We thank Reviewer 1 for the careful and constructive assessment of our manuscript. In response to the reviewer's comments, we have explained the differences between particle ensembles used for hybrid BNN training, strengthened the discussion about the morphology-resolving ADDA physics and the advantage of the hybrid BNN, and added further analysis and discussion of model uncertainty and sensitivity. Our response to each comment is provided below in blue text, while reviewer comments are reproduced in black.

Reviewer 1 Assessment:

This study introduces a hybrid Bayesian neural network (BNN) model trained on discrete dipole approximation (DDA) simulations for black carbon (BC)-containing particles. The description of the proposed model is generally clear. A total of eight input features are considered: five particle-related inputs (number of BC particles, coating amount, fractal dimension, volume-equivalent size parameter, and imaginary refractive index of the coating) and three additional inputs derived from homogeneous-sphere Mie calculations (extinction coefficient, single-scattering albedo, and asymmetry parameter). The three model outputs consist of optical properties (extinction coefficient, single-scattering albedo, and asymmetry parameter) from DDA simulations. In addition, the treatment and decomposition of aleatoric and epistemic uncertainty are generally well explained.

Overall, the manuscript is well written, well organized, and easy to follow, except for some details in the machine-learning-related sections where I am not an expert. The scope is well addressed, and the general methodology and evaluation approaches appear reasonable. However, several aspects would benefit from additional clarification. Therefore, I recommend publication after the authors address the comments listed below.

Specific Comments:

1. Different assumptions between Ensemble I and II (Table 1 and related discussion): Why do differences exist with respect to wavelength, coating refractive index, AAE of the coating, and other parameters?

   The differences between Ensembles I and II reflect differences in the underlying physical model that was used to generate the BC core. Ensemble I was generated before a physical BC restructuring model was available, and was used to develop scaling laws for BC absorption as a function of phase shift parameter, coating amount, wavelength, and coating absorptivity (Beeler and Chakrabarty, 2022). Discrete values for wavelength, coating absorptivity, etc. were used to aid in fitting of the scaling laws developed in Beeler and Chakrabarty (2022). These simulations were used in this work to represent edge cases for BC shape, coating amount, and wavelength-dependence of coating refractive index.

   On the other hand, Ensemble II was generated using an observationally-constrained BC compaction model (Beeler et al., 2025), where the fractal dimension of the BC core increases as a function of coating amount. The ADDA simulations of these particles were performed for the purposes of training the BNN model, and therefore we varied the input parameters in a more systematic fashion. The advantage of including both ensembles in the training/validation datasets is that the effects of each input feature on the predicted optical properties can be discerned because the input features have differing correlations with one another between the two ensembles. We have added more information on the distinction between Ensembles I and II in Sections 2.1 and 2.2 of the revised manuscript.

2. Figure 1: What is the distinction between the blue and brown particles? I assume the partially transparent regions indicate coatings, but it was not entirely clear whether the figure simply represents uncoated versus coated particles.

The reviewer is correct that the partially transparent regions indicate coating material, while the blue represents the BC core. We have added a legend to the revised Figure 1 to clarify this distinction.

3. How is the core fractal dimension ($D_{f,c}$) calculated? It would be helpful to provide at least a brief definition or formulation in the Appendix. In addition, the definition or relationship of $D_{f,c}$ with other parameters appears to differ between Ensembles I and II, which was somewhat difficult to follow.

   We now clarify in lines 81-84 and 101-104 of the revised text that the core fractal dimension of each particle in the training ensembles is calculated using the structure factor approach of Sorensen (2001). This calculation is the same for both ensembles, which we now make clear in lines 81-84 and 101-104 of the revised text; what differs between the two ensembles is how the particles are generated. In Ensemble I, core morphology and coating amount were varied independently, without accounting for core restructuring as particles are coated; in Ensemble II, the particle-generation procedure couples the core morphology to V/V_0 and N_pp using Equations 1–3, which were derived from laboratory measurements and detailed models of coating-induced capillary compaction, as described in Beeler et al. (2025). We have revised Figure 1 to show that $D_{f,c}$ and V/V_0 are statistically independent in Ensemble I, denoted by $D_{f,c} \perp\!\!\!\perp V/V_0$. We also show that $D_{f,c}$ and V/V_0 are not statistically independent in Ensemble II, denoted by $D_{f,c} \not\!\perp\!\!\!\perp V/V_0$.

