In this version, the main modifications include: 1. The use of two algorithms, Multi-Layer Perceptron (MLP) and Random Forest (RF). Different combinations of hyperparameters were tested to analyze their impact on model performance, and the best combination was selected. 2. The time range of the dataset is from August 2018 to July 2019. The ratio of various cloud fraction in the training set is 0: 0.16: 0.33: 0.5: 0.76: 0.83: 1= 5: 1: 1: 1: 1: 1: 5. After these adjustments, the model's performance showed significant improvement.

Refereer3 unresolved previous comments:

The manuscript primarily discusses the method, but the abstract and the content focuses only on the results. There is insufficient explanation about the method (machine learning and correction), its characteristics, and why it leads to improvements.

Answer: Based on your feedback, a more detailed introduction to machine learning methods has been provided, along with a comparison of the impact of different methods on retrieval.

Line 237-356

The manuscript lacks a discussion about the impact of different machine learning structures on the retrieval of cloud fraction. Please provide more results regarding this.

Answer: Based on your feedback, we have added comparisons of different machine learning algorithms and tested various combinations of hyperparameters to assess their impact on machine learning performance.

The data resolution of CloudSat and CALIPSO is not consistent with AGRI. It is crucial to discuss this dataset uncertainty and its impact on the retrieval.

Answer: The data resolution of CloudSat and CALIPSO is not consistent with that of AGRI, so when creating the training dataset, it is necessary to perform spatiotemporal matching of the data. Ensure that the time range of the matched data from the two instruments is within 15 minutes and the distance range is within 1.5 km. Additionally, within the AGRI pixel, at least two CloudSat and CALIPSO pixels should be covered. After matching, the cloud fraction detected by CloudSat and CALIPSO can better represent the actual cloud fraction within the AGRI pixel.

However, the errors in the matched dataset are unavoidable. The AGRI scanning method operates from left to right and top to bottom. Each complete scan of the full disk takes 15 minutes and generates a dataset. It is impossible to determine the exact moment of a specific point within the full disk. This limits the time range for matching datasets to within 15 minutes. However, in areas with higher wind speeds, clouds can move a significant distance within that 15-minute window. Therefore, errors arising from timing issues cannot be avoided.

Line 187-199

Refereer2 comments:

Given these results, I think that readers will need to be convinced that you have selected a reasonable algorithm with reasonable hyperparameters. Given the wide difference in results between the two algorithms so far, I don’t think you are reaching the ceiling of possible model skill given your data.

- I think you need to experiment with more hyperparameters for the random forest (and report those experiments here) and/or compare with some other baseline algorithms with reasonable hyperparameter choices.
- Given that your RF does not consistently outperform a 3-hidden-unit MLP (an incredibly small network), I’d be surprised if it significantly improves upon a linear regression model or a model that simply outputs the mean of the training data. These would be useful tests for you, and they would help to understand the behavior of your evaluation metrics.
- I expect you can achieve better performance with larger trees / forests, or with an MLP with more than 1 layer and more than 3 hidden units per layer. I think you should try both. Trying more hyperparameters for your RF implementation should be very simple.
- If you already have code that can train an LSTM on your data, it should not be too complicated to try something like a 3-layer MLP with, for example, 16 hidden units per layer. This is still a very simple network, but much more expressive than the 3-unit MLP you tried before. I also think a well-designed convolutional neural network would significantly outperform your current results.

Answer: Two algorithms, RF and MLP, were used, and through experimentation, the optimal hyperparameter combinations for both were determined.

In my previous review, I suggested you use data with broader seasonal coverage. You responded: “A single observation from FY4A AGRI, the northern and southern hemispheres contain data from different seasons, climates, and surface types. Therefore, the training dataset required for training the model does not need to cover a long period of time.” I disagree. While your data covers both hemispheres, there is a relationship between latitude, seasonality, and clouds. I see no reason not to use data from more times of year. If there is too much raw data, you can randomly sample the same amount of data you currently have, but from more seasons / years.

Answer: The time range of the dataset has been extended to one year, with 80% allocated for training and 20% for testing.

Line 224-232

Finally, you need to discuss the numbers achieved by other methods which attempt the same or similar tasks. For cloud detection, your highest recall (POD) is about 0.9, and your highest precision (1 – FAR) is about 0.82, but you also get values as low as 0.67. This seems low to me, especially when these numbers aren’t far from multilayer cloud detection numbers using similar inputs and 2B-CLDCLASS-LIDAR labels (Ding et al 2022. Multilayer detection is a (much) harder task. Still, it’s hard to compare, as you only report precision/recall (POD / FAR) while other papers seem to report accuracy. You should include accuracy alongside POD / FAR in your results. As for the ME / RMSE values, these seem high to me. ME is a little misleading when compared to RMSE, as your positive / negative errors cancel out, but the values are still quite high. Just because your method outperforms the operational product does not mean it is competitive, it just means the operational product is bad. The results (especially in light of the inconsistency between the MLP results and the RF results) suggest to me that you have more work to do when it comes to selecting and training the model.

Answer: For the characterization of cloud detection accuracy, it has been changed to directly display the confusion matrix (as shown in Figure 1 of the manuscript). Figure 1 includes recall rate, false alarm rate, and accuracy. The values for ME and RMSE have significantly decreased. The value of ME can be misleading due to cancellation of positive and negative values, so the Mean Absolute Error (MAE) has been added.

Line-by-line comments:

130: I can't find Qu et. al. in your citations. I stumbled upon this, and I have not checked all other citations.

Answer: I have added this reference to the citation.

139: I tried to look up the Hu et. al. paper, but the DOI links to a different author / title than the one you list.

Answer: doi:10.27248/d.cnki.gnjqc.2020.000625

307: You cite Quesada-Ruiz et. al. 2022, which is a very specific type of random forest algorithm called AFGRRF. Do you use their method, or a more general random forest algorithm? If it is the former case, you need to explicitly state that you use this method as well as briefly summarize the method (here you are only summarizing the general random forest algorithm). If instead you use a more general random forest, you should probably cite a seminal random forest paper, a survey paper, and/or a paper that applies (regular) random forests to similar data. I've noticed that Quesada-Ruiz et. al. have a publicly available implementation of AFGRRF in R. There are plenty of other publicly available implementations of random forests, so I'm curious why you use this method. It is important that you justify the choice of method in the text.

Answer: I used a more general random forest model. I have already made changes to the citations in the manuscript.

Line 254

316-332: Citations are needed throughout this section to back up your claims. This is especially true in the absence of experiments showing that your hyperparameter choices are competitive.

323-325: If you're using sqrt(M) for your Mtry parameter, shouldn't you be using 4 and 3 instead of 3 and 2? Sqrt(14) is closer to 4 than to 3, and sqrt(8) is closer to 3 than to 2.

Answer: This part of the content has been changed.

324: Which channels are not available at night? You should include this in Table 1.

Answer: The first six visible light channels have no values at night, meaning that channels with a central wavelength less than or equal to 2.225 are unavailable during nighttime.

Line 131-133