Decision
In this new version of the manuscript, the authors have addressed part of my concerns by documenting more carefully the observational datasets and by clarifying the goal of their study. However, they still compare directly the model outputs to satellite observations in section 3.6 - making it an apples-to-oranges comparison - and draw wrong conclusions based on this. Therefore, I would recommend revisiting this section and updating the conclusions accordingly – along with a few minor things – before publication.

Main comments:
There is a confusion in the paper between optically thick and geometrically thick although I reckon that 500mb thick clouds may also correspond to optically thick clouds. This should be clarified in the manuscript.

Additionally, my main concern is about Section 3.6. While I acknowledge that the authors produce a lot of efforts to improve this section, it is still confusing for the reader and needs further work. In addition, I disagree with some conclusions made by the authors that are not supported by evidences.
At the beginning of the section, the authors state that the comparison is done using the simulator. However, the first figure doesn't show any simulator results and most of the conclusions are (wrongly) made based on this particular figure. It's even more confusing as the authors state "here we use COSP simulator…" right after mentioning Fig. 7. The reader is left thinking that he can compare all plots in Fig. 7 in a consistent way.
Yet, the differences shown here are due to an apples-to-oranges comparison: "mass phase ratio vs. frequency phase ratio". The satellite observations should be compared with the model output+ simulator only. The authors have the model+simulator outputs so they should add a row to Fig. 7, reproduce the top row using the model+simulator outputs and move the observations to that bottom row (and delete fig. 8). In addition, as mentioned in the previous round of review, the authors should treat the undefined-phase clouds as being mixed-phase clouds. Their lidar return as well as their mixing ratio are dominated by the ice phase, which is why they should be accounted for in the ice type of clouds (e.g., Cesana et al., 2016).

From this section and section 5, the authors conclude that "*the formation pathways revealed that most of the simulated ice clouds are in the blind spot of the lidar in the lower part of optically thick clouds*". This statement is not supported by evidences. If the authors want to find out about this, they should i) filter out these clouds of the statistics and ii) select them only to reproduce fig. 7 in both the model outputs and model+simulator outputs. Only then, the authors would be able to conclude what is the effect of these clouds on the MPR/FPR vs T relationship. However, these conclusions would be valid only for this particular model.
At the moment, nothing allows one to conclude that these thick clouds (less than 30% of all clouds) are responsible for the differences between model outputs and model+sim outputs and even less that the bottom part of these clouds is.

To further show that these clouds probably do not affect much the FPR-T relationship I provide below a brief analysis using the observation. CALIPSO-GOCCP now offer new information about the opacity of the column. It is possible to discriminate lidar shots that reach the ground from those completely attenuated (Guzman et al., 2017), the latter being less frequent than the former. After computing the ratio of thin cloud to thin+opaque clouds, I reproduced Fig. 8 for all clouds (black-dash line) thin-cloud dominated gridboxes (ratio>0.5, magenta dash-line) and opaque-cloud dominated gridboxes (ratio<0.5, green-dash line). Undefined-phase clouds are threated as ice clouds (as mentioned earlier). This highlights the very small impact of *optically thick* clouds on the FPR-T relationship.

