We appreciate the effort and time the reviewer has invested in reviewing our manuscript. We are grateful for the constructive feedback, which have improved the quality of our research. Please find our response below with revision details.

The topic of the paper is very important and interesting. The paper is, in general, well-written and quite easy to follow. I only have a few comments, including some technical ones. PXLX below indicates page X and line X.

Thanks for the positive comments.

P5L11 The atmosphere and ocean system is assumed to be a four-layer system. Is there any coupling effect between the layers being taken into account?

Yes, the radiative transfer simulation (RTSOS) fully considers the coupling between these four layers. We added the following information:

Original:

> "The atmosphere and ocean system are assumed to be a four-layer system."

Revised:

> "The atmosphere and ocean system are assumed to be a four-layer system and **the radiative transfer interactions among them are fully considered in the RTSOS model.**"

P5L15 The layer extends from the ocean surface to a height of 2 km. Which profile shape is used?

The aerosol layer is assumed to be uniform, and the molecular profile follows the US standard atmospheric profile. We revised our manuscript as follows:

Original:

> "…The third layer is an aerosol layer mixed with Rayleigh scattering. The layer extends from ocean surface to a height of 2km. The last layer contains atmospheric molecules from 2 km to the top of the atmosphere."

Revised:

> "…The third layer is an aerosol layer mixed with Rayleigh scattering. This layer extends from ocean surface to a height of 2km **with a uniform aerosol vertical distribution.** The last layer contains atmospheric molecules from 2 km to the top of the atmosphere. **The US standard atmospheric constituent profile is used to describe the molecular distributions (Anderson et al., 1986).**

P5L27, ‘An accuracy of less than 1% for reflectance and less than 0.003 for DoLP has been achieved’ What is the uncertainty for all instrument-related issues for AirHARP and HARP2?

The instrument uncertainty in reflectance is 3% for both AirHARP and HARP2. For DOLP, the uncertainty is 0.01 for AirHARP and 0.005 for HARP2. These information is provided in a latter location in Section 3.1 :

> “Correlated errors for both the AirHARP and HARP2 instruments are generated according to the same 3% uncertainty for reflectance, but 0.01 in DoLP for AirHARP and 0.005 in DoLP for HARP2.”

We made the following revision to introduce these uncertainties early:

Original:

> “The accuracy of the NN forward model is examined with an independent synthetic measurement dataset not used in training. An accuracy of less than 1% for reflectance and less than 0.003 for DoLP has been achieved (Gao et al., 2021a).
> “

Revised:

> “The accuracy of the NN forward model is examined with an independent synthetic measurement dataset not used in training. An accuracy of better than 1% for reflectance and better than 0.003 for DoLP has been achieved (Gao et al., 2021a).**The uncertainties of the NN forward model are less than the instrument uncertainties of AirHARP and HARP2 (~3% in reflectance, 0.01 in DoLP for AirHARP, and 0.005 in DoLP for HARP2).**
> “

P6L2 what is t in ρ_t and P_t

Here we use ‘t’ to indicate “total” signals, as this signal involves contributions from aerosol scattering, Rayleigh scattering, ocean surface contribution, etc. We revised the sentence as follows:

Original:

> “Measurement vector m includes both reflectance (ρ_t) and DoLP (P_t) with the total number of measurements of N, which has been used in previous studies (Gao et al., 2021a).”

Revised:

> “Measurement vector m includes both reflectance (ρ_t) and DoLP (P_t) **where the subscript t indicates the total signal measured by the instrument. The total number**

of measurements, N, at each pixel includes contributions from both reflectance and DoLP, which has been used in previous studies (Gao et al., 2021a, 2021b).”

P6L22, AR(1) works well for most cases. Under what conditions will AR(1) have potential problems?

Thanks for the question. We have revised the sentence to make it more clear. Note that AR model is only briefly mentioned here and more detailed analysis will be followed:

Original:

> “However from our analysis of AirHARP measurements in Sec. 4.3, AR(1) works well for most cases.”

