Reviewer 1 Response

The authors utilize remote sensing observations and a regime-based approach to isolate the effects of varying aerosol index on cloud microphysical (1st indirect effect) and cloud macrophysical properties (adjustments). The authors utilize regimes of above cloud RH and stability. LWP is binned to account for variations in cloud state in each regime. The results show that in some regions adjustments and the first indirect effect have opposing signs. The authors also show that as LWP increases the radiative response to AI saturates. The analysis presented here satisfies the important problem of separating variability due to meteorology from aerosol-cloud interactions (aci). The authors find a relatively weak ERFaci from warm-topped clouds over oceans, which appears to be due to dimming in regions in the equatorial Atlantic and Indian ocean.

We would like to thank the reviewer for taking the time to read our manuscript and provide feedback and comments.

While I appreciate that the authors are applying the methodology developed in a previous study, it is hard to understand what is being done and I think the authors could briefly summarize their methodology to allow readers to more efficiently refer to DL19. The description of the observational data sets could be much more substantial. It is confusing what observational and modeling data is being used for what. In some cases it appears that observational data sets that are not appropriate are being used, but it is hard to confirm this from the data section. One solution that might make this un-ambiguous would be to create a table of variables and data sources.

We are only using satellite observations and reanalysis data intended to be paired with satellite observations (MERRA-2). To clarify what observations we are using, we have added to section 2.1 Data:

> *"Collocated satellite observations of cloud shortwave forcing, cloud fraction, and aerosol index are obtained by NASA A-Train satellites Aqua, CloudSat, and The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) from 2007 to 2010. The NASA A-Train is configured to maximize the synergy between different satellite products to improve our understanding of clouds, aerosols, and the environment (L'Ecuyer et al. 2011)."*
> *"2B-CLDCLASS-LIDAR combines CloudSat's CPR with CALIPSO lidar observations in order to discern even the thinnest clouds."*
> *"To broadly characterize large-scale environmental conditions, MERRA-2 temperature and humidity profiles are collocated by taking the environmental profile within 30 minutes of a CloudSat overpass and within ~1/2 degree latitude and longitude"*

A critical issue with this paper is use of area-mean LWP (in-cloud LWP*CF) from microwave when the authors imply they are using in-cloud LWP based on wording in the paper (ln 153). From reading the discussion in DL19 I believe that scene-mean LWP from AMSR is just used to filter data into rough bins, and does not play a role in the analysis beyond this. While this is probably not a big problem, the authors may want to clarify what the footprints of the different data sets are that they are using, possibly with a diagram overlaid over an actual satellite image to allow readers who are less familiar with remote sensing to contextualize what is being shown, especially because the authors are using active instruments averaged along track with passive instruments. In particular, in this regard I am confused how the authors are overlapping along-track averaged CF from Cloudsat-CALIPSO with AMSR LWP and a diagram might be helpful. A nice image of the actual cloud field from MODIS on the background would be helpful to readers trying to contextualize the retrievals in terms of cloud features.

We have added the caveats of the footprint discrepancies along with how close geometrically the footprints are. Added to section 2.1 Data:

> *"While the footprints of CloudSat and AMSR-E do not perfectly overlap, the AMSR-E LWP is used to establish a scene based constraint on the clouds in order to better consolidate our observations into regimes. AMSR-E's footprint is within ~2.5 km of CloudSat's track, meaning both sensors are observing the same, liquid clouds (Lebsock et al. 2014)."*

CloudSat observations are often combined with AMSR-E scene averaged LWP in a number of cloud and aerosol studies (such as L'Ecuyer et al. 2009 and Chen et al. 2014). Our study does not aim to understand how the LWP responds to aerosol, only to use LWP as a higher level constraint in order to partition warm clouds into characteristic regimes.

The authors need to either apply their analysis in a GCM simulating PI and PD (Gryspeerdt et al., 2017; Gryspeerdt et al., 2016; McCoy et al., 2019; Costa-Surós et al., 2019) to make sure that their analysis methodology has predictive power, or examine the response of cloud to some sort of transient change in aerosol (Malavelle et al., 2017; Toll et al., 2019) and make sure that their analysis trained over different data can predict the response to the transient change in AI. Without these falsification tests of their predictions, it is unclear what predictive use their correlation model has in that there is no way to falsify their predictions. Even an approximate calculation using model LWP, CF, SW, and AI without any complex output along the satellite overpass (which doesn't appear to be a major source of error compared to problems from low aerosol amount as shown in Ma et al. (2018)) would provide a much more powerful validation of what the authors are hypothesizing is the ERFaci.

