We would like to thank the reviewer again for his/her comments. Our response is shown in blue color.

This is the second round of review.
First I want to remind the authors that as a reviewer, I'm required to give my opinion on certain criteria, including the novelty of a study and whether its length is appropriate or not, which I'm doing here based on my (imperfect) knowledge of the literature and regardless of what other reviewers may say. However, it doesn't mean that I find the study irrelevant or not good.

That being said, the authors addressed most of the points I rose in the first round of review but some concerns remain and need to be addressed before I recommend the paper for publication. These are listed below with additional comments regarding some author arguments.

Comments that do not need to be further addressed
1) Novelty of the study
I just want to remind the author that previous literature already largely investigated the cloud phase characteristics globally (e.g., Cesana et al., 2015; Hu et al., 2010; Li et al., 2017; Matus and L'Ecuyer, 2017; Yoshida et al., 2012). Some of that literature also studies the link with radiation for different overlap using the very same product as the authors (e.g., Matus and L'Ecuyer, 2017). Focusing on a specific region doesn't make a study novel, but again it doesn't mean it's not worth being published. However, I acknowledge that the spatial heterogeneity component of the study ¬¬–and its link with cloud phase– is quite new and interesting.

We thank the reviewer for this comment. We are also aware that a series of research has focused on the characteristics of cloud phase (with no overlap information), cloud overlap (with no cloud phase information) and their linkage with radiation (largely broadband not spectral). Paragraph 2 of Introduction reviews this literature and states our motivation to investigate the characteristics of cloud phase overlap and its link to spectral and spatial heterogeneity signatures.

Many references mentioned here are already in the paper. We add Cesana et al. 2015, and Yoshida et al. 2012 to Introduction section to make the literature review more complete. Thank you.

2) Length of the study
While I appreciate the author efforts to shorten the manuscript, I still find the study quite long, but it doesn't bar it from being published.

We thank the reviewer.

Concerns that need to be addressed
1) Uncertainties and caveats related to the observational product
- The authors now better mention the caveats and uncertainties of the product, which is good, in particular the reduction in confidence of the diagnostic when the lidar is completely attenuated. However, they fail to mention how often this happens in their study, a breakdown depending on the category would be helpful (i.e., how often the diagnostic relies on radar only by category).

This region is dominated by deep convection and therefore I would expect most of the observations to be radar only, which is why it has to be quantified, it's essential information for the reader.

Quantifying how often the cloud classification relied on radar-only signals is helpful in certain studies, but not so here. First, it is true that deep convection is active in this region. But, the most common cloud type is cirrus, some of which are spawned by deep convective clouds. Table 3 shows that the combination of ice-only, ice-over-liquid, and ice-over-mixed represent ~71% of all cloud categories. These ice-containing categories occur predominately for cirrus with optical depth < 3.0 (Figure 6), as they require lidar beam penetration for classification. For mixed-clouds, which includes deep convective clouds, the detection of ice at the top with the aid of the lidar is sufficient for classification (even though the lidar beam gets completely attenuated), since the 2B-CLDCLASS-LIDAR only provides the information for cloud layers, not for each radar range gate – so, deep convective clouds are classified as mixed-phase cloud, though there could exist only water at the bottom of the cloud (as discussed in Section 2.1). This last point may well be why the 2B-CLDCLASS-LIDAR product doesn't archive the information when lidar/radar signal is available.

- The authors say "the most comprehensive cloud phase and overlap information to date" p4 L37 I disagree with that statement. There is no paper that supports this statement to the best of my knowledge. The authors themselves stated that they are not aware of any kind of validation of the 2B-CLDCLASS-LIDAR product, which makes it difficult to conclude on whether this product is the most comprehensive to date. There are at least 2 other LIDAR-RADAR cloud phase datasets out there using different methods (DARDAR and Kyushu University products) as well as 3 lidar-only cloud phase products (CALIPSO-ST, GOCCP and Kyushu University), some of which have been validated against ground-based or in-situ measurements contrary to 2B-CLDCLASS-LIDAR product. Please, rephrase.

We thank the reviewer for this comment. What we want to express is that the combined radar-lidar measurements are able to provide comprehensive cloud information including cloud phase and cloud overlap.

The original statement is now rephrased as:
Despite these limitations, the combined radar-lidar measurements provide comprehensive cloud phase and overlap information.

- Finally, the method used by the authors to account for the cloud fraction (i.e., cloudy profile each time the lidar cloud fraction within the cloudsat volume is greater than 0) leads to an overestimate of the cloud fraction in regions of fractionated clouds such as the trade winds and this should be explicitly mentioned in the manuscript (e.g., Cesana et al., 2019; Marchand et al., 2010).

We do not calculate the cloud fraction, such as in Marchand et al. (2010). Instead, we calculate the occurrence frequency of the radar column containing some cloud, even if not fully cloudy. When the lidar cloud fraction is greater than zero, we count the radar column as containing some cloud. A threshold similar to Cesana et al. 2019 is not adopted since we wanted to include small-size clouds detected by the lidar in our analysis (e.g., Section 3.1.5) and avoid issues with traditional resolution and thresholding effects on cloud fraction as discussed in Marchand et al. (2010) and many of the references we cite.

In P4 Line 39-40:
We revised the statement by adding 'in order to include small size clouds':
Four years of 2B-CLDCLASS-LIDAR data, version P1_R05 (2007-2010) with lidar cloud fraction greater than zero are used in order to include small size clouds.