Revised:

> “However from our analysis based on the retrieval results from real AirHARP measurements, **AR(1) works well for most cases. Detailed analysis can be found in Sec 2.5 for theoretical basis, Sec 4.3 for real data applications, and Sec 5 for general discussions.**“

Then for Sec 2.5, we made the following edits:

Original:

> “
>
> However, the mean values and variance in the fitting residuals often vary with respect to the angular grids. This type of signal is classified as non-stationary and difficult to model (Priestley, 1983). To overcome this issue, the original residual data y is processed by removing its mean and normalizing by its standard deviation.
>
> “

Revised:

> “
>
> However, the mean values and variance in the fitting residuals often vary with respect to the angular grids. This type of signal is classified as non-stationary and difficult **to study by the AR models** (Priestley, 1983). To overcome this issue, the original residual data y is processed by removing its mean and normalizing by its standard deviation.
>
> “

Furthermore, we have detailed discussion in Section 5 with several potential challenges in using AR model and other correlation analysis tools (no revision here):

> "…
> 1. The retrieval is based on a forward model which also has uncertainties, a portion of which may be correlated. This uncertainty will contribute to the fitting residuals and may impact correlation analysis, but it is difficult to quantify.
> 2. The fitting residuals are often not stationary with uniform mean and variance. To reduce this issue, the residuals are normalized, but it would be valuable to analyze how the mean value and variance depend on the angle, as this may provide insight into the modeling uncertainties.
> 3. Some residuals are not continuous with angle due to removed cirrus clouds, which may reduce the correlation.
> 4. Synthetic data analysis has demonstrated that the retrieval is likely to overfit the data when the correlation is strong.
> 5. The angular grids for HARP measurements are slightly non-uniform, which is likely to further reduce the correlation strength from auto-correlation analysis. To evaluate impacts of this feature, an uncertainty model considering the impact of the real angular grids need to be built. But since the variation of the angular grids are less than 1 (670nm band) or 2 (other bands), which may impact more the cases with small correlation angles.
>
> "

P7L15, how uncertain is such an assumption ,the retrieval parameters successfully converged to the global minima ' and what is the potential impacts on the error propagation and also on the retrieval results?

Since our retrieval algorithm involves many retrieval parameters, it is always a challenge to ensure that all parameters converge to the global minima. If the retrievals converge to a local minimum, the retrieval parameter can be less accurate, and the derivatives around the local minima may not represent the actual values near the global minima. We made the following edits:

Original:

> "The retrieval uncertainties estimated by error propagation (hereafter called theoretical retrieval uncertainty) as shown in Eq. (6) represent the optimal scenarios, with limitations such as the assumption that the retrieval parameters successfully converged to the global minima (more discussions in Sayer et al. (2020); Gao et al. (2022))."

Addition:

> "…**Both the retrieval results and associated Jacobians can be less representative to the truth values and therefore lead to inaccurate error propagation and uncertainty estimation**."

P9L4-5, what is the value of a typical correlation angle for AirHARP and HARP2 near 660 nm?

Thank for the questions. Since the correlation properties for HARP instrument have not been characterized before, there is less quantitative information on a typical correlation angle value. However, in our analysis based on the retrieval analysis, the correlation angle for reflectance is likely larger than 10-20°. Since, there are also possible chances of underestimation as discussed in Sec 4.2, we choose a wider range of correlation angle in our study from 0 to 120 °. In this section, we use 10° and 60° as representative value of correlation angle to demonstrate the properties of correlated errors.

Furthermore, Knobelspiesse et al (2012) has assumed a correlation parameter of 0.9 (corresponding a correlation angle of 10 °) in the study of the angular correlation of RSP instrument. This is another reason, we select 10° as an example.

We have revised as follows:

Original:

> "The correlated error samples with correlation angle of θc = 10° (r = 0.9) and correlation angle of θc = 60 ° (r = 0.98) are shown in Fig. 2 (a) and (c)."

