Response to re-review 1

Dear Professor Garrett,

Thank you for giving us the opportunity of further responding to comments made by referee 1. Please note, in the uploaded re-revised manuscript, changes to the original manuscript that correspond to the first review process are highlighted in red. Changes to the manuscript resulting from the re-review are highlighted in blue. It is the changes in blue that correspond to the re-review.

The reviewer is thanked for taking the time to re-review our manuscript and for providing a diligent review process, which we believe has further improved the paper. However, we find the reviewer does make subjective comments in his responses to our initial rebuttals (most of which he accepted). For instance, the second reviewer suggested only minor revisions, and in our opinion we did answer all the second reviewer's comments. In fact, the second reviewer gave the manuscript an excellence marking. It is not for this reviewer to comment on behalf of the second reviewer. Also, it is up to the community to decide whether we are "misdirecting future research". Of course, the reviewer is entitled to their opinions but they cannot speak for the community as a whole. However, having said all that, the reviewer still has legitimate scientific concerns. The last two remaining concerns are the issue of underlying water cloud and the method of employing minimised rmse (root mean square errors) to select a model type for each of the PARASOL twelve pixels under dispute, which indicated non-fully randomized phase functions. We thank the reviewer for bringing these up, because this gives us an opportunity to improve the article and refine the data analysis.

We have now performed substantial re-analysis on both these issues and our findings are presented below. We first present our results for the issue of underlying water cloud. In our initial response to review 1we presented range corrected lidar images, which gave the impression that the lidar signal was saturated in the vicinity of the 12 pixels under question. We have re-examined this image and have now applied even higher resolution, but first, we show the aircraft path at the time between 13:00:00 and 14:00:00 UTC, shown in the figure below.

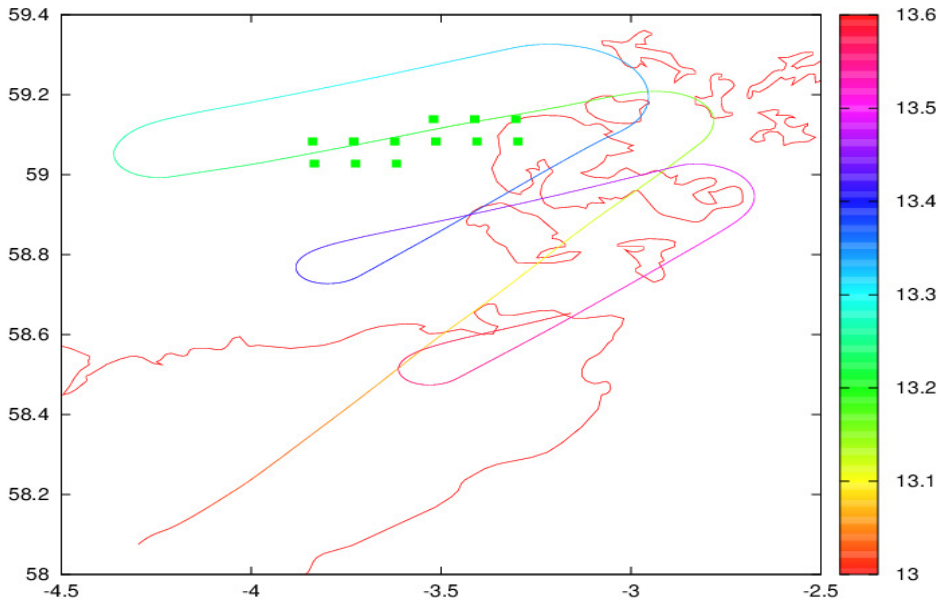

Figure 1. The colourbar on the right-hand side indicates the time in UTC in hrs:min.

From Figure 1 it can be seen that the aircraft overflew the twelve pixels between about 13:20 to 13:30 UTC, the corresponding range corrected lidar image is shown below, in Figure 2, covering the same time. This is a higher resolution lidar image than previously used but also has a dark background. At the time the aircraft over flew the PARASOL pixels under question, at between 13:20 to 13:30 UTC, from Figure 2 it can be seen that there is **NO** evidence for underlying water cloud. What can be seen in the image is sea surface reflection. Since the lidar receives a signal from the underlying surface, we can confidently state that it would also have received a signal from a water cloud, had there been one; the cloud being closer to the instruments (higher) and generally brighter than ocean surface. The water cloud that is present, at less than 2 km in altitude, occurred outside these times and where **NO** PARASOL pixels were included in our analysis. A further point the reviewer raised was why only in the manuscript we included the lidar image (Figure 7a) showing the retrieved vertical extinction profile at altitudes greater than 6 km. The difference between Figure 2 below and the image shown in Figure 7a is that Figure 7a is a retrieval of the vertical profile of volume extinction coefficient, whilst the image below is a "raw" range corrected backscatter lidar image. In regions, where only clear air is apparent, the retrieval equation becomes unstable. Below 6 km, no meaningful retrievals of volume extinction coefficient could be found. This is why only retrievals greater than 6 km are shown in Figure 7a of the manuscript. *This point has been noted in the re-revised manuscript on page 24 lines 10-14.*

