We are deeply grateful to the reviewer for their constructive and pertinent comments on the manuscript, which help to greatly improve the paper.
Please find hereafter our point-by-point responses to the comments and suggested corrections.
Comments are in black, our responses in blue and the text modifications in green (in bold when only one part of the sentence was modified).
Note that the indicated line numbers correspond to the line numbers of the "track changes" version.

RC1:
Manuscript number: amt-2021-414
Full title: Cloud optical properties retrieval and associated uncertainties using multi-angular and multi-spectral measurements of the airborne radiometer OSIRIS
Author(s): Matar et al.

The paper investigates the retrieval uncertainties of liquid cloud properties, specifically cloud optical thickness (COT) and droplet effective radius (Reff), associated with various error sources using an optimal estimation (OE)-based retrieval procedure. The retrieval error covariance matrix is decomposed into the one originating from individual sources of errors, which makes it possible to evaluate the quantitative retrieval uncertainty estimation for each error source, including measurement error, model parameter errors, and error due to the assumptions in the forward model. The authors apply this framework to multi-angular and multi-spectral measurements made by OSIRIS airborne radiometer with a focus on liquid clouds. The results indicate the forward model assumptions that do not consider in-cloud heterogeneous profiles and 3D radiative transfer effects induce the largest uncertainties in the retrievals, followed by the measurement errors. While the model parameter errors provide the least impact on the retrieval uncertainties that are less than 0.5% of the retrieval quantities. These error estimates and retrieval procedures will be useful for the further 3MI observations.
While this paper is well written in the introduction and methodology part, I am concerned about the representativeness of the results and the analysis flow of this work. Please find my comments below.
The topic presented in this paper is suitable for Atmospheric Measurement Techniques. I recommend Major Revisions to reconsider the manuscript for publication.

Major comments

Representativeness
First of all, the paper discusses only the relative standard deviations (RSD) of the retrieval variables. Why don't the authors discuss the retrieval biases associated with the measurement errors, model parameters, and assumptions in the forward models? In particular, the vertical heterogeneity and 3D radiative effects would induce substantial biases in the retrievals. The different magnitude of errors among observational bands and angular directions would cause impacts on the cloud property retrievals (through changing the sensitivity weights).
However, the relative standard deviations of the retrieval variables do not tell you how much the retrieval variables are biased.

The work presented in this paper is based on the optimal estimation method. The principle of the method is to determine the most probable state knowing that measurements have been performed with uncertainties represented by the PDF of the measurements. The reasoning is based on the Bayesian formalism with a-priori and a-posteriori probability densities which allow linking the space of the states to the space of the measurements. The retrieved state is then the most likely consistent state according to the available information. In the optimal estimation method, all the PDF are assumed to be Gaussian PDFs. This implies that the uncertainties can be estimated by the covariance matrix of the parameters, and then by the relative standard deviation (RSD).

However, we agree with the reviewer that the use of Gaussian PDFs can have an impact in the

assessment of the uncertainties depending on the type of errors : the measurements errors are usually, at first order, considered as random errors. They can consequently be modeled by a Gaussian distribution. The non-retrieved parameters and the forward model parameter errors are derived following Rodgers, 2000 (chapter 3). The computation of these two errors lies on the assumption that the forward model can be locally linearized about the state vector value. If the non-retrieved parameters have been properly retrieved, their own uncertainties are unbiased and can be modeled by a Gaussian PDF. With the assumption of a linear model, the resulting errors on the state vector are unbiased and can also be represented by a Gaussian PDF.
The forward model error is obtained (Eq. 21) from the difference between the results of the simplified model F used for the retrieval and the realistic model F' multiplied by the Gain matrix, which represents the sensitivity of the retrieval to the measurements. A bias between the results of the two models will then be included in the Gaussian pdf representing the forward model errors, which tends to overestimate the errors.

To clarify these different points and clearly state the different assumptions made. We add several paragraphs or sentences:
In section 3.1, we add a paragraph explaining the Bayesian formalism and the assumption of the Gaussian distribution, line 25 :
"To achieve it, a Bayesian probabilistic approach is applied. Before the measurements, an a priori knowledge of the state vector can be described by a probability density function (PDF). Once the measurements have been carried out, this knowledge can be described by the posterior PDF of the state, which is a conditional probability (probability of having given that is true). The posterior PDF of the state vector can be related to its a priori PDF by the Bayes' theorem:

$$P(x|y)=\frac{P(y|x).P(x)}{P(y)} \tag{2}$$

Where P(y) is the PDF of the measurements including the uncertainties and P(y|x) is the PDF of the measurements given that we know the state vector.