Revised:

> "**To demonstrate how angular correlations impact the errors,** the correlated error samples with correlation angle of θc = 10 ° (r = 0.90) and correlation angle of θc = 60 ° (r = 0.98) are shown in Fig. 2 (a) and (c). **A value of r=0.9 has been assumed in the study of RSP angular correlation by Knobelspiesse et al (2012)**."

On a related subject, there are several works study the correlation in different domain (retrieval parameters, spectral, spatial). We have summarized related work in the introduction and adding more reference below:

Original:

> "Retrieval algorithms that exploit correlation information in retrieval parameters and measurement uncertainties have shown benefits in improving remote sensing capabilities. The Generalized Retrieval of Aerosol and Surface Properties (GRASP) algorithm retrieves multiple pixels simultaneously, while considering the spatial correlation of the retrieval parameters (Dubovik et al., 2014, 2021). Xu et al. (2019) developed a correlated multi-pixel inversion approach (CIMAP), which further considers the correlation between different retrieval parameters. Theys et al. (2021) developed a Covariance-Based Retrieval Algorithm (COBRA) based on an error covariance matrix estimated from

measurements with spectral correlation, applied their approachto sulfur dioxide (SO_2) retrievals from the TROPOspheric Monitoring Instrument (TROPOMI) data, and demonstrated improved retrieval performance. "

Addition:

> **… To accurately evaluate pixel-level uncertainty in ocean color retrievals, spectral correlation associated with the uncertainty in top-of-atmosphere reflectance are also accounted for OLCI (Lamquin et al, 2013) and MODIS (Zhang et al, 2022) in the uncertainty propagation."**

Added reference:

- **N. Lamquin, A. Mangin, C. Mazeran, B. Bourg, V. Bruniquel, and O. F. D'Andon, "OLCI L2 Pixel-by-Pixel Uncertainty Propagation in OLCI Clean Water Branch," (ESA, 2013), p. 51.**

- **Zhang, M., A. Ibrahim, B. A. Franz, Z. Ahmad, and A. M. Sayer. 2022. "Estimating pixel-level uncertainty in ocean color retrievals from MODIS." Optics Express, 30 (17): 31415 [10.1364/oe.460735]**

P9L6, Errors start to form a longer range of correlation with smoother variations. Is the smoother pattern caused by the relatively small magnitude of the correlation angle of 60 degrees, or is it really true that an increase in the correlation angle leads to a decrease in the magnitude and pattern of the errors with respect to the viewing angles? Why did the author limit the viewing angles to 25, rather than 60?

Thank you for the question. A larger correlation angle indicates a longer range of correlation and appears with smoother variation also shown in Fig 2. However, the increase of correlation does not necessarily mean a decrease of the magnitude. As shown in Fig 2(b), the errors with stronger correlation will start to move together, but the overall magnitude can vary in a wide range.

We added more discussion here:

Original:

> "With larger θc the errors start to form longer range of correlation with smoother variations."

Revised:

> "With larger θc the errors start to form longer range of correlation with smoother variations. **Note that the overall magnitude of the errors can vary within the full range as described by the calibration uncertainties"**

Regarding the viewing angle at the right side of the plot in Fig 2, since there is strong glint in the right side, which cause issues in training the neural network models, we have removed the sunglint as shown in Fig 4, which corresponds to the partial removal of the angles in the right side.

The following sentence is added to the Fig 2 caption:

> **".. The right side of viewing angle ends around 25^{o} due to the removal of sunglint as shown in Fig 4."**

P11L6, Chla to Chl-a

Corrected into Chl-a. And checked the whole document.

P11L8 The range of [0.01, 0.5] for AOD sounds reasonable. However, as there are many plumes along the coastal regions, will such a restriction of 0.5 lead to a too-small error estimation for real measurements?

Thanks for the question. The range of [0.01, 0.5] is the nominal range of aerosol loading for ocean color remote sensing. The development of neural networks for cases with larger AOD will be a subject of future work.