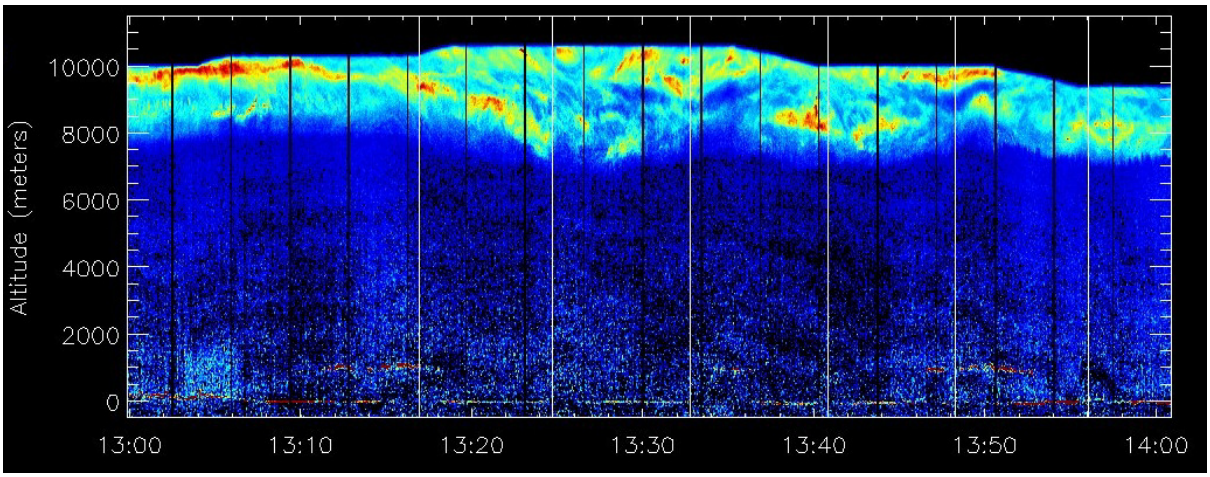

Figure 2.

A further useful comparison between the lidar and PARASOL, in regard to this question, is the averaged retrieved optical thickness obtained from the 12 pixels. Figure 7 (a) from the manuscript is pasted below, for ease of comparison, shown here as Figure 3.

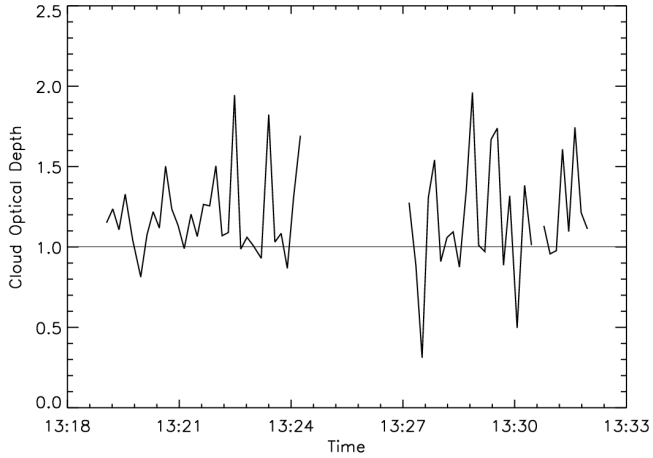

Figure 3 (manuscript Figure 7a).

From Figure 3 it can be seen that the lidar retrieved visible optical thickness of the cirrus, between the times of about 13:20 to about 13:30 UTC, varies generally between about 0.5 to at most 2.0. The averaged retrieved PARASOL optical thickness (averaged over the 12 pixels), from Figure 1 above, is 1.81±0.32. Moreover, the retrieved PARASOL optical thickness over the same 12 pixels assuming the fully randomized phase function is 1.52±0.26. Both retrievals are within the range of the lidar estimates. The point is, if there were underlying water cloud, then, of course, the retrieved PARASOL optical thickness would be outside the range of the lidar retrievals. The consistency between *two independent* measures of optical thickness is further evidence that there was no underlying water cloud beneath the PARASOL retrievals of cirrus properties. *Again, this point has been noted in the revised manuscript on page 31 lines 10-21 in the re-revised manuscript.* Please note also that in the caption of Figure 6 in the first revision we stated the time as 13:21:00 UTC. This in fact was not correct and the time should have been 13:33:00 UTC. This has been corrected in the re-revised manuscript and throughout the paper wherever it was mentioned.

