Reviewer #1

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*.

The manuscript attempts to answer the following questions: How many different cloud types co-exist within a particular area? What cloud type mixtures are more prevalent? How do answers to the above two questions depend on area size? (side question that emerges: at what spatial scale does one encounter the greatest diversity of distinct cloud type mixtures?). These all sound kind of philosophical questions, but the authors find practical relevance (at least for the first question) for AIRS (and AMSU) scales cloud retrievals. The link to AIRS allows the authors to make one the major compromises of the study: only cloud type identification of the topmost cloudy layer matters because that's where AIRS is most sensitive even though the data source identifying cloud type provides vertical profile information. The other major compromise is that when identifying cloud mixtures, the frequency of occurrence of each cloud type does not matter, in other words cloud mixtures consisting of the same cloud types are treated as equivalent even if the contributions of a cloud type are different. These two simplifications, along with an additional one where the spatial arrangement of the cloud types is ignored allow the authors to reduce the dimensionality of the problem and make the analysis tractable.

We appreciate the reviewer's synthesis of the paper and agree these are the major thrusts of the first aspect of the paper.

This is overall quite a difficult paper to read, but I find the results of the first part quite fascinating (I was less excited about the implications for AIRS retrievals–although I understand that these findings are important for understanding the quality of the AIRS retrievals), so I recommend acceptance of the article to AMT. As you can see below, I have some inquiries some of which are also of the philosophical kind I'd like the authors to consider.

We have made some edits and changes that follow from the reviewer suggestions that we hope make the manuscript more readable. We do appreciate the reviewer's point of view regarding how this paper may be viewed in two distinct pieces. However, our end goal of this work was to show practical relevance to the cloud scene variability and ultimately establish why it is important to describe cloud scenes on a pixel-by pixel basis. The reasons are hopefully clearer in the revised version as we have more carefully documented AIRS retrievals of cloud phase and ice cloud properties as a function of whether the scene contains one cloud type or multiple cloud types, or whether the scene is completely or partly cloudy, instead of reporting them in a slightly convoluted manner. These changes are closely coupled to those arising from reviewer #2's comments (please refer to replies to reviewer #2 for further detail).

– What does the cloud type from 2B-CLDCLASS mean? The names of cloud types are the same as the ones used by surface observers, but are they related? Some description of the physical meaning of the cloud types given their method of identification by the 2B-CLDCLASS algorithm is needed.

The 2B-CLDCLASS algorithm is described in Sassen and Wang (2005, 2008) that is referenced in the manuscript. As stated in Sassen and Wang (2008), the algorithm is based on earlier work by the same authors and combines the measurements of ground-based multiple remote sensors. They report having tested the results against surface observer cloud reports. We have included some additional text at the beginning of Section 2.2 to clarify:

"*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. Geometric quantities such as cloud base, top, and horizontal extents are present in decision trees (Sassen and Wang, 2005).*"

I'm sure the authors are aware that another version of the product currently exists, 2B-CLDLASS-LIDAR where the CALIPSO lidar assists in the identification of the cloud type. Why was this newer product not used? (I suspect the authors may have started the work before this product was released). If the authors were to use 2B-CLDCLASS-LIDAR and the results changed in a major way, how would that undermine the fundamentals and motivation for the first part of the study?

This is a good question and a fair one to ask. We are motivated by the results of Kahn et al. (2018) that suggests larger particle sizes in convective clouds compared to thin cirrus. The 2B-CLDCLASS product is better suited for differentiating cloud types other than small particle thin cirrus, in which 2B-CLDCLASS-LIDAR would excel. While AIRS is very sensitive to thin cirrus, the sensitivity of AIRS ice cloud particle size most strongly responds to clouds around an optical thickness on the order of 1.0. If we used 2B-CLDCLASS-LIDAR, the statistics would be weighted to the detection of vast areas of thin cirrus in layers above optically thicker clouds and elsewhere in absence of other cloud types. The Ci classification dominates in the 2B-CLDCLASS-LIDAR data set and will blur the signals of cumulus and deep convective cloud types capped by thin cirrus. As we describe in the paper, the simplifications required to make the approach tractable require us to quantify cloud type at cloud top, and thus we would lose the discriminatory ability for clouds that occur just below a thin layer of cirrus. We expect that the mixtures would be profoundly different in 2B-CLDCLASS-LIDAR, with much less ability to demonstrate AIRS' skill at obtaining larger particle sizes at convective cloud tops (Kahn et al. 2018). Ultimately, we argue that 2B-CLDCLASS is a more appropriate tool for this work. We have included some additional text in Section 2.2 to clarify:

