Response to Interactive comment by E. Jensen for 'Application of infrared remote sensing to constrain in-situ estimates of ice crystal particle size during SPartICus' by S. J. Cooper and T. J. Garrett

We thank Dr. Jensen for his critical comments.

Direct comparison of satellite remote-sensing measurements of clouds is generally challenging, and such comparison for cirrus clouds is particularly problematic. Aircraft sampling essentially provides a pencil of measurements through the atmosphere. Cirrus clouds are typically very highly structured, both horizontally and vertically. Surface area density and ice water generally vary by orders of magnitude over horizontal distances of just a few km, and effective radius can vary by more than a factor of two over these spatial scales. In agreement with past studies, SPartICus measurements indicate that effective radius often increases systematically with decreasing height in cirrus (Lawson, 2011). This vertical structure is expected due to differential sedimentation speeds of small versus large crystals. An example from tropical anvil cirrus indicated effective radius increasing from '30 _m to 80 _m as the sampling aircraft descended from 12 to 9 km (Lawson et al., 2010). Aircraft necessarily provide a very limited view of the vertical variability in cloud microphysical properties, and biases toward the upper or lower parts of cirrus could result in misrepresentation of the vertically averaged cloud properties. Although the comparisons presented in this manuscript focus on moderate optical depth cirrus, this does not imply that the cirrus were necessarily vertically thin nor does it imply a lack of vertical variation in re.

In terms of difficulties in matching satellite with in-situ measurements, we repeat much of the answer we gave to the anonymous reviewer. We agree with the reviewer in that there is no way to exactly match remote sensing observations with in-situ sampling volume. Although again, such an argument could be used to negate any attempt to validate cloud properties from remote sensing efforts or vice versa. For example, given a 2D-S sample volume of order 10 L per second, and assuming a total of 10,000 hours of flight time in cirrus that has been accomplished in all cirrus to date, this would still only be a total volume sampled of 0.001 kilometers cubed, just one ten millionth of a single 100 km squared cirrus cloud, and one hundred billionth of global cirrus cloud volume. Perhaps the argument could be that it is pointless to study cirrus from aircraft. The reason cirrus programs are useful is because the small pencil that is measured by aircraft is high correlated with surrounding cloud, due to turbulent mixing and similar physical formation mechanisms. Still, the reviewer is correct that there is spatial and temporal variability that nonetheless needs to be considered. Like numerous other techniques/ papers in the literature, we tried our best to account for these issues. Again, we tried to minimize the effects of the sampling issue (horizontally) by selection of our test cases. We only selected test cases where the Learjet was in large homogenous areas of high BTD cirrus clouds at the time of the MODIS overpass. So, regardless of the exact plane location relative to the timing of the overpass, the Learjet was in thin cirrus clouds that were dominated in a radiative sense by small ice crystals. (Alternately, for one case we find the Learjet in a large area of low BTD cirrus indicating the dominant radiative presence of large ice crystals.)

Dr. Jensen is, of course, correct that the Learjet could be in a small patch of large ice crystals for these thin cirrus clouds dominated radiatively (optical depth) by small crystals (or vice versa), e.g. resulting from a vertical gradient of effective radius. Even though we 'see' the entire cloud for these thin clouds from the split-window perspective, there could always be a layer of very small optical depth cirrus that has a different effective radius than the cloud-layered average. This does not seem to be the case for our cases, as we find in general good agreement between our results and in situ measurements. Furthermore, it is possible to examine co-incident CALIPSO lidar measurements/ retrievals in context of flight path to help examine our results. Unfortunately, only 1 of 3 test cases (March 17 large crystal case) directly underflew CALIPSO/ CloudSat. For this large crystal case, however, the Learjet was located near the bottom of the cirrus clouds, see Figure 1 below. Such a scenario makes the suggestion that the Learjet was in a thin layer of small ice crystals for a cloud dominated radiatively by large high crystals very unlikely. Furthermore, the 2C-ICE combined CloudSat- CALIPSO approach also found large particles much greater than 20 μm (personal communication with Dr. Min Deng, U of Wyoming).

