixing state in ambient aerosols. This large, physically constrained dataset is then utilized to train the Deep-GF-PRM model for learning the end-to-end mapping between the HTDMA responses and the GF-PDF modal parameters.

2.4 Optimization and assessment of Deep-GF-PRM

To quantify the performance of Deep-GF-PRM in reproducing the ground-truth GF-PDF and corresponding modal parameters, three different metrics are used: mean squared error (MSE), Wasserstein distance, and first-moment of GF-PDF. Together, these metrics evaluate the numerical accuracy, geometric fidelity, and physical consistency of the model predictions. The MSE is the most fundamental measure of the regression accuracy and is defined as

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=1}^{N}(c_{j,\mathrm{pred}} - c_{j,\mathrm{true}})^2 \tag{8}$$

where $c_{j,\mathrm{pred}}$ and $c_{j,\mathrm{true}}$ denote the reconstructed GF-PDF from model-predicted modal parameters (hereafter predicted GF-PDF) and the true GF-PDF, respectively. N is the number of points corresponding to fixed growth factors. The one-dimension Wasserstein distance, also known as Earth Mover's Distance, is utilized to evaluate the overall geometric similarity between the predicted and true GF-PDFs (Saleh et al., 2025), and is defined as:

$$W_1 = \int_0^{+\infty} \left|C_{\mathrm{pred}}(g) - C_{\mathrm{true}}(g)\right| \mathrm{d}g \tag{9}$$

where $C_{\mathrm{pred}}(g)$ and $C_{\mathrm{true}}(g)$ are the cumulative distribution functions of the predicted and true GF-PDFs. Unlike the MSE which emphasizes point-to-point agreement, Wasserstein distance quantifies the minimal cost required to transform one distribution into the other, and thus provides an effective measure of the geometric discrepancy between the predicted and true GF-PDFs. A smaller value of W_1 indicates a higher degree of similarity between the predicted and actual GF-PDFs. Beyond purely statistical metrics, we also calculated the first moment of the GF-PDF, which represents the mean hygroscopic growth factor (hereafter $\mathrm{GF_{mean}}$) of the aerosol population and provides a physically interpretable quantity for evaluating the model performance.

The Deep-GF-PRM model is trained and optimized through a two-stage Bayesian optimization. In the first stage, an initial search is performed to optimize the model architecture, including the number of hidden layers (i.e., residual blocks), number of neurons per layer, the choice of activation function, and the optimizer type. Candidate architectures are trained for 20 epochs using 3-fold cross-validation, and the model parameters corresponding to the lowest validation loss are selected. In the second stage, based on the selected architecture, finer optimization of the learning rate, batch size, and weight decay coefficients is conducted. The Adam optimizer (Adaptive Moment Estimation), an adaptive learning rate algorithm that combines momentum with root-mean-square propagation, was adopted with an initial learning rate of 1×10^{-4} and a batch size of 512. All experiments were conducted on a workstation equipped with an AMD Ryzen 5600 CPU and an NVIDIA GeForce RTX 4060 Ti (16 GB) GPU.

To compare the performance of Deep-GF-PRM against the conventional approach, the state-of-art nonparametric inversion, i.e., Twomey's method, are applied to invert the GF-PDFs (Zhang et al., 2022). The inverted GF-PDFs are then fitted with least-squares minimizations to retrieve the modal parameters. According to Zhang et al. (2022), the HFIMS/HTDMA kernel function is the key to the nonparametric inversion, which maps the GF-PDF to the instrument response through a matrix with a dimension of $I \times J$. Here, the number of GF bins (J) and the number of D_{p}^{*} bins (I) are set to 20 and 30, respectively, to achieve the optimal solution of the inverted GF-PDF. To evaluate whether the proposed model significantly improve prediction accuracy, three widely used machine learning models were also trained for comparison: Linear Regression (LR), K-Nearest Neighbor (KNN), and Random Forest (RF). Statistical comparisons between the model and these benchmark methods are presented in Sect. 3.2.

2.5 Real-world aerosol hygroscopic growth measurements

To evaluate the application of Deep-GF-PRM to real-world measurements, we apply the trained model to two independent field datasets collected by HFIMS and HTDMA under distinct atmospheric conditions. The first dataset was

obtained from the HFIMS measurements conducted during the Tracking Aerosol Convection Interactions ExpeRiment (TRACER) campaign in Houston, Texas, from 8 July to 5 September 2022. Owing to the high temporal resolution of HFIMS (Zhang et al., 2021), size-resolved hygroscopic growth measurements were conducted at six diameters (35, 50, 75, 110, 165, and 265 nm) under three RH levels (75 %, 85 %, and 90 %). A full measurement cycle, consisting of 18 independent measurements, took approximately 24 min. The GF-PDF was inverted from the HFIMS measurements using nonparametric inversion methods with Twomey's iterative regularization, following the procedure described in Zhang et al. (2022).

The second dataset was obtained from the HTDMA measurements conducted during the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition (Rabe et al., 2022). Size-resolved aerosol hygroscopic growths were measured at 85 % RH for five dry particle diameters (50, 100, 150, 200, and 250 nm), and the measurement at each dry size took approximately 16 min. However, the ARM data repository provides the measured humidified particle size distributions ($\mathrm{d}N\,/\,\mathrm{dlog}D_\mathrm{p}$) rather than the raw instrument response R required for nonparametric inversion (Eq. 4). We therefore derived the GF-PDF using a simplified approach: (1) the humidified particle size distribution $\mathrm{d}N\,/\,\mathrm{dlog}D_\mathrm{p}$ was converted to particle number concentration $\mathrm{d}N$ in each size bin; (2) the humidified diameter D_p was mapped to growth factor $\mathrm{GF} = D_\mathrm{p}/D_\mathrm{dry}$, and $\mathrm{d}N$ was assigned to the GF domain to obtain $\mathrm{d}N\,/\,\mathrm{dGF}$; (3) $\mathrm{d}N\,/\,\mathrm{dGF}$ was normalized by the total particle concentration to yield the number probability density function of GF, i.e., $(\mathrm{d}N\,/\,\mathrm{dGF})\,/\,N$. This approach neglects the finite width of the DMA transfer functions and is therefore an approximation of the GF-PDF inversion. However, the 10 : 1 sheath-to-aerosol flow ratio ensures that the first DMA transfer function is sufficiently narrow that this approximation is adequate. We refer to it as the "measured GF-PDF" to distinguish it from the inverted GF-PDF from the HFIMS measurement.