"*Lastly, the results of Kahn et al. (2018) suggest larger ice cloud particle sizes occur at convective cloud tops compared to thin cirrus at the same cloud top temperature. Given the key assumption of cloud typing only at cloud top, the 2B-CLDCLASS product is better suited for identifying convective clouds in AIRS apart from stratiform clouds, the latter of which are dominant in 2B-CLDCLASS-LIDAR. If 2B-CLDCLASS-LIDAR was used, the statistics would be weighted towards the detection of vast areas of cirrus in thin layers above and in proximity to convective clouds. The Ci classification dominates in 2B-CLDCLASS-LIDAR at cloud top and will blur the signals of underlying cumulus and deep convective cloud types that are capped by thin cirrus*."

What if a completely different cloud type product was used, e.g., based on passive satellite observations where cloud type is identified by location in a cloud-top- pressure/cloud-optical-thickness joint histogram (the authors briefly touch on this in the last paragraph, but only with regard to the AIRS application – I'm more interested in the cloud scene climatology aspects)?

As the reviewer certainly knows, how one goes about "typing clouds" is not a settled research topic and is sensitive to the instrument sampling, radiance sensitivity, wavelength, underlying assumptions, and so forth. In the conclusions, we touch on 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. We expect that the CloudSat radar will have more skill in discriminating convective clouds from stratiform clouds than passive sensors, as these two types of clouds show strong differences in the AIRS microphysical retrievals. For many cloud types, the detection is similar between passive and active; please see Wang et al. (2016) for specifics.

– It seems to me that the results depend completely on how frequently 2B-CLDCLASS identifies certain cloud types based on its internal definitions. Yes, the authors do not often find mixtures containing stratus (St) simply because St is extremely rare in 2B-CLDCLASS, probably unrealistically so given other methods identifying St (I mean, cloud types will always be loosely defined).

We agree. Please refer above for our generalized perspective. It is well established that 2B-CLDCLASS contains very little stratus because of ground clutter in the bottom 3-4 bins.

I think one figure that the paper needs to include is the global frequency of the different cloud types according to 2B-CLDCLASS at its native resolution. This will give immediately clues on why certain cloud type mixtures (scenes) will be rare right off the bat (the authors kind of bring this this up already in some instances, e.g., p. 6, line 4). With DC, Cu, and St being rare according to 2B-CLDCLASS, one would expect that scenes containing those will also be rare.

With regard to plan view maps and zonal averaged plots of cloud type frequencies from 2B-CLDCLASS, these have been reported in the literature, in particular we are referring to Sassen and Wang (2008), their Figure 1 for the plan view and their Figure 2 for the zonal average. We do not see a need to repeat these results in this paper. With regard to the relative histogram counts of cloud type as a function of length scale, these are depicted in Figure 6 in the manuscript. The percentages of cloud type and cloud scene frequencies are also reported in Tables 3-6 in the revised manuscript.

We have added the following statement in Section 2.2 to refer the reader for more detail on cloud type frequencies: "*Plan view and zonal average frequencies of 2B-CLDCLASS cloud types at its native resolution are reported in Sassen and Wang (2008)*."

It is unfortunate that the abbreviation for certain cloud types changes throughout the text, tables, and figures: As becomes AlSt, Ac becomes AlCu, Ci becomes ci, DC becomes Dc, and so forth. Please fix and make consistent throughout!

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.

I don't understand panel d in Fig. 2. Whatever it depicts, it does not appear to have a very interesting pattern!

We agree that these are confusing and have removed panels b and d in figure 2 as they are not key pieces of information for the manuscript.

I recognize that the authors make a valiant effort in section 3.3, but that part of the paper remains a hard read. In this section, line 8 of p. 8 indicates that 200 possible mixed scenes were identified which seems to contradict the 194 figure quoted earlier (p. 5, line 23). Are these numbers for areas of different size (e.g., a third figure of 210 different scenes emerges for 105 km). Please clarify, 194, 200, and 210. Moreover, I found odd that the authors state (p. 7 line 2) that "The maximum number of observed cloud scenes (210) at a particular horizontal scale (105 km) remains unexplained" when the section that immediately follows tries to explain exactly that. Am I missing the subtle distinction? Section 3.3 tries to explain why the maximum number happens at 105 (not sure it succeeds), but why this maximum number is 210 remains as the unexplained mystery?

