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 #2:

General Comments:

This paper showcases a useful deep learning method to retrieve growth factor probability density functions (GF-PDF) from humidified tandem differential mobility analyzer (HTDMA) measurements. The authors train a deep learning model to map the instrument response directly to the modal parameters of the GF-PDF, using synthetic responses generated through the known instrument kernel. The deep learning method is compared against a conventional approach, nonparametric Twomey inversion followed by multimodal Gaussian fitting, on the synthetic test dataset, supplemted by two field campaign measurements. This retrieval approach is particularly useful as it maps the instrument response directly to modal parameters, circumventing the parametric fitting process of the conventional method, which is time-consuming and subject to convergence issues.

The manuscript is overall well written and easy to follow. The deep learning method proposed might be a good alternative to the conventional one considering its accuracy and efficiency. But some clarifications on concept and methodology should be addressed before I can recommend publication in ACP.

Responses: We sincerely thank the reviewer for the constructive and insightful comments that have greatly guided us to improve the manuscript. We have carefully addressed every comment and revised the manuscript accordingly. Point-by-point responses are detailed as follows, and all modifications are integrated into the revised manuscript.

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 weights (f_{cond}), 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 0.089), and the agreement in GFmean values improved correspondingly for both synthetic and field datasets. These changes have been integrated into the revised manuscript.

Major Comments:

1. The authors claim the deep learning framework is physics-informed as it is based on a physics informed loss defined as the discrepancy between the true GF-PDFs and the GF-PDFs reconstructed from the predicted modal parameters. More accurately, this term serves as an extra constraint on PDF shape rather than as physical information such as a conservation law, a governing equation, or any physical processes. Because the true GF-PDF in this loss is reconstructed from the true modal parameters via Equation 5, its target adds no information beyond the ground-truth parameters the network already learns. The real physics within the framework is the synthetic data generated through the instrument kernel, which alone is not sufficient to justify calling this a physics-informed deep learning framework. I therefore suggest the authors soften the "physics-informed" terminology when naming the method.

Responses: We sincerely thank the reviewer for pointing this out. We fully agree that the GF-PDF reconstruction loss in our framework serves as a shape constraint on the output probability density, rather than encoding a governing equation, a conservation law, or a physical process in the sense of standard physics-informed neural networks (PINNs). Our original use of "physics-informed" was intended to highlight two aspects: (1) all training data are generated through the instrument kernel function derived from first-principles aerosol physics (DMA transfer functions, charging efficiency, flow rates, and etc.), and (2) the loss function enforces consistency between the predicted GF-PDF shape and the true GF-

PDF. We acknowledge that the terminology is imprecise and may mislead readers who expect a standard PINN formulation.

Following the reviewer's suggestion, we have removed the terminology "physics-informed" throughout the revised manuscript.

2. Some clarifications are needed regarding the datasets. How were the training and testing sets split? Since both come from the same synthetic generator, is the test set independent and free of near duplicates of the training samples? The manuscript also reports only test set metrics, and showing the training vs. testing error would help confirm that overfitting is not an issue. Additionally, the framework is built for one, two, and three modes: how well do these synthetic GF-PDFs represent real-world GF-PDFs, and have the authors compared the synthetic PDF shapes against the field measurements?

Responses: We thank the reviewer for these detailed and constructive questions. First, we would like to clarify that the training and test sets are generated independently. Although both share the same parameter space, i.e., GF-PDF parameters randomly sampled from $f \in [0.05, 1.0]$, $G \in [0.9, 1.8]$, $\sigma \in [0.015, 0.15]$, the parameter combinations are re-randomized independently. Even if two samples happened to share identical GF-PDF parameters, their instrument responses (R) would differ substantially because each sample receives: (1) a unique random total particle count $n_{tot} \in [100, 1500]$, (2) independent Poisson noise (via poissrnd.m), (3) independent multiplicative Gaussian noise (via randn.m). Two samples with the same GF-PDF shape but different n_{tot} and noise realizations produce meaningfully different input vectors. In addition, the test dataset spans 6 distinct sizes corresponding to different instrument kernel functions (i.e., $\mathbf{M}$), while the training uses only one. We have examined the training and testing datasets, and found that ~11% of them share the same parameter label (exclusively 1-mode GF-PDFs), however the corresponding R have no similarities. We also note that we intentionally included multiple variants for each parameter combination in both the training and testing datasets, i.e., for every set of modal parameters, six discrete particle count levels (n_{tot}, spaced from 100 to 1500) are generated, and each is then perturbed by independent Poisson and Gaussian noise, and therefore the corresponding instrument responses R are different. From the learning perspective, this also serves as an implicit regularization, i.e., the model is forced to learn a mapping from noisy responses to parameters that is robust to noise perturbations, rather than memorizing the noise pattern of a particular realization.

For training the Deep-GF-PRM model, we used a 3-fold cross-validation procedure to monitor overfitting. The training data are split into training and validation folds. An early-stopping strategy (patience = 5 epochs) terminates training when the validation loss ceases to improve, and the model with the lowest validation loss across all folds is retained as the final model. As shown in Fig. R4, the training vs. validation loss curves closely aligned throughout training with no systematic divergence. The final test loss, evaluated on the fully independent testing dataset, is consistent with the cross-validation loss, and the model delivers good accuracy on the field data. Together, these results confirm the absence of overfitting.

