Douglas and L'Ecuyer review

General comments:

In general, I like the paper's approach to the critical problem of understanding aerosol impacts on clouds, and how they carefully dealt with co-varying meteorology. I also like how they combine measurements from different sources to get a broader picture of the system and to constrain the observations better. This paper has potential to be very important and useful for/referenced by a host of other studies using a similar methodology in the future.

That said, I had four main concerns with the paper:

1) Substantial extra work is required to clarify and explain the methodology. Another person attempting to replicate this study could probably not do so, as it is written now, and this makes it difficult for a reviewer to fully judge the value of the work. Unfortunately, by the time I reached section 3.5, the cumulative uncertainty I had with the methodology was large enough for me to have strong reservations about my ability to judge the meaningfulness of this section. That said, I do believe that if the authors answer the questions in the specific comments fully, I can better judge of the work in the next revision.
2) There is a possibility that serious errors were made regarding Fig. 5 that might have a large impact on the results. For the reasons listed in the specific comments below, I request that the authors double-check their results carefully, and if necessary, re-run the analysis.
3) A variety of confounding factors were ignored here, including the difficulties in co-locating aerosols and clouds based on AI, errors and biases in near-cloud satellite aerosol detection, and potential confounding influences of aerosol semi-direct and direct effects. These should be addressed and acknowledged.
4) Throughout the paper the conclusions tended to be a bit overstated. It should be made clearer in the manuscript that the study only focuses on a subset of data, and that this subset is not necessarily broadly representative of all conditions.

Specific comments:

Title, abstract, and conclusions: I suggest that the title and abstract better clarify the focus on warm marine clouds, and contain some stronger hints of the large remaining uncertainties (e.g., by adding "**Better** Understanding Aerosol-…." to the title). The reason for this suggestion is that large groups of clouds were excluded in this study, and there were some fairly major inherent uncertainties in the methodology. The study focused on daytime, single-layer, warm clouds, and as best I can tell, it only includes clouds with latitudes < 60^{o} and over the ocean. LW forcing at night was excluded entirely. Therefore, the study cannot address several important complex cloud-radiation interactions related to aerosols. For example, the method presented here would not address ice nucleating effects or seeding effects in multi-layer clouds, which can be quite complex. The results may also not be applicable to terrestrial areas where, for example, diurnal effects of heating can be much more variable.

Introduction/Methods: Please clarify to the readers why RH and LWP, which are not independent variables, are considered separately, and essentially independently, in this study.

Methods: Please add a table where readers can quickly find what subset of clouds were included in the study. From Figures 8 and 9, it appears that terrestrial clouds and clouds poleward of 60° were eliminated. However, this is not explicitly stated in the paper. A concise central location to find which latitudes, LWPs, and temperature levels, etc. for the subset of clouds assessed in this study would be useful.

Section 2.2 or 3.5: A map of the frequency of observations of the subset of clouds compared to all clouds observed in the region would be very helpful for interpreting the relevance of this study. Are the types of clouds studied here more common in some locations than in others, and is there any geographic bias in Figures 8 and 9?

p. 4, l. 27: "*An along satellite track cloud fraction is determined by finding the average number of warm cloud pixels that satisfy these criteria (seen by CloudSat or CALIPSO, below freezing level, and LWP greater than 20 g m^{-2})*" What is meant by "below freezing level"? Is that determined from MERRA2 temperature profiles below 0 °C? Please clarify. As I interpret it, this suggests that the altitude or pressure level where the results are obtained varies by profile, and that this altitude would probably vary quite a bit over latitude and surface type. If this is the case, how do cloud altitudes in the study vary? Can we rule out that vertical variation in the clouds being studied would not add substantial error or bias the results (e.g., by introducing different aerosol types at different levels, or horizontal/vertical winds, etc.)?

p. 5, l. 3: "*The shortwave cloud radiative effect (CRE) is then defined in terms of the all sky and inferred clear sky forcings from CERES and cloud fraction from CloudSat*." How is CloudSat cloud fraction defined? Which conditions are included in "all sky" conditions? From what I understand, situations when there are multi-layer clouds, and clouds below freezing temperatures, etc. are excluded. If this is correct, then the term "all sky" may be a little confusing, and perhaps other wording would be better.

p. 5, l. 15: Which version of MODIS is used, and why? What is the resolution of these data and how does that relate to the cloud resolution?