A next step will be to find these same signals within output from a GCM, however that is beyond the scope of the current study. This study intends to only document how the observed brightness and extent of clouds respond as aerosol concentration increases, and how these signals depend on the environment and cloud state. These responses *are then used to derive an estimate of ERFaci that is consistent with the specific observations used. While similar methods can be applied to GCM output, the results do not provide a stringent test on the methodology since model responses will depend strongly on how the underlying processes are represented in the model. Non-linearities or stronger/weaker dependencies on environmental state may yield vastly different results that do not provide a useful assessment of the validity of the decomposition approach. Furthermore, any meaningful comparison of GCM output against observations is severely limited by mismatches in resolution between the large GCM gridbox and the fine-scale satellite observations (e.g. Kay et al, 2019).*

In principle, results from this study can be used to assess how well GCMs recreate the derived linearized relationships between aerosol, cloud brightness, and cloud extent under different environmental regimes but such an evaluation requires considerable additional effort and requires close cooperation with modeling groups to ensure appropriate interpretation of the results. It is acknowledged in the manuscript that our study merely aims to document the observed relationships in present climate, not to predict how these may have changed since pre-industrial conditions. Our study provides a benchmark of regimes to be used to evaluate how well updated parameterizations capture current signals.

Within section 2.4 Decomposing the ERFaci we point out that we do not use the lowest 12% of aerosol indices in order to reduce biases in regimes where the correlation between our aerosol proxy and CCN is expected to be weak.

The authors ultimately present a correlative study to predict ERFaci (or at least ERFaci for warm-topped clouds over oceans- see comments below). Characterizing covariance is important but does not guarantee an accurate prediction. In the case of aerosol-cloud adjustments in particular, there is not a unique causality flowing from aerosol to cloud (Wood et al., 2012; Gryspeerdt et al., 2019). In this context, and because their ERFaci is rather weak compared to other studies it seems possible that their analysis conflates aci with precipitation scavenging and other confounders (Gryspeerdt et al., 2019), which would tend reduce correlation strength between aerosol and cloud amount (eg precipitation scavenging is strongest when there is a lot of cloud and there tend to be less cloud and more aerosol off the coast of continents).

It should first be noted that our estimate of the warm cloud ERFaci is within the limits of uncertainty (±0.16 Wm^{-2}) of other observation based estimates such as Christensen et al. (-0.36 Wm^{-2}). To address potential biases due to scavenging effects, we explicitly control for precipitation using CloudSat observations that represent the most sensitive satellite-based metric for precipitation occurrence (Haynes et al, 2009). Separating precipitating from non-precipitating clouds in order to understand how precipitation scavenging and other processes that differ between the two alter their ERFaci reduces our decomposed estimate from -0.21 to -0.207 Wm^{-2}. If our estimates were highly affected by precipitation scavenging of aerosol, we would expect the difference between these estimates to be greater.

We acknowledge that our regimes do not capture all signals of covariability between the environment and aerosol and have added to section 2.2 Regimes:
"Using EIS and RH_{700} does not guarantee to limit all covariability between the environment, aerosols, clouds, and their interactions. Some covariability may still exist, such as surface winds affecting both clouds and aerosol (Nishant et al. 2017)."

The authors need to refer to their ERFaci as ERFaci_liquid-topped_over_oceans (or at least that is my take from the methodology and Eq 9).

We have changed all mentions of ERFaci to $ERFaci_{warm}$, RFaci has become $RFaci_{warm}$, and CA has become CA_{warm} in order to remind the reader these results only apply for warm-topped clouds. We have added mentions of our observations being limited to only marine warm clouds throughout section 2.1 Data.

Specific changes:
Pg 1 ln 3: ERFaci is a combination of microphysical (RFaci) and macrophysical changes (adjustments) and the latter could be further split into changes in extent and thickness (Gryspeerdt et al., 2019; Gryspeerdt et al., 2017; Gryspeerdt et al., 2016). As written this implies that thickness stays constant and the only possible adjustment is CF. I understand now that this is more like the intrinsic extrinsic separation in other studies (Christensen et al., 2017), but this would be better to clarify in the abstract.

We have added to the abstract intrinsic and extrinsic to make the connection to the study by Chen et al. 2014 and Christensen et al. 2017 adding next to the RFaci term intrinsic and the cloud adjustment term extrinsic.

Pg. 2 ln 40: The goals of DL19 overlap a lot with the goals of the present study. A sentence like 'The present study expand on DL19 in the following ways:' would be helpful. I believe the primary difference between these studies is the inclusion of adjustments, but it would be helpful to state that explicitly for readers to rapidly ingest what is happening.
We have added to section 1 Introduction:

> *"The present study expands upon work done in DL19 by specifying what aspects of the cloud lead to changes in the CRE, whether that be the brightness or cloud extent or both, and whether these changes can negate each other, such as when a cloud shrinks but the brightness increases."*

Pg. 3 ln 85: It would help readers to quickly process what data sets are being used to describe what variable to use subheaders here (2.1 Data, 2.1.1 Warm cloud fraction). This is stylistic, but I found it hard to understand where precipitation measurements were coming from. I think that it would help a lot to have a table of what the precise data sets used are, especially since some of the remote sensing data sets being used may be inappropriate, but it is unclear if they are actually used (eg AMSR rain rates, although I believe these are not used despite being mentioned).

