

## REPLY to REVIEWER 2

We thank the reviewer for their review and their detailed comments. Below you will find the reviewer's comments in bold and our replies.

**Section 2: I'm not sure I agree with the methodology used to calculate the expected number of calibration points per unit time presented in the manuscript, and will outline my concerns here, though I welcome input from the authors clarifying if I have missed a step or am misunderstanding their approach. The first point I would like to clarify is that the quantity referred to as "monthly intersections" throughout the paper is in fact more aptly described as the monthly-average number of instrument footprints satisfying the coincidence criteria gridded at 2 x 2 degree resolution. That is to say, for a single intersection of the orbit tracks of the two spacecraft (assuming nearly temporal coincidence), there are many points that will satisfy the spatiotemporal criteria and be logged as many "intersections" in a single 2 x 2 deg grid cell. If this is true, the problem for me arises when this "weekly/monthly intersections" variable is multiplied by the "mean number of ice layers" variable to get the number of weekly/monthly calibrating points. Specifically, I interpret the distributions in figure 5 as the mean number of ice cloud range bins per 2 x 2 degree grid cell, and not the mean number per horizontal radar footprint gridded at 2 x 2 degrees. Thus, to properly calculate the number of coincident ice cloud range bin detections within a given 2 x 2 degree grid cell, I would calculate it as: (mean number of ice cloud bins per 2 x 2 deg)\*(number of footprints within the 2 x 2 deg grid cell satisfying the coincidence criteria per unit time)/(number of footprints per 2 x 2 deg grid cell).**

Yes we agree with the reviewer and indeed we have used a misleading terminology. With "intersections", or "coincidences", indeed we mean the monthly-average number of instrument footprints satisfying the coincidence criteria gridded at 2 x 2 degree resolution. They are computed over a year time span and are averaged over a different time units. E.g. in figure 3 we plot the monthly-average number of instrument footprints satisfying the coincidence criteria gridded at 2x2 degrees.

Figure 5, as the reviewer has noticed, depicts the mean number of ice cloud range bins with a resolution of 500 m per 2 x 2 degree grid cell. This means that a number of 2 corresponds to an area where, on average, there are ice clouds with a vertical thickness of 1000 meters. Then, we multiplied each pixel of the grids reported in Figure 3 for the corresponding pixel of the grids reported in Figure 5 in order to obtain similar grids for calibration points. So, basically, we weight each intersection point on the number of calibrating target we have at that point (e.g. if an intersection occurs where no calibration target is present, that intersection can't be a calibration point; if it occurs where two calibration targets are present, we have two possible points that can be used to calibrate the radar. So, we computed the distribution of calibration points as: (mean number of ice cloud bins per 2 x 2 deg)\*(number of footprints within the 2 x 2 deg grid cell satisfying the coincidence criteria per unit time). We didn't divided for the "(number of footprints per 2 x 2 deg grid cell)", because we are interested in having an absolute number of calibration points. Then, we summed all the elements of the grid containing the number of calibration points, to obtain the total number of calibration points per unit time.

**Lines 130-135: The authors state that due to the low attenuation in ice at Ka and W-band, the measured reflectivity of a given ice cloud target will be nearly identical from near-nadir and large off-nadir viewing geometries. This seems to neglect the possibility of backscatter dependence on angle for oriented ice crystals, which are known to exist in certain stratiform regimes. The potential impact of this effect on the proposed calibration method should be addressed in this manuscript.**

In general the reflectivity may be slightly different in presence of non-spherical particles which are preferen-

tially oriented. However for the calibration we do use reflectivity values that are large, thus corresponding to generally more randomly oriented aggregate. These scatterers tend to have backscattering cross sections independent from incidence angle. With respect to the attenuation the Hitschfeld–Bordan attenuation-correction could be implemented to account for the differential signal between the two viewing directions.