Section 2.3: MODIS has a variety of known issues with reliably detecting aerosols near clouds. What kind of cloud screening was used, and how sensitive are the results to this choice? It would probably be useful to note in the paper that binning the data by RH conditions could create some biases, due to aerosol swelling near high-humidity conditions typical near clouds.

Section 2.3: A major issue with using the MODIS AI, which is a column-averaged value, is its inability to co-locate aerosol layers with clouds. In some cases, for example, there may be long-range transported aerosols at high altitudes that lead to high AI values that are not representative of the aerosol conditions that the low clouds of focus in this study experience. This is of particular concern over the low-altitude marine regions of focus in this study, which are often quite clean toward the surface. Uncertainties in aerosol-cloud co-location have high potential for biasing these results in time and space, and potentially may lead to incorrect conclusions. Therefore, I recommend that the authors not only acknowledge this uncertainty here, but that they also at minimum include a short literature review of whatever is known about when and where the biases would be most likely to occur (e.g., based on CALIPSO and/or model output) for the regions of focus in this study. If the authors have the resources to do extra analysis to clarify or reduce this uncertainty as it applies to their results specifically, that would make this contribution much stronger.

Equations 2-5: On the first read-through, I was quite confused about the upper limits of summation (e.g., the number 7 in equation 2). Justifying the specific choice of those numbers might make more sense in a later section (e.g., section 2.5) than in section 2.4, where they first come up. Therefore, I suggest the authors make the equation more generalizable by having the upper limit of summation be a variable, to be assigned a value later when explanation for that value can be more logically provided. This might also help if others want to cite this method in future work, but want to use different numbers of states for their specific application. Also, please specify earlier on in the text what the upper limit of summation represents, as this was not clear in these equations and in section 2.4 in general. Moving the following text from p. 7 into section 2.4 where these limits are first introduced could help: "*The regime bounds depend on the resolution used, which is varied to establish the degree to which environmental factors must be constrained to accurately characterize sensitivity*".

Equations 2-5: As written, it is unclear how the weighting was done. Does $N_{i,j,k,l}$ imply that the data were weighted one time, or four separate times with each summation compared to N_k? Others may be able to get this information immediately from looking at the equation, but for readers like myself, the clarification is important.

p. 6, l.5, "*N_k is the fraction of clouds that fall into LWP state k.*" Did the authors mean the "fraction of cloud profiles" instead of the "fraction of clouds"?

p. 6, l. 6: How is estimated inversion strength calculated? (note, some information is provided later, on p. 7, l. 2, but this information is not fully descriptive).

P. 7, l. 4: "*The relative humidity at 700 mb is used as a measure of the effect of entraining free tropospheric air.*" As I understand it, the RH at on vertical level is assumed to be representative of the whole vertical column up to the freezing point, or at least to provide important information for the whole column. However, RH at 700mb will be most relevant for clouds in that general altitude range. Will this bias the results, or add error? What is the variability in cloud locations? This was not quantified. Why not just use RH at the appropriate altitude ranges where the cloud layer is found?

p. 7, l. 5: "*All observations within the 5% - 95% percentiles of both EIS and RH are partitioned into regimes.*" As I understand it, one nice thing about taking the weighted mean is that you can use all of the data, and still get representative results. Thus, I don't understand why these data were excluded in the first place? (from the above statement, I believe the excluded data would equal between 10-20% of their subset of observations?) Was a similar procedure was not followed for LWP, and if so, why not?

p. 7, l. 5: "*Environmental regime limits are defined such that there are the same number observations within each percentile of either EIS or RH. The regime bounds depend on the resolution used, which is varied to establish the degree to which environmental factors must be constrained to accurately characterize sensitivity.*" I am very confused by this part of the methodology. Please define in the text what is a bound and what is a limit. I am guessing that the "regime bounds" are the same thing as the upper limits of summation in equations 2-5? Is the "regime limit" the lengths of the [i,j,k, or l] bins in equations 2-5 or something else? I also think the wording of "each percentile," which in general usage implies 1 of 100 equal groups in a dataset, may also be incorrect because it does not seem consistent with the rest of the sentence and equations 2-5. Did the authors mean bins instead of percentiles? If so, what are these bins, specifically, how were they chosen, and how does the choice of spacing affect the results? I was also confused as to why one would want to group the same number of observations within each EIS or RH percentile [or bin?], if the results are going to be weighted later?

p. 7, l. 19: Did the authors mean "…*warm**, single-layer, marine** cloud SW*…"? It doesn't appear that they looked at terrestrial clouds, and they stated that they excluded multi-layer cloud cases.

