Response to Review

General comments

This paper presents useful insights into the quality of the two most commonly used global cloud climatologies, from MODIS and ISCCP. The results reported here should be of great interest to those using MODIS and ISCCP cloud products. The paper is generally well written, though a number of improvements are needed:

The introduction should be improved. It does a good job of citing the literature but lacks a clear organization and contains information which is not relevant to the manuscript, especially in the first two paragraphs. For example, it mentions contrails twice but the manuscript does not report results for contrails. It discusses cloud trends, but can a study looking at a single year tell us anything about trends? Many studies of global cloud cover have been performed over the years. The inclusion of so many different results makes it difficult to follow. The introduction should tell us what motivated this study and in what way the results reported here will improve our knowledge of global cirrus. Lines 51 – 78 exhibit better organization, describing the evolution of techniques to observe clouds and some of the difficulties in characterizing them.

Thank you for your valuable feedback. In response to your comments, we have made the improvements to the introduction. Regarding the mention of contrails and cloud trends: we have removed the reference to contrails. We have eliminated the cloud trends section. Furthermore, we have reorganized the introduction to more clearly present the motivation for this study.

A major reason for the discrepancy of high cloud cover between CALIOP and MODIS is the greater detection sensitivity of CALIOP to optically thin cirrus. This can be seen in the large discrepancy in cloud cover in the upper tropical troposphere, which is dominated by optically thin cirrus. It would be very useful to the community to examine the extent to which the superior detection sensitivity of CALIOP can explain the differences with MODIS. Repeating these comparisons after CALIOP cirrus is filtered by removing optically thin cirrus layers (optical depths less than 0.1, for example) could provide important insights.

This issue has been re-examined and is now addressed in the discussion section.

Specific comments

The abstract should be a little more clear on the motivation for this research. It says the usefulness of active sensor data in climatological studies is limited. Limited in what way and in what way are active sensor data useful for this study?

Corrected, as suggested.

Line 18: It is useful to point out that it has been known for decades that clouds are radiatively significant, but global net cloud forcing is now estimated to be closer to -20 W/m^2. It would be good to add a more recent estimate.

Corrected, as suggested.

Lines 35-36 state that cirrus occurrence is between 28% and 42%. A citation should be given for these estimates. A few lines later, cirrus cover is given as 17%, from Sassen (2009), based on

CALIOP observations. Why the discrepancy? Sassen is using the WMO definition of cirrus (optical depth less than 3.6 and above 440 hPa, as in Figure 1). Are other studies using a different definition, or have difficulty in determining cloud altitude?

Line 80: The Introduction discusses "high level clouds" as composed of Cirrus, Cirrocumulus, and Cirrostratus. It should be made clear at this point what is meant when the text says "cirrus".

Corrected, as suggested.

Line 82: MODIS is more properly referred to as a multi-band radiometer than a spectroradiometer

We agree with your suggestion and have revised the text accordingly. In the main text, we now refer to MODIS as a "multi-band radiometer" to better reflect its functionality. However, we retained "spectroradiometer" in the expansion of the acronym ("Moderate Resolution Imaging Spectroradiometer") to align with its official designation. We believe this approach balances accuracy and consistency with the instrument's official nomenclature.

Line 83: Was there a reason that 2015 was picked as the year of study? It should be pointed out somewhere in the paper, perhaps in the discussion in Section 5, that the ISCCP statistics presented in the paper are not representative of the early years of the ISCCP climatology, which relied on polar orbiting data from AVHRR rather than MODIS.

There was no specific reason for choosing 2015 for the study. We only required reasonably large sample of CALIPSO-MODIS match-ups for various seasons and locations, hence one full year of global observations. 2015 was an arbitrary choice.

The symbols used in the math expressions in lines 127 and 131 should be explained. I am not familiar with these.

Corrected, as suggested.

Line 143-145: This description of ISCCP should be moved to the Introduction. The Introduction should also discuss the significance of the ISCCP project and the resulting climatology, as the manuscript reports many results for ISCCP.

Corrected, as suggested.

Line 152-157 present two different definitions of 'cirrus'. When cirrus statistics are presented later, it should be made clear which of these definitions is being used. Further, regarding Figure 1: The current version (Version 4) of the CALIOP retrieval algorithm does not report optical depths larger than about 10 (see the right panel of Figure 3). Thus the optical depth reported by CALIOP will be less than 23 whether the actual cloud is Cirro-stratus or Deep Convection, as defined in Figure 1. The manuscript needs to be more clear on what classes of high cloud are included in the various occurrence statistics which are reported.