2) Climate model evaluation argument
I appreciate the effort of the authors to clarify how to use their results to inform model simulations. However, I'm still not convinced by their explanation. The 2B-CLDCLASS-LIDAR cloud ice and liquid frequency cannot be used to evaluate climate models. There are no cloud ice or liquid frequency in the models to compare with. Also, if such diagnostic was available, it would still be not consistent to directly compare the observations with the models without using a method that takes into account the inherent biases of the instruments. For example, one should use a forward simulator that reproduce the 2B-CLDCLASS-LIDAR product process and biases to compare with the models (see for example Masunaga et al., 2010 and Hashino et al., 2013 referenced by the authors, and many other not referenced here). Such simulator doesn't exist.
Additionally, a quick look at Loveridge and Davies –referenced by the authors as an example of how to use their heterogeneity index for model evaluation– also shows that they use a simulator in their study to reproduce MISR and MODIS quantities, then compute their Hindex and evaluate the heterogeneity parametrization. A GCM grid box is typically on the order of hundreds of kilometers with the most recent one being on the tens of kilometers, which is still far larger than the 1km pixel size used in MODIS observations. This is why it can't be used directly to evaluate a GCM (although not true for finer scale models).
I understand how it could be useful for observations as explained in the paper, but in its actual state, these observations cannot be used for pure model evaluation. Therefore, I'm still recommending to remove these statements of the manuscript.

We appreciate the reviewer's concerns. However, satellite cloud product summaries have been used to gauge model performance for more than 40 years. True, there are many issues in doing so, but the practice continues. The use of forward model simulators, in part, attempts to address some, but not all, of the issues. Just because a forward simulator for the 2B-CLDCLASS-LIDAR product doesn't exist, it doesn't mean that someone won't build one in the future. Moreover, our paragraph emphasizes the forward simulation of radiances form MODIS as the basis of comparison.

With the reference to Loveridge and Davies work, they did not use H_σ for direct evaluation of any quantities in the model. Instead, the H_σ was used to interpret the remote sensing data, arguing that differences in cloud optical depth (for example) between satellite and model should be considered in context of subpixel heterogeneity information such as H_σ. So H_σ is used as part of the model evaluation process (not direct comparison) to identify high confidence observations that act as strong constraints on the model. This is also the point that we emphasize: ' careful comparisons between model and observations can use $H\sigma$ as a measure of departure from the plane-parallel assumption in a manner similar to (Loveridge and Davies, 2019), where they used $H\sigma$ within their analysis in examining GCM clouds in different sectors of southern hemisphere cyclones.'

While we disagree with the reviewer and are confident that our statement is on target, we have modified our original statements as (P20 L13-19):

Finally, careful comparisons between model and observations can use $H\sigma$ as a measure of departure from the plane-parallel assumption in a manner similar to (Loveridge and Davies, 2019), where they used $H\sigma$ within their analysis in examining GCM clouds in different sectors of southern hemisphere cyclones. $H\sigma$ can also be used to gauge biases in other satellite products that are used in model evaluation (e.g. Gettelman et al., 2015; Song et al., 2018), such as cloud optical depth and effective radius, whose biases have been noted to covary with H_σ (Fu et al., 2019; Zhang et al., 2016).

References

Cesana, G., Waliser, D. E., Jiang, X., & Li, J. L. F., (2015), Multimodel evaluation of cloud phase transition using satellite and reanalysis data, Journal of Geophysical Research, 120(15), 7871–7892. https://doi.org/10.1002/2014JD022932

Cesana, G., Del Genio, A. D., & Chepfer, H., (2019), The Cumulus And Stratocumulus CloudSat-CALIPSO Dataset (CASCCAD), Earth System Science Data Discussions, 2667637(November), 1–33. https://doi.org/10.5194/essd-2019-73

Li, J., Lv, Q., Zhang, M., Wang, T., Kawamoto, K., Chen, S., and Zhang, B.: Effects of atmospheric dynamics and aerosols on the fraction of supercooled water clouds, Atmos. Chem. Phys., 17, 1847–1863, https://doi.org/10.5194/acp-17-1847-2017, 2017.

Hu, Y., Rodier, S., Xu, K. M., Sun, W., Huang, J., Lin, B., et al., (2010), Occurrence, liquid water content, and fraction of supercooled water clouds from combined CALIOP/IIR/MODIS measurements, Journal of Geophysical Research Atmospheres, 115(19), 1–13. https://doi.org/10.1029/2009JD012384

Marchand, R., Ackerman, T., Smyth, M., & Rossow, W. B., (2010), A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS, Journal of Geophysical Research Atmospheres, 115(16), 1–25. https://doi.org/10.1029/2009JD013422

Matus, A. V., & L'Ecuyer, T. S., (2017), The role of cloud phase in Earth's radiation budget, Journal of Geophysical Research, 122(5), 2559–2578. https://doi.org/10.1002/2016JD025951

Yoshida, R., Okamoto, H., Hagihara, Y., & Ishimoto, H., (2010), Global analysis of cloud phase and ice crystal orientation from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) data using attenuated backscattering and depolarization ratio, Journal of Geophysical Research Atmospheres, 115(16), 1–12. https://doi.org/10.1029/2009JD012334

Koren, I., Oreopoulos, L., Feingold, G., Remer, L. A. and Altaratz, O.: How small is a small cloud?, Atmos. Chem. Phys., 8, 3855–3864, doi:10.5194/acp-8-3855-2008, 2008.