The authors are very grateful for the helpful comments provided by both reviewers. Our point-by-point responses are provided below in blue directly after the reviewers' comments.

To summarize, we made 5 main changes in the manuscript:

1) we now use the latest 2b-cldclass-lidar release (R05) in the case studies and statistics since the changes compared to R04 are substantial –although the conclusions remain unchanged–, which is not the case for the RL-GeoProf (reviewer 1).
2) we further justify future use of CASCCAD dataset for model evaluation in the conclusions (reviewer 1).
3) we now compare the CASCCAD datasets with the passive-sensor clustering datasets in addition to the COT-CTP technique (reviewer 1 and 2).
4) we further define the content of the CASCCAD dataset at the beginning of the dataset section (reviewer 2).
5) finally, we slightly modified the CASCCAD algorithm to keep track of the cloud type called Cu with stratiform outflow, which was accounted for as pure Cu before. As a result:
    a. the algorithm chart flow (fig. 2) has been updated as well as the cartoon explaining the VCF and horizontal continuity test logic (Fig. 5).
    b. the case studies have been updated although the changes are minor
    c. the global statistics are slightly changed in RL-GeoProf CASCCAD (a little less/more Cu/transitioning cloud) while it is negligible in GOCCP-CASCCAD.
    d. these changes do not impact the conclusions of the study

Reviewer 2

This paper describes a new global cloud dataset built from active sensor CALIPSO and CloudSat observations. The distinguishing feature of the dataset is that cloud profile data is used to discriminate between stratocumulus and shallow cumulus, rather than relying on cloud-top altitude and optical depth derived from passive sensors. The motivation for this effort is nicely summarized in the introduction. The paper is well written and the work is a useful contribution to the field. The paper deserves publication but would benefit from some additional discussion and clarification of a few points.

1. The algorithm used to identify and discriminate cumulus and stratocumulus is described well, but the paper needs some description of the actual CASCCAD data set:
Is this a Level 2 or Level 3 dataset? What is the spatial resolution of the product? What parameters are reported – are cloud base and top altitudes included?
Section 5 says CASCCAD dataset is formed using both GOCCP and RL-Geoprof. How? Are these two datasets merged together somehow, or are they included as two separate elements of CASCCAD?
We acknowledge that the description of the dataset was somewhat ambiguous in the previous manuscript. The CASCCAD datasets are actually two distinct datasets using the same algorithm but applied to two different level 2 observational datasets, which are CALIPSO-GOCCP and RL-GeoProf. There are level 2 and level 3 CASCCAD files for both datasets. The level 2 files use the native resolution of the GOCCP and RL-GeoProf datasets (i.e., 333 m x 480 m for GOCCP and ~1.1 km x 240 m for RL-GeoProf). The level 3 files are global statistics over a 2.5˚x2.5˚ grid over 40 480 m levels for each month between 2007 and 2016, which are then averaged over the time dimension. The monthly means will be made available once the manuscript is accepted. The parameters reported are 2D and 3D cloud fractions of low, Cu, Sc and transitioning clouds for level 3 files and cloud mask profiles of Cu, Sc and transitioning cloud types for the level 2 files. While the cloud top/base is not reported explicitly as a variable for each profile, it can be derived from the cloud mask contained in level 2 files.

We added a full paragraph at the beginning of section 2 –before section 2.1– to address the reviewer's concerns and describe what we report above. Additionally, we make it clear throughout the manuscript that the GOCCP and RL-GeoProf CASCCAD datasets are two distinct datasets based on the same algorithm but using GOCCP native level 2 observations and the combined CloudSat-CALIPSO RL-GeoProf level 2 observations, respectively. For example, in the last paragraph of the introduction, the first of section 4.1 and the first paragraph of section 5, we now specify that the algorithm is apply "separately" to both GOCCP and RL-GeoProf observations. Finally, in the conclusion, when the CASCCAD datasets are mentioned we now refer to "distinct GOCCP and RL-GeoProf CASCCAD datasets" rather than only "CASCCAD datasets".