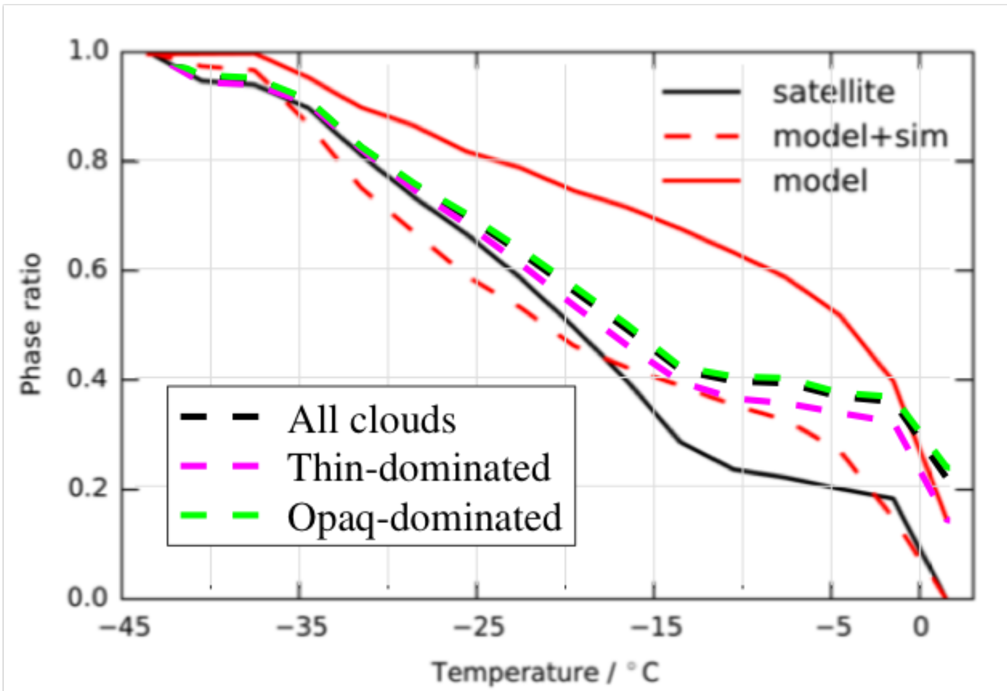

Figure 1: Reproduced Fig. 8 with CALIPSO-GOCCP FPR for all (black-dash), thin-dominated (magenta-dash) and opaque-dominated (green-dash) clouds. Note that the solid black line is different from the black-dash line because the undefined-phase clouds are treated as ice clouds: FPR is the ratio between ice+undefined-phase clouds to ice+undefined-phase + liquid clouds. This plot was created using 2x2˚ monthly CALIPSO-GOCCP observations version 3.1.2, nighttime only, for the period 2007-2016 (Map_OPAQ and 3D-CloudFraction_Temp files).

Minor comments:

P1 L17: The authors do not demonstrate this in their study, they just show that the mixing phase ratio is substantially different from the frequency phase ratio in their model, and that roughly 30% of the clouds are “geometrically” thick clouds in their model. This doesn’t mean that the differences MPR/FPR are due to the bottom of the geometrically thick clouds (called optically in the abstract).

P2 L3: Community ➔ Cloud

P9 L13: make “use” of…

P9 L20: from 2006-2012? In the caption of Fig. 3 it is written 2008-2014.

P9 L24-28: anomalies ➔ bias
I guess you refer to the change from blue to red while stating “distinct jump”. It might be good to state this in the text.

P9 L28 “is differ less”

P9 L29: This actually shown in Cesana and Waliser, 2016. If the overlap were in better agreement with the observations, the cloud amount overestimation would be substantially larger, such as that found in Fig. 3.

Fig 4 & 5: the 2M lines are difficult to spot because all other lines are on top of it. It might be worth it to replot this with 2M on top.
When I suggested to add the +/- 1STD I meant annual STD, so that the frequency of annual oscillation could be taken into account when compared to the model. I believe the authors show here a monthly interannual STD rather than a yearly interannual STD. I’d recommend that the authors change the monthly interannual STD to yearly interannual STD in these figures. I did it myself for the LW/SW CRE and the results is far smaller than what shown here.

P13 L13: it is not based “on a threshold value for the backscattered polarization ratio to separate…” but on the cross-polarized attenuated backscatter and the total attenuated backscatter and a discrimination line. Note that this process is reproduced in the simulator
P13 L14-L15: The lidar spatial resolution is actually around 70m while the distance between lidar shots is ~333m.

P13 L33: They are more frequent in the 2M simulation than the REF and it should be stated. While not comparable, it is also much more frequent in the observations.

P14 L1: I don’t understand this statement.

P19 L17-18: This is wrong, it is due to comparing two different quantities, which do not take into account the instrument peculiarities. One should only conclude based on comparison between obs and model+sim outputs.

P19 L18-19: As explained before, this is not what’s shown in the paper.

P19 L19-20: I completely disagree with this statement, which is not supported by any evidence. Even if it was shown that the PR-T definition is substantially different in ECHAM6 for thick and thin clouds (which is not performed so far), it would only suggest that the phase simulated with the lidar simulator cannot be used to assess the PR-T relationship of all clouds in this particular model.