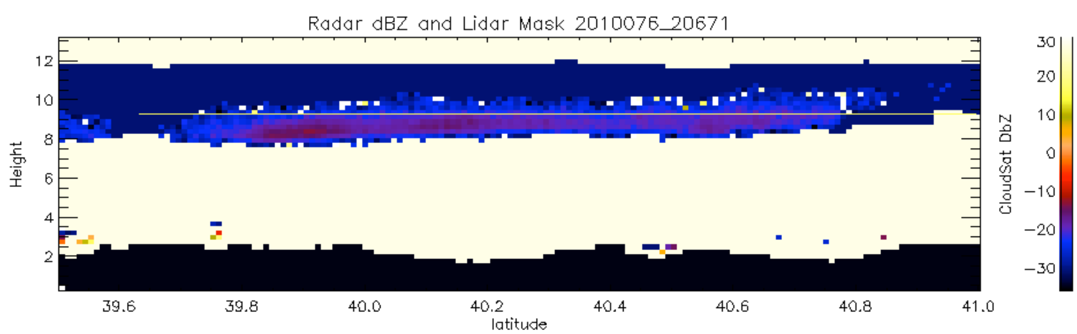

Fig 1: White line indicates flight path of Learjet relative to cloud dBZ/ lidar mask.

But again, it IS possible that there could always be a layer of very small optical depth cirrus that has a different effective radius than the cloud-layered average. We must acknowledge this in the manuscript. But we also feel that the possibility of the vertical gradient of effective radius does not necessarily limit the utility of our technique. Instead, it simply highlights the need for a priori consideration of our technique in in-situ campaign design. Ideally, our infrared technique would be applied in concert with an instrument with high temporal and spatial resolution such as the MODIS Airborne Simulator both to maximize the number of cases and to minimize sampling issues. These high resolution measurements would also allow for examination of the vertical profile of effective radius to alleviate Dr Jensen's concerns given a corresponding flight leg design. But in general, if there is a repeated discrepancy of in-situ measurements and cloud radiometric signature, there is a problem. Either flight leg design needs to be re-considered (otherwise we do not get a true measure of cloud properties and then why are we even bothering with in situ measurements?) or we need to re-examine the self-reported collection efficiencies and shattering algorithms of the in-situ instrumentation.

Therefore in the final paragraph of the conclusion section, we add

Ideally, our infrared technique would be applied in concert with an instrument with high temporal and spatial resolution such as the MODIS Airborne Simulator both to maximize the number of cases and to minimize sampling issues. These could be combined with airborne vertical profiles to better assess the impact of the vertical variability of effective radius on intercomparisons.

The authors state that in the case studies chosen the Learjet was located in relatively homogeneous areas of cloud. However, examination of the MODIS images (Figs 2-4) suggests considerable horizontal inhomogeneity in the cloud fields where the Learjet was sampling. In fact, casual examination of satellite imagery indicates that homogeneous cirrus clouds are a very rare exception. The authors focus on 5-10 minute average values of effective radius, and little or no discussion of variability is included.

Dr. Jensen is correct that there is considerable horizontal inhomogeneity in the cloud fields (as indicated by brightness temperature and BTD in Figs 2-4) where the Learjet was sampling. It is important to remember, however, that BTD is a function of both cloud effective radius and cloud optical depth. This means that a cloud field composed uniformly of 'small' particles (e.g. all exactly 13.2 μm) would still show great variability in the BTD field, if cloud optical depth varied from near 0 to about 4 (See Figure 1). So, variations in BTD in the boxed areas of Fig 2-4 do not necessarily indicate variation in cloud effective radius, i.e. the presence of horizontal patches of small and large crystals.

Dr. Jensen's concern is not applicable for the 'large' crystal case of Figure 3, as the entire box is filled with low BTD values that could not be associated with small ice particles given atmospheric conditions. For the 'small' crystal case in Figure 2, it is important to realize that the area of low BTD and high brightness temperature in the lower left corner of the box simply correspond to clear-sky or very thin cirrus, as indicated by a corresponding lack of in situ airplane measurements. The remaining areas of the cloud indicate high BTD. Note that the edges of the cloud are slightly lower in BTD than the center of cloud bands, a finding consistent with lower optical depths expected near cloud edge. Likewise for the 'intermediate' crystal case in Figure 4, the boxed area is dominated by high BTD cirrus clouds, with a small clear patch suggested by low BTD and very high brightness temperatures. Again, the edges of the cloud are slightly lower in BTD than the center of cloud bands, a finding consistent with lower optical depths expected near cloud edge.