4. L79: Is it reasonable to assume a single fixed value for the BC refractive index? Does the BC refractive index have a known uncertainty range or variability that should be considered?

   The reviewer is right that the refractive index of BC has been found to vary from less refractive ($1.7+0.64i$) to highly refractive ($2.0 + 1.0i$). In Section 7 of the original manuscript, we evaluated the sensitivity of BNN predictions to different assumed values of the BC refractive index (panels B, E, H of Figure 6). We show that prediction errors for Q_{ext} increased when the BC refractive index departed from the value used in the training set ($1.95 + 0.79i$), but the error produced by the hybrid BNN remained comparable to those from core-shell and homogeneous sphere approximations. We also found that the hybrid BNN is more accurate than core-shell and homogeneous sphere approximations for SSA and g across the tested BC refractive indices. This analysis indicates that the homogeneous-sphere optical descriptors allow the hybrid BNN to retain useful predictive skill for BC refractive indices not explicitly represented in the training set, provided that $Q_{ext,\mathrm{hs}}$, $\mathrm{SSA}_{\mathrm{hs}}$, and g_{hs} are recalculated using the desired BC refractive index. Additional details on the sensitivity analysis can be found in Section 7. We have also expanded the discussion of the sensitivity of hybrid BNN predictions to assumed BC refractive index, coating refractive index, and monomer size in the discussion section (lines 474-482).

5. L130: Naming of input properties. For clarity, it may be helpful to explicitly distinguish the two groups of input features. For example, the first five inputs could be described as "microphysical/morphological descriptors," while Q_{ext}, SSA, and g could be referred to as "optical-property descriptors" or "physics-based optical descriptors." The authors do not necessarily need to follow these specific suggestions, but clearer categorization of the input features would improve readability. In particular, I found it somewhat difficult to follow the transition from optical-property-based inputs to the more general term "particle properties."

   We agree that describing the input features as "microphysical descriptors" and "homogeneous sphere optical descriptors" will help make the discussion more clear. We have added these labels to Figure 2 of the revised text and changed references to these variables where appropriate.

6. Sections 5 & 6: These results appear to demonstrate the superiority of the proposed hybrid BNN over the homogeneous-sphere and core-shell assumptions. However, this outcome seems somewhat

expected because the DDA simulations are treated as the reference truth, and the hybrid BNN is directly trained on those DDA-generated optical properties. Therefore, Figures 3–5 may reflect the superiority of morphology-resolving DDA physics itself rather than the unique advantage of the hybrid BNN architecture. I wonder if the authors could clarify more explicitly what aspects of the improvement should be attributed to the Bayesian/surrogate framework versus the underlying DDA-based physical representation.

We agree that the accuracy of hybrid BNN architecture is reflective of the accuracy of the underlying ADDA-generated dataset that was used to train the model, and therefore the superiority of BNN-predicted optical properties is rooted in the morphology-resolving ADDA physics. The advantage of the hybrid BNN model is that it is able to accurately predict the optical properties that would be calculated using ADDA without running the computationally-expensive ADDA code with associated uncertainty. We have explained this nuance in more detail in the revised manuscript on lines 265-273.

7. Figure 5: The SHAP analysis suggests that the homogeneous-sphere optical-property inputs contribute substantially to the hybrid BNN predictions. Since the framework conceptually combines homogeneous-sphere optical information with morphology-aware particle descriptors, it would be very helpful to more directly quantify the relative influence of these two groups of inputs on the final predictions. For example, an analysis comparing the hybrid BNN predictions with the corresponding homogeneous-sphere optical properties could help clarify the extent to which morphology-dependent particle information contributes beyond the baseline optical physics.