In the optimal estimation method, the previous PDFs are represented by Gaussian distributions, assuming that the errors of the measurements, the errors related to the non-retrieved parameters and the errors of the forward model are normally distributed around a mean value. In other words, we assume that the model can be linearized around the most probable state vector.

We also add clearly the assumption of linearity, section 3.1, line 280 :
Assuming the model is nearly linear around a given state vector,...

The choice of RSD to characterize the errors derives from the assumption of Gaussian distribution. We specify line 350 :

The use of Gaussian PDF leads to compute the uncertainty on a particular parameter x_k as the square root of the corresponding diagonal elements **of the covariance matrix** $\sigma_k=\sqrt{S_{xkk}}$, where k is the index of the parameter in the state vector x.

We add later in section 3, a sentence for each errors indicating that we assume a Gaussian distribution and linearity of the model about the state vector value :

- Line 381, for the measurements errors: "The **uncertainties** of the measurements **remaining after the calibration processes are assumed**, random, uncorrelated between channels and **can be consistently approximated as a Gaussian** probability density function over the measurement space"
- Line 400, for the errors related to the non retrieved parameters: "These errors are considered

to be independent and random under the assumption of linearity of the radiances around the non retrieved parameters "

- Line 448 , concerning the errors related to the forward model: "The simplified model used for the retrieval can lead to biased retrieved parameters. In this case, the bias will be included in the Gaussian PDF width, resulting in an overestimation of the uncertainties."

If the authors want to solely focus on the retrieval RSD for each error source, then it can be more appropriately achieved through numerical experiments based on synthetic simulations of observational signals based on this framework or through incorporating the root-mean-square error (RMSE) that includes the retrieval bias information and can be theoretically derived from the bias and RSD.

The aim of our paper is not to give an exhaustive view of the possible errors concerning optical thickness and effective radius retrieval. It is to present a method to derive, from real data separately the different sources of uncertainties and to evaluate them in one example. The idea is thus not to have a general representativeness of all the possible cases, which is a huge work but more to focus on an airborne data campaign and shows how from multi-angular measurements, it is possible to derive the usual optical thickness and effective radius parameters with the partitioning of their uncertainties. To clarify our objectives, we add, as mentioned later, a paragraph at the end of the introduction section (line 141-146).

Alternatively, the authors may consider to additionally perform cloud property retrievals based on error covariance matrix that considers ALL sources of errors (i.e., measurement errors, model parameter error, and forward model errors) to see how different the cloud property retrievals are compared to measurement error only cases.

We agree that the method suggested by the reviewers is the ideal way to assess the uncertainties of the retrieved parameters but, in reality, it is impossible to implement for computational cost reasons because 3D RT simulations are much too long.
In addition as the uncertainties due to the forward model are large, the optimal estimation algorithm will tend to converge at the first iteration as the simulated radiances will be included in pdf defined by their measurements and covariance matrix.

In addition, the uncertainties and fixed parameters of the source of errors seem to be determined in different ways. The cloud top height value and its uncertainty are derived based on collocated lidar observations (i.e., representing the CALIOSIRIS campaign), the effective variance seems to be arbitrarily determined, and the wind speed value and uncertainty may represent global statistics. This gives me a question on the representativeness of this study.

As explained above, the study does not aim to be representative of all cloudy situations. The fixed parameters and their uncertainties are chosen according to the experimental set-up of the airborne campaign. During the CALIOSIRIS campaign, a Lidar was on-board the airborne, which allowed us to have an accurate value with low uncertainties for the cloud top. Obviously, if less accurate information is available, the uncertainties due to cloud top retrieval could be higher.

The effective variance was chosen according to the number of supernumerary bows visible in the polarized measurements. At the time of the CALIOSIRIS campaign, polarized radiances had calibration issues and were not usable for retrieval. They were, nevertheless, used to find the effective variance that best fits the number of the supernumerary cloud bows. Figure 1 represents the averaged polarized radiances for a transect obtained with OSIRIS and simulated radiances using an effective variance of 0.02. As the used effective variance was based on polarized radiances of the measurements, we decided to assume weak uncertainties for Veff, by choosing a standard deviation of 15% .