Line 166 mentions the CALIOP cloud subtype flag. In computing statistics in this manuscript, is this flag being used to define "cirrus" (category 6) as observed by CALIOP? It is not clear how the CALIPSO-based cloud mask is constructed.

Added, as suggested.

Lines 185-190: More details on spatial matching of CALIPSO and MODIS observations should be given. There are between one and three CALIOP lidar shots within each 1-km MODIS pixel, depending on the exact alinement of the two satellites. What criteria was used to define a 'match'? Also, the CALIPSO orbit was offset from the orbit of Aqua. At the equator, the view angle of MODIS to the CALIOP footprint at the Earth surface was about 17 degrees, which introduces parallax. Was this considered in the spatial matching? If so, how?

Clarified and explained, as requested.

Lines 205-230: I had a hard time remembering what all the 2- and 3-letter abbreviations for the statistical parameters mean (ROP, POD, OA, etc). Listing these in a table would be helpful.

Added, as suggested

Lines 237-248: If bootstrapping is really necessary to avoid biased results, more detail is needed here as I'm not aware that bootstrapping has ever been used in previous studies of global cloud cover. Bootstrapping is often applied in situations where the number of samples is small but in this case the number of samples seems large enough that bootstrapping is not necessary. Is the bootstrapping needed for estimating cloud cover, or only for the performance statistics (POD, FAR, etc). Please consider providing a simple example to illustrate the bias that bootstrapping avoids.

Thank you for your feedback. We appreciate your concern regarding the necessity of bootstrapping, especially considering the relatively large dataset. However, the primary purpose of bootstrapping in this context is to address the issue of class imbalance, which can significantly bias the performance evaluation of models, even when the number of samples is seemingly large enough. While bootstrapping is often applied in situations with small sample sizes, its application in this case is critical for ensuring a fair and accurate assessment of model performance.

To clarify the need for bootstrapping, we would like to provide a simple example illustrating the potential bias in performance evaluation when class imbalance is not accounted for. Consider a dataset with 100 observations, where 15 represent cirrus clouds (positive class) and 85 represent non-cirrus clouds (negative class). In such an imbalanced dataset, a naive model that predicts only the majority class (non-cirrus) can achieve high overall accuracy (OA) by simply classifying all instances as non-cirrus. In this case, the model's accuracy is 85% (OA = 85%), as it correctly classifies all negative cases but entirely ignores the minority class (cirrus clouds). This results in a misleadingly high accuracy metric, which does not reflect the model's true performance, especially in detecting the minority class.

When bootstrapping is applied, however, we resample the dataset with replacement to create a balanced set of positive and negative instances (e.g., 15 cirrus and 15 non-cirrus). Using this resampled dataset, the same naive model achieves only 50% accuracy (OA = 50%) because it is now evaluated on a more balanced distribution of both classes. This exposes the model's true limitations in detecting the minority class (cirrus clouds), which would otherwise be overlooked in the original, imbalanced dataset.

The key benefit of bootstrapping in this context is its ability to reduce the bias caused by the dominance of the majority class in the original dataset. Without bootstrapping, performance metrics like POD for cirrus clouds could be skewed, as the model might appear to perform well overall while failing to detect cirrus clouds effectively. By resampling the dataset to balance the classes, bootstrapping ensures that both classes are fairly represented in the evaluation, providing a more accurate picture of the model's true performance, especially for detecting rare events like cirrus clouds.

Therefore, bootstrapping is not only necessary for improving the reliability of performance statistics, but it also helps avoid the bias of under-representing the minority class. This results in a more realistic evaluation of the model's capabilities, ensuring that metrics reflect the model's ability to detect both the majority and minority classes fairly.

Lines 258-259: These numbers for cirrus coverage are lower than I would expect from CALIOP observations and the difference between day and night is larger than I would expect. How is "cirrus cloud" being defined here? Is additional filtering being done besides CAD score greater than 80? Is bootstrapping being used to compute cloud cover here?