Original:

> "The same sampling approach discussed in Gao et al. (2022) is conducted assuming that the aerosol optical depth (AOD) and fine mode volume fraction are uniformly distributed within [0.01, 0.5] and [0,1], respectively."

Addition:

> **"… A larger range of AOD values will be needed for applying this algorithm to cases of smoke and plume events."**

P13L5-7 The real uncertainties in both the root mean square error (RMSE) and the mean average error (MAE) are larger when uncertainty is correlated (comparing (b) and (a)) , it seems for (b), the real one is smaller (0.029 vs 0.03) as compared to the theoretical one?

Thank you for noticing the difference. Since the analysis is based on Monte Carlo sampling, there could be statistical fluctuations. To reduce the impact of such fluctuation (and corresponding uncertainties), in our latter discussion, we sampled 10 times, and take their average. As shown in Fig 6, the average real uncertainties are still mostly larger than theoretical uncertainties. Please find more detailed discussion for this technique in Gao et al (2022) (https://doi.org/10.5194/amt-15-4859-2022) . We have revised our discussion as follows to be more precise:

Original:

> "The real uncertainties in both the root mean square error (RMSE) and the mean average error (MAE) are larger when uncertainty is correlated (comparing (b) and (a))."

Revised:

> "The real uncertainties in both the root mean square error (RMSE) and the mean average error (MAE) are **mostly** larger when uncertainty is correlated (comparing (b) and (a)) **with exceptions possibly due to statistical fluctuations in the Monte Carlo sampling."**

P14 L1, Both real errors and theoretical uncertainties have occasional outliers with large values due to poor convergence. Is there any link due to the assumption that the retrieval parameters successfully converged to the global minimum? (P7L15)

Yes, we also agree with the reviewer. The outliers could be related to the poor convergence not reaching the global minimum. We revised as follows:

Original:

> "Both real errors and theoretical uncertainties have occasional outliers with large values due to poor convergence, and this has large impacts on the RMSE values."

Revised:

> "Both real errors and theoretical uncertainties have occasional outliers with large values possibly due to convergence **to local minima instead of global minima,** and this has large impacts on the RMSE values."

P16 Title of Fig7, change zero to ‚0'

Corrected.

P16L4 change (Gao et al., 2021b) to Gao et al (2021b)

Corrected.

P17 L1, but show significant impacts (as small as 0.5) for $\theta c = 60^{\circ}$. Is this due to the small value of theoretical uncertainty (red in Fig.6), which leads to small ratio? If so, I would suggest the author put more effort to explain Fig. 6.

Yes, we agree with the reviewer. The theoretical uncertainty decreases faster with stronger correlation. We added more discussion here:

Original:

> "The impacts on the remote sensing reflectance are generally smaller for θc = 10 º, but show significant impacts (as small as 0.5) for θ^c = 60º."

Revised:

> "The impacts on the remote sensing reflectance are generally smaller for θc = 10 º, but show significant impacts (as small as 0.5) for θ^c = 60º. **This is because the theoretical uncertainty with correlation angle decreases faster than the real uncertainties (see Fig. 6).**"

P19L3 overfitting of the data, have you check this issue with , test set '

Yes, this study is based on the synthetic simulation data with controlled error added to the simulation. To understand this overfitting effect we also computed the standard deviation of the retrieval residuals with results shown in Fig 11, which also shows smaller values than expected uncertainty model.

Original:

> "One cause of the shift is due to overfitting of the data, which results in smaller residuals and larger real retrieval uncertainties as shown in Fig. 6."

Revised:

> **"Smaller cost function values indicate smaller retrieval residual, which may be caused by overfitting of the data, and also possibly lead to the larger real retrieval uncertainties as shown in Fig. 6."**

More discussions are provided in the next section.

P19 L6, a degree of freedom of 40 and 20 are found to better fit the cost function histogram with θc = 10◦ and θc = 60◦, as compared to θc = 0◦? Or the comparison between (a) and (b) in Fig. 10?