We have added an additional paragraph in section 2.1 Data to clarify how we separated precipitating and non-precipitating clouds exactly.

> *"Clouds are separated into precipitating and non-precipitating regimes using CloudSat's 2C-PRECIP-COLUMN precipitation flag. Clouds with a 0 precipitation flag, no precipitation detected, are designated as non-precipitating. Precipitating clouds are separated using flag 3, where rain is certain (Haynes et al. 2009). Our precipitating clouds include a majority of the drizzling cases, as CloudSat's 2C-PRECIP-COLUMN's threshold for drizzle is -15 dB, which should capture all but the lightest drizzling clouds (Stephens et al. 2007)."*

Pg4 ln 124: is the material not shown in the citation? If it's in the citation no need to put not shown here.

The material is shown within the citation, we meant to say that we do not show these results within the current paper. We have removed not shown.

Pg 4 ln 125: Swelling is a key issue in trying to understand adjustments. I believe that swelling is not an issue for SPRINTARS because the model can be internally consistent, but an additional comment is needed about MACC aerosol swelling. It's unclear that MACC can fully correct for swelling given the very complex way that swelling occurs (Christensen et al., 2017; Twohy et al., 2009). This needs to be explained and caveated. Also, why mix MACC aerosol and MERRA-2 meteorology? MERRA2 produces a very similar aerosol reanalysis to MACC and this would avoid confusing MERRA2 meteorology with aerosol reanalysis in a different framework. Also- how are SPRINTARS and MACC not sensitive to precipitation scavenging? Presumably both data sets have a precipitation sink of aerosol otherwise it would be very hard to maintain realistic aerosol.

Our results shown do not include any MACC aerosol products. We removed the reference to MACC aerosol in order to not confuse the reader. We have done the same regime analysis with MACC and SPRINTARS AOD for the same time period in order to validate the sign of the regime signals derived here. We have removed the precipitation scavenging mention since SPRINTARS does include some type of precipitation sink for aerosol.

We have added to section 2.4 Decomposing the ERFaci:
"Aerosols swell in the vicinity of clouds, which increases their size and therefore affects the MODIS retrieval AI (Christensen et al. 2017). To assess how significantly this may affect results we have randomly added errors of 10% to our AI estimates and re-derived all signals with all regime constraints. Even with extreme amounts of error in AI, the signals within our environmental and LWP regimes are robust.

Pg.5 Ln140: This methods section is really short. I understand that the authors refer to DL19, but I think it would help readers evaluate this paper more quickly if a paragraph or so was taken to summarize DL19.

We have added to the methods:

> *"In DL19, environmental and cloud state regimes were imposed on a regional basis in order to identify regime specific behavior of aerosol-cloud-radiation interactions. Within each regime, we regressed the cloud radiative effect (CRE) against AI in order to find the susceptibility of warm cloud radiative properties to aerosol. We use these same susceptibilities within section 3.1 to quantify the total warm, marine ERFaci. DL19 found that the susceptibility varies regionally and by regime, however the $ERFaci_{warm}$ depends on the magnitude to which aerosol has increased since pre-industrial times. Further, the $ERFaci_{warm}$ does not diagnose what characteristics of the cloud are causing the effect, prompting us within this paper to decompose the $ERFaci_{warm}$ into the effects on the albedo and the effects on cloud extent."*

Pg. 6 Eq3-6: how do the authors account for CF being bounded between 0-1 in this calculation?

Our cloud fraction is the fraction of a 12 km x 1 km along track region covered in clouds according to CloudSat's 2B-CLDCLASS-LIDAR, which includes even the thinnest clouds not captured by CPR. Therefore, our cloud fractions should be between 0 and 1.

Pg. 8 ln 221: The authors assert that by binning LWP they reduce the chances of buffering. One thing that should be mentioned in this study is that AI and LWP will naturally anti-correlate due to precipitation and scavenging correlating with cloudiness (eg LWP or CF) (Wood et al., 2012) and due to air mass history leading to both drier and more aerosol-laden air (Gryspeerdt et al., 2019). These non-causal relationships are not meaningful to ERFaci, but can substantially affect the covariability of cloud macrophysical properties and aerosol, and thus the inferred aci strength (McCoy et al., 2019). It is possible that the LWP binning and precipitation stratification reduce this effect. However, the authors must show some demonstration of the predictive ability of this method by either (1) applying it to GCM data (in this case SPRINTARS) and showing that their methodology when applied in a GCM can accurately reproduce the GCM response to enhance aerosol as in Gryspeerdt et al. (2016) or McCoy et al. (2019) – or – (2) examining one of the transient aerosol emissions identified in recent studies (Malavelle et al., 2017; Toll et al., 2019) and see if their characterization of sensitivity of cloud to aerosol has some predictive ability. Without this sort of test there is no guarantee that the inferred ERFaci_warm-topped_oceanic is accurate.