   We thank the reviewer for this thoughtful suggestion. Whereas Figure 5 identifies the influence of individual features, it does not directly quantify the relative contributions of the two feature groups. We, therefore, expanded the discussion in Section 6 to compare the relative importance of the combined microphysical descriptors with the homogeneous sphere optical descriptors on the target variables. We now report the aggregated mean absolute SHAP values for (1) the five microphysical descriptors and (2) the three homogeneous-sphere optical descriptors in lines 388-395 of the revised manuscript, which reflect the total importance of each group on the model predictions.

8. Figure 6: In panels B, E, and H, the hybrid BNN appears to exhibit relatively larger errors within portions of the parameter range represented in the training dataset. I may be misunderstanding the interpretation of this figure, but I initially expected the prediction errors to be minimized near the training-domain conditions. Additional clarification on how these perturbed-parameter experiments relate to the original training manifold versus the shaded "training range" would help improve interpretation of the figure.

   We would like to thank the reviewer for bringing our attention to Figure 6. The original version of Figure 6 incorrectly duplicated panels D, E, and F in panels G, H, and I. The revised manuscript has rectified this error and panels G, H, and I now correctly show errors in calculated g. Nevertheless, panel E still shows a local increase in the SSA error at the RI_{BC} value used in the training dataset. The perturbed cases were drawn from the held-out evaluation set, and the model is trained over a multidimensional parameter space. Each curve in Figure 6 is based on ten held-out particles for which one physical parameter was systematically perturbed. The shaded region indicates the RI_{BC} value used during training, but it does not indicate that the complete combination of particle size, morphology, coating amount, coating absorptivity, and wavelength lies near a densely sampled region of the full multidimensional training space. Model performance at a given RI_{BC} value, therefore, also depends on the specific combination of other particle properties represented by the selected particles. The local variation in panel E likely reflects the particular combinations of particle properties represented by this relatively small evaluation ensemble rather than a systematic reduction in performance at the training value. Further, we note that, although the SSA error in Figure 6E is relatively high at the training value of RI_{BC}, its magnitude remains below 0.05 and is smaller than the corresponding errors from the homogeneous-sphere and core-shell approximations.

Thus, although this local variation is worth noting, the BNN maintains relatively small SSA errors across the tested RI_{BC} values.

9. Uncertainty decomposition and Monte Carlo sampling: The uncertainty decomposition is an important component of the proposed hybrid BNN framework, but the manuscript provides relatively limited examples and interpretation of its behavior across different particle conditions. Are there general patterns or tendencies in the aleatoric versus epistemic uncertainty components with respect to particle size, morphology, coating amount, or wavelength? In addition, how sensitive are the uncertainty estimates to the Monte Carlo sampling sizes used during inference? Some discussion regarding convergence behavior and the choice of sampling size would help clarify the robustness of the uncertainty characterization.

   Figure 7 shows epistemic uncertainty as a function of N_{pp} and V/V_0 for wavelengths that are commonly used in optical instruments. We have added a similar plot to the appendix which instead shows aleatoric uncertainty (Figure K1). The aleatoric and epistemic uncertainties follow broadly similar regions of elevated uncertainty, but the magnitude of aleatoric uncertainty is generally smaller than the epistemic uncertainty.

