Zamora et al. ACPD Reviewer Responses

<u>Response to reviewer #1</u>

GENERAL COMMENTS:

Much of this analysis is interesting and relevant. Nonetheless I have some concerns with the paper as it is that primarily relate to the extremely intensive filtering of the data that has happened. The authors are left with only a handful of cases (It's not clear to me exactly how many, maybe thousands) from two years worth of satellite data. Generally the utility of satellite data is found in the large sample volume; an advantage eliminated in this study. Even after throwing out most of the data the authors then proceed to estimate the effect of the aerosol on cloud longwave radiation over the whole arctic. It is very difficult to believe that the handful of cases examined here can be representative of meteorological conditions over the entire arctic and throughout the year. It is not clear to me that this extensive filtering is necessary or even useful. In fact it may introduce undesirable sampling biases.

An example of this over-filtering of the data is eliminating clouds that are detected by CALIPSO but undetected by CloudSat. Many of these clouds will be shallow liquid-only clouds with small drop sizes (exactly the cloud type purported to be studied here) and yet they are thrown away. Wouldn't one interesting test be to determine if these clouds are more prevalent in the polluted conditions. This might be expected from the authors hypothesis. As another example what sense does it make to require the cloud optical depth to be less than 3. Once again don't we want to know if there are changes in the relative frequency of occurrence of these optically thick clouds in the presence of aerosol.

In short I found the justification for the filtering methodology to be lacking and I would really encourage either a convincing justification for why most of the data is thrown out or more appropriately just include all cloud in the analysis. Finally, the authors really want to get at the impact of aerosol on the cloud longwave effect. The CloudSat data products (2B-FLXHR-LIDAR) have already calculated clear/cloudy fluxes for every pixel using combined input from CloudSat and CALIPSO cloud and aerosol profiles. The authors have put a good bit of work into identify clean and polluted conditions. All of this could be put together to simply calculate the aerosol effect for all cloud conditions without all of the filtering. Some more specific comments are below.

We thank the reviewer for their helpful comments on our manuscript, which have improved the paper.

1) In the general comments above, the reviewer suggested that we provide, "either a convincing justification for why most of the data is thrown out or more appropriately just include all clouds in the analysis." We have addressed this suggestion/concern in three ways:

i. At the request of the reviewer, we now have expanded our analysis to compare MOONLiT clouds with the broader group that they best represent: all nighttime optically thin, predominantly liquid clouds. Within this subgroup of clouds we still have fairly high confidence in the aerosol conditions surrounding the clouds. The nighttime criterion is kept to simplify the assessment of indirect effects (as opposed to semi-direct or direct effects) and to help identify aerosol conditions from the lidar with greater confidence (i.e., higher signal/noise). Because we still have higher confidence in classifying aerosol conditions for the MOONLiT subset than in this new group, the old data are still presented, although now mostly in the supplementary material. Clouds with very different characteristics and less certain aerosol conditions (i.e., daytime clouds, optically thick clouds, icy clouds) were not included in this analysis. For more discussion on our reasoning, please see response #3iii below.

ii. We have made a concerted effort to better clarify our goals and methods, which we think will also help address some of the reviewer's concerns (please see response #2 below).

iii. We have stressed more clearly that our conclusions are limited, and are only fully substantiated for the subset to which they pertain.

2) The methods have been clarified as follows:

i. "The authors are left with only a handful of cases (It's not clear to me exactly how many, maybe thousands) from two years worth of satellite data."

The confusion on sample number might have arisen due to a typo in the formatting of Table 2 where the sample numbers were presented. The sample numbers (labeled in the "n" columns) for the "all cloud" cases were truncated, so that a sample number of clouds over sea ice that was actually 4579, for example, appeared as 45 and 79 separately. We have corrected this error in the new draft.

ii. "It is very difficult to believe that the handful of cases examined here can be representative of meteorological conditions over the entire arctic and throughout the year."