We have added to section 3.2 Impact of LWP within the results:

> *"While regime constraints on LWP do reduce the covariability between aerosol-cloud interactions and the role LWP plays in buffering these interactions, it does not remove all sources of covariability between LWP, aerosol, the*

environment, and cloud properties. Aerosol has been shown to negatively correlate with LWP (Gryspeerdt et al. 2019). It is possible that this relationship, and the inherent relationship between the environment and LWP, could affect results shown."

Future work is planned to evaluate how regime-specific relationships compare to those derived via application of similar methods to GCMs, however, as noted above, uncertainty in the parameterization of aerosol-cloud interactions and their regime-dependence preclude drawing concrete conclusions regarding the validity of the methodology from such analyses. More importantly, the resolution of today's GCMs is not sufficient to accurately emulate the distributions of clouds and aerosols on the same scales as the observations so considerable thought and effort will be needed to ensure that the methods can be applied within a model framework in a meaningful way. We agree that the observations have caveats, which we have acknowledged within our manuscript, but we have thoroughly documented our methods, the underlying datasets used, and the analysis approach. As with any study, these choices can be debated and improved upon but the results presented here are (a) an accurate representation of the correlations that exist in the datasets employed; (b) reproduceable; and (c) accompanied by an appropriately large error bar. We believe these data and the analysis method described here represent the current state of the art given current Earth observing capabilities but acknowledge that these estimates will likely be refined in the future.

Pg 10 ln 300: An alternative explanation of the weakening precipitation effect in clouds with higher LWP may be that precipitation increases with LWP, which means that precipitation scavenging becomes larger, which in turn means that the true adjustment strength is obscured by non-causal covariance between aerosol and cloud macro physics (see discussion in McCoy et al. (2019)).

We have added to section 3.2 Impact of LWP:

"An alternative explanation is that thicker clouds with larger LWPs are more likely to precipitate, scavenging aerosol and weakening the susceptibility. Aerosol-cloud-precipitation interactions complicate cloud adjustment processes in higher LWP clouds; the true susceptibility may be masked by covariance between aerosol and precipitation in these clouds (McCoy et al. 2019)."

Figure 7 and ln 456: The authors find a large ERFaci in the SH, which is really surprising given the very small change in anthropogenic aerosol in these regions. Figure 1 shows change in AI, but it is a bit hard to distinguish small changes from zero and the authors may want to consider some sort of log normalization to their color scale. However, strong ERFaci exists along a line around 40◦S, which is hard to square with studies examining pristine days in the PD (Hamilton et al., 2014). That is to say, the pattern of ERFaci in this study is dramatically different than the RFari shown in, for example, aerocom (Myhre et al., 2013).

Since our estimates of the ERFaci are weighted by occurrence, regions with the highest occurrence of warm clouds will have larger ERFaci. The southern hemisphere is known to have the largest occurrence of warm cloud decks, therefore our weighted ERFaci from observations will weight the southern hemisphere over the northern hemisphere. Further, the southern ocean may have a higher susceptibility due to their usual pristine conditions making them primed and highly sensitive to any changes in aerosol.

Figure 7: While I think it's good to pursue analysis to its conclusion by applying it to all data, I am surprised at the positive RFaci and CA in the tropics. Can the authors comment on whether their analysis is sensitive to retrieval errors in convective cloud? In particular, a positive forcing due to RFaci is quite unusual- while it may be due to biomass burning aerosol above cloud in some regions via semi-direct effects or blocking reflective light (so not really aci) (Bellouin et al., 2019), the appearance of a positive RFaci seems to be more related to SST, than aerosol type given its appearance over the tropics, and far away from strong aerosol sources.

A limitation of our data is that the cloud radiative effect can be reduced due to semi-direct effects not constrained by our environmental or LWP limits.
We have added to address that the semi-direct effect is not accounted for by our methodology and may result in a reduced RFaci, in Results and Discussion section 3.3. Constrained by local meteorology

> *"It is also possible lambda_RFaci is impacted by some semi-direct effects by smoke aerosol which would lead to a cloud dimming and positive susceptibility. Semi-direct effects are not accounted for by our methodology, however aerosol within the cloud layer could lead to cloud breakup processes, a dimmer albedo, and changes to the local environment by the absorbing aerosol."*