**I find the description of the Z-PDF method at the beginning of Section 2.3 somewhat confusing. My interpretation of the procedure is that a correlated pair of PDFs is calculated for 5-km-long along-track swaths of radar reflectivity profiles that are separated by  $\Delta s$ , and this pair could be labeled with an index “i”, say  $f_i(Z)$  and  $g_i(Z)$ . The JS distances are calculated for these specific PDFs  $f_i(Z)$  and  $g_i(Z)$  with index i and then an ensemble of such realizations is made by finding N ( $i = 1, 2, \dots, N$ ) pairs of such spatially separated PDFs. In lines 185-188, the single-swath, spatially separated PDFs are defined, but then it is said that the PDFs are accumulated up to a specified sample size before the JS distance is calculated between “the two PDFs”. It isn’t totally clear what is meant by this accumulation, and could be interpreted as accumulating counts across many spatially separated observation points to form a single PDF consisting of reflectivity observations from N swaths. This is confusing since it seems to be the case that the JS distance can only be meaningfully calculated on isolated pairs of single swath PDFs. I think cleaning up the wording and notation in this section would make the manuscript much more easily readable.**

We agree with the reviewer that the description was confusing. The key of the methodology here presented is to build correlated pair of PDFs which should simulate the PDFs as collected by the calibrating and by the to-be-calibrated radar. For each of the quasi-coincidence criteria we have constructed these PDFs by using points collected by CloudSat or GPM that are separated by a distance  $\Delta s$  like illustrated in Fig.4. In each of the 5-km thick layers only a limited number of calibrating points is present. In order to smooth the pdfs a large number of calibrating points is needed. This is achieved by putting together 5-km thick belonging to different orbits. By so doing two PDFs  $f_1(Z)$  and  $g_1(Z)$  are obtained with a characteristic number of sampling points (higher number of points obviously requiring larger number of orbits). This number of sampling points is increased to assess the impact that the number of calibrating points has in reducing the sampling noise. The exercise is then repeated multiple times (thus involving different orbits or different segments of the same orbit) to find an ensemble of such pair PDFs. We name these realizations  $f_i(Z)$  and  $g_i(Z)$  with  $i = 1, 2, \dots, N$ . For each pair, the Janson-Shannon distance  $d_{JS}[i]$  is computed. If N is large enough the pdf of all  $d_{JS}[i]$  gives an idea about the expected distance between 2 PDFs and its natural variability associated to the fact that in our calibration methodology clouds not from coincident but from quasi-coincident overpasses are sampled.

#### Minor comments

**Line 31: It is not clear why conical scanning permits larger domain coverage, unless the statement is more about the difficulty of reaching the same large incidence angles using a cross track technique vs using a conical scanning approach.**

Yes we agree with the reviewer. In principle conically scanning radars can achieve larger domain coverage because they can reach larger incidence angles, but yes in principle also cross-tracking radars they have the same capability. We will make this clear in the revised version.

**Line 31: The considerations around surface clutter at large off nadir vs near nadir incidence are very different in nature, and it is not a given that one offers lower clutter free heights in general than the other. Specifically, beam width determines the clutter height at large off nadir angles, and the problem is essentially independent of the radar waveform parameters/ the range weighting function. The situation is completely different for near nadir scattering**

where the beam width has little to no impact on the clutter free height, and it is the range weighting function as well as the transmitter phase noise (for high time-bandwidth product PC systems) that determines the clutter. This statement on line 31 should be clarified or removed.

Agreed. We will remove the statement relative to the clutter.

**Line 4 and Line 35: 10 degrees vs 12 degrees for surface scattering-based calibration?** Thanks for pinpointing at this discrepancy. The sweep spot is actually at  $11^\circ$ , see Tanelli et al CloudSat's Cloud Profiling Radar After Two Years in Orbit: Performance, Calibration, and Processing IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 46, NO. 11, NOVEMBER 2008. we will amend this.