For the GF-PDF derived from both the HFIMS and HTDMA measurements, we use up to three Gaussian distributions to fit each GF-PDF spectrum and obtain the modal parameters corresponding to the lowest fitting residuals. For Deep-GF-PRM, the HFIMS/HTDMA responses were normalized and provided as input to the model, which directly output the predicted modal parameters without explicit GF-PDF inversion or curve fitting. For both the TRACER and MOSAiC datasets, the Deep-GF-PRM predictions were compared against the results obtained using the conventional approach, i.e., nonparametric inversion followed by a multimodal Gaussian fitting. The agreement between these two approaches was evaluated by comparing both the reconstructed GF-PDF spectrum and the first moment of GF-PDF.

3 Results and discussions

3.1 Model validation using synthetic data

The performance of Deep-GF-PRM was first evaluated using an independent testing dataset containing 27 000 samples, generated following the procedure detailed in Sect. 2.3. As a benchmark, the conventional approach was applied to the same dataset, to invert GF-PDFs from the synthetic HTDMA responses and fit the corresponding modal parameters. The MSE, Wasserstein distance, and the first moment of the GF-PDF were used to assess the accuracy of the model predictions, as shown in Fig. 3. The model predictions were compared with those inverted using Twomey's method, as well as those reconstructed from the fitted modal parameters.

As shown in Fig. 3a, the MSE between model predicted GF-PDFs and the ground-truth is generally low, with a mean value of 0.089 ± 0.418, lower than that of the inverted GF-PDFs using Twomey's method (0.149 ± 0.413). In contrast, the GF-PDFs reconstructed from fitted modal parameters yields relatively larger errors, with an MSE of 0.219 ± 0.721, indicating that additional uncertainties are introduced during the parametric fitting of the inverted GF-PDF. It is worth noting that the initial guesses were optimized by trial and error until the fitting residuals were minimized; and thus, the fitted modal parameters represent the best possible solution for parameterizing the true GF-PDFs using the conventional approach. As shown in Fig. 3b, the predicted GF-PDFs also have a lower Wasserstein distance than the inverted GF-PDFs (0.082 ± 0.065 vs. 0.103 ± 0.069), demonstrating that the model accurately captures the overall shape and modal structure of the true GF-PDFs. These results confirm that Deep-GF-PRM successfully learns the implicit mapping between the instrument response and the GF-PDF modal parameters, effectively emulating the conventional nonparametric inversion and mode-fitting procedures with a data-driven framework.

The first moment of the GF-PDF, i.e., $\mathrm{GF_{mean}}$, was further examined to evaluate whether the model preserves the physical consistency. As shown in Fig. 3d–f, the $\mathrm{GF_{mean}}$ derived from the inverted GF-PDFs (inverted $\mathrm{GF_{mean}}$) and the $\mathrm{GF_{mean}}$ from the predicted GF-PDFs (predicted $\mathrm{GF_{mean}}$) exhibit almost identical agreement with the true $\mathrm{GF_{mean}}$, with an R^2 value of ≥ 0.99 and a slope of ~ 1.0. In contrast, the $\mathrm{GF_{mean}}$ derived from the GF-PDFs reconstructed using fitted parameters (fitted $\mathrm{GF_{mean}}$) demonstrates a weaker alignment with the true value, with an R^2 of 0.917 and a slope of 0.981, consistent with the largest MSE shown in Fig. 3a. The deviations of the predicted and inverted $\mathrm{GF_{mean}}$ relative to the true values are negligible, while the fitted $\mathrm{GF_{mean}}$ is systematically underestimated, with an average bias of approximately -1.64 %. This again indicates that the parametric fitting step introduces non-negligible errors when parameterizing the inverted GF-PDFs, whereas the model yields superior accuracy in retrieving the GF-PDF modal parameters.