We agree with this comment, and addressed it in several ways. First, we have changed the title to more clearly emphasize that our samples cover only nighttime data. This was previously discussed in the manuscript (e.g., p3, lines 1, 14, and 16; p 4 line 12; p 17 line 31 in the ACPD paper), but the change to the title will hopefully make it clearer to the reader that our data are relevant to mainly wintertime (and some spring and fall) cases, and not to conditions throughout the year.

Next, we now better clarify in the title, abstract, and in numerous places throughout text that the meteorological regimes discussed were only for regions over the Arctic Ocean and not for the entire Arctic. For example, please see p. 2, l. 20 & 24; p. 3, l. 24; p. 4, l. 26, etc.

Thirdly, as previously stated in the ACPD paper (e.g., p. 17, l. 27-35), our original conclusions were intended to only represent a subset of optically thin, predominantly liquid clouds, and not all nighttime Arctic Ocean clouds. We now more strongly emphasize this point throughout the text (e.g., first paragraph of the new section 3.2).

Lastly, regarding the general representativeness of our data, we have conducted a series of additional analyses to place our results in context of how much of the total Arctic region they cover (p. 10, l. 5-13; p. 16, l. 5- 12). The results of these analyses have helped narrow our maximum regional estimates of radiative impact to more precise levels, and are also used to more clearly stress the limitations of our conclusions in the abstract and elsewhere (e.g., p. 19, l. 28; p. 10, l. 5-13).

iii. "An example of this over-filtering of the data is eliminating clouds that are detected by CALIPSO but undetected by CloudSat;" and "By limiting analysis to cases where both CALIPSO and CloudSat identify approximately the same cloud height the authors throw out a great number of cases where clouds may have radar reflectivites below the detection threshold of the radar but are the thin liquid clouds of interest to the study.... I can't reconcile this with the statement that 95% of the data are included in the analysis."

Clouds that were detected by CALIPSO and not detected by CloudSat actually were included in the analysis. As can be seen in Table 2, the total data sample number is higher for parameters obtained from CALIPSO (e.g., cloud base height) than for parameters derived from CloudSat (e.g., cloud droplet effective radius), since there were no trustworthy CloudSat data in clouds that CloudSat did not detect.

The confusion here likely arose from hangover information from a previous draft that should have been deleted. The third row in the Table 1 CloudSat criteria has now been deleted. To improve clarity, we also added a footnote to Table 1 for the lines with CloudSat data, as follows:

"*as available for clouds with radar reflectivity above the detection limit of -29 dBZ."

Our wording in section 2.1 has also been edited for clarity on this issue (new text in bold).

"~~To ensure comparability of clouds measured with both instruments, only clouds for which the reported cloud top height was within 0.4 km in both instruments were included (i.e., ~95% of the data).~~ Because the CloudSat radar does not accurately estimate cloud properties below ~0.7-1 km agl (Huang et al., 2012; Mioche et al., 2015), we focused on clouds with bases ≥ 1 km agl. We recognize that many Arctic clouds lie below this altitude (Devasthale et al., 2011a; Shupe et al., 2011) and that these low-level clouds have important radiative impacts. However, we still chose to focus on clouds at these higher levels to obtain higher certainty in the data. **Also,**

some of the very thin clouds detected by CALIPSO had radar reflectivities that were too low to be detected by CloudSat, and CloudSat may sometimes mistakenly assign precipitating ice as a cloud (de Boer et al., 2008). Therefore, radar reflectivity data and CloudSat reflectivity-derived cloud parameters, where available, were obtained from the height bins closest to where CALIPSO detected a cloud."

The text stating that ~95% of the clouds had cloud top heights within 0.4 km of each other was a related error. The 95% number was calculated from the final dataset made after figuring out that it was best to approximate CloudSat cloud height bins from the CALIPSO cloud height bins; presenting this calculation was inconsistent with the proceeding sentence. So what the 95% number signifies is that in the cases where both CALIPSO and CloudSat observed a cloud in approximately the same height range, ~95% of the time, the observed cloud top height bin from CloudSat was within 0.4 km of the cloud top height bin that would have been expected based on the CALIPSO data. To avoid confusion, we removed this from the new text. Thanks for drawing our attention to this.