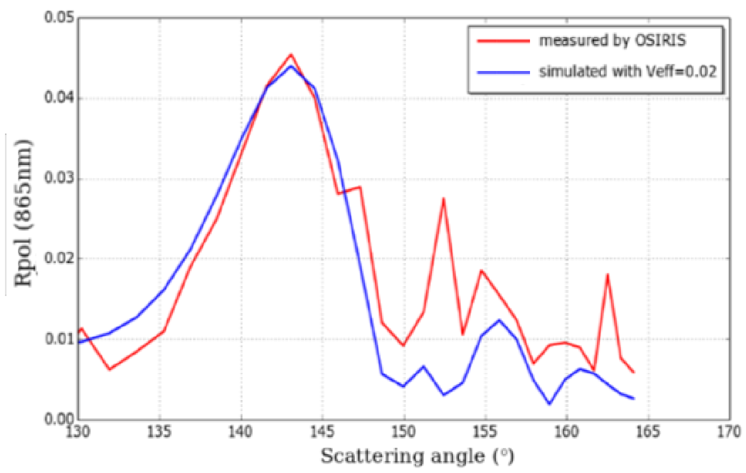

Figure 1 (not included in the paper) : Averaged polarized radiances measured by OSIRIS for a transect in the middle of the central image of CALIOSIRIS scene and simulated polarized radiances with an effective variance of water droplet distribution equal to 0.02 (in blue), as a function of the scattering angles.

Concerning the wind speed, as no direct measurements were available, we used the value given in the NCEP reanalysis of the National Oceanic and Atmospheric Administration (NOAA) database for the day of the campaign. As we do not know the uncertainties of this value, we chose to use 10% of uncertainties for the wind speed, a value that can be adjusted if the real uncertainty value becomes known.

To clarify the choice of the fixed parameters, a sentence line 418:
"The values and the uncertainties of the fixed parameters are chosen according to the experiment setup of the campaign."

For the cloud top altitude, we modified the paragraph line 420 to 425 (modifications are in bold):
"To estimate the uncertainties originating from the fixed cloud altitude, we used **the opportunity of having the LIDAR-LNG on board the aircraft, which gives** the backscattering altitudes signal obtained around the case study of CALIOSIRIS. From 11:01:06 to 11:03:06 (time where the same cloud scene is apparent), it varies between 5.57 and 5.73 km in our cloud scene. **For practical reasons due to the radiative transfer code, we use a value of 6 km for the cloud top altitude with** a standard deviation of $\sigma_{alt} = 0.16$ km (3% of the cloud altitude). **This value is low thanks to the knowledge provided by the Lidar.**"

For the choice of effective variance, we just complete the following sentence, line 428 : "...., we fixed a value of 0.02 based on the **number** of supernumerary bows in the polarized radiances **(not shown)"**

and add line 430. As the value of V_{eff} was fixed using the polarization measurements of OSIRIS, this uncertainty is weak and not representative of all clouds"

We add also in the analysis of the non retrieved parameters, section 4, line 577 : "We remind that we fixed the value of veff using multi-angular polarized measurements of OSIRIS, which leads to choose a weak uncertainty of v_{eff} (15%). However, if no information on veff is available in the measurements, the uncertainty should be higher and thus the errors due to the non-retrieved effective variance. Platnick et al. (2017) obtain 2% and 4% uncertainty for COT and Reff respectively for a standard deviation from veff between 0.05 and 0.2"

Concerning the wind speed, we add line 328 : "…with a fixed ocean wind speed based on NCEP reanalysis of the National Oceanic and Atmospheric Administration (NOAA)"

If the authors make the results useful to the future 3MI mission, the uncertainty (specifically, the standard deviation of a model parameter) of each error source (in particular, the model parameter errors) should represent those of global climatology. Specifically, I feel that the cloud top height and effective variance uncertainties may be a bit too small, so the retrieval RSD might be underestimated (although the general conclusion the little contribution of model parameter uncertainties to retrieval RSD may not change).

The reference to the 3MI instruments concerns the use of OSIRIS data, which is a prototype of 3MI and the methodology rather than the results and analysis of our studies. Obviously, for 3MI, the methodology has to be adapted. For example, the uncertainties on cloud top would be larger, as the cloud top pressure uncertainties from the O2 band are larger (Desmons et al., 2013). For the effective variance, it will be possible to add this parameter in the state vector using a combination of the polarized multi-angular and shortwave infrared measurements in the measurements vector.