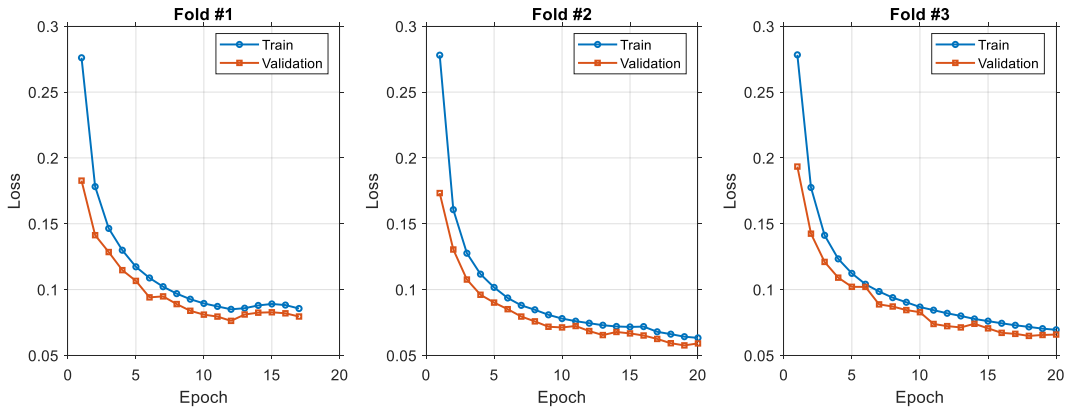

Fig. R4. Training and validation loss curves for the Deep-GF-PRM model trained on HTDMA synthetic data. A 3-fold cross-validation procedure was employed with early stopping (patience = 5 epochs). Loss curves for training (blue) and validation (red) are shown for each fold.

To assess how well the synthetic GF-PDFs represent real-world GF-PDFs, we examined the distribution of GF-PDF modal parameters of the TRACER HFIMS dataset and compared against the parameter ranges used to generate the synthetic training data, as shown in Fig. R5. Across 39,532 TRACER samples, the modal parameters fall within the training ranges for 96.9% (1-mode), 97.9% (2-mode), and 95.8% (3-mode) of all mode instances. The small fraction outside the training range (<5%) consists predominantly of modes with σ slightly below 0.015 (extremely narrow near-hydrophobic peaks) or G marginally exceeding the upper bound, and these outliers are rare and isolated.

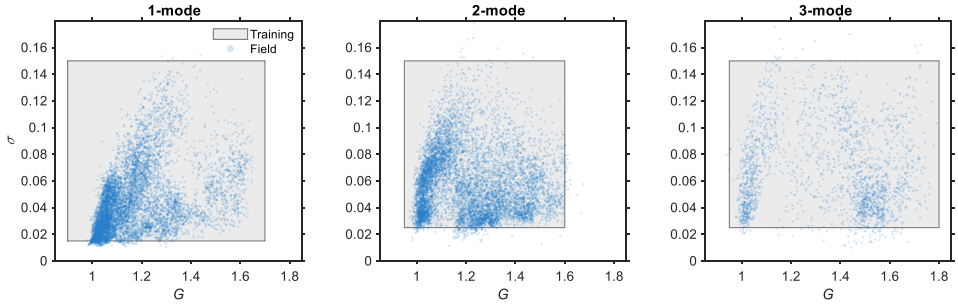

Fig. R5. Scatter plot of modal parameters (G vs. σ) of the TRACER field dataset, compared against the synthetic training parameter ranges (gray rectangles). Panels correspond to 1-, 2-, and 3-mode predictions.

3. Is it possible to quantify the prediction uncertainty regarding model structure and parameters? If practical, the authors may perturb the model layers or other structures and random seedings to quantify the uncertainties of the model. It would provide the helpful reference for the users who use the model to retrieve the GF-PDF modal parameters.

Responses: We thank the reviewer for this constructive suggestion. We fully agree that quantifying prediction uncertainty would be valuable for users of the model, and we have carefully considered this point.

First, in the current framework, the dominant source of prediction uncertainty for a given measurement is the counting statistics of the instrument, which directly controls the signal-to-noise ratio. We have quantified this effect explicitly in Sect. 3.2 (Fig. 4): by training models on data synthesized at n_{tot} = 100, 1500, and 100–1500 and evaluating them on a test set spanning the full range, we demonstrated how the GF-PDF retrieval error scales with SNR. This provides users with a practical uncertainty reference based on the counting statistics of their own measurements.

Second, we acknowledge that a systematic ensemble analysis over random seeds and network architectures, as suggested by the reviewer, would provide a more complete characterization. In our current training workflow, model selection is already regularized by (i) 3-fold cross-validation during hyperparameter optimization and (ii) early stopping, which limits the sensitivity of the final model to initialization. Furthermore, we observed that the test performance of the final model is stable and reproduces the cross-validation results closely, suggesting that the uncertainty is relatively small. However, we acknowledge that this is an indirect assessment. A formal ensemble-based uncertainty quantification, e.g., training multiple models with perturbed seeds and architectures and reporting the

spread of predicted modal parameters, would be a valuable extension, and we have noted it as important future work.

Minor Comments:

L42: Revise "the particle" to "the humidified particle"

Responses: Revised.

L59: Please add citations for "Tikhonov regularization and Twomey's iterative regularization".

Responses: Added.

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, Atmospheric Measurement Techniques, 15, 2579-2590, 10.5194/amt-15-2579-2022, 2022.

L93: The theoretical functions are useful but look overwhelmed. Please add a short statement of what each transfer function (omega1, omega2) physically represents, which would aid readability.

Responses: We have revised it in the revised manuscript, and it now reads:

"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.".

L189: Please elaborate on how the two loss terms are balanced (the value/selection of lambda) and its influence on the prediction results.

Responses: We have revised it in the revised manuscript, and it now reads:

"The weight of each loss (λ_{param}= 0.3, λ_{phys}= 1.0, and λ_{sp}= 0.01), were determined through Bayesian hyperparameter optimization with cross-validation. ".

L385: "accurate" to "accurately"

Responses: Corrected.

L453: “reginal” to “regional”

Responses: Corrected.