3) The reviewer voiced concern over the sample selection criteria. To paraphrase, they wanted to know why very rigorous cloud selection criteria are useful and necessary (i), especially given: (ii) that stricter criteria comes at the tradeoff of sample representativeness, (iii) that we might obtain other useful information by loosening the criteria to include various products or data, and (iv) that filtering may introduce sampling biases.

As previously mentioned, at the reviewer's suggestion, we have expanded the dataset to be more comprehensive and so hopefully some of these concerns are already addressed. However, even in the expanded dataset we still had to make choices on which clouds to include in the analysis, and so in this context we will respond to individual points raised by the reviewer.

i) Regarding why the rigorous sample selection criteria are useful:

The quantifiably correct identification of a subset of clean background clouds is a crucial step in our method, which we now more clearly state in the paper (p. 2, l. 26-31). Any expansion in the scope of the study comes at the cost of higher and less quantifiable error in our results. For example, we now include clouds below another cloud layer, but uncertainty in the aerosol classification for these cases from the lidar is much higher. Thus, air mass classification for those clouds is now more reliant on the model, making errors less quantifiable, which we now discuss. As another example, our data in section 3.1 suggest that at night, CALIPSO without FLEXPART misses ~33% of dilute aerosol layers. If we had included daytime samples where 2B-FLXHR-LIDAR estimates are more reliable, CALIPSO would have missed ~60% of dilute aerosol layers based on preliminary analysis. Ultimately, the criteria we chose were selected to balance representativeness with data quality, erring on the side of data quality.

ii) As the reviewer mentioned, the tradeoff of improved certainty in aerosol conditions is reduced sample representativeness. The reviewer questioned the utility of the study based on the low sample representativeness.

> This is a good point that we have worked hard to address. Besides including more samples and more quantitatively discussing representativeness of the samples in the new section 3.2, we now more clearly state both the limitations and the utility of our results (as previously mentioned, p. 19, l. 28; p. 10, l. 5-13). Also, all trends within the cloud subsets are discussed in the context of both statistical significance (which incorporates sample number as a factor) and meaningfulness (as discussed on a case-by-case basis in the text).

iii) The reviewer suggested expanding our dataset to include a) clouds with higher optical depth, b) cloud radiative forcing from the CloudSat 2B-FLXHR-LIDAR product, and c) clouds below 1 km.

> One of the goals of this work is to provide a foundation for expanding the study to other cloud types that were too complex to include in one manuscript. Although we have made an effort to include a larger and more representative cloud dataset in our revised analysis, understanding the aerosol indirect impacts on different types of clouds with lower confidence in the assignment of clean background conditions is quite complicated. In many cases, we felt it would best be done in a separate paper that can be fully dedicated to the substantial amount of additional caveats, analysis, and discussion required for exploring those data. Please see responses to specific reviewer comments below.

> a) Specifically regarding the following reviewer question: "…what sense does it make to require the cloud optical depth to be less than 3? …[D]on't we want to know if there are changes in the relative frequency of occurrence of these optically thick clouds in the presence of aerosol?"

> COD is required to be < 3 to enable the lidar to detect aerosol layers below the cloud.

> We agree that it would be interesting to test this in our dataset, for example by seeing if there is a difference in the relative frequency of optically thick clouds in cases where high aerosol was seen above the clouds and the model indicated high aerosol should have been present below the cloud. We plan to explore this topic as part of continuing work, but it is beyond the scope of the current paper. These other cloud types are not otherwise discussed in the paper, and the different analytical approach required and the much less-certain results that would be provided would result in a longer discussion with results not central to the rest of the paper.

> For the reviewer's interest, as mentioned in the paper, daytime MODIS COD observations from Coopman et al., 2016 in liquid phase Arctic clouds suggest that a frequency difference similar to what the reviewer mentions should be observable.

> b) The reviewer also wondered why we didn't use the CloudSat 2B-FLXHR-LIDAR

product.