Given the fundamental difficulties in precisely extrapolating MODIS overpass pixel-level observations to in situ flight leg measurements, we choose to focus on only those cases where the Learjet was in large homogenous areas of high or low BTD cirrus in terms of 'small' or 'large' as described above. This severe constraint, in fact, is the primary limiting factor in finding good cases from the SPartICus campaign. There are many cases where large areas of high BTD cirrus existed during the campaign, yet the Learjet would be hundreds of miles away in a highly heterogeneous cloud field (Dr. Jensen is correct on the somewhat ubiquitous presence of heterogeneous cirrus). So, we could not use these cases. Similarly, the earlier Terra MODIS overpass on April 28 showed huge areas of exceedingly high BTD (much greater than in Figure 2) indicating small particles, a finding corroborated by SGP ARM ground-based instrumentation. Yet, the Learjet was not sampling these (or any) clouds at the time. Since it is ice cloud particle size that dominates the BTD signal to the first order for thin ice clouds, it would seem wise to try to use this infrared information (GOES, MODIS Airborne Simulator) in design of campaign flight plans.

To address Dr. Jensen's concerns about the heterogeneity of cloud fields in Figures 2-4, we add detailed information in the Figure captions pointing out the clear-sky regions within the black boxes. Such information should eliminate any confusion for those readers unfamiliar with infrared radiative transfer as to why there are areas of low BTD in these boxes where we claim 'small' particles. Specifically, the following changes were made:

Fig 2. Figure shows the MODIS 11.0 μm brightness temperature (top panel) and the 11.0 μm minus 12.0 μm BTD (bottom panel) for the April 28 1925 GMT overpass. The SPEC 25 Learjet was located in thin cirrus in north central Oklahoma as shown by the high BTD areas within the black box in the bottom panel. High values of BTD for these cirrus indicate clouds dominated radiatively by small crystals. The low BTD areas in the lower-left portion of the black box and corresponding high 11.0 μm brightness temperatures indicate regions of clear-sky or very thin cirrus clouds.

Fig. 3. Figure shows the MODIS 11.0 μm brightness temperature (top panel) and the 11.0 μm minus 12.0 μm BTD (bottom panel) for the March 17 2030 GMT overpass. The SPEC 25 Learjet was located in thin cirrus in western Colorado as indicated by the low BTD areas in the black box. Low values of BTD for these cirrus suggest clouds dominated by large crystals for this case.

Fig. 4. Figure shows the MODIS 11.0 μm brightness temperature (top panel) and the 11.0 μm minus 12.0 μm BTD (bottom panel) for the June 7 2015 GMT overpass. The SPEC 25 Learjet was located in thin cirrus along the Front Range of Colorado as shown by the high BTD areas within the black box in the bottom panel. High values of BTD for the cirrus suggest clouds dominated radiatively by small crystals. The low BTD areas in the lower-right portion of the black box and corresponding high 11.0 μm brightness temperatures indicate regions of clear-sky or very thin cirrus clouds.

In terms of variability, we did mention that the choice of averaging time did not affect our results in terms of 'small' or 'large'.

Given the unavoidable problems and limitations with comparison of in situ measurements and satellite retrievals of cirrus microphysical properties, it would seem that solid conclusions could only be drawn if a large number of cases were included in the analysis. Unfortunately, that does not seem to be possible here given the limitations of the BTD threshold technique approach and the limited number of satellite/aircraft coincidences.

We wholeheartedly agree we need more test cases to more definitively evaluate instrument performance. Again, the purpose this paper is to show how our bi-spectral technique can be applied to in-situ campaigns through use of SPartICus Spec Inc. provided data and to provide a first order estimate of instrument/ algorithm performance. It is not to present the definitive study on the veracity of either instrument results or Spec Inc. cloud property algorithms. We, of course, would be interested in pursuing such a paper. But we simply had too few good cases during the SPartICus campaign for more definitive conclusions. Ideally, our infrared technique would be applied in concert with an instrument with high temporal and spatial resolution such as the MODIS Airborne Simulator both to maximize the number of cases and to minimize sampling issues. Input on flight plans and campaign instrumentation as well as access to all in-situ data would be greatly beneficial for a more definitive work. To stress the fact that we do not aim to rigorously define instrument performance, we added the following sentence in the last paragraph of the introduction,

'However, given the limited number of good test cases for our technique during the campaign, and the fact that our infrared technique was not considered for design of the campaign, we cannot present either a broad characterization of SPartICus cloud properties or a definitive analysis of in situ instrument performance.'

