Dear reviewer,

Thank you for your time and efforts in reviewing this manuscript. We really appreciate your constructive comments, which are very helpful to improve the clarity of the manuscript. We have addressed every point in the revised manuscript, which are detailed below in red.

**RC2**: 'Comment on egusphere-2023-1843', Anonymous Referee #2, 18 Sep 2023 reply

**General comments:**

This study trained a new NN model through measurement uncertainty-aware training and Training data augmentation. The new NN model was used to generate pseudo HARP2 observations and retrieve both aerosol and ocean properties. The methods and results are reasonable, and the manuscript is well written. I have only a few confusions that needs to be clarified.

Thank you for the positive feedback.

**Specific comments:**

1. For Equation (3), some terms are not explained. I think the terms with f superscript is NN simulation and the terms without f superscript are pseudo-observations. Please confirm it or correct me.

   You are right that f indicates the forward model, which in this case is represented by the NN. The term without f represent satellite observations, which in this case are the synthetic data, or the pseudo-observations as you referred. We have revised the manuscript as follows:

   > "…where $\rho_t$ and $P_t$ are measurements and $\rho_t^f$ and $P_t^f$ are the corresponding quantities computed from the forward model..."

2. For Equations (5) and (6), every term should be explained. Is the uncertainty of DoLP a constant (0.005)? If so, what is difference between Equation (6) and conventional MSE cost function. It seems Equation (6) is just a conventional MSE cost function multiplied by a constant. If uncertainties of DoLP are not a constant in Equation (6), how they are quantified?

   Thank you for the questions. Equations (5) and (6) are defined similarly to the retrieval cost function in Eq. (3) with the same definition between the reflectance and DoLP uncertainties. The use of reflectance uncertainty of 3% in Eq. (5) is in a percentage form, which can efficiently incorporate sunglint without being impacted by its large magnitude.

   You are right that in the DoLP part in Eq (6) a constant value of DoLP uncertainty is used. The main difference between Eq(6) and MSE cost function is only in scaling. The scaled MSE provides a convenient way to compare with the measurement uncertainty and decide when the training is sufficient. For future applications when there is more sophisticated DoLP, which can be directly applied in Eq (6).

We have revised the manuscript as follows:

> "…where $\rho_t$ and $P_t$ indicate training data, and $\rho^{NN}_t$ and $P^{NN}_t$ indicate the NN predictions. N in the denominator is the batch size used in the training (taken as 1024 here). The same total uncertainty of $\sigma_\rho = 0.03\rho_t$ and $\sigma_P = 0.005$ as in Eq. 3 are used here. Therefore, $\chi^2_{NN,\rho}$ represents the percentage error of the NN predictions, which can effectively incorporate the sunglint signals without directly impacting by its large magnitude. Since a constant value of $\sigma_P$ is used, $\chi^2_{NN,P}$ is equivalent to a scaled MSE cost function. Polarization signal is better constraint within 0 and 1 for all viewing geometries and therefore its training performance less affected by the sunglint. This new cost function is a convenient and meaningful extension to the conventional MSE cost function applied on a set of normalized training data especially for reflectance (e.g. Aggarwal (2018); Fan et al. (2019); Gao et al. (2021a); Aryal et al. (2022); Stamnes et al. (2023)). We found the NN training hyperparameters (such as learning rate, batch size, etc) reported by Gao et al. (2021a) still work well for the new cost function. The resulting training process is aware of the measurement uncertainty and therefore optimizes in a way more relevant to the retrieval's operation."

3. The performance of the NN model is not well validated. Figure 2 has shown the cost function of training and validation, but readers cannot tell if the accuracy of the NN model is sufficient for simulation and retrieval. In this study, observations are generated by the NN model and the NN model is used for retrieval. Thus, it is important to compare the performance of the NN model with that of the radiative transfer model.

   Thank you for the suggestions. The accuracy of the NN is evaluated through the test data set which is not used in the training process as detailed in Appendix 1. The accuracy of the radiative transfer simulations are discussed in Gao et al 2021a, where an accuracy much higher than both measurement uncertainty and NN uncertainties are used. Since this study employed the same accuracy of the radiative transfer model in the training data. We expected a similar accuracy of the radiative transfer model.

   Since the simulation involves 10 million pixels, it is not practical to generate such a large amount of simulations using the full radiative transfer model. For the application of the NN model in real measurements, as demonstrated by field AirHARP measurement (Gao et al 2021a), we need to add the NN uncertainty, RT uncertainty into the total uncertainty model.

We revised our manuscript at the end of section 2.3 NN training and performance analysis:

> "Note that to ensure the high accuracy of the NN models, the RT simulations with a numerical accuracy much higher than the measurement and NN models are used to generate the training data as discussed in Gao et al. (2021a). For the application to real field measurements, the uncertainties including the NN models, RT simulations and the measurement uncertainties need to be considered."

**Technical corrections:**

Line 116: σm. m should be subscript.

Corrected.

Line 189: (Gao et al., 2021a) -> Gao et al. (2021a)

Corrected.