2. As reported here, classifications based on observed 3D structure appear to be better than classifications based on passive retrievals of cloud altitude and optical depth, which are subject to ambiguities. It would be useful to cite Mace and Wrenn (2013) who discuss a comparison in the eastern Pacific
Mace and Wrenn, 2013: Evaluation of the Hydrometeor Layers in the East and West Pacific within ISCCP Cloud Top Pressure-Optical Depth Bins Using Merged CloudSat and CALIPSO Data, J. Climate, doi:10.1175/JCLI-D-12-00207.1
Thank you for pointing out this relevant reference. We added it to the manuscript at the beginning of section 4.2.1 where the comparison between passive and active is performed:
In addition, (Mace and Wrenn, 2013) showed that except for thin cirrus and Sc cloud types, the COT-derived cloud types are mostly mixtures of different cloud types in two regions of the Eastern Pacific.

3. Referee #1 dismisses the utility of CASCCAD for model evaluation for reasons which are not clear. CASCCAD appears quite useful for model evaluation to me. Some additional discussion of this would be useful, especially to clarify differences in using GOCCP (which is based on a simulator approach) versus RL-Geoprof (which is not).
We agree with reviewer 2 and we added some additional discussion –as requested by reviewer 2– to better explain how we believe CASCCAD can be used for model evaluation in the conclusion. A point by point answer is given to reviewer 1 above (however, we added our answers to reviewer 1 with respect to the use of CASCCAD for model evaluation at the end of this response for reviewer's 2 convenience). Based on the case studies, profile and zonal means presented in the paper, the biggest differences between CASSCAD versions for the two datasets mentioned above are the tendency for the RL-GeoProf version to diagnose more transitioning clouds than the GOCCP version poleward of 40˚, because i) RL-GeoProf is less affected by overlapping mid- and high-cloud attenuation and better captures the full vertical extent of low clouds and ii) RL-GeoProf contains larger cloud clusters than GOCCP, making the continuity test more efficient . We now mention this in the second paragraph of section 4.2.1.

4. Finally, I agree with Referee #1 that it would be interesting and useful for the authors to say something about the weaknesses and strengths of CASCCAD vs. weather state approaches to analyzing passive observations.
The WS and CR approaches do not exactly classify clouds by cloud type or by low, middle and high layers although they provide CTP-tau histograms for each WS/CR. Therefore they represent mixtures of cloud types among which one cloud type is often prevalent. As such it is not exactly consistent to directly compare with the CASCCAD datasets, however, we agree with the reviewers that showing the CR and WS more recent approaches would provide a broader context and additional information as their horizontal and time sampling and their time record are better than CloudSat-CALIPSO.
For this reason, we added the ISCCP WSs and MODIS CRs observations in section 4 of the analysis.

Minor points:

The last sentence of the first paragraph of the Introduction repeats an earlier sentence.
The first occurrence of this sentence has been removed.

To document limitations of CloudSat in sensing shallow clouds it would be useful to cite: Liu, D., Q. Liu, L. Qi and Y. Fu, 2016: "Oceanic single-layer warm clouds missed by the Cloud Profiling Radar as inferred from MODIS and CALIOP measurements", JGR, doi:10.1002/2016JD025485.
Thank you for bringing this reference to our attention. We modified the sentence at L26 of P3 to include that reference: "*Although the radar-only product extends over a longer time-period (for daytime only, see section 2.2), the CPR is less sensitive to fractionated and thin shallow cumulus clouds than the CALIPSO lidar and its ground clutter prevents cloud detection below 1 km, which preclude the detection of a large amount of marine low-level clouds (mostly Sc and Cu clouds, Liu et al., 2016).*".

"CALIPSO" is mis-spelled in the caption of Figure 6:
This was corrected

Responses to reviewer 1 with respect to model evaluation using CASCCAD:

-Page 10, line 29: I would remove this statement unless you want to elaborate. I can't see how you could use this data to evaluate a GCM in anything but a qualitative way. There is generally nothing in the model called Sc and Cu to compare against. Also, there is no way to apply an instrument simulator to recreate the data using the model output because the categories depend on spatial continuity.
We disagree with the reviewer's comments. Models make Sc and Cu explicitly, based on specific but different physical mechanisms. We understand that some models are now using unified turbulence schemes of one type or another, but these schemes still have to produce Cu and Sc and the transition between them under appropriate conditions. We discuss in response to a later comment why CASSCAD is valuable for the evaluation of such parameterizations. We also note that Reviewer 2 feels that CASSCAD is useful for model evaluation.
It is true that it would be difficult to recreate the same Cu Sc diagnostic in the simulator as in the CASCCAD algorithm, however:

1) The simulator sums up the convective and stratiform cloud fraction before computing the diagnostics and therefore one could just separate their contribution rather than summing them up.
2) A simulator is not necessarily needed for this particular type of comparison, because we identify the different cloud modes explicitly and we can select regimes in which lidar attenuation is negligible (e.g., ω500 > 0 hPa/day, Cesana et al., 2019a)

We expanded the last paragraph of the conclusion to discuss this: "*Finally, one of the reasons we developed CASCCAD is to provide an improved observational constraint for low-level cloud feedbacks in GCMs. Although the CASCCAD DA cannot be implemented in a lidar simulator (Chepfer et al., 2008), it is still possible to use CASCCAD datasets for model evaluation because i) both the convective and stratiform cloud fraction are provided as inputs to the lidar simulator and could be easily saved separately rather than summed up; and ii) a simulator is not necessarily needed for model-to-obs comparison of Cu and Sc clouds over the tropical oceans, because we identify the different cloud modes explicitly and we can select regimes in which lidar attenuation is negligible (e.g.,* $\omega_{500} > 0$ *hPa/day, Cesana et al., 2019a).*"

-Page 123, line 20: Again, I don't know how you use this to evaluate a model. Just because the model has some arbitrary distinction between a boundary layer parameterization and a shallow convection parameterization doesn't mean that this is the same as stratocumulus and cumulus as you have defined them from the observations. Furthermore, newer parameterization are beginning to 'unify' these distinct regimes (e.g. CLUBB, EDMF). I think the dataset is interesting enough in its own right without having to sell it as a model evaluation tool.

Here we also disagree with the reviewer's comment. As stated above, the models do actually simulate Sc and Cu based on physical mechanisms. Existing GCMs fall into two classes: (1) Those that use separate parameterizations for cumulus and stratocumulus clouds, for which CASSCAD is directly applicable as an evaluation tool to determine whether biases are attributable to process assumptions made about one cloud type vs. the other. (2) Those that use unified turbulence parameterizations that are intended to represent the full spectrum of boundary layer clouds. Even for the latter class of parameterizations, though, the models are in effect capturing the two cloud types and the transition between them via different segments of the pdf and making assumptions about when each type of cloud is active, either in the physics itself or in the diagnosis of the model behavior. In both cases, the distinction between Cu and Sc is made crudely using criteria such as inversion strength or 500 mb vertical velocity as the "definition" of one vs. the other cloud type. For example, Koehler et al. (2011, QJRMS), in an EDMF scheme, use inversion strength as a "decoupling criterion" to determine when they should turn off the shallow convection (MF) component of EDMF. Bogenschutz et al. (2013, J. Climate) use inversion strength and vertical velocity to define "stratocumulus," "transition," and "cumulus" regimes for the purpose of understanding where CLUBB is performing well vs. poorly. We feel that a dataset that directly diagnoses the clouds themselves is ultimately a more reliable indicator than assumptions about cloud types based on large scale environmental properties. CASSCAD in fact allows the relationship between Sc, Cu, and environmental state to be diagnosed directly rather than being assumed. Given the large impact that schemes like CLUBB have on climate sensitivity (Gettelman et al. 2019, GRL), there is an increasing need to use metrics beyond mean state biases to decide what is and is not realistic. Note that we did not explain this in the previous manuscript. We now specify this in the description of the algorithm, section 3.1: *"However, one cannot separate clouds according to the mechanisms that form them as GCMs do using different PBL and convective parameterizations, which is why we choose to use the morphology to discriminate cloud types in this study."*

Based on our study, it seems that CASCCAD is doing better at identifying Cu-Sc clouds than the other methods –while it is not perfect– and therefore CASCCAD looks like a good candidate for model evaluation.

We now specify this in the last paragraph of the conclusion: *"Finally, … to judge model realism and fidelity".*