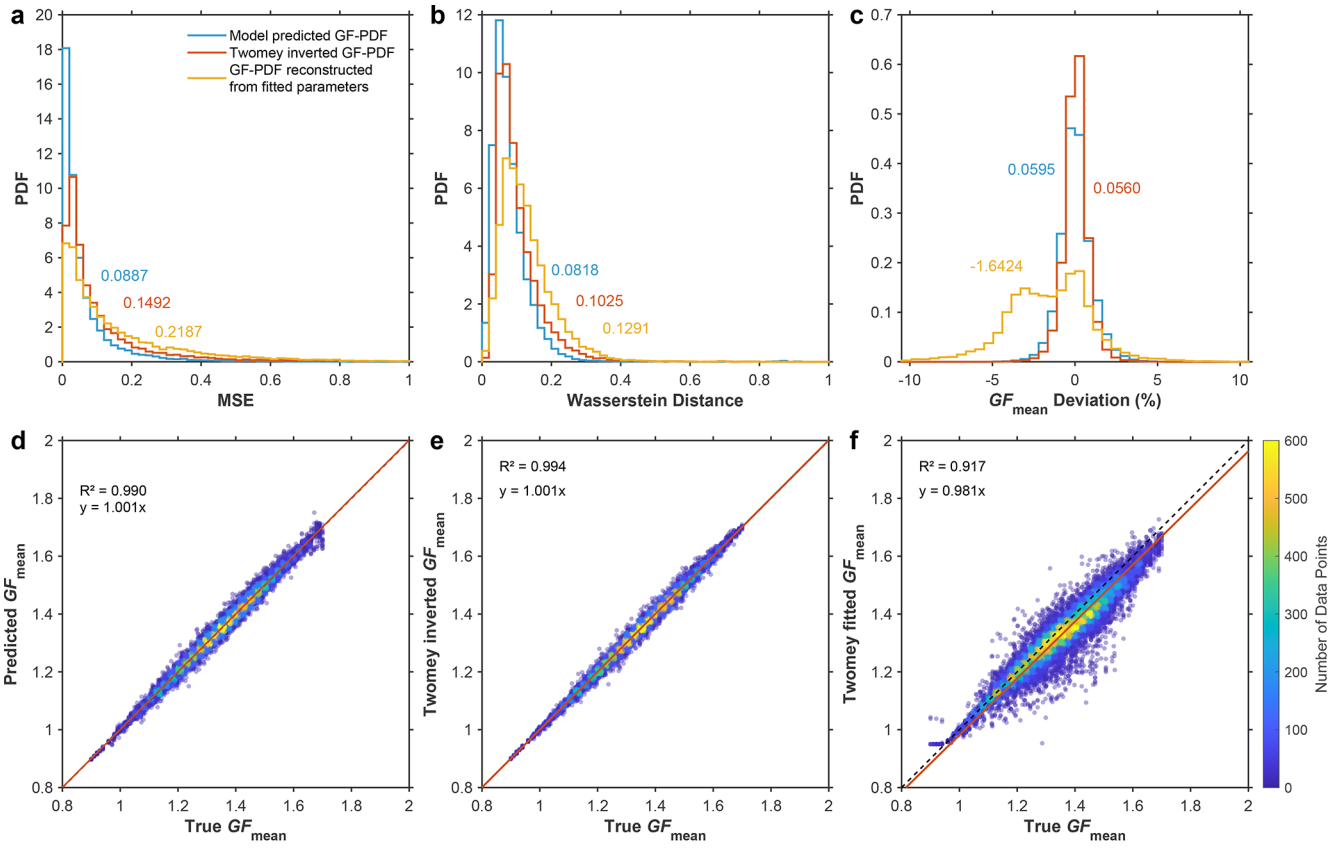


Figure 3. Model performance on the testing dataset. **(a)** Probability density functions (PDF) of mean squared error (MSE) between true GF-PDFs and predicted GF-PDFs (blue), inverted GF-PDFs (red), and GF-PDFs reconstructed using fitted modal parameters (yellow). **(b–c)** same as **(a)** but for Wasserstein distance **(b)** and deviations between first moment of GF-PDFs and true GF_{mean} **(c)**. **(d–f)** Density scatter plots comparing the true GF_{mean} with those predicted by Deep-GF-PRM **(d)**, derived from the inverted GF-PDFs **(e)**, and derived from the GF-PDFs reconstructed using fitted modal parameters **(f)**.

We also evaluated the model's ability to recover the individual modal parameters, including the number of modes and f, G, σ of each mode. As shown in Fig. S1 in the Supplement, Deep-GF-PRM identifies the correct number of modes with an overall accuracy of 96.6 % on the test dataset, compared to 61.2 % for the conventional Twomey + Gaussian fitting approach. The per-parameter MSE is consistently lower for Deep-GF-PRM across all mode counts (Table S1 in the Supplement). A parameter-level comparison is detailed in the Supplement.

3.2 Comparison versus other machine learning methods

The performance of Deep-GF-PRM is also evaluated through comparison with other widely used machine learning models. Here, Linear Regression (LR), K-Nearest Neighbor (KNN), and Random Forest (RF) are trained with the same training dataset, and the model predictions on the testing dataset are compared with that of Deep-GF-PRM. Table 1 summarizes the MSE, Wasserstein distance, and R^2 of GF_{mean} resulted from different models. Since all models were trained to predict the same set of modal parameters, the MSEs of individual modal parameters, i.e., f, G, σ, were directly compared as a measure of numerical model accuracy.

The LR model exhibits the poorest performance among all tested models, indicating that the mapping between instrument responses and GF-PDF parameters is highly nonlinear and therefore cannot be represented by a linear function. Owing to their nonlinear architectures, both RF and KNN achieve improved accuracy. The performance of RF model can be optimized by increasing the number and depth of decision trees; however, a more complex ensemble also leads to increased training and prediction time. Among the three reference models, KNN yields the lowest MSE and Wasserstein distance for the GF-PDF, representing the best achievable performance for a purely data-driven model. Nevertheless, KNN requires traversing the entire training samples during prediction, making it computationally inefficient (e.g., approximately 1 h for the test dataset in this study).

In contrast, Deep-GF-PRM achieves the best overall performance across all metrics. In addition to the lowest MSE

and Wasserstein distance for the predicted GF-PDF, the model also attains the lowest numerical MSEs for the modal parameters, i.e., 0.029, 0.055, and 0.011 for f, G, and σ, respectively. These results demonstrate that Deep-GF-PRM outperforms other machine learning models in predicting the modal parameters and reconstructing the GF-PDF spectrum.

3.3 Application to real-world data

The model performance was further evaluated using real-world HFIMS and HTDMA measurements. The HFIMS measurements were conducted at multiple diameters and RH levels during the TRACER campaign. Since the true GF-PDFs are not available from real-world measurements, the GF-PDFs retrieved from Deep-GF-PRM predictions are compared against those inverted using Twomey's nonparametric method. Figure 4a, b presents the temporal evolution of the predicted and inverted GF-PDFs of 35 nm particles at 85 % RH. The model predicted GF-PDFs show excellent agreement with those inverted directly from the HFIMS measurements, capturing both the multimodal structure and temporal variability. The difference between the inverted and predicted GF-PDFs, i.e., $|\text{GF-PDF}_{\text{inv}} - \text{GF-PDF}_{\text{pred}}|$, is minimal (Fig. 4c), confirming the strong consistency between the inverted and predicted GF-PDFs. In addition, the first moments of the inverted and predicted GF-PDFs, i.e., predicted and inverted GF_{mean} are highly consistent with each other (Fig. 4d).

