RC1: 'Comment on amt-2022-112', Feng Xu, 06 Jun 2022

The paper by Gao et al. validates a Bayesian uncertainty propagation model against the *real* uncertainty estimate from analyzing synthetic retrievals. The simulation study indicates that the theoretical uncertainties of the retrieved pixel aerosol quantities basically reproduce the real retrieval uncertainties in most cases, though with certain degree of underestimates. This is a major finding and should be useful for aerosol and ocean color remote sensing using PACE polarimeters. Another valuable observation is about the importance of sample size for uncertainty estimate. The authors shows the increased robustness of their uncertainty analysis when the sample size increase from 50 to 1000. Following the validation, retrievals were performed using the real observations from AirHARP field measurements and the Bayesian uncertainty estimate model. In algorithm development, the authors deployed the FastMAPOL approach, which couples NN based RT calculation and the automatic difference method for Jacobian evaluation. These two elements ensure the efficiency of the uncertainty model assessment in the present work.

We appreciate your time and efforts in reviewing this work. The comments and questions are valuable in improving the clarity of this manuscript. The responses are below (in red) with manuscript revised accordingly (also uploaded).

Overall, the paper was well-written. I have the following four comments for the authors to consider to clarify their approach:

Thank you for the positive comments on our work.

1. Do I understand correctly that the retrieval results for statistical analysis (e.g. those in Figs. 5-6) subjected to certain the convergence criterion ? For example, the cost function needs to be less than or equal to the metric unit when the retrieval is flagged to be successful so that the results are further used in the uncertainty analysis. Associated with this question, what is the success rate of the retrieval ?

   The reviewer is correct. We have chosen a maximum cost function value of 1.5 for the analysis of the synthetic data retrievals. However, this corresponds to a very high success rate of 96% for AirHARP and 93% for HARP2. In this way, only outliers are removed, and the main statistical distribution are maintained in the analysis. The cost function histogram is shown in Fig. 3, also copied below:

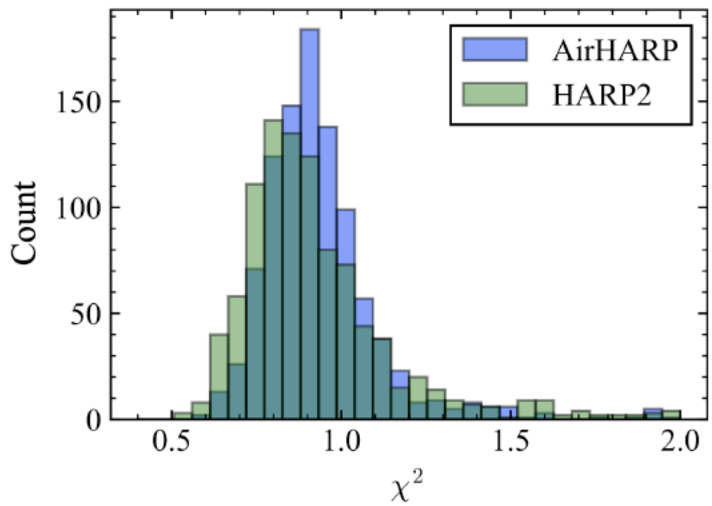


Figure 3. The histogram of the cost function values for the synthetic retrievals.

We added the following in our manuscript:

"

To reduce the impact of outliers, we choose a maximum χ^2value of 1.5 in this study as shown in Fig. 3, which corresponds to a success rate of 96% for AirHARP cases and 93% for HARP2 cases. "

2. Table 1 is commendable as it lists the range of 11 retrieval parameters which further decides the a priori matrix used in retrieval. Could the authors comment on whether there is potential impact of the a priori on the conclusion ? For example, if we relax the upper bound of the imaginary part of refractive index to be larger (e.g. 0.1 or larger for some strongly absorbing aerosols), will the Bayesian model based uncertainty still mimic the real uncertainties, or there might be additional underestimates of uncertainty ?

This is an excellent question and will be a very interesting topic for future study.

When increasing the bound of the retrieval parameters, it is likely to have more challenges to correctly retrieve the corresponding parameter. 1) This may relate to the accuracy of the forward model and the neural network model used to represent the forward model. For example, more sampling point may be required to generate the neural network training data, to accurately represent the forward model, and its Jacobians. 2) There may be also impact of the local minima of the cost function around the new territory of the parameter. Therefore, it is important to quantify the difference between the theoretical uncertainty and the real uncertainty when new parameter range are used. Although, our current study only focuses on weakly absorbing aerosols (with imaginary refractive index <0.03), but the approach proposed in this study can be a useful tool to access such impacts in a future study when more complex aerosols are presented.

We added the following discussion in the manuscript in Sec 2.2:

"In this work, we only consider weakly absorbing aerosols with mi<0.03. It will be a subject of future studies on how the theoretical uncertainties represents the real uncertainties for more complex aerosol models, following the approach discussed in this study.
"

3. Eq.(5): Is the modeling error (e.g. five size components of aerosols, Cox-Munk ocean surface, etc) excluded or included in the piece of VRTM model uncertainty "sigma_{VRTM}" ? I'm curious in this aspect since in real data retrieval (e.g. the demonstrated AirHARP retrieval), one of the major error sources is the modeling errors. To enhance the connection of synthetic retrieval and AirHARP retrievals, it would be great if modeling error is included in the Bayesian model via RT simulation uncertainty "sigma_{VRTM}" and then via Eqs. (5)-(6).