Yes, this refers to the Fig 10 (a) and (b). Here we are trying to find an approximated degree of freedom which can fit the cost function histogram when correlations are presented. We find that we have to reduce the original degree of freedom of 150 (150 measurements used) to a value of 40 and 20 for correlation angle of 10 and 60 degrees.

Original:

For example, in the right panel in Fig. 10, a degree of freedom of 40 and 20 are found to better fit the cost function histogram with θc = 10◦ and θc = 60◦,

Revised:

For example, in the right panel in Fig. 10, **chi^2 distribution with a degree of freedom of 40 and 20 are found to better fit the cost function histogram with θc = 10◦ and θc = 60◦, comparing with results using all the measurement degree of freedom (150).**

More information is also provided in the Fig 10 caption:

“The red line indicates the chi2 distribution with a degree of freedom of 2Nv = 150. The green and blue lines indicate the chi2 distribution with a reduced degree of freedom of 40 and 20 fitted to the corresponding histogram.”

P19L14, This behaviour indicates overfitting, where the uncertainties are partially removed as real signals. It is removed or considered?

Thank you for the suggestion. Here the errors is treated as real signal, and removed from the retrieval residuals by the model. We believe you can also say they are “considered” in the model. We revised the paragraph:

Original:

“This behavior indicates overfitting, where the uncertainties are partially removed as real signals.”

Revised:

“This behavior indicates overfitting, where the errors are partially removed as real signals **and lead to reduced residuals**.”

P22L17 Partial autocorrelation for reflectance showed similar results for from the synthetic data in Fig. 3 (b) with only the first order term prominent, which suggest that the AR(1) model is sufficient to describe the fitting residual for reflectance. However, we can see clear differences in the dependence on angular step k. Why is this?

The difference is likely due to the small higher correlation terms. The AR(1) model capture the major correlation behavior in the data. The higher order contributions are mostly small as suggested by the partial autocorrelation function.

Original:

"Partial autocorrelation for reflectance showed similar results for from the synthetic data in Fig. 3 (b) with only the first order term prominent, which suggest that the AR(1) model is sufficient to describe the fitting residual for reflectance."

Revised:

"Partial autocorrelation for reflectance showed similar results for the synthetic data in Fig. 3 (b) with only the first order term prominent, which suggest that the AR(1) model is sufficient to describe the fitting residual for reflectance, **with higher order contributions negligible**."

After reading the whole manuscript, I am thinking maybe the author should make a more detailed summary at P5L3-8 because there are many citations of their previous work, which requires quite some effort to check those very relevant publications. But I leave this comment open to the authors.

Thank you for the suggestion. The paragraph in page P5L3-8 is aimed to provide a high-level summary to our previous work, which focusing on the development of the retrieval algorithm and application of neural networks to improve speed and accuracy. The study in the error correlation is tested using the results from the retrieval algorithm, but not limited to a specific algorithm, and should be general in treating any other error correlations. Since we relied on the uncertainty quantification method developed in Gao et al. (2022), we provided more details and examples in Sec. 3.2. More details on the multi-angle cloud masking used in this study to reduce impact of the cirrus cloud is discussed in Sec. 4.3. We hope those discussions are sufficient for the readers to follow our work.

At the end of Sec 2.1 we added:

"In this study, we will discuss the retrieval uncertainty and performance in aerosol properties, ocean surface wind speed, and Chl a in the ocean, as well as water leaving signals based on the retrieval parameters. The water leaving signal refers to the remote sensing reflectance (Rrs), which is the ratio of the upwelling water leaving radiance and the downwelling solar irradiance just above the ocean surface (Mobley 2022). Rrs can be estimated through the atmospheric correction process which removes the contribution from the atmosphere and ocean surface from the total measurements at the sensor and additional BRDF correction to reduce the dependency on the solar and viewing directions. Both atmospheric and BRDF corrections with their associated uncertainties are implemented using neural networks as discussed in Gao et al., (2021a, b) and followed by this study."