Figure 5 compares the predicted and inverted GF_{mean} of multiple particle diameters (35, 50, 75, 110, 165, and 265 nm) from a total of 39 532 data points. The predicted and inverted GF_{mean} show a strong linear correlation (R^2 of 0.994), indicating excellent agreement across different sizes. The mean deviation between predicted and inverted GF_{mean} is ~ 0.24 %, indicating that the model predictions align well with the nonparametric inversion in quantifying the aerosol hygroscopic growth factor. It is worth noting that these results are based on a model trained on synthetic HTDMA datasets, suggesting that Deep-GF-PRM can be directly applied to real-world HFIMS measurements without retraining, and the model accurately reproduces the GF-PDFs and generalizes robustly across different particle sizes.

Figure 6 shows the temporal evolution of the model predicted GF-PDFs and those derived from the MOSAiC HTDMA measurements (measured GF-PDF), which exhibit good agreement in both the mode structures and the temporal patterns. The absolute differences between the measured and predicted GF-PDFs (Fig. 6c) remain small, and the discrepancies are mainly due to slight deviations in the magnitude of the predicted GF-PDFs, which can be partially attributed to neglecting the finite widths of the two DMA transfer functions in deriving measured GF-PDF. As shown in Fig. 6d, the first moments of the measured and predicted GF-PDFs, i.e., measured and predicted GF_{mean}, are highly consistent throughout the entire 1-month period, indicating that the model successfully tracks the change of aerosol hygroscopic GF-PDF without systematic bias.

Figure 7 compares the measured and predicted GF_{mean} of multiple particle diameters (50, 100, 150, 200, and 250 nm) from a total of 11 914 data points. The results show a strong linear correlation (R^2 of 0.965), demonstrating consistent agreement across different sizes. However, the predicted GF_{mean} are slightly smaller than those derived from measured GF-PDF (~ 3.36 %). We suspect that it is partially due to the stronger uncertainties of HTDMA measurement affected by noise, resulting in the measured GF-PDF falling outside the model's training dataset. It is worth noting that HTDMA data generally exhibit stronger measurement uncertainties than HFIMS data, as reflected in the noisier GF-PDFs derived from HTDMA measurements. This is due to the distinct principles of the two instruments: HTDMA measures one particle size at a time, with relatively lower counting statistics, whereas HFIMS simultaneously measures particles across the entire size range, significantly improving counting statistics and signal-to-noise ratio. Overall, the Deep-GF-PRM predictions agree reasonably well with the GF-PDF derived from the HTDMA measurements. These results demonstrate that the model can be successfully applied to the real-world HTDMA and HFIMS measurements, for retrieving the GF-PDFs and their modal parameters.

4 Summary

This study proposes and demonstrates the concept that nonparametric inversion and parameterization of aerosol hygroscopic growth factor probability density functions (GF-PDFs) can be emulated by a physics-constrained neural network model. Deep-GF-PRM directly maps the instrument responses to the GF-PDF modal parameters, without explicit GF-PDF inversion or parametric curve fitting. Trained with synthetic HTDMA responses generated under diverse modal combinations and noise conditions, the model demonstrates excellent capability in reproducing the true GF-PDFs, achieving accuracy comparable to that of regularized nonparametric inversion. The modal parameters predicted by Deep-GF-PRM exhibit smaller GF-PDF reconstruction errors than the fitted modal parameters using conventional fitting approaches. When applied to the field HTDMA and HFIMS measurements, the model predictions show strong agreement with GF-PDFs retrieved using conventional inversion methods, therefore indicating that the model can be successfully applied to the real-world aerosol hygroscopic growth measurements for retrieving the GF-PDFs and their modal parameters.

The development and application of Deep-GF-PRM have broader implications. First, Deep-GF-PRM achieves fully automated GF-PDF retrieval from raw instrument responses, without requiring any parametric assumptions or manual intervention for mode fitting. This allows for consistent

Table 1. Comparison of performance among different machine learning models.

Model	MSE of GF-PDF	Wasserstein Distance	R^2 of GF_{mean}	MSE of modal parameters		
				f	G	σ
Linear Regression	2.666	0.3262	−0.4772	0.1436	0.4333	0.0778
Random Forest (trees = 100, depth = 12)	0.8349	0.2040	0.8125	0.0383	0.0953	0.0176
Random Forest (trees = 200, depth = 16)	0.5707	0.1745	0.8879	0.0254	0.0621	0.0116
KNN ($k = 100$)	0.3567	0.1353	0.9124	0.0334	0.0891	0.0160
KNN ($k = 200$)	0.4221	0.1547	0.9018	0.0383	0.1012	0.0182
Deep-GF-PRM	0.0887	0.0818	0.9904	0.0288	0.0545	0.0110