Thank the reviewer for the interesting question which is important for the application to real measurement. We do not consider explicit modeling errors in the total uncertainty model in this work, but we have taken efforts to reduce the impact of modeling uncertainties.

For the use of Cox-Munk ocean surface model, we do observe discrepancy in fitting the sunglint signal, which may relate to wind direction or ocean swell. Therefore, we have removed the sunglint signal within 40 degree with respect to the spectral reflection direction to minimize its impacts. The following discussion is included in the manuscript:

> "Strong sunglint is excluded here by removing view angles within 40^{o} of the specular reflection direction due to the challenges to represent the sunglint signals from ACEPOL field campaign using the isotropic Cox and Munk model (Gao et al 2020, Gao et al 2021a)."

For the use of five size aerosol model, it provides a robust approach to retrieve aerosol size parameters, which have also been demonstrated by Dubovik et al., 2006 and Xu et al., 2016. Moreover, for a more general study, Fu and Hasekamp discussed the representation of aerosol size distribution through various numbers of sub-modes and also found that a similar five-mode approach can provide good retrievals for most aerosol parameters (Fu and Hasekamp, 2018). We have revised our discussed as follows:

> "The aerosol size distribution is assumed as a combination of five lognormally-distributed aerosol sub-modes, each with prescribed mean radii and variances; the five volume densities (*Vi*) are free parameters (Dubovik et al., 2006; Xu et al., 2016). **The five-mode approach is found to provide good retrievals for most aerosol parameters (Fu and Hasekamp, 2018).**

Moreover, the modeling error is very challenging to quantify, since it depends on the model itself used in the RT simulation, and the real measurements selected for retrieval, which can vary pixel by pixel. For example, when there is contamination of the thin cirrus cloud in real data, the aerosol only model will correspond to a large uncertainty (even "wrong" results when the cirrus cloud signal is strong). We have been using a data screening approach to reduce the discrepancy between the model and data used in retrievals, and we also expect the uncertainty quantification approach proposed in this study can be also used to provide a tool in understanding and accessing modeling errors. Here are more discussions:

1) To minimize the impact of the scenarios with insufficient models in this study, we are using the data screening approach developed in Gao et al 2021b, Frontiers. The measurement which cannot be fitted well by the forward model are removed based on the ratio of fitting residual and an uncertainty model. This approach is applied adaptively by conducting retrieval several times. Through this approach, the forward model is more likely to be applied on the measurement proper for its design and therefore reduce the modeling error.
2) The uncertainty quantification approach itself can be also used to analyze the modeling error. We can analyze the retrieval residuals after data screening, and compare its statistics directly with the assumed uncertainty model. If we find there are differences, it is likely due to the impact of modeling error.

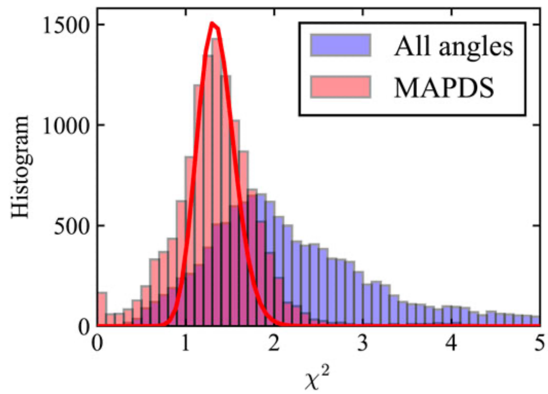
We added the following sentence in Sec 2.3

> **"As discussed in Sec. 1, an adaptive data screening method is used to remove the real measurements which cannot be fitted well by the forward model (Gao et al., 2021b). In this way, the impact of forward model uncertainties can be reduced. We do not consider additional forward model uncertainties in this study.**
> **"**

4. What is convergence metric for AirHARP's real data retrieval ? Is it consistent with the those used in the synthetic retrievals ?

For the study of real data, we consider all the retrieval cases. This is because the data screening approach has been applied on the real data retrievals, which removed the measurement cannot be fitted well by the forward model. The maximum cost function is often less than 3. Therefore, we do not define a maximum cost function to remove the outliers. For the measurement cannot be fitted by the forward model at all, the number of valid angle Nv will be zero, such as in Fig 10, panel d, some pixels in the center of the image are removed completely.

As an example, the cost function histogram for scene2 before and after data screening have been shown in Gao et al 2021, Frontiers paper as copied below. The red histogram is after data screening. Note that the center of the histogram is slightly larger than 1 which may relate to the contribution of modeling error.



We revised the following discussion in the manuscript:

> "The adaptive data screening method of (Gao et al., 2021b) was applied on all these scenes to mask out viewing angles contaminated by cirrus clouds, ocean surface floating objects, or other irregularities that could not be represented adequately by the current forward model. **The resulted cost function histogram is much better described by the χ^2 distribution using the assumed uncertainty model (Gao et al., 2021b).** "