Reviewer #2

We thank the reviewer for the insightful comments (**bold font**) and we have replied to each of the comments and queries below each comment (regular font) and modifications to the manuscript are in *italic font*.

In this paper AIRS measurements are combined with Cloudsat observations to obtain statistics of the variation of cloud types within the AIRS field of view. These are then used to study the variation of cloud properties per cloud scene and the sensitivity of ice cloud properties retrievals to the variation of cloud types within the pixels. The study is interesting and the methods are generally sound. However, some issues with the presentation and the analysis need to be addressed before the paper can be accepted for publication. Below the major and minor comments are listed:
1) The Cloudsat cloud types are used. However, these type names are not very quantitative. Please indicate how the different types are defined and summarize how they are derived from Cloudsat observations. The reader should not have to dig through other papers to be able to understand the data presented here. For instance, what are the altitude boundaries distinguishing low, middle and high clouds? It is also not clear to me how cumulus and stratus can be distinguished using a single radar profile.

According to Sassen and Wang (2005), the cloud classification occurs in two steps. First, a clustering analysis is performed to group cloud profiles into cloud clusters. Secondly, classification methods are used to classify clouds into different cloud types. The decision trees guiding the classification are complex and based on 23 variables derived from the clustering analysis of the first stage. Conditions upon geometric quantities such as cloud base, top, and horizontal extents are present in decision trees (page 32 to 35 in Sassen and Wang, 2005).

One of our responses to reviewer #1 is relevant here and we added text to the manuscript (repeated from response to reviewer #1): "*The CloudSat 2B-CLDCLASS product is used in this work and the algorithm is described in Sassen and Wang (2005, 2008). As summarized in Sassen and Wang (2008) and previous works, the algorithm uses methods developed from ground-based multiple remote sensors that have been tested against surface observer-based cloud typing reports. The cloud classification occurs in two steps. First, a clustering analysis is performed to group cloud profiles into cloud clusters. Secondly, classification methods are used to classify clouds into different cloud types. The decision trees guiding the classification are complex and are based on 23 variables derived from the clustering analysis of the first stage. Conditions upon geometric quantities such as cloud base, top, and horizontal extents are present in decision trees (Sassen and Wang, 2005).*"

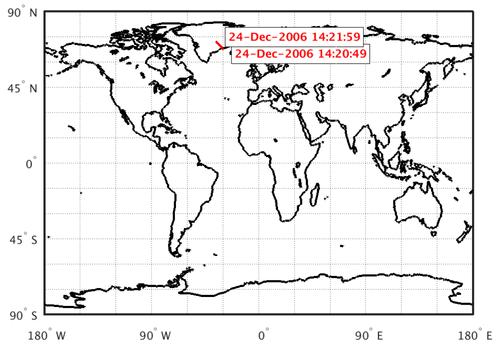
In the conclusions, we describe the results of Wang et al. (2016) where comparisons of CloudSat cloud types as used here in this work are compared to ISCCP-like categories derived from the MODIS imager, which could be used in place of the CloudSat cloud typing. For many cloud types, the detection is similar between passive and active; please see Wang et al. (2016) for specifics.

2) Please be consistent with cloud type names throughout the paper. For example, sometimes “As” is used and sometimes “AlSt”. Sometimes “clear sky” is used, sometimes “nc”. Also, mixtures are denoted with commas, although clear-sky is left out in this notation. I suggest to indicate mixtures of cloud types and clear sky consistently, e.g., “As, nc”. To give an example how this is confusing: Looking at Fig. 6, It was unclear how cumulus can have a length scale of 400 km, but I guess that’s possible because clear-sky is mixed in. This can be made clearer by naming this mixture “Cu, nc” (This then excludes pure “Cu” cases. If these are included too, then I suggest using something like “Cu + Cu, ns”).

We have checked the notation throughout the revised manuscript and fixed any inconsistencies in the labeling. We intended to follow the notation cumulus (Cu), stratocumulus (Sc), stratus (St), altocumulus (Ac), altostratus (As), nimbostratus (Ns), cirrus (Ci), deep convective (Dc) clouds and a ninth classification of clear sky designated no cloud (nc). In addition to text changes, we corrected figure 1, 5, 6 and 7.

Cumulus lengths reported in Figure 6 are for pure Cu and not a blend of clear and Cu (Cu,nc). Cumulus clouds of any length are possible according to the decision tree of low cloud classifier (page 34 of Sassen and Wang, 2005) when a cluster cloud fraction is smaller than 0.25. Sassen and Wang (2005) do not indicate that a maximum cluster length is prescribed, so long cumulus clouds might result from the unlikely situation where a long cumulus cloud is part of a cluster with cloud fraction less than 0.25.