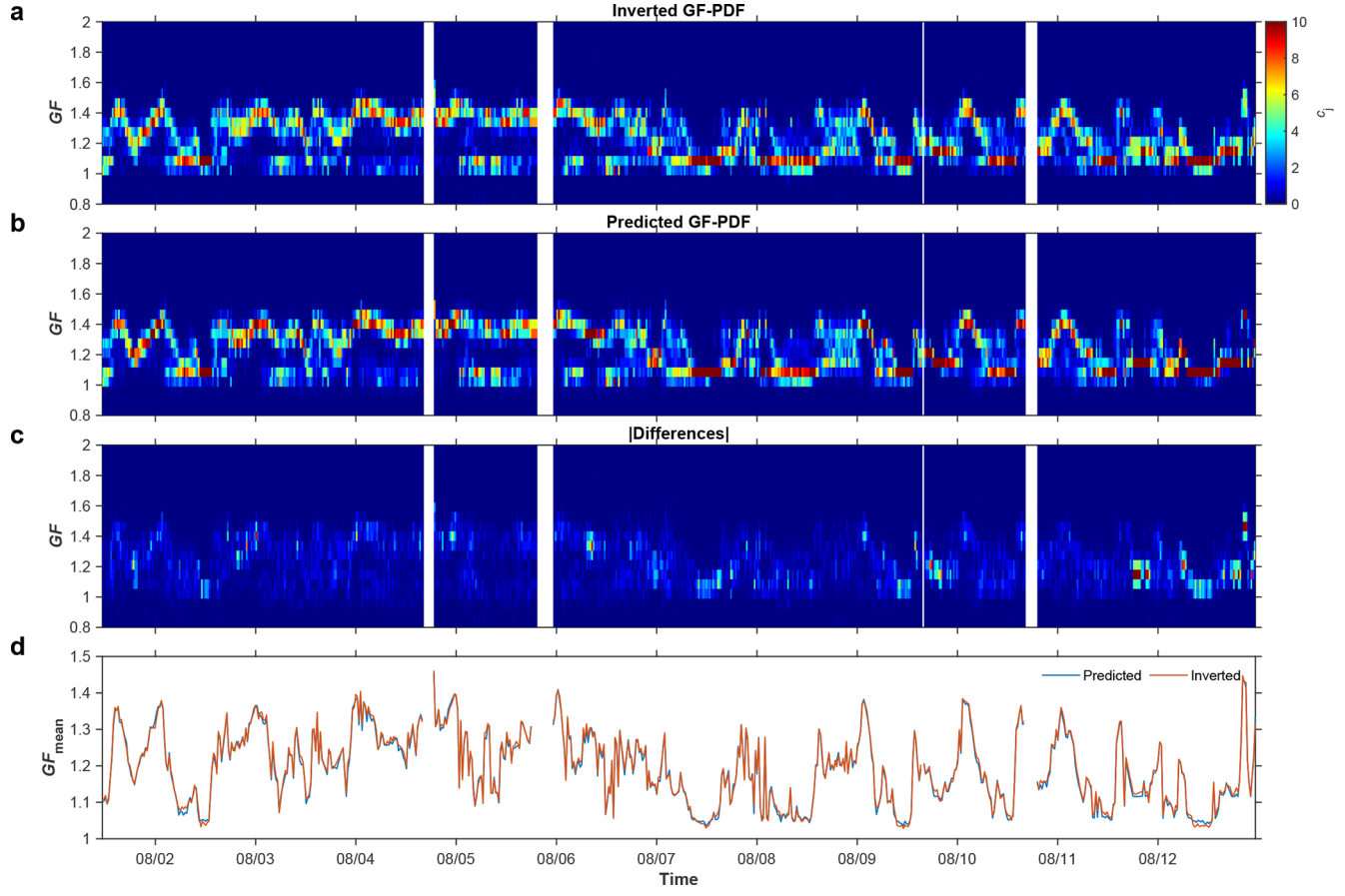


Figure 4. Comparison of GF-PDFs derived from HFIMS measurements and corresponding Deep-GF-PRM predictions. **(a–b)** Temporal evolution of GF-PDFs of 35 nm particles under 85 % RH inverted using Twomey's method **(a)** and predicted by Deep-GF-PRM **(b)**. **(c)** The difference between inverted and predicted GF-PDFs. **(d)** Time series of mean growth factor, i.e., first moment of GF-PDF, derived from inverted and predicted GF-PDFs. c_j denotes the probability density at the jth growth factor.

and objective processing of diverse observational datasets across different instruments. Second, owing to its fully parallelized architecture, Deep-GF-PRM demonstrates a superior computing efficiency compared to the serial inversion-fitting workflow. For example, while the conventional approach typically requires ∼ 1 h to perform GF-PDF inversion (mainly inversion matrix derivation) and mode fitting on the 27 000-sample testing dataset, the model predictions can be implemented instantaneously (∼ 0.1 s on GPU and ∼ 10 s on CPU). This efficiency makes Deep-GF-PRM particularly suitable for processing large-scale and long-term datasets. With the combination of high predictive accuracy and computing efficiency, Deep-GF-PRM provides a promising solution for processing global HTDMA observation compilations, which will facilitate the integration and assimilation of measured aerosol hygroscopicity (Pöhlker et al., 2023) and

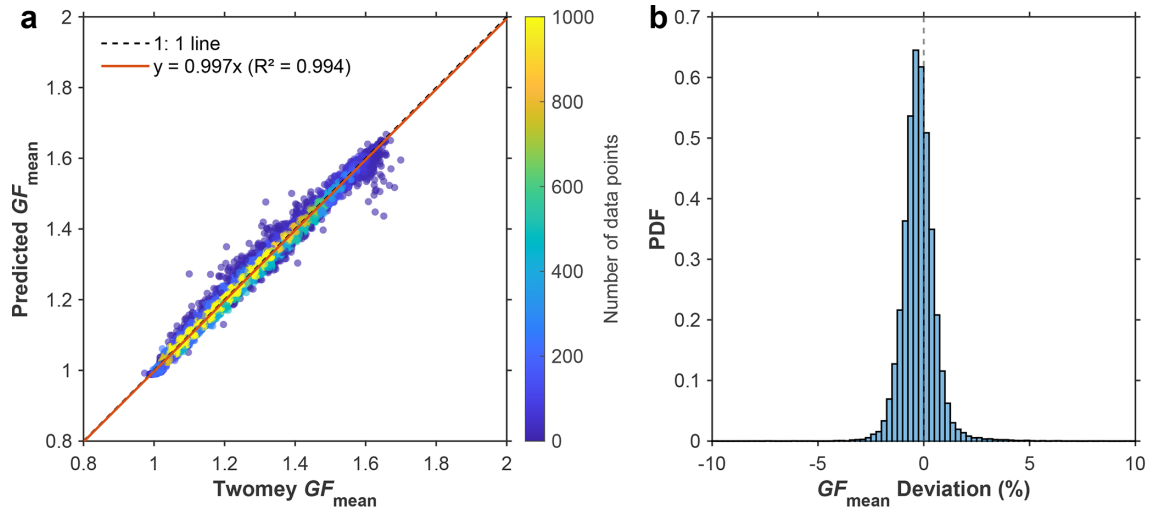


Figure 5. Comparison of predicted and inverted GF_{mean} for all HFIMS measurements. **(a)** Density scatter plots comparing GF_{mean} derived from the inverted and predicted GF-PDFs. **(b)** Probability density functions (PDF) of deviations between GF_{mean} derived from the inverted and predicted GF-PDFs.

