Manuscript No.: egusphere-2026-848

Title: Technical Note: Deep-GF-PRM - A physics-informed deep learning framework for parameterizing aerosol hygroscopic growth factor probability density function

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 study by Lu et al. developed the Deep-GF-PRM framework for parameterizing GF-PDF in aerosol hygroscopicity studies. Trained by large synthetic instrument responses, this model is shown to quickly retrieve modal parameters that performs no worse than the conventional fitting approaches. Especially, this method advantages in its quick responses and ensured convergence, which is of urgent need for large-scale online data analysis. Overall, this study is well conducted in innovative ways, clearly presented, and presented a promising and powerful new tool. I've only a few minor concerns.

Responses: We sincerely thank the anonymous referee for the valuable and constructive comments on our manuscript. We have thoroughly revised the manuscript accordingly; our point-to-point responses are provided below.

In addition, we refined the Deep-GF-PRM model through two coordinated changes. First, the loss function was restructured: the original SmoothL1 loss on the 9-parameter output was demoted to an auxiliary term (weight = 0.3), while the physics loss—computed directly on the reconstructed GF-PDF curves—was elevated to 1.0 as the dominant training signal. Second, an L1 sparsity penalty (weight = 0.01) was added on the predicted mode fraction (f_i), which encourages the model to prefer the sparsest modal decomposition that reproduces a given GF-PDF shape (i.e., penalizing unnecessary modal complexity). Together, these refinements effectively mitigate the one-to-many mapping ambiguity, whereby a single broad mode and two narrow, closely spaced modes can reproduce nearly identical GF-PDF curves. As a result, we retrained the model with expanded σ range for unimodal cases (0.015–0.15, previously capped at 0.05) to cover the full range observed in field data. With the updated model, the overall GF-PDF MSE on the synthetic test set decreased substantially (from 0.187 to 0.09), and the agreement in GF_{mean} values improved correspondingly for both synthetic and field datasets. These changes have been integrated into the revised manuscript.

Detailed Comments:

1. In this study, the GFmean is set as the major target of comparison. How about the other parameters (e.g., mode numbers, and detailed f, G, σ for each mode), especially when more than one GF-PDF modes are present due to the external mixing, etc.?

<u>Responses</u>: We thank the review for the insightful comment. In addition to GF_{mean}, we have systematically evaluated how well Deep-GF-PRM recovers the modal parameters, including the number of modes and f, G, σ of each mode, and compared its performance against the conventional Twomey inversion with multimodal Gaussian fitting. Note that this analysis is based on the synthetic testing dataset, as the ground-truth modal parameters are unknown for field dataset.

As shown in Fig. R1, Deep-GF-PRM identifies the true number of modes with 99.0%, 95.9%, and 96.6% accuracy for 1-, 2-, and 3-mode cases, respectively. In contrast, the conventional method (i.e., Twomey inversion + multimodal fitting) achieves only 20.8%, 46.8%, and 87.6% for the same cases, and its accuracy is lowest for the simplest single-mode case. This reflects a systematic tendency of the parametric fitting to overestimate the number of modes: 53.7% of the true 1-mode samples are misclassified as 2-mode and 25.5% as 3-mode, and 50.4% of the true 2-mode samples are misclassified as 3-mode. Deep-GF-PRM, by contrast, shows no systematic bias toward over- or undercounting modes.