The fact is that we observed long clouds in each cloud category of the 2B-CLDCLASS CloudSat product. One may quibble about this categorization but the algorithm most closely associates that cloud with Cu. For example, we found that the longest cumulus cloud is located in granule 03498 of the 2B-CLDCLASS product, 2006358131321_03498_CS_2B-CLDCLASS_GRANULE_P_R04_E02.hdf using the cloud type defined at the cloud top as described in the manuscript. The very long Cu cloud starts at profile # 25681 and ends at profile # 26118. It therefore has length of 438 CloudSat profiles or about 482 km. The location of this long cumulus cloud is shown in the figure below:



This fact emphasizes that no classification is perfect, that some misclassification is inevitable and that some cloud physical states are ambiguous for a given classification scheme. We used the 2B-CLDCLASS CloudSat product for our study but a comprehensive critical analysis of this product is beyond the scope of our paper.

3) Two different footprint sizes are considered, namely the AIRS/AMSU footprint of ∼45 km and the AIRS footprint of ∼15 km. It is a bit unclear throughout the paper which analysis is applied to which of the two footprints. In any case, the two different scales need to be addressed more consistently throughout the paper (in addition to the abstract). For example, Section 3.1 focuses on the AMSU footprint but not on the AIRS footprint. If I'm not mistaken, the AIRS cloud retrievals are performed on the AIRS ∼15 km footprint, so the last section focuses only on that scale (If so, state this clearly in the paper.) Please expand the discussion in section 3.1 to the AIRS footprint as well and add a figure similar to figure 1 for the AIRS footprint. Furthermore, I assume Fig. 2 is showing statistics for the AMSU footprint. If this is true, please state that in the paper. I suggest adding figures similar to Fig 2. for the AIRS footprint, or at least discussing the (lack of) differences between AIRS and AMSU global distributions. In fig. 3, 4 and 6, indicate the size of AIRS and AMSU pixels.

We have added a second panel to Fig. 1 for the AIRS FOV scale and have labeled each panel clearly with the two scales depicted. Note that the number of scenes that explains >90% of the observed scenes is different between the two scales.

We describe in section 2 that the analysis is first conducted at the AMSU FOR spatial scale and that this scale is 45 km. We replaced the initial notation "AIRS/AMSU FOR" with "AMSU FOR" throughout the manuscript to avoid confusion between the two different resolutions and their corresponding scales. After discussing scene observations at this scale in section 3.1, we extend the analysis as a function of resolution in sections 3.2 and 3.3. After showing the cloud scene resolution dependence in Section 3, we now describe the dependence of AIRS cloud properties on the different types of cloud scene in section 4. We added a sentence at the beginning of section 4, to emphasize that the resolution used for cloud scenes determination, in this section is AIRS, i.e. about 15 km: "*In this section, the scenes are determined at the AIRS FOV resolution (approximately 15 km)*."

We agree with the referee and added a sentence in the caption of figure 2 to specify the resolution scale: "*These scenes were observed at the AMSU FOV resolution (~45 km)*." We note that the AIRS version of Figure 2 is nearly identical and are not included. We have added the following text to the Fig. 2 caption: "*Similar plots of the AIRS FOV resolution (~15 km) are nearly identical (not shown)*."

We think it is a good idea to indicate the AIRS and AMSU resolutions in figures 3 and 6. New versions of these figures with vertical lines indicating both resolutions are added to the revised manuscript.

4) If my interpretation is correct, one goal is to compare AIRS cloud retrievals for cases with a single cloud type in the footprint with clear sky mixed in versus without clear sky. However, a comparison is made between 1) cases with a single cloud type in the footprint and no clear sky and 2) cases with a single cloud type in the footprint either with and without clear sky. Thus, the case 2 set also contains the case 1 set in addition to the mixture of clear sky and a single cloud type. Therefore, the differences in properties listed in the paper are not representing the differences between cases without clear sky mixed in versus those with clear sky mixed in. From the information in the paper, we cannot deduce the relative number of cases per cloud type with versus without clear sky mixed in. If, for example, the number of cases without clear sky mixed in is much larger than those with clear sky, then the small differences shown between table 2 and 3 and figures 7 and 8 are surely expected. I suggest the following: 1) Include a table or figure showing the relative number of single type cases with and without clear sky mixed in. 2) For the single types listed in table 3 and figure 8 include only cases that also include clear sky. (Or add a table and figure showing this, leaving table 3 and fig. 8 as is.) To include a complete comparison between table 2 and 3 and figures 7 and 8, I suggest to also include the deep convection type (mixed with clear sky) in table 3 and figure 8. 3) Adjust the discussion of the differences accordingly.