We mention the adjustments that have to be made for applications of the method to 3MI measurements in the conclusion section, line 740 to 746 :
"Note that, since information provided by Lidar or polarized measurements was used, the uncertainty for the non-retrieved parameters was chosen to be low. For applications to other cases without these available information, errors would be higher. If the method is applied to 3MI for example, the errors related to the cloud top altitude would be higher as the O2-A band leads to cloud top pressure uncertainties between 40 and 80hPa depending on the cloud types (Desmons et al. 2013). A more complex algorithm could also be used with a measurement vector including O2-Aband radiances and multi-angular polarized radiances to have information on and to add the cloud top altitude and the effective variance (Huazhe et al. 2019) in the state vector."

For these reasons, I suggest the authors clearly state what are the focuses of this paper in the last paragraph of the introduction and reconsider the experimental design of this work through, but not limited to, using numerical experiments, adding the retrieval biases to the current analysis, or other appropriate methods.

Using numerical experiments can also be a good way to assess the uncertainties of the retrieved parameters. However, doing it for different types of clouds, different geometry conditions (solar and view) is a huge work, which is beyond the scope of this paper.
To clarify our objectives, we add a paragraph in the end of the introduction section (line 141-146) :
"The aim of this paper is not to give an exhaustive view of the possible errors concerning optical thickness and effective radius retrievals but to simply introduce a method to derive the different sources of uncertainties from a specific case of data acquired during an airborne campaign. Uncertainties due to error measurements, to non retrieved parameters but also to the assumed forward model are considered. If generalized to several cloudy scenes, the partitioning of the errors can help to understand if and which non-retrieved parameters or forward model need to be optimized in order to reduce the global uncertainties of the retrieved cloud parameters."

We agree that a study based on simulated data can validate our framework by showing that errors on retrieved parameters are included in the uncertainties obtained using the methodology presented.
We add in the conclusion section, line 767 to 769.
"A way to check the consistency of the method and the validity of the uncertainty ranges would be to simulate radiances using Large Eddy Simulation model with realistic cloud physical description, add noise for the errors measurements and derive the cloud parameters and their uncertainties."

Analysis flow

While I appreciate the authors for describing the solid mathematical basis of the analysis procedure, I may have an argument on the use of COT and Reff retrievals obtained based only on measurement noise in the error covariance matrix for the following analysis. This can overfit the retrieval variables to obtain an optimal solution (i.e., J ~ ny) because radiative signal perturbations (here what I mean is any of radiative perturbation induced by atmospheric-cloud properties naturally occur in the real world) originating from other sources of errors are partly explained with the biased retrieval values. If such perturbations are significant (and yes, it is significant particularly for vertical heterogeneity and 3D effects), then the sensitivity of the other error sources to the retrieval quantities are obtained from the sensitivities to the retrieval variables at the biased cloud property conditions. Ideally, numerically generated cloud property fields (such as Large Eddy Simulations) would provide datasets for the evaluations of the retrieval uncertainties based on the authors' flamework. As the true cloud properties are not available in the observed cloud field, it is not possible to address this issue based on the given observations. Therefore, at least, the authors should discuss a potential bias of the retrieval uncertainty evaluations based on this framework.

Indeed, the optimal estimation method gives the state vector that best matches the measurement vector under the assumption of a transfer function to pass from the state vector to the measurement vector. If the transfer function is false or biased, the retrieved parameters will obviously be biased. The uncertainties obtained for the non-retrieved parameters or to the forward model will also be incorrect if the variations predicted by the model about the retrieved (biased) values are different from those about the true value.

We add a sentence to raise this issue in the section 4, line 533 to 538 :
"The parameters retrieved in the first step may be biased, in particular due to the use of a simplified cloud model to connect the state vector to the measurements. The estimation of the uncertainties performed in the second step assumes that the variations predicted by the simplified and the realistic models around the retrieved values (potentially biased) and around the true values are identical. This is correct with a linear forward model but can be a too strong assumption in cloud retrieval regarding the non linearity of the relationship of the radiances in function of cloud parameters. A way to test this assumption would be to use numerical experiments"