Table S1 compares the mean squared error (MSE) of the retrieved f, G, and σ between the two approaches, categorized by the number of modes. For single-mode samples, where only one well-resolved peak needs to be fitted, the conventional approach recovers the modal parameters as accurately as, or slightly better than, Deep-GF-PRM (MSE of f: 0.0249 vs. 0.0252; G: 0.0002 vs. 0.0010; σ: 0.0003 vs. 0.0057). Note that the MSE is calculated only for the true modes (each true mode is matched to the predicted/fitted mode with the closest G). The spurious modes produced by the conventional fitting method are excluded here. Therefore, the low MSE for 1-mode samples does not reflect the over-prediction of modes (Fig. S1). As the number of modes increases, however, the conventional MSE rises sharply, especially for G (2-mode: 0.0142 vs. 0.0008; 3-mode: 0.0906 vs. 0.0013) and σ (2-mode: 0.0023 vs. 0.0003; 3-mode: 0.0059 vs. 0.0005), whereas Deep-GF-PRM remains stable and is consistently more accurate for f, G, and σ across all multi-mode cases. The inaccurate modal parameters retrieved by parametric fitting for multi-mode samples therefore lead to degraded GF-PDF reconstruction and systematic underestimation of GF_{mean}, as shown in Fig. 3 in the main text.

These results indicate that, although the multimodal Gaussian fitting is a widely used approach for resolving distinct hygroscopic modes, i.e., from the GF-PDF spectrum, it is considerably less robust when applied to noise-corrupted instrument responses in practice. As a result, empirical quality control (QC), which can be subjective sometimes, is often required to yield a physically meaningful modal decomposition, such as merging closely spaced modes (e.g., $\Delta G < 0.1$) or discarding modes with a small *f*, and etc. In contrast, Deep-GF-PRM performs this decomposition end-to-end without such manual intervention, which explains its higher accuracy and consistency, particularly for the more challenging multi-mode cases relevant to externally mixed aerosol populations.

We have added the analysis to the supplement information of the revised manuscript, and added a brief in the main text, and it now reads:

"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, 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). A parameter-level comparison is detailed in the Supplement."

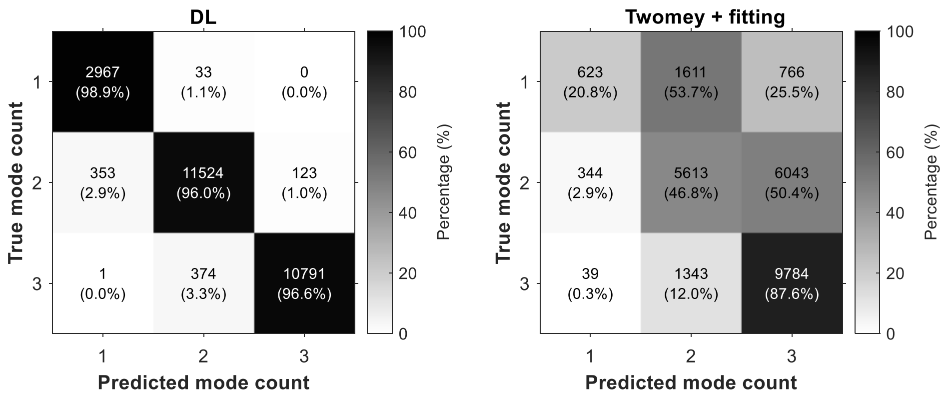

Fig. R1. The confusion matrix between the true and predicted number of GF-PDF modes on the synthetic test dataset for (**a**) the Deep-GF-PRM (DL) model and (**b**) the conventional Twomey inversion with multimodal Gaussian fitting. Rows correspond to the true mode count and columns to the predicted mode count; each cell reports the row-normalized percentage (color scale) and the corresponding number of test samples (n).

Table R1: MSE of retrieved modal parameters between Deep-GF-PRM (DL) and conventional inversion + fitting methods (conv.).

MSE of modal parameters	**1-mode**		**2-mode**		**3-mode**	
	DL	conv.	DL	conv.	DL	conv.
$\boldsymbol{f}$	0.0252	0.0249	0.0202	0.0449	0.0104	0.0453
$\boldsymbol{G}$	0.0010	0.0002	0.0008	0.0142	0.0013	0.0906
$\boldsymbol{\sigma}$	0.0057	0.0003	0.0003	0.0023	0.0005	0.0059

2. Can the authors briefly state / predict the applicability of this methods？ For example, when the deviations tend to be larger (except when total counts are low and therefore low SNR)? What kind of parameters / instrumental design changes, or under which scenarios the applicability of this method needs special attention and / or to be re-examined first?