Fig. 1: Please describe in the Figure caption what the red lines and blue dots represent (blue dots are presumably λ, but it is best to be completely clear). In the caption, λ is referenced. To avoid confusion, please state which λ is being referenced (so far λ_0, λ_{LWP}, λ_{ENV}, λ_{BOTH}, and λ_{ALL} have been defined, but no λ without a subscript). Please also specify which λ is being discussed in the rest of the paper as well as the symbol is used frequently. Please clarify that the R^2 value in Fig. 1 is describing the blue points and not the underlying distributions of all the data, because the largely overlapping red bars would suggest that in fact the correlation of the more raw data before that averaging happens is much smaller. Attaching a p-value to this and other similar figures appearing later in the paper seems appropriate.

p. 7, l. 22: "*The indicated variation of SW CRE within each ln(AI) (red bars) bin alludes to variation in the overall effect not captured by a single linear regression*." Please describe exactly how these red bars were calculated – without knowing that, it is very hard to interpret the information the red bars are intended to convey.

Section 3.1: Many others have provided similar values to the -12.81 value provided here. It might be good to compare this finding with findings from previous works.

Figure 2a: What cloud states are included here, and how were they derived and chosen? Some explanatory information is provided in section 3.2, but only after this Figure is referenced, which makes things confusing for the reader. It would be best if the figure could be a standalone item without requiring substantial reference to the text. The y-axis label for Fig. 2a seems to be missing, and only the units are provided. The text explaining Fig. 2a is not explicitly identified in the caption.

p. 7, l. 29: It might be helpful to reference which equation was used to derive the -13.12 value.

p. 8, l. 4: "*Constraints on LWP limit these influences.*" This is already a well-known phenomenon. The authors should probably credit here some of the other work that has previously established this finding.

p. 8, l. 6: why were 3,7,11, and 23 divisions chosen?

Section 3.2: The term "cloud state" is commonly used throughout the paper, and it is the focus on section 3.2. However, cloud state is not explicitly defined in the paper, as far as I can tell, and this is very confusing for the reader. The data in section 3.2 mostly revolve around clouds binned by LWP. Is it possible to just use LWP, instead of "cloud state"? Another minor suggestion: the authors might consider changing "cloud regimes" to "LWP bins" (if this is correct). That would be a lot easier for a casual reader of the paper to understand.

Fig. 3a: Where is the caption text describing Figure 3a? Please clarify where the -11 value in the Fig. 3 caption comes from in relationship to these figures. Is it based on the weighted mean of the data in Fig. 3a? Where is the label for the z-axis in Fig. 3a?

Figs. 3b and 3c: Please state in the caption how moist and dry environments are defined. Are these figures examples of data within individual grid cells from Fig. 3a? If so, please state that. The red bars seem to suggest that there may be no significant differences between any of the ln(AI) values within Fig. 3b or Fig. 3c, including at very high ln(AI) values and very low ln(AI) values?

p.10, last line: "*To account for the local meteorology, warm clouds are separated into 100 environmental regimes…*" This method seems to closely parallel the methodology of previous work (e.g., Chen et al. (2014)). It would be appropriate for the authors to reference such work here.

Section 3.3: To support the conclusions in this section, I strongly recommend that more information be provided on which colors of the grid cells of Fig. 3a are significantly different from zero. To address the signficance, the authors might consider obtaining a confidence interval around the slope for each grid cell (preferably a bootstrapped confidence interval, as that would be valid even if the data don't meet all the assumptions of a normal linear regression analysis). That would indicate whether the slope in each individual grid cell is significantly different from zero, and this information could be conveyed in the figure with markings on the grids.

