Evaluation

The focus of this analysis was to evaluate the accuracies of commonly used infrared-based deep convective cloud (DCC) detection methods, with additional comparison to ISCCP scheme for cloud classification and detection. These methods were evaluated based on how well they detected CloudSat-CALIPSO identified DCCs. The main finding was that the brightness temperature difference method performed the best out of the three infrared-based detection methods, and that a fixed 11-micron IR threshold should not be used for detecting DCCs. This study is important given the robust literature on deep convective system climatologies from a passive remote sensing perspective that often use single 11-micron IR thresholds for distinguishing DCCs. Furthermore, the data sets used and accuracy tests done in this study seem to be thorough, and limitations were addressed in great detail. I recommend the manuscript for publication after addressing the comments below.

Specific Comments

L149: What was the reasoning to use the -10 K threshold as opposed to the stricter -5 K? Was it to ensure sufficient sampling? And was the -10 K threshold applied as a filter before seeing if CloudSat-CALIPSO flagged any DCCs?

L211-212: Could you please expand upon what exactly was repeated? Do you subsample the DCC sample as well as the non-DCC sample? Since you repeat it 1000 times, do each of your subsamples contain unique cases (i.e. they don't contain the same cases as the other iterations?)

L234-236: Interesting, why do you think the TROPO approach improved while the other methods were less accurate for a smaller domain?

L243—"derived accuracy measures for zones at 5° latitude": How did you do this? Did you calculate confusion matrices for a range of thresholds and then define which one was optimal? If so, how did you select the range of thresholds?

L308-310: Please clarify explanations, specifying what "change" and "increase" / "decrease" is referencing (i.e. what are you comparing to?). Also, the "two exceptions" are not clearly defined—exceptions to what?

L315—"Antarctica": I would suggest specifying that the DCC occurrence is over the Southern Ocean (not Antarctica). Also, do the authors think that high DCC frequency over this region is real? Or is the fixed threshold perhaps detecting another type of high cloud? It would be beneficial if the authors could comment on this point, especially if it is an issue with using a fixed threshold because it can misclassify DCCs.

L316—sentence beginning with "The latter finding": This sentence might be more impactful if it was rephrased to explain whether an increased DCC frequency over the mountainous regions is valid.

L344-345: Since this whole section does not discuss the accuracy of CloudSat-CALIPSO, perhaps restructure this paragraph and the next by joining the paragraphs. Also, what other inaccuracies were there other than the potential misclassification of Ns that was then stated not be a significant issue in L355?

L348-349: Are there references stating that CloudSat is overrepresenting As and Ns clouds? It would be particularly interesting to share those studies since CloudSat and CALIPSO are generally considered to have a better representation of clouds than ISCCP.

L357-363: Is it possible to quantify the significance of the inaccuracies of misclassifying clouds due to multi-layer scenarios in each of the methods, including in the ISCCP data? If not, do you have a qualitative sense of how significant the misclassification is (i.e. does misclassifying multi-layer cloud scenarios lead to the most inaccuracies in detecting DCCs)?

L430-431: It is unclear what the authors mean in this sentence. Are the authors suggesting that it is likely that DCCs occur as frequently in higher latitudes as they do in the ITCZ? Also, what is the "misclassification of cold Ns and As" referring to: the initial $Tb_{WV} - Tb_{IR} > -10$ K screening, or CloudSat-CALIPSO potentially overestimating As and Ns clouds?

L469-470: What specifically are the uncertainties in CloudSat-CALIPSO cloud classification, other than the As/Ns overrepresentation that was deemed insignificant in L355?

Technical Comments

L111: "referring" instead of "refeering"

L113: It is unclear what the text in parentheses is referring to

L151-153—sentence beginning with "Aqua…": I think this information should go at the beginning of the section before introducing the individual data products to give some context on the observations.

L178-179: what does "the first tropopause" mean?

L246-247: If I understand correctly, this statement can be generalized to say that overall accuracy increases by using multiple TB thresholds that change with latitude?

L285—sentence beginning with "It is important…": Perhaps rephrase as it is unclear what is meant by this sentence.

L296: omit "to a hemisphere; more specifically," as this explanation is not quite accurate.

L340: Discussion section is section 4 (not 5)

Figure captions: Perhaps specify what each row and column correspond to, as well as the colors of each line, to make it quicker to study each figure.