Thank you for your insightful comment. Cirrus clouds are defined here as Category 6 in the CALIPSO cloud class. No additional filtering was applied beyond the CAD score criteria, and bootstrap methods were not utilized to compute cloud coverage. However, we note that similar results have been reported in the literature, supporting the consistency of these findings. According to (Sassen et al., 2008), the total frequency of cirrus clouds from 15 June 2006 to 15 June 2007 was reported as 16.7%, compared to 18.7% observed in our study for 2015. Nevertheless, the day-night difference observed in their study was smaller than in ours, with values of 15.2% during the day and 18.3% at night, compared to 13.2% and 23.3%, respectively, in our analysis. As added in the manuscript, the Cirrus cloud mask (Ci) was generated by applying a condition that classified each 4-degree pixel based on the proportion of observations identified as Cirrus. Specifically, the number of Cirrus observations (nCi) and non-Cirrus observations (nNONCi) within each pixel were counted. The percentage of Cirrus observations (CiCoverage) for a given pixel was calculated using the formula:

CiCoverage=nCi/(nCi+nNONCi)*100

Line 265: Regarding figure 3, I find cumulative distributions useful but difficult to interpret without also showing the frequency of occurrence, which in this case would show the difference in the day and night

cumulative distributions to be due to the detection of many more low optical depth clouds at night. I suggest adding a figure showing the two frequency-of-occurrence distributions. By the way, the paper should point out that the major reason for the CALIOP day-night difference in cirrus occurrence is better detection sensitivity at night.

Caption of Table 2: what does "precluded the use of the test" mean? Which test?

By the phrase "precluded the use of the test," we meant that the specific indicators in question reach values that, in our judgment, make it impossible to use these tests directly for identifying Cirrus cloud masks. More specifically, the values of these indicators are such that they do not provide reliable or clear discrimination in the context of Cirrus clouds, thus preventing their straightforward application for this purpose.

Line 290: what does "physical properties of the respective radiation range" mean? Please reword or explain.

By 'respective radiation range,' we are referring to the different wavelengths of radiation used by the individual channels of the instrument. The variation in cirrus detection statistics across latitudes can be attributed to factors such as varying illumination conditions due to the Earth's axial tilt, as well as the presence of phenomena like the polar day and night. These conditions affect the effectiveness of each channel and its corresponding wavelength range, meaning that not all channels can be applied uniformly or with the same level of effectiveness across different latitudes. This part in the manuscript has been revised.

Figure 4: I can't tell the difference between the curves for ATC and ISCCP3.6 here, or between BT6.7 and BT1.38. Switching to colored lines would make this more legible. I have similar difficulties with Figure 5.

Corrected, as suggested.

Summary: Due to the importance of the ISCCP cloud climatology, results related to the ISCCP evaluation should be summarized in Section 5.

Added, as suggested.

Figure 6: It seems the choice of color bar could be improved for 6i and 6j.

Thank you for your suggestion regarding the color bar for panels 6i and 6j. While the range of classes may appear broad, we chose to divide the coefficient into three classes based on the values it assumes in our analysis. This division is sufficient to capture the variability observed in the data while maintaining clarity and interpretability of the figure. However, we are open to further refining the color bar if additional feedback suggests that a more detailed classification would enhance understanding.

Line 374: The "CALIOP data cirrus mask" isn't really described in Section 3. Some description is necessary (in Section 3) as there are many ways the data might be used to create a mask.

Added, as suggested.

Technical corrections

Line 158: "for example CALIOP" rather than "in example CALIOP"

Corrected, as suggested.

Line 226: I don't think PE is ever defined

Corrected, as suggested.

Line 269: I think maybe "4.20 at night" is a typo and should be "0.42 at night"?

After careful consideration, we have determined that the referenced CALIOP "all cloud" COT values are not directly meaningful to our study. To address this, we have removed the sentence as suggested.

Figures 2 and 3: the captions should state that these statistics are based on CALIOP data.

Corrected, as suggested.

Line 273: "table" should be spelled out (not tab.2), here and elsewhere

Corrected, as suggested.

Line 297: Should be “A similar pattern ...”

Corrected, as suggested.

Line 302: I think “Figure 5” here should be Figure 4

Corrected, as suggested.

Line 326: “notably the ATC test”

Corrected, as suggested.

Line 329: Should be “An increasing number of ...”

Corrected, as suggested.

Line 334: “IGBP groups were aggregated ...”

Corrected, as suggested.

Line 374: “The CALIOP data mask ...” would be better

Corrected, as suggested.