   The reviewer is correct that the number of Monte Carlo samples affects the uncertainty estimates. We have added Figure D1 to the Appendix showing the convergence of aleatoric and epistemic uncertainty (σ_a and σ_e, respectively) for different values of the nested Monte Carlo sample sizes S and L. We quantify convergence by comparing the estimated σ_a and σ_e values at each combination of S and L with the corresponding converged estimates obtained using large sample sizes ($S = L = 1000$). At the default values used in the main text, $S = L = 50$, the deviations from the $S = L = 1000$ reference case are within 4% for all target parameters and both uncertainty components. During inference, the computational cost scales as $O(S \times L)$. Therefore, increasing $S = L$ from 50 to 1000 would increase computational cost by approximately a factor of 400, while changing the uncertainty estimates by less than 4%. The inference-throughput measurements reported in Appendix E further quantify the computational cost of the chosen default. We therefore retain $S = L = 50$ during inference as a cost-accuracy compromise. We have added this reasoning to lines 238-242 of the main text and pointed readers to Figure D1 and Table D1 in the appendix.

10. Section 8: Since epistemic uncertainty is interpreted as uncertainty associated with limited training coverage, it would be interesting to examine how the epistemic and aleatoric uncertainty components change when the size of the training dataset is artificially reduced (e.g., using half of the current training samples). Such an experiment may help demonstrate whether the epistemic uncertainty behaves consistently with the proposed interpretation, and it may be possible without requiring additional DDA simulations.

    We agree that directly testing the proposed relationship between training coverage and epistemic uncertainty is important. We have therefore added Appendix Figure J1, which shows the results of a data-restriction experiment designed to test whether epistemic uncertainty responds to a reduction in local training coverage. We retrained the hybrid BNN after removing approximately 12% of the data, corresponding to the region $70 \leq N_{\mathrm{pp}} \leq 180$ and $8 \leq V/V_0 \leq 25$. Test particles within this masked region were retained for evaluation but were not used during retraining. For these particles, epistemic uncertainty increased by 14.4%, 28.6%, and 11.9% for Q_{ext}, SSA, and g, respectively, relative to the baseline model trained using the full dataset. For test particles outside the masked region, the corresponding increases were 3.9%, 5.0%, and −1.4%. Thus, the increase in epistemic uncertainty was larger inside the masked region than outside for all three target variables, supporting the interpretation that epistemic uncertainty responds to reduced training coverage. We also examined the corresponding changes in aleatoric uncertainty, which increased by 6.9%, 27.9%, and 3.6% inside the masked region for Q_{ext}, SSA, and g, respectively, compared with increases of 0.8%, 8.7%, and −0.4% outside the region. Because the aleatoric and epistemic components are learned jointly through the same predictive model, changes in the training distribution can affect both components. We therefore interpret this experiment primarily as a test of whether the epistemic component responds to reduced training coverage. The localized increase in epistemic uncertainty is evident for all three target variables, although the magnitude of the response is target-dependent. We have added this discussion to Section 8 of the revised manuscript (lines 425-429) and pointed readers to Appendix Figure J1.

11. From the perspective of future radiative transfer and remote sensing applications, it would be interesting to know whether the framework could eventually be extended to predict more detailed angular scattering information beyond the asymmetry parameter, such as single-particle phase functions or phase-matrix-related quantities. Although the current study focuses on single-particle optical properties, such extensions could further increase the applicability of the framework to aerosol optics studies.

    The scattering phase matrix of particles was also saved during DDA calculations for this work. Therefore, future versions of the hybrid BNN could be extended to predict angular scattering, Legendre polynomials, or depolarization. Such extensions are planned for future work. We have noted in the discussion section that such extensions are possible for future model versions (lines 490-491).

References

Beeler, P. and Chakrabarty, R. K.: Constraining the particle-scale diversity of black carbon light absorption using a unified framework, Atmospheric Chemistry and Physics, 22, 14 825–14 836, 2022.

Beeler, P., Corbin, J. C., Sipkens, T. A., and Fierce, L.: A Framework for Quantifying the Size and Fractal Dimension of Compacting Soot Particles, Environmental science & technology, 59, 5994–6003, 2025.

Sorensen, C.: Light scattering by fractal aggregates: a review, Aerosol Science & Technology, 35, 648–687, 2001.