And we add in the conclusion section, line 763 to 768 :
"The method was applied to real data, which means that the true cloud parameters are unknown. Consequently, it is not possible to know if real errors on the retrieved parameters are included in the uncertainties given by the method presented here. One reason that can lead to an erroneous assessment is that the estimations of the uncertainties are done around the retrieved values than can be biased. A way to check the consistency of the method and the validity of the uncertainty ranges would be to simulate radiances using Large Eddy Simulation model with realistic cloud physical description, add noise for the errors measurements and derive the cloud parameters and their uncertainties."
"

Minor comments

1. Title: Actually, the authors use only two bands for the cloud property retrievals, and therefore, "bispectral" measurements would be more descriptive rather than "multispectral" measurements?
Agree, it was done

2. Line 241, Eq. (7): Why do the authors use a linear scale of the cloud optical thickness for a state vector element, not a logarithmic scale? Although the relation between solar reflectance and COT is linear at very small COT conditions (t << 1), it is in general quite non-linear over most of the COT range, which makes the retrieval process slow and may degrade the convergence, and also may limit the representativeness of this results to optically thin cases.
We agree that using a logarithmic scale could help the convergence but we do not use it.
We mention in the text, that it could accelerate the convergence, line 306 :

"It can be noted because the relationship between radiances and optical thickness has a logarithmic shape, using *log(COT) i*nstead of COT in the state vector may accelerate the convergence."

3. Section 3.3: I do not find the values of a priori error covariance matrix in the manuscript. Please briefly mention what values are chosen.
It is an omission in the text, we precise the a priori vector and Sa value, line 309 : "The a priori state vector was set to [10,10µm] and the a priori covariance matrix Sa was set to 108. The latter was chosen very large in order to favor the measurements in the determination of the state vector."

4. Line 338 "$\boldsymbol{\sigma}$alt = 0.16" :: This represents only the cloud properties observed during the CALIOSIRIS campaign.
As explained above, this corresponds to the specific case of the CALIOSIRIS campaign. If the same methodology is followed for another campaign or another experimental setup (for example without LIDAR), the value can differ. To avoid misunderstanding, we add in line 423: "This value is low thanks to the knowledge provided by the Lidar."

5. Line 333 "$\boldsymbol{\sigma}$veff = 0.003" :: How did you get this value? It seems too small. Please site references that support this quantity.
See the answers and modifications made in the general comments response.

6. Lines 372–376: In addition, the collision-coalescence process can provide a larger droplet radius at the lower part of clouds, which are observed from CloudSat. As vertical heterogeneity is important to the cloud property retrieval, the authors may consider a better representation of cloud profiles using a better cloud profile parameterization (e.g., Saito et al., 2019).
We add it in the introduction section, line 114:
"Saito et al. (2019) propose a method to retrieve the vertical profile using Empirical Orthogonal Function (EOF) to reduce the degrees of freedom of the droplet size profile"

And in section 3.3.3 line 463:
Depending on the maturity of the cloud, turbulent and evaporation processes can reduce the size of droplets at the top of the cloud **and collision and coalescence process can increase the size of the droplets in the lower part of the clouds as observed by Doppler Radar (Kollias et al., 2011). The profile used in this study aims to represent the case of droplet size reduction at the top of the cloud but other and more sophisticated and representative profiles can be used (e.g. Saito et al., 2019).**

7. Lines 416–420: As the authors assume the flat cloud top, which reduces some of cloud 3D effects such as illuminating and shadowing effects, I had an impression that the authors may focus on lateral photon transport effects here. If so, it would be better to rephrase 3D with lateral photon transport or state "3D" regarded as the lateral photon Transport.
In the paragraph describing the 3D radiative transfer simulations, we refer to 3D and 1D radiative transfer, so we decided to keep the term "3D" but add a sentence to express that differences between the two are related to the heterogeneity along the lines of sight and lateral photon transport, line 519 :
"The differences are thus mainly due to the lateral photon transport which tends to smooth the radiances fields compared to their 1D counterpart (Davis et al. 1997) and to the cloud heterogeneity along the line of sight (e.g. Fauchez et al., 2018)."