Responses:

First, we would like to clarify that the refined Deep-GF-PRM model greatly improves the prediction of the GF-PDF and its modal parameters. The average GF-PDF MSE has a significant decrease from 0.187 to 0.089. However, as the reviewer pointed out, the model predictions can be sometimes largely deviated, corresponding to the tail of the MSE distribution. To identify the dominant factors that limit the model prediction accuracy, we systematically analyzed test samples with GF-PDF MSE exceeding the 75th percentile (P_{75} = 0.073), and identified two independent factors affecting the accuracy of model predictions, i.e., the narrowest mode width (min σ) and the counting statistics (n_{tot}, a proxy for SNR).

As demonstrated in Fig. R2, the fraction of samples with large MSE rises sharply as the narrowest mode width drops below ~0.05, exceeding 60% for 1-mode cases with $\sigma \leq 0.03$, compared to <10% for $\sigma \geq 0.10$ — a factor of >6. For SNR, the dependence is weaker than that of σ. The high-MSE fraction declines from ~35–40% at $n_{tot} \approx 150$ to ~15–20% at $n_{tot} \approx 1400$, i.e., a factor of ~2. These results indicate that narrow peak width ($\sigma \leq 0.05$) is the dominant factor for large GF-PDF MSEs, with low SNR a secondary but compounding contributor.

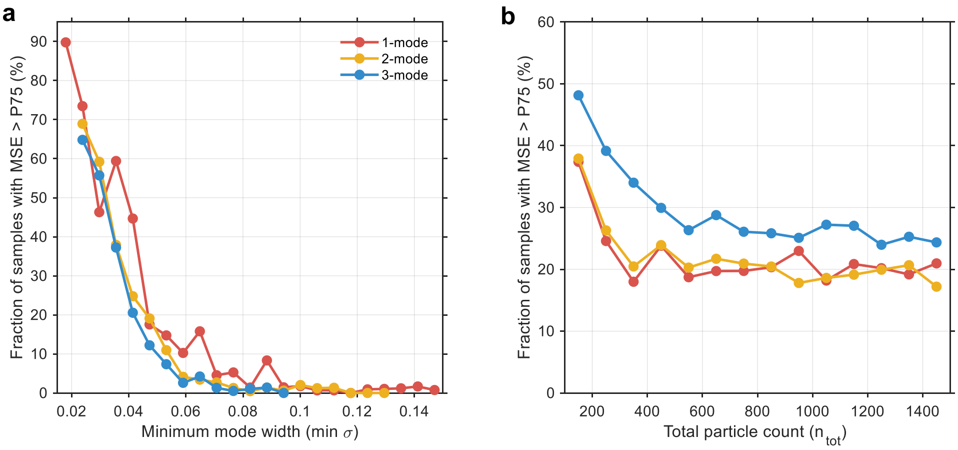

Fig. R2. Fraction of test samples with GF-PDF MSE exceeding the 75th percentile (P_{75}= 0.073), shown as a function of (**a**) the narrowest mode width (min σ) and (**b**) the total particle count (n_{tot}, a proxy for SNR). Curves are stratified by true mode count (1-, 2-, and 3-mode).

This mechanism is illustrated by two representative cases in Fig. R3. For case A with low total count (n_{tot} = 122), Deep-GF-PRM fits the noisy response well (R^2 = 0.96) yet underestimates σ by ~24% (0.033 → 0.025), giving a GF-PDF MSE of 0.85. Twomey performs comparably poorly (MSE = 0.93) but produces a fake, narrow peak near GF = 1, likely due to the artifact of parametric mode fittings. For case B with the narrowest mode width (σ = 0.015), Deep-GF-PRM broadens the width by ~2-fold (σ = 0.032) and thus results in a very large MSE of 6.52. Note that in both cases G is recovered accurately (with ~1% bias), and the error is confined almost entirely to σ.