Comment 1: The focus of the paper seems to be on evaluating effective radii determined from measurements made with the 2D-S probe and from traditional FSSP probes with inlets and no corrections for shattering. However, in the most interesting "intermediate" case, the authors switch over to using CDP measurements instead of FSSP measurements. There is no discussion of why this is done, and the manuscript discussion and conclusions seem to imply that the two probes are equivalent. However, as the authors acknowledge earlier in the manuscript, the CDP has no shroud or inlet and therefore is likely much less susceptible to particle shattering compared to the FSSP. The authors should therefore acknowledge that the "intermediate" case presented has no relevance to the evaluation of effective radii determined from FSSP probes.

FSSP data was not available from Spec Inc. for that case. (It would appear the FSSP was not flown.) We do not imply in any manner that conclusions can be drawn on FSSP behavior for that case.

We added in a parenthetical saying that '*FSSP data not available'*.

Comment 2: At the end of the abstract, the authors state "There is no evidence to support that an FSSP-100 with unmodified inlets produces measurements of re in cirrus that are strongly biased low, as has been claimed." They should also acknowledge that the evidence presented here does not convincingly demonstrate that FSSP-100 probes with unmodified inlets do not produce measurements of re in cirrus that are strongly biased low. The manuscript only provides one extreme small particle case study and one extreme large particle case study for comparison with the FSSP measurements. The BTD approach only provides a somewhat qualitative comparison (re larger or smaller than '20 um). As discussed above, the problems associated with comparisons leave open the possibility that the results are affected by sampling biases in the aircraft measurements.

In agreement with past studies, the results presented here indicate that effective radii determined from FSSP measurements in cirrus are considerably lower than those determined from 2D-S measurements (e.g. Lawson, 2011). The authors acknowledge that the comparisons with the BTD retrievals do not definitively indicate that either is incorrect. Korolev et al. (2011) presented comparisons between measurements made with a standard FSSP and an FSSP with the inlet and shroud removed. The comparison indicated that in ice clouds the standard FSSP response was overwhelmingly dominated by shattering artifacts. For a balanced presentation, the authors should acknowledge the results from the (Korolev et al., 2011) study.

Based upon this and other reviewer comments, we now realize we included an unnecessarily argumentative sentence in our abstract in regards to previous work involving shattering in the FSSP: 'There is no evidence to support that an FSSP-100 with unmodified inlets produces measurements of *r*e in cirrus that are strongly biased low, as has been claimed.' This sentence gives the un-intended impression that we feel previous efforts may have been wrong. In our paper, we had previously pointed out the Korolev and Isaac (2005) demonstrated that ice particles will shatter in the FSSP for some test conditions. We simply meant that we found no definitive evidence for shattering based upon our bi-spectral technique and given SPartICus test cases. We do not doubt the findings of previous works for their specific test cases.

So we have re-written this offending sentence, adding in 'For our test cases,' and removing 'as has been claimed', …

'For our test cases, there is no evidence to suggest that an FSSP-100 with unmodified inlets produces measurements of re in cirrus that are strongly biased low.'

Comment 3: Table 3 provides values of re determined from 2D-S measurements with and without shattering artifacts removed. The values are nearly identical, and the apparent point of showing this is to demonstrate that shattering has negligible impact on determination of effective radii. As discussed above, there is every reason to believe that the shattering problem is much more severe for FSSP probes with a shroud and inlet than for 2D-S probes that are designed to limit the possibility of shattering artifacts reaching the sample volume. The authors should acknowledge that the comparison presented in Table 3 is not relevant for the issue of shattering artifacts in FSSP datasets. A related issue is that accurate measurements of ice concentration are important. Effective radius is an important measure for determining cloud radiative effects, but knowledge of ice concentration is needed for understanding cloud nucleation processes as well as for predicting how the cloud will evolve over time. Even for the 2D-S probe, shattering can significantly affect ice concentration (Lawson, 2011). Note that the relative impact of shattering on 2D-S ice concentrations depends strongly on the concentration of natural small crystals (Jensen et al., 2010).

We simply presented the results we found for our limited test cases. We did not imply broad applicability to all other FSSP test cases past or present. Again, as Dr. Jensen correctly points out, we would need to do need more test cases to more definitively evaluate instrument performance.