Section 3.3: Since some of the cells in the figure probably have much greater sample sizes in the natural environment than others, to me, the weighted mean is probably more meaningful than the findings of individual grids, and I think it would be appropriate to stress this more in the paper.

p. 11, l. 3: "*The highest sensitivity is observed in stable regimes (EIS > 5.0) with a moderately dry free atmosphere*." And p. 11, l. 8: "*Above 1 K, λ increases with increasing RH, while in less stable environments, RH plays only a secondary role in modulating the sensitivity*." In Fig. 3a, I don't see evidence so far of there being higher sensitivity in drier environments, or of the latter statement at all. Were Figs. 3b,c supposed to be referenced here? Please provide more information to substantiate these statements.

Fig. 4: In the caption, sensitivity of what? Again, here, I think it would be really useful to note which of the grid cells are significant. Sample number in each grid cell will go down as resolution increases, and that would presumably impact the weighted mean values presented and discussed with respect to this figure, so significance would be a useful metric to help evaluate these results.

Fig. 5: This figure seems very likely to have an error. I do not see how the clouds at two extreme EIS and RH values can have the highest frequency of occurrence. If the x- and y-axis ranges were selected appropriately, one would expect the points approximately in the middle to be most frequent, and the points at the edges to be least frequent. Also, why is there such a strong mirror-like diagonal pattern in the plot? Natural data rarely show such a distinct pattern unless the x and y variables are highly related to each other. Please check that the data plotted here are correct.

The last 2 paragraphs of section 3.3 could probably go better in supplementary material.

Fig. 6: Again, please define in the caption which λ is being presented. Please show significance of each grid cell. Please clarify what the weighted summed sensitivity is (is it for plots a-g?). What is plotted in Fig. 6h? What meteorological constraints are plotted for each subpanel in this figure (please label that in both the caption and on the subpanels). I am still confused about what "cloud state" even is. Is it LWP? Why were the ranges shown in this figure chosen?

p. 14, l. 3: "*Overall, the largest ln(AI) sensitivity is seen in stable, dry environments (Figure 6h)*." I don't see this shown in that figure.

Section 3.4, paragraphs 1-3: It seems to me that the information most relevant to the conclusions in this section is the weighted mean, and not the data shown in Figure 6. As presented, the differences between the Fig. 6 subpanels are difficult to distinguish from each other. They are also difficult to interpret, especially without information on their individual frequency of occurrence within each LWP bin and without data on their individual significances. Even if that information were presented, the figure would probably still be confusing for readers to interpret. For these reasons, the authors might want to consider changing the figure to show only the weighted means (or perhaps weighted means of the quandrant of the figures if they are trying to compare differences at high and low RH and EIS). The current plots could still be presented, e.g., in supplemental material. As an example, compare how the example Figure 1 in Zamora et al. (2018), which is analogous in many ways to the panels shown here, is simplified later in their Figure 3.

Fig. 7: Again, please double check that the frequencies of occurrence are correct.

p. 16, l. 5: "*In the absence of constraints (top), λ exhibits larger variations in magnitude and sign than when cloud, environmental, or cloud and environmental constraints are in place (panels b and c and Figure 9).*" Was Fig. 8 supposed to be referred to here? I don't see a panel b and c in Fig. 9, but these trends are not evident in Fig. 8….

Section 4: How do the values in Table 1 compare to other literature values? What other studies have looked at aerosol-meteorology co-variation and found similar results as here?

Figs. 8 and 9: Which of the pixels shown here are significantly different from zero?

p. 18, l. 1: It would be useful to also mention earlier on (e.g., methods?) that there were 1.8 million observations in the study.

Technical comments:

p.7, l. 6: "same number **of** observations"?

Fig. 3: Is the 11 supposed to have a period after it?

References

Chen, Y.-C., Christensen, M. W., Stephens, G. L. and Seinfeld, J. H.: Satellite-based estimate of global aerosol-cloud radiative forcing by marine warm clouds, Nature Geosci, 7(9), 643–646, doi:10.1038/ngeo2214, 2014.

Zamora, L. M., Kahn, R. A., Huebert, K. B., Stohl, A. and Eckhardt, S.: A satellite-based estimate of combustion aerosol cloud microphysical effects over the Arctic Ocean, Atmospheric Chemistry and Physics, 18(20), 14949–14964, doi:https://doi.org/10.5194/acp-18-14949-2018, 2018.