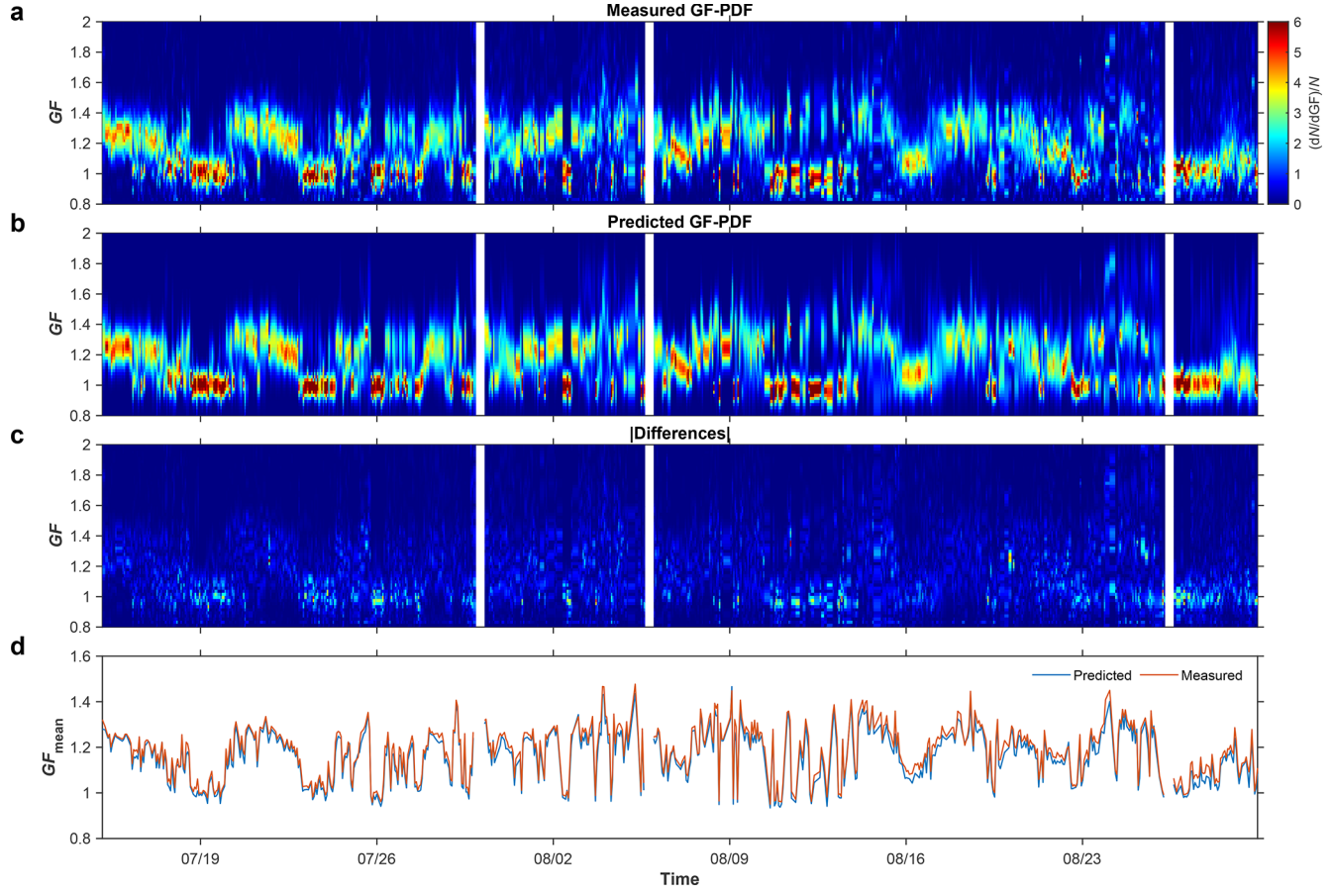


Figure 6. Comparison of GF-PDFs derived from HTDMA measurements and corresponding Deep-GF-PRM predictions. **(a–b)** Temporal evolution of GF-PDFs of 50 nm particles under 85 % RH derived from the HTDMA measured humidified size distribution **(a)** and predicted by Deep-GF-PRM **(b)**. **(c)** The difference between measured and predicted GF-PDFs. **(d)** Time series of mean growth factor, i.e., first moment of GF-PDF, derived from measured and predicted GF-PDFs.

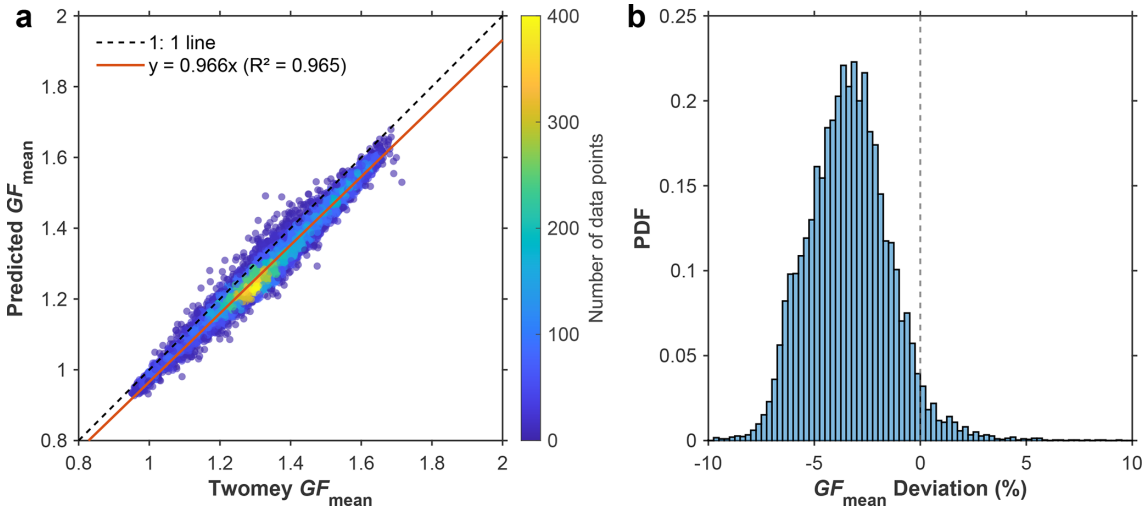


Figure 7. Comparison of predicted and measured GF_{mean} for all HTDMA measurements. **(a)** Density scatter plots comparing GF_{mean} derived from the measured and predicted GF-PDFs. **(b)** Probability density functions (PDF) of deviations between GF_{mean} derived from the measured and predicted GF-PDFs.

mixing state (Su et al., 2010) into regional or global models. Overall, this study highlights a paradigm shift in the retrieval of aerosol hygroscopic GF-PDF from conventional physics-based inversion to data-driven deep learning, providing an efficient, robust, and generalizable framework for large-scale data analysis and model constraint.