These are very helpful comments by the reviewer and we have made some significant changes in the manuscript to make the categorization of cloud scenes, and the differences in the cloud phase and ice microphysical retrievals clearer.

First, we have added a new Table 2 and a new (brief) Section 4.1 that summarizes the new breakdown in cloud scenes: "*Table 2 summarizes five types of scenes at the 15-km AIRS FOV scale: (i) clear sky, (ii) cloudy sky with one cloud type, (iii) partly cloudy sky with one cloud type, (iv), cloudy sky with multiple cloud types, and (v) partly cloudy sky with multiple cloud types. The raw counts and the relative percentages for the two-year observing period are shown. The dominance of clear sky (~31%) at 15-km is apparent and is consistent with an absence of thin cloud features in the 2B-CLDCLASS data set. Cloudy sky cloud scenes amount to less than 1% of all observed cloud scenes, with a vast majority as partly cloudy, and a majority of those with only one cloud type. This value is higher than that reported by Yue et al. (2011) for which the 45-km AMSU scale was used but is consistent with the factor of 3 difference in scale*."

Second, we have reworked the original Tables 2 and 3 into four different tables (3-6) that show the ice cloud property statistics for the scene classification (2), (3), (4) and (5) listed above. In the case of (1), we have included clear sky in (2).

Third, we have similarly modified Figures 7 and 8 in the original manuscript to Figures 7-10 in the revised manuscript. They follow exactly the same scene classification divisions illustrated in Table 2.

We have added text in the manuscript that describe these new tables and figures and more clearly delineate the performance of cloud phase and ice cloud property retrievals by whether the cloud scene is completely versus partly cloudy, and whether the cloud scene contains a single cloud type or multiple cloud types. We refer the reviewer to the revised version of the manuscript to see every change made, but the most substantial for Section 4.2 and 4.3 follow:

The first paragraph of Section 4.2 that addresses cloudy sky with one cloud type now reads as: "*The occurrence frequency histogram of the sum of all thermodynamic phase tests is shown for cloudy sky with one cloud type in Fig. 7. While these clouds account for a small percentage of the total number of AIRS FOVs, homogeneous cloud scenes serve as an ideal point of reference for establishing cloud phase sensitivity benchmarks. Overall, there is strong differentiation in the cloud thermodynamic phase among cloud scenes with single cloud types. Ice tests dominate Ci, Ns, Dc, and As, while liquid and undetermined tests dominate Ac, Sc, and Cu*."

The fourth paragraph of Section 4.2 that addresses partly cloudy sky with one cloud type reads as: "*The occurrence frequencies of cloud phase for partly cloudy sky with one cloud type are shown in Fig. 8. The biggest change is the relative ordering of the ranks among cloud scene types between Figs. 7 and 8. Ac is now more common than As, Dc ranks higher than Ci as it is typically horizontally more extensive (Miller et al. 2014; Guillaume et al. 2018). There are more subtle changes in the cloud phase histograms that are consistent with partly cloudy sky. A weaker spectral signature for partly cloudy scenes results in slightly greater counts of unknown phase and also subtle shifts in liquid and ice phase tests in Fig. 8 compared to Fig. 7. In the Ac cloud scene histograms, there is a small but discernible increase in ice tests in Fig. 8 compared to Fig. 7. Horizontally heterogeneous Ac appears to have more frequent ice detection than horizontally homogeneous Ac*."

There is a new and abbreviated fifth paragraph: "*The nine most frequent partly cloudy scenes with multiple cloud types are shown in Fig. 10. The biggest change is the relative ordering of the ranks among cloud scene types between Figs. 9 and 10. Furthermore, there are additional (yet subtle) changes in the phase test histograms for the cloud scenes that are common between Figs. 9 and 10*."

The fifth paragraph of Section 4.2 that addresses partly cloudy sky with multiple cloud types reads as: "*The nine most frequent partly cloudy scenes with multiple cloud types are shown in Fig. 10. As with the differences between Figs. 7 and 8, the biggest change is the relative ordering of the ranks among cloud scene types between Figs. 9 and 10. Furthermore, there are additional (yet subtle) changes in the phase test histograms for the cloud scenes that are common between Figs. 9 and 10*."