8. Figure 4: The readers cannot recognize if there are optimal/non-optimal solutions from these plots. If it is non-optimal, a set of cloud retrievals may not adequately explain the measured signals. I suggest the authors add the cost function distributions in addition to these two plots.
We add the normalized cost function and the convergence type in Figure 4 and Figure 10. Values less or equal to one indicate a convergence of type 1 represented in green in Figure 4 and Figure 10. If

convergence of Type 1 does not occur, the iteration can stop with convergence of Type 2 when the difference of the state vector between two successive step are less than n_X

We add in the comment of Figure 4, line 542 to 550: "Figure 4c presents the normalized cost function, which is less or equal to one when the retrieval successfully converges according to Eq. 6 (convergence of Type 1). In case of multi-angular measurements, the normalized cost function is often above one meaning that the simulated radiances do not fit the measurements while considering the measurements error covariance only. This comes from the attempt to fit the measured radiances from all the available viewing directions with a too simple forward model far from reality. The retrieval stops thus mainly according to Eq. 7 (convergence of Type 2) indicating that the state vector remains almost constant between two successive iterations. When neither Eq. 6 or Eq. 7 are achieved the retrieval fails. For the whole scene, failed retrievals account for 3.3% of the pixels. The failure may be associated with pairs of radiances outside the LUT that can occur for several reasons well documented in Cho et al. (2015)."

And in the comment of Figure 10, line 677 :

"**A normalized cost function value (Figure 10c) less or equal to one** is not necessarily an indication of an accurate retrieval, but only that a fit occurred."

9. Lines 483–484 "3D effects due to solar illumination do not appear in the retrieved cloud properties" :: This is an obvious statement as the cloud top is assumed to be flat, which removes the shadowing and illuminating effects. Please state that this error evaluation focuses solely on the lateral photon transport effect.

Even if illumination and shadowing effects are also present to a lesser extent for flat cloud top (Varnai et Davies 1999), we remind that we assume a flat cloud top that minimizes the solar and shadowing effects, line 612 : "However, in this work, we are dealing with flat cloud tops that induce weaker 3D effects than bumpy cloud tops (Várnai and Davies, 1999)"

10. Lines 539–540: The cost function divided by the number of measurement signals (J/ny) is a comparable quantity among mono-angular- and multi-angular-based retrievals.

We agree that it is comparable but we chose to compare RSD to be consistent with the rest of the paper. We add that we can also compare this quantity, section section 5, line 693:

"To compare the uncertainties of the two retrievals , we use the relative standard deviation (RSD) **to be consistent with the previous results**"

11. Lines 559–560" Why are these uncertainties reduced for multi-angular cases?

Multi-angular measurements provide more information to constraint the state vector, especially in the cloud bow regions, leading to a reduction of the RSD. This was already reported in line 669:

"Clearly, the multi-angular measurements contain more information and allow to resolve the problem encountered with the mono-angular bispectral method which is also clear in the reduction of the failed convergences from 7.6% to 3.3%."

12. Line 610: If the authors state that "the uncertainties related to the measurement errors is implementable in an operational algorithm," the uncertainty evaluations should be based on global climatology of cloud, surface, and atmospheric properties. A limited case (i.e., a granule of a cloud scene) may not be adequate to state so.

In this sentence, we refer to the account of the measurement errors that are included in the PDF of the measurement vector through their standard deviation. For the other types of errors that are currently hardly implementable in an operational algorithm due to computational cost reasons, using climatology can be a good solution. We complete the sentence about these others sources of uncertainties, section 6, line 775:

"The second step that consists of computing the uncertainties resulting from the non-retrieved parameters and from the forward model is more computationally expensive but could also be included. The uncertainties related to the non-retrieved parameters, in addition to the one related to measurement errors, have already been implemented since Collection 5 in MODIS operational algorithm through the computation of covariance matrix where Jacobian are derived from look-up table and was completed for Collection 6 (Platnick et al. 2017). Concerning the forward model errors, the method cannot be implemented as in this work in an operational algorithm because of the prohibitive computation time. A climatology based on several cases studies, depending of the type of clouds, land or ocean surface flag for example could be used in order to obtain a distribution of the errors according to the scene characteristics"

13. Typo and grammatical errors: Please proofread the main body of the manuscript again. I have found several grammatical errors, e.g., Lines 557, 571, and possibly more.
We apologize for these typos and grammatical errors and have again done a careful proofreading, hoping to have almost removed the typo and grammatical errors.

Reference

Saito, M., P. Yang, Y. Hu, X. Liu, N. Loeb, W. L. Smith Jr., and P. Minnis, (2019) An efficient method for microphysical property retrievals in vertically inhomogeneous marine water clouds using MODIS–CloudSat measurements, J. Geophys. Res. Atmos., 124, 2174-2193.