Code and data availability. The Python and MATLAB code for the automation script and the model workflow, together with the testing datasets, is publicly available on Zenodo (https://doi.org/10.5281/zenodo.21934344, Lu and Zhang, 2026). The HTDMA data from the MOSAiC expedition are available from the ARM Data Center (https://doi.org/10.5439/1095581, Uin et al., 2012).

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/acp-26-16653-2026-supplement.

Author contributions. Conceptualization: J.S. Zhang and H. Lu; Investigation: H. Lu and J.S. Zhang; Methodology: H. Lu, J.S. Zhang, and J.Y. Zhang; Writing – original draft: J.S. Zhang and H. Lu; Writing – review and editing: H. Lu, J.S. Zhang, J.Y. Zhang, S.Q. Zhou, Y. Wang, J. Liu, and H. Gui; Funding acquisition: J.S. Zhang and H. Gui. Project Administration: H. Gui and J. Liu.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. We acknowledge the Atmospheric Radiation Measurement (ARM) User Facility, a US Department of Energy (DOE) Office of Science User Facility, for providing the HTDMA data used in this study. We thank the editor and the three anonymous referees for their constructive comments, which helped improve the manuscript.

Financial support. This research has been supported by the National Key Research and Development Program of China (grant no. 2023YFC3705100) and the National Natural Science Foundation of China (grant nos. 42675160 and 42375124).

Review statement. This paper was edited by Allison C. Aiken and reviewed by three anonymous referees.

References

Alonso-Blanco, E., Gómez-Moreno, F. J., and Artíñano, B.: Size-resolved hygroscopicity of ambient submicron particles in a suburban atmosphere, Atmos. Environ., 213, 349–358, https://doi.org/10.1016/j.atmosenv.2019.05.065, 2019.

Cai, S., Wang, Z., Wang, S., Perdikaris, P., and Karniadakis, G. E.: Physics-Informed Neural Networks for Heat Transfer Problems, J. Heat Transfer, 143, 060801, https://doi.org/10.1115/1.4050542, 2021.

Cubison, M. J., Coe, H., and Gysel, M.: A modified hygroscopic tandem DMA and a data retrieval method based on optimal estimation, J. Aerosol Sci., 36, 846–865, https://doi.org/10.1016/j.jaerosci.2004.11.009, 2005.

Fu, D., Shi, H., Gueymard, C. A., Yang, D., Zheng, Y., Che, H., Fan, X., Han, X., Gao, L., Bian, J., Duan, M., and Xia, X.: A Deep-Learning and Transfer-Learning Hybrid Aerosol Retrieval Algorithm for FY4-AGRI: Development and Verification over Asia, Engineering, 38, 164–174, https://doi.org/10.1016/j.eng.2023.09.023, 2024.

George, I. J. and Abbatt, J. P. D.: Heterogeneous oxidation of atmospheric aerosol particles by gas-phase radicals, Nat. Chem., 2, 713–722, https://doi.org/10.1038/nchem.806, 2010.

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

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27–30 June 2016, 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Holmgren, H., Sellegri, K., Hervo, M., Rose, C., Freney, E., Villani, P., and Laj, P.: Hygroscopic properties and mixing state of aerosol measured at the high-altitude site Puy de Dôme (1465 m a.s.l.), France, Atmos. Chem. Phys., 14, 9537–9554, https://doi.org/10.5194/acp-14-9537-2014, 2014.

Huang, Y. and Seinfeld, J. H.: A Neural Network-Assisted Euler Integrator for Stiff Kinetics in Atmospheric Chemistry, Environ. Sci. Technol., 56, 4676–4685, https://doi.org/10.1021/acs.est.1c07648, 2022.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, Nature Reviews Physics, 3, 422–440, https://doi.org/10.1038/s42254-021-00314-5, 2021.

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, Atmos. Environ., 12, 99–104, https://doi.org/10.1016/0004-6981(78)90192-0, 1978.

Lu, H. and Zhang, J.: Code and data for "Deep-GF-PRM – A deep learning framework for parameterizing aerosol hygroscopic growth factor probability density function", Zenodo [data set] and [code], https://doi.org/10.5281/zenodo.21934344, 2026.

Molina Catricheo, C. A., Lambert, F., Salomon, J., and van 't Wout, E.: Modeling global surface dust deposition using physics-informed neural networks, Commun. Earth Environ., 5, 778, https://doi.org/10.1038/s43247-024-01942-2, 2024.

Petters, M. D.: Revisiting matrix-based inversion of scanning mobility particle sizer (SMPS) and humidified tandem differential mobility analyzer (HTDMA) data, Atmos. Meas. Tech., 14, 7909–7928, https://doi.org/10.5194/amt-14-7909-2021, 2021.

Pilinis, C., Pandis, S. N., and Seinfeld, J. H.: Sensitivity of direct climate forcing by atmospheric aerosols to aerosol size and composition, J. Geophys. Res.-Atmos., 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 Sci. Tech., 51, 1212–1222, https://doi.org/10.1080/02786826.2017.1338664, 2017.