This is a good question. The 2B-FLXHR-LIDAR version R04 product is a general product useful for the many conditions and cloud types across the world. However, it also has some limitations for the specific types of clouds that are discussed in this study, particularly for its COD and r_{el} input information. It was important to us to obtain the most accurate estimate of r_{el} available because, based on Twomey effect expectations, we would expect that r_{el} is one of the most important parameters affected by aerosol indirect impacts. When clouds are detected by CloudSat, the 2B-FLXHR-LIDAR product assigns cloud r_e and COD from the CloudSat 2B-CWC-RO product. LWC is then estimated from these parameters for input into the model. As discussed in section 2.1.2, the CloudSat 2B-CWC-RO product is associated with a lot of errors, particularly for r_{el}. We reduced some of those errors in our study by only focusing on predominantly liquid clouds, and specifically selecting the 2B-CWC-RO subproduct that assumes droplets were liquid. That was not done to our knowledge in the 2B-FLXHR-LIDAR version R04 product.

In clouds where no CloudSat data were available, the 2B-FLXHR-LIDAR product estimates r_e based on temperature; above -20 ^{o}C clouds are assumed to be liquid and assigned a r_e of 13 μm. Below that level clouds are assumed to be ice and assigned a r_e of 30 μm (Henderson et al., 2013). Since all of our clouds were predominantly liquid-containing, but some of them reached temperatures well below -20 ^{o}C (Table 2), the 2B-FLXHR-LIDAR r_e temperature-based estimates are likely to contain higher errors than in our method for our specific cloud subset of interest. Unlike in the 2B-FLXHR-LIDAR product, we did not attempt to estimate r_e in clouds where no CloudSat data were available. Instead, our radiative flux calculations were based on the average for clouds with slightly thicker CODs where CloudSat data were detected (please see Table 2).

Another difference is that the CloudSat 2B-CWC-RO product determines COD when those clouds are detected by CloudSat, and otherwise it uses the CALIPSO COD values (at least for nighttime data when MODIS data are unavailable). Because our particular sample set contained a relatively high fraction of cases with no CloudSat data, we thought it best to use CALIPSO CODs across all clouds to obtain a more internally comparable dataset.

Other than a few other minor differences, the radiative transfer calculations in the 2B-FLXHR-LIDAR calculations are not expected to be very different from our own. Like us, their model relies of surface conditions detected by passive microwave, and like us, they conducted two additional sets of flux calculations that are performed with all clouds and all aerosols removed, respectively. For these reasons, we chose to calculate radiative fluxes using our own method in this particular study. We agree with the reviewer that the 2B-FLXHR-LIDAR product could be very useful in future studies that cover a wider group of Arctic cloud types (for example, daytime clouds), and in future work we hope to incorporate this product more.

c) The reviewer suggested we include clouds in the lower 1 km of the atmosphere.

At the reviewer's suggestion we now include clouds with bases down to 200 m above the ocean surface in the new analysis, with some caveats. As mentioned in the paper, "the CloudSat radar does not accurately estimate cloud properties below ~0.7-1 km agl (Huang et al., 2012; Mioche et al., 2015)." Thus, the CloudSat data for shallow clouds only represent clouds > 750 m asl. We elected to keep the 1 km criterion in the MOONLiT portion of our study in order to base our reflectivity, r_{el}, and radiative transfer conclusions on a dataset in which we have higher confidence. We do not include clouds with bases below 200 m to avoid fog and to enable detection of below-cloud aerosol.

iv) The reviewer also expressed concern about the biases potentially induced by our sample selection criteria. We discussed these biases in detail in the former section 2.3 (now section 3.2). For the reasons discussed above, for the purposes of the present study, the cumulative errors induced by biases related to the sample selection criteria are likely to be much smaller than the error added by including cloud subsets with poor quality data or with uncertain aerosol influence. We now note this at the end of section 2.3.

4) Regarding clouds that are detected by CALIPSO but undetected by CloudSat, the reviewer asked: "Many of these clouds will be shallow liquid-only clouds with small drop sizes ... Wouldn't one interesting test be to determine if these clouds are more prevalent in the polluted conditions? This might be expected from the author's hypothesis."