In Section 3.1, we describe the statistics of cloud scenes observed at the AMSU resolution at 45-km scales. There are 194 observed cloud scenes at this resolution. In section 3.2, we quantify the scale dependence of cloud scene statistics. The number of cloud scenes is shown to first increase with resolution then decrease as scales get large. The maximum number of cloud scenes is 210 and is observed at a resolution of 105 km. Both of the numbers 194 and 210 appear as 2 points on figure 3a. The former is indicated with the intersection of the red line and the curve, and the latter is just the value at the maximum.

The third number mentioned by the reviewer is 200 and it is a result of the procedure described in section 3.3. We shortened the name of this section 3.3 "Generalizing to all scales" as we understood from the reviewer's comment that the title was potentially confusing or misleading. The number of 200 mixed scenes is identified *independently of a grid resolution*.

Therefore, there are 194 cloud scenes observed at a resolution of ~45 km, 210 scenes observed at a resolution of 105 km and 200 scenes identified with a procedure independent from a regular grid.

We have added the following text at the start of Section 3.2 to clarify: "*In Section 3.1, the relative frequencies of cloud scenes were derived for exact matches of AIRS and AMSU observations to the CloudSat ground track. As the CloudSat ground track can oscillate over several AIRS FOVs across a scan line within a given orbit, the numbers of coincident CloudSat profiles matching to AIRS and AMSU will vary. Below, cloud scenes are derived independently of the specific AIRS and AMSU collocation geometry*."

As far as the quote mentioned by the reviewer on p. 7 line 2, that was intended to be a transition from Section 3.2 to 3.3 and as motivation for why we calculated the scale dependence of cloud scenes independently of a particular grid resolution. We have changed the text to the following to clarify: "*The reasons for the maximum number of observed cloud scenes (210) at a particular horizontal scale (105 km) are not immediately clear*." Then we have changed the text on p.7 lines 4-5 to further clarify: "*A simple conceptual model is described below that is able to approximate the results of Fig. 3 and offers some insight for the observed maximum frequency of cloud scenes and the spatial scale at which it occurs*."

We have rewritten some of the paragraph on p. 8 lines 11-20 as follows: "*A total of 200 out of 247 possible mixed scenes were identified. The minimum and maximum length occurrence frequencies of five cloud scenes (Ac,Sc), (As,Sc,Cu), (Ci,As,Cu,dc), (As,Ac,Ns,dc) and (Ci,As,Ac,St,Sc) selected randomly from the 200 present in the two-year record, are shown in Fig. 4a and 4c, respectively. Recall that the maximum length is defined from relation (2), while the minimum length is defined from relation (3), with an illustrative example shown for (Ac,Sc). From top to bottom, their respective ranks are 1, 26, 51, 76 and 101. It is striking that each frequency histogram in Fig. 4a and 4c is not monotonic and displays a frequency maximum between 100 and 1000 km. Consequently, the sum of all (200) observed mixed scenes across length scales will result in a curve with a maximum and these are shown in Fig. 4b and 4d. Both curves are very similar to Fig. 3a and have maxima for about 180 observed scenes at 77 km and 174 km, respectively. Using the methodology outlined in (1) to (3), the scale dependence of the number of observed scenes shows that the maximum will occur somewhere between 77 and 174 km.*"

– Can the same scale be used for the y-axis of Figs. 7 and 8? You say that that the common panels of these two figures (single cloud type scenes) should look very similar (inclusion of clear-sky notwithstanding), but the comparison is hampered by different y-axis range.

We regenerated figures 7 and 8 with the same ordinate scales. However, the reviewer should keep in mind that each bin in Figures 7-10 (revised) are normalized by the total number of cloud scenes (within an AIRS footprint) for single or multiple cloud types, or partly cloudy or fully cloudy scenes. We have added the following text to the caption of Figure 7: "*Each histogram sums to 1.0 and does not show the numbers of counts relative to another histogram. Relative counts could be inferred from the percentages listed in the 2nd to left column of Table 3*."

– Somewhere in section 2 mention what the maximum optical thickness retrievable by AIRS is.

We added the following sentence on line 12 page 5: "*The AIRS sampling includes nearly all ice clouds with $\tau i > 0.1$, while the maximum values of τi asymptote to values near 6–8 (e.g. Kahn et al., 2015)*."