We now examine further independent evidence obtained from satellite remote sensing from the NASA/Langley cloud imagery and cloud product website constructed by Pat Minnis, which is located at the following address http://www-pm.larc.nasa.gov/

The images below were obtained from the above site at the times of relevance to this paper. The first image, shown as Figure 4, shows the retrieved likelihood of multiple cloud layers at 13:00 UTC (closest time available to the PARASOL overpass, which was at about 12:50 UTC). The retrievals below use data from the European METEOSAT satellite, these four channel split-window retrievals (0.65, 3.9, 10.8, and 12 μm) are explained in a series of papers, available from the website (http://www-pm.larc.nasa.gov/real-time/real-time_refs.html)

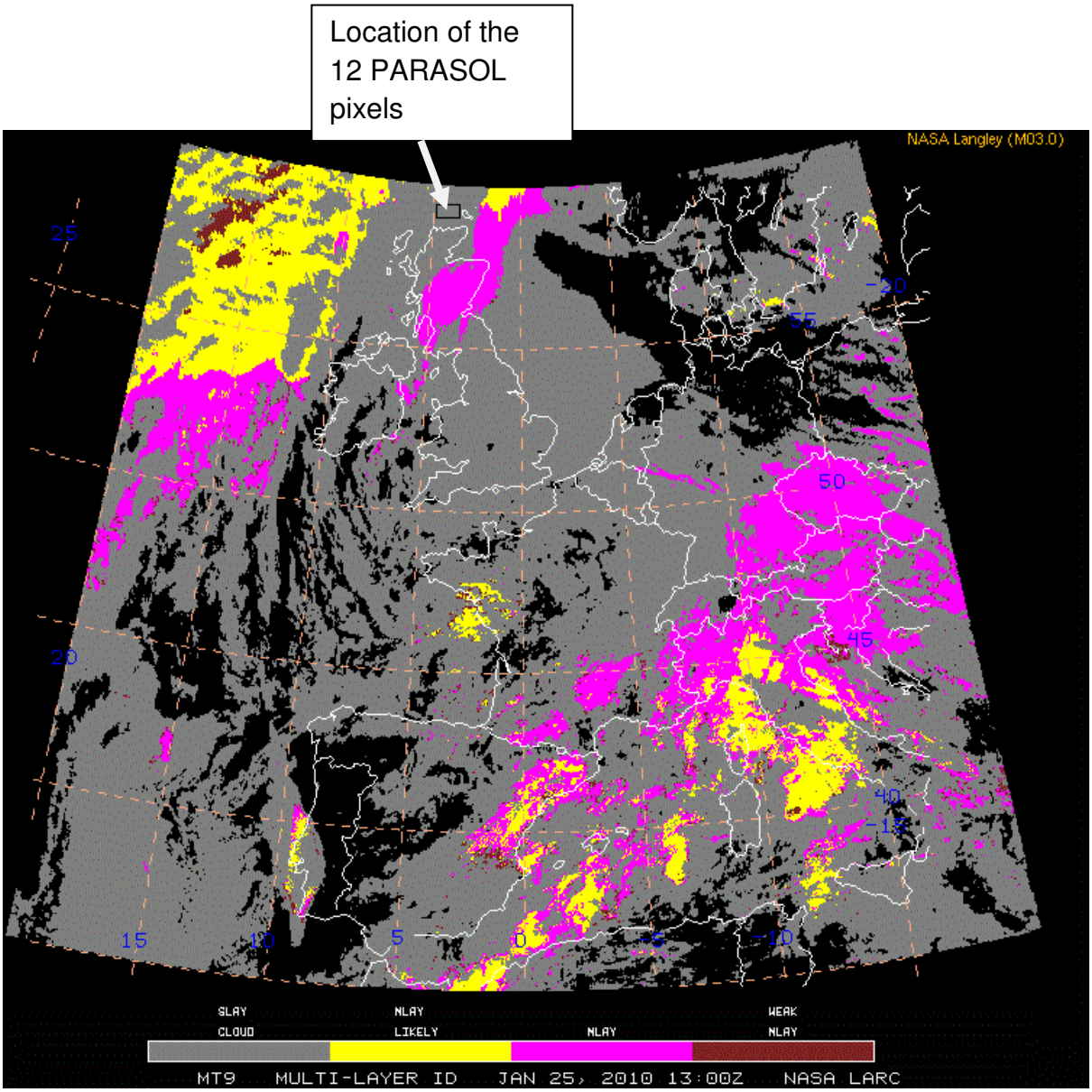

Figure 4.

The grey scale indicates single layer cloud, yellow likely multi-level cloud, purple multiple layer cloud, and brown weak likelihood of multiple layer cloud. The 12 pixels in question are in the region of single layer cloud, and the location of these pixels is shown by the superimposed box shown in the image. The corresponding retrieved cloud-top pressure, for the same time, is shown in Figure 5 below.