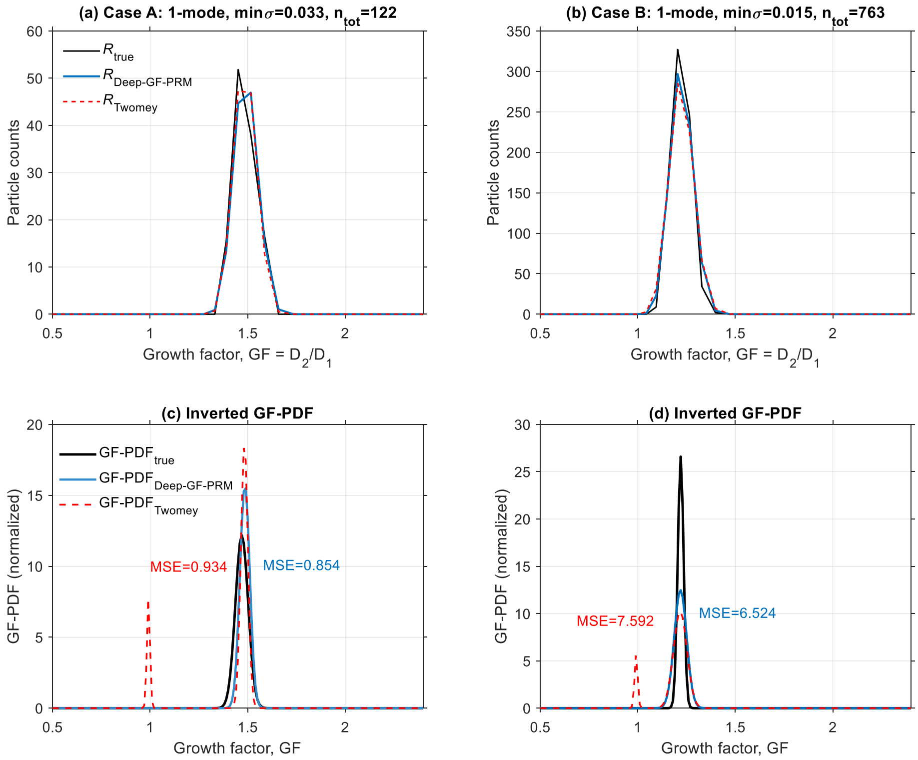

Fig. R3. Two controlled representative cases with large deviations. The instrument response (R) and inverted GF-PDF of two representative cases with large deviations. (**a, c**) case with low counting statistics (n_tot = 122, σ=0.033). (**b, d**) case with extremely narrow mode (n_{tot} = 763, σ = 0.015). Upper panels show the instrument response (*R*) (true vs. reconstructed from Deep-GF-PRM and Twomey GF-PDFs); lower panels show inverted GF-PDFs vs. true GF-PDF.

The underlying mechanism is that a narrow mode occupies only a few channels of the instrument response spectrum, meaning the signal that constrains σ is inherently sparse. Therefore, it is the most fragile quantity in the inversion, susceptible to both the measurement noise and the finite GF resolution. Therefore, a detailed re-examination is necessary if the measured instrument response contains very narrow peaks and/or poor counting statistics. From the measurement perspective, increasing the total sampling time or particle flow rate, and refining the GF resolution (e.g., denser voltage scan steps) in the narrow-mode regime would most directly mitigate these errors.

In terms of practical application, as we stated in Sect. 4 in the revised manuscript, Deep-GF-PRM is applicable to the rapid, parallelized processing of aerosol hygroscopic measurements from instruments such as HTDMA and HFIMS, directly predicting parameterized hygroscopic modal parameters from instrument responses. It has three main advantages: 1) Deep-GF-PRM achieves a good accuracy of recovering GF-PDF spectrum from instrument responses that matches conventional non-parametric inversion; 2) the direct mapping avoids the parametric fitting conventionally required to resolve GF-PDF modes, thereby eliminating associated fitting errors; 3) the parallelized architecture makes Deep-GF-PRM particularly efficient for unified, high-throughput processing of large-scale field datasets.