We thank the reviewer for the good suggestion. The reviewer is correct that there is a significantly higher probability of clean background clouds being detected by CloudSat than in all clouds and in aerosol-impacted clouds. We now include this information in the text.

SPECIFIC COMMENTS:

1) Section 2.1.2: By limiting analysis to cases where both CALIPSO and CloudSat identify approximately the same cloud height the authors throw out a great number of cases where clouds may have radar reflectivites below the detection threshold of the radar but are the thin liquid clouds of interest to the study. Eliminating clouds that have a base height greater than 1 km further aggravates this situation. In fact the authors have chosen a sampling strategy that minimizes the data availability from either instrument because it will be infrequent that clouds have optical depth less than 3 but still have a radar reflectivity above the -28 dBZ CloudSat sensitivity. This is why it looks likethere are maybe only a few hundred points on figure one. I can't reconcile this with the statement that 95% of the data are included in the analysis. How many pixels are included in the analysis? How many total pixels are there over the two year period?

This comment has already been addressed above. We actually did include clouds that CALIPSO observed but CloudSat did not (see general response #2iii). The sample

numbers for Figure 1 are noted in Table 2. Please see the response to general comment #3iv for more discussion on the inclusion of data below 1 km.

As can be seen in Table 2, we noted the relative percent of precipitating clouds that otherwise met our sample criteria in each of the different air mass types, and there was an apparent aerosol influence. However, all other cloud characteristics listed in Table 2 are for clouds with no observed precipitation. To better explain this in the paper, we have added the following to the methods section:

"CloudSat may sometimes misclassify precipitating ice as part of the cloud (de Boer et al., 2008), which can lead to overestimation of r_{el}. Quality-flagged data were excluded, such as observations from precipitating clouds, as determined from the CloudSat 2B-CLDCLASS-LIDAR version R04 product. Note: although we counted the number of cases where precipitation occurred for comparison at a different step, precipitating cases were otherwise excluded **from most other derived cloud parameters in the analysis. These cases were excluded in order to obtain comparable data across cloud characteristics, which was particularly important for the longwave emissions calculations detailed in section 2.2 that included the r_{el} as one of several input parameters.**"

And the following to footnote B in Table 2:

"Precipitating clouds were included in [the % precipitating] metric only; for all other attribute classifications, clouds were required to have no observed precipitation **in order to be comparable with r_{el} estimates that were most reliable in non-precipitating clouds**."

As mentioned in general comment #2iii above, clouds not detected by radar were included in the original analysis. We also have supplied more information on our reasons for excluding clouds with high optical depth in our response to general comment #3iiia above. For our response on biases, please see the answer to general comment #2iii above.

As mentioned in the paper, "NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, version 2 data (Meier et al., 2013; Peng et al., 2013) were used to approximate the fractional sea ice cover over ocean at the specific month and location

of each profile. A sample was classified as being primarily over sea ice or open ocean when the sea ice fraction at the given location and month was > 80% or < 20%, respectively."

5) Fig 3. Does this map include only the filtered data points shown in Fig 1.

Yes, it was obtained from the sea ice concentration below each cloud point in the study during the month that cloud was sampled. This has now been clarified in the Figure caption.

6) Page 11, Line 11: How is precipitation determined? Which product?

As mentioned in the methods section: "Cloud phase and precipitation occurrence were acquired from 2B-CLDCLASS-LIDAR version R04 estimates (Wang, 2013)."

7) Page 11, Line 22: I see Fig 1 differently. To my eye there is a clear clustering of the data with substantially more aerosol cases north of Europe and relatively more clean cases north of Siberia and North America. This statement is not justified by the analysis.

We have removed this sentence. Note: regional clustering of aerosol-influenced cases is more apparent in the larger dataset (see the new Fig. 1).

8) Page 12, Line 8: It is fairly obvious that you won't find an optical thickness difference when you have artificially limited the range of optical thicknesses to less than 3.

This sentence is no longer relevant in the new analysis.