The first paragraph of Section 4.3 that addresses cloudy sky with single cloud type ice property retrievals reads: "*The mean ice cloud property retrievals are summarized in Table 3 for cloudy sky with one cloud type only for the ice only portions of the cloud phase histograms depicted in Fig. 7. Scenes identified as clear sky exhibit properties of a small population of thin cirrus detected by AIRS (Fig. 7) with mean values of* τ_i *=0.78 and* r_{ei}*=20.9* μ*m (Table 3). The AKs are notably lower and the relative error for* τ_i *is higher than other cloud scenes. The corresponding cloud phase histograms are shown in Fig. 8.*"

The sixth paragraph of Section 4.3 that addresses partly cloudy sky with single cloud type ice property retrievals reads: “*The mean ice cloud property retrievals are summarized in Table 4 for partly cloudy sky with one cloud type with cloud phase histograms depicted in Fig. 8. The biggest difference between Tables 3 and 4 is the relative frequency of occurrence with large differences between cloud scenes with or without clear sky. Another significant change is an overall reduction in AKs and magnitude of* τ_i*, with an increase in* χ^2 *in Table 4, consistent with partly cloudy scenes. The changes in* r_{ei} *AKs, magnitudes, and error estimates between Tables 3 and 4 are smaller than those for* τ_i*. Overall, the differences between Tables 3 and 4 are reassuring in that the AIRS retrieval is responding to partly cloudy scenes by reducing information content and the magnitude of* τ_i*, while* χ^2 *residuals are increasing somewhat*.”

The second to last paragraph summarizes the differences between Tables 5 and 6: “*The ice cloud property retrievals for cloud scenes that contain multiple cloud types are summarized in Table 5 for cloudy scenes and Table 6 for partly cloudy scenes. These tables list the nine most frequent cloud scene types as depicted in Figures 9 and 10. Seven of the nine cloud scenes are common between Tables 5 and 6. There is a general tendency for reductions of* τ_i*, increases in % relative error, and slight reductions in AKs in Table 6 for the seven common cloud scenes in Table 5. Changes in* r_{ei} *related variables are smaller than changes in* τ_i *related variables*.”

The last paragraph synthesizes Tables 3-6 and is: “*To summarize Tables 3-6, larger differences in ice cloud property retrievals are found between different cloud types than between cloudy and partly cloudy scenes. However, the differences between cloud scene types are the sharpest for the subset of cloudy scenes with one cloud type (Table 3). The AIRS cloud property retrievals are not greatly impacted by mixtures of cloud types within the AIRS footprint, and ice cloud property differences among cloud scenes are broadly consistent with the expected performance of infrared retrievals among these cloud types*.”

Minor comments:

At the start of Page 11 it is stated that “the AIRS footprint (∼ 15 km) is commensurate with the dimension of a single cloud so that the most frequent observations involve the characteristics of one cloud.” What is meant by this statement? Many clouds, for example cumulus, have typical scales much smaller than that.

We agree that this statement could be confusing and we have removed the following from the revised manuscript: “*These differences between single and mixed cloud scenes can be explained by the fact that the AIRS footprint (~ 15 km) is commensurate with the dimension of a single cloud so that the most frequent observations involve the characteristics of one cloud. Cloud length statistics, such as median lengths and median absolute deviations calculated from the data shown in Fig. 6, corroborate this statement*.”

Table 1 and 2: Mean optical thicknesses are given here, but these are biased low because AIRS is not sensitive to any variation past an optical thickness of ∼5. Please make this clear in the text. The true mean optical thickness of most types will be larger than the AIRS-retrieved means shown here.

We added the following sentence line 12 of page 5: "*The AIRS retrieval sample includes nearly all ice clouds with* $\tau_i > 0.1$*, while the maximum values of* τ_i *asymptote to values around 6-8 (e.g. Kahn et al., 2015).*"

We discovered an error in Table 1, which are now fixed, and now the values on the third row actually report the median absolute deviation.

I think the caption of figure 4 is missing the word "scenes" at end of first sentence. It is noted in the text that Fig. 4b and 4d are similar to Fig. 3, but it is not directly clear what the difference in the calculations are. Please explain.

We added the word "*scene*" at the end of the first sentence of the figure 4 caption. Figure 3 shows the number of scenes observed at different FOR scale whereas Figure 4b and 4d represents the number of scenes as a function of a cloud scale (not as a function of FOR scale).

I noted that the statistics listed in Table 3 are different than in Fig. 1, and this might be due to the different scales of AIRS vs AMSU. (See main comments.) However, for the ice properties to be retrieved, the phase needs to be identified as ice, so are these proportions for ice clouds only? Is that also explaining the difference in statistics compared to figure 1? Please explain in the text.

The referee is correct in that these are only for ice clouds (the blue bars in Figures 7-10). The statistics in Table 3 are obtained at the AIRS FOR whereas statistics in Figure 1 were obtained at the AMSU FOR.