**Line 79: More justification is needed for this particular choice of statistical metric, the JS distance, in assessing the similarity between two PDFs. What are the strengths/sensitivities of this metric to PDF structure?**

The Jensen-Shannon distance is a method, commonly used in statistics, to measure similarity between two probability distributions. Given that the base 2 logarithm is used in the definition, it is bounded by 1 with the maximum obtained for non-overlapping distributions which is a clear advantage over the standard L2 distance. Moreover, the JS distance is equal to 0 if and only if two PDFs are equal.

**Line 42: The authors highlight the statistical cross-calibration approach of Protat et al. (2009) that utilizes CFADs to perform a calibration comparison including information about the vertical distribution of observed reflectivity, but the approach implemented in this work collapses the distribution into a 1-D PDF with no vertical information retained. This difference should be addressed with at least some discussion in the manuscript.**

The method from Protat makes some qualitative consideration based on CFADs of reflectivities which could also be repeated in our approach. However when it comes to make quantitative assessment Protat et al., uses the mean vertical profiles of reflectivities (i.e. they condense the 2D distribution in a 1-D PDF by getting rid of the Z-dimension). We decided to get rid of the vertical dimension keeping the dependence on the reflectivities.

**Line 132: The authors state that only ice clouds away from deep convection are used in this analysis, but do not state how the observational data sets have been filtered to ensure such ice cloud detections are excluded. Please include these details.**

We used the CloudLayerType field from the 2B-CLDCLSS CloudSat data product (<https://www.cloudsat.cira.colostate.edu/data-products/2b-cldclass>) to identify the cloud type, and then exclude the deep convective clouds. Similarly for GPM we have excluded convective pixels. We will include these details in the manuscript.

**Section 2.2: the definition and description of statistical quantities that are intended to quantify the intersection occurrence rate could be improved. Most importantly, the gridding resolution of 2 deg x 2 deg seems rather arbitrary, and for instance, the "mean number of layers" plots in fig would scale (likely proportionally) to this choice of angular area. Furthermore, it seems slightly misleading to refer to this quantity as a number of ice cloud layers, when my interpretation is that this is actually referring to the mean number of radar range bin volumes (e.g., 5 x 5 x .25 km for GPM) that have returns meeting the ice cloud criteria laid out in the paper. It would seem more appropriate to use for this statistical quantity a mean area density of ice cloud bins that would not depend so strongly on the choice of grid resolution. Furthermore, this would allow for a more straightforward calculation of expected mean number of intersection bins per unit time by multiplying this density by the mean inter-satellite swath overlap and the mean rate of overpass "coincidence" for that lat/lon (see major comment 1 above).**

Assuming my interpretation of the calibrating target definition in Section 2.2 is correct, I suggest that the ice cloud layer variable be renamed to something like “ice cloud radar bins” to ensure that the reader is not confused by this. The term ice cloud layer evokes an extended cloud morphological feature that would nominally encompass a very large number of radar range bin volumes. If my interpretation of this “ice cloud layer” definition is wrong, I suggest that the explanation in this section be reworded for clarity.

Here the terminology we have used was misleading as well. Indeed what is plotted are ice cloud radar 500-m thick bins. The selection of the  $2 \times 2$  gridding is done with the goal of reducing the sampling noise associated to CloudSat and GPM observations. The results are expected to be independent to that resolution because we have normalised the number of detected ice cloud radar 500-m thick bins by the total number of radar footprints at each grid point.

**The vertical black and blue dashed lines in Fig. 4 (right panel) are difficult to distinguish**  
We will use a continuous line in correspondence to the black line to make the two lines better distinguishable.

**Section 2.1: More details are needed on how orbits are propagated for the purposes of calculating intersection points**

The orbits have been propagated analytically using the trajectory equation obtained with the integration of the equation of motion for the restricted two-body problem. The initial orbital parameters used in the propagation are reported in Table 2 and 3. The J2 perturbations only have been taken into account.