Pöhlker, M. L., Pöhlker, C., Quaas, J., Mülmenstädt, J., Pozzer, A., Andreae, M. O., Artaxo, P., Block, K., Coe, H., Ervens, B., Gallimore, P., Gaston, C. J., Gunthe, S. S., Henning, S., Herrmann, H., Krüger, O. O., McFiggans, G., Poulain, L., Raj, S. S., Reyes-Villegas, E., Royer, H. M., Walter, D., Wang, Y., and Pöschl, U.: Global organic and inorganic aerosol hygroscopicity and its effect on radiative forcing, Nat. Commun., 14, 6139, https://doi.org/10.1038/s41467-023-41695-8, 2023.

Rabe, B., Heuzé, C., Regnery, J., Aksenov, Y., Allerholt, J., Athanase, M., Bai, Y., Basque, C., Bauch, D., Baumann, T. M., Chen, D., Cole, S. T., Craw, L., Davies, A., Damm, E., Dethloff, K., Divine, D. V., Doglioni, F., Ebert, F., Fang, Y.-C., Fer, I., Fong, A. A., Gradinger, R., Granskog, M. A., Graupner, R., Haas, C., He, H., He, Y., Hoppmann, M., Janout, M., Kadko, D., Kanzow, T., Karam, S., Kawaguchi, Y., Koenig, Z., Kong, B., Krishfield, R. A., Krumpen, T., Kuhlmey, D., Kuznetsov, I., Lan, M., Laukert, G., Lei, R., Li, T., Torres-Valdés, S., Lin, L., Lin, L., Liu, H., Liu, N., Loose, B., Ma, X., McKay, R., Mallet, M., Mallett, R. D. C., Maslowski, W., Mertens, C., Mohrholz, V., Muilwijk, M., Nicolaus, M., O'Brien, J. K., Perovich, D., Ren, J., Rex, M., Ribeiro, N., Rinke, A., Schaffer, J., Schuffenhauer, I., Schulz, K., Shupe, M. D., Shaw, W., Sokolov, V., Sommerfeld, A., Spreen, G., Stanton, T., Stephens, M., Su, J., Sukhikh, N., Sundfjord, A., Thomisch, K., Tippenhauer, S., Toole, J. M., Vredenborg, M., Walter, M., Wang, H., Wang, L., Wang, Y., Wendisch, M., Zhao, J., Zhou, M., and Zhu, J.: Overview of the MOSAiC expedition: Physical oceanography, Elementa, 10, 00062, https://doi.org/10.1525/elementa.2021.00062, 2022.

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

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Saleh, E., Ghaffari, S., Curtis, J. H., Patel, L., Bosler, P., Riemer, N., and West, M.: Generative Modeling of Aerosol State Representations, arXiv [preprint], https://doi.org/10.48550/arXiv.2510.10361, 2025.

Sharma, H., Shrivastava, M., and Singh, B.: Physics informed deep neural network embedded in a chemical transport model for the Amazon rainforest, npj Climate and Atmospheric Science, 6, 28, https://doi.org/10.1038/s41612-023-00353-y, 2023.

Stolzenburg, M. R. and McMurry, P. H.: 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 Sci. Tech., 42, 421–432, https://doi.org/10.1080/02786820802157823, 2008.

Su, H., Rose, D., Cheng, Y. F., Gunthe, S. S., Massling, A., Stock, M., Wiedensohler, A., Andreae, M. O., and Pöschl, U.: Hygroscopicity distribution concept for measurement data analysis and modeling of aerosol particle mixing state with regard to hygroscopic growth and CCN activation, Atmos. Chem. Phys., 10, 7489–7503, https://doi.org/10.5194/acp-10-7489-2010, 2010.

Swietlicki, E., Hansson, H. C., Hämeri, K., Svenningsson, B., Massling, A., McFiggans, G., McMurry, P. H., Petäjä, T., Tunved,

P., Gysel, M., Topping, D., Weingartner, E., Baltensperger, U., Rissler, J., Wiedensohler, A., and Kulmala, M.: Hygroscopic properties of submicrometer atmospheric aerosol particles measured with H-TDMA instruments in various environments – a review, Tellus B, 60, 432–469, https://doi.org/10.1111/j.1600-0889.2008.00350.x, 2008.

Tang, M., Chan, C. K., Li, Y. J., Su, H., Ma, Q., Wu, Z., Zhang, G., Wang, Z., Ge, M., Hu, M., He, H., and Wang, X.: A review of experimental techniques for aerosol hygroscopicity studies, Atmos. Chem. Phys., 19, 12631–12686, https://doi.org/10.5194/acp-19-12631-2019, 2019.

Uin, J., Cromwell, E., Hayes, C., and Salwen, C.: Humidified Tandem Differential Mobility Analyzer (AOSHTDMA), 2019-10-11 to 2020-09-30, ARM Mobile Facility (MOS), MOSAIC (Drifting Obs – Study of Arctic Climate); AMF2 (M1), Atmospheric Radiation Measurement (ARM) User Facility [data set], https://doi.org/10.5439/1095581, 2012.

Wang, J., Flagan, R. C., Seinfeld, J. H., Jonsson, H. H., Collins, D. R., Russell, P. B., Schmid, B., Redemann, J., Livingston, J. M., Gao, S., Hegg, D. A., Welton, E. J., and Bates, D.: Clear-column radiative closure during ACE-Asia: Comparison of multiwavelength extinction derived from particle size and composition with results from Sun photometry, J. Geophys. Res.-Atmos., 107, AAC 7-1–AAC 7-22, https://doi.org/10.1029/2002JD002465, 2002.

Wang, Y., Zheng, G., Spielman, S., Pinterich, T., Hering, S., and Wang, J.: Retrieval of High Time Resolution Growth Factor Probability Density Function from a Humidity-controlled Fast Integrated Mobility Spectrometer, Aerosol Sci. Tech., 53, 1092–1106, https://doi.org/10.1080/02786826.2019.1628917, 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, https://doi.org/10.5194/amt-14-5625-2021, 2021.

Zhang, J., Wang, Y., Spielman, S., Hering, S., and Wang, J.: Regularized inversion of aerosol hygroscopic growth factor probability density function: application to humidity-controlled fast integrated mobility spectrometer measurements, Atmos. Meas. Tech., 15, 2579–2590, https://doi.org/10.5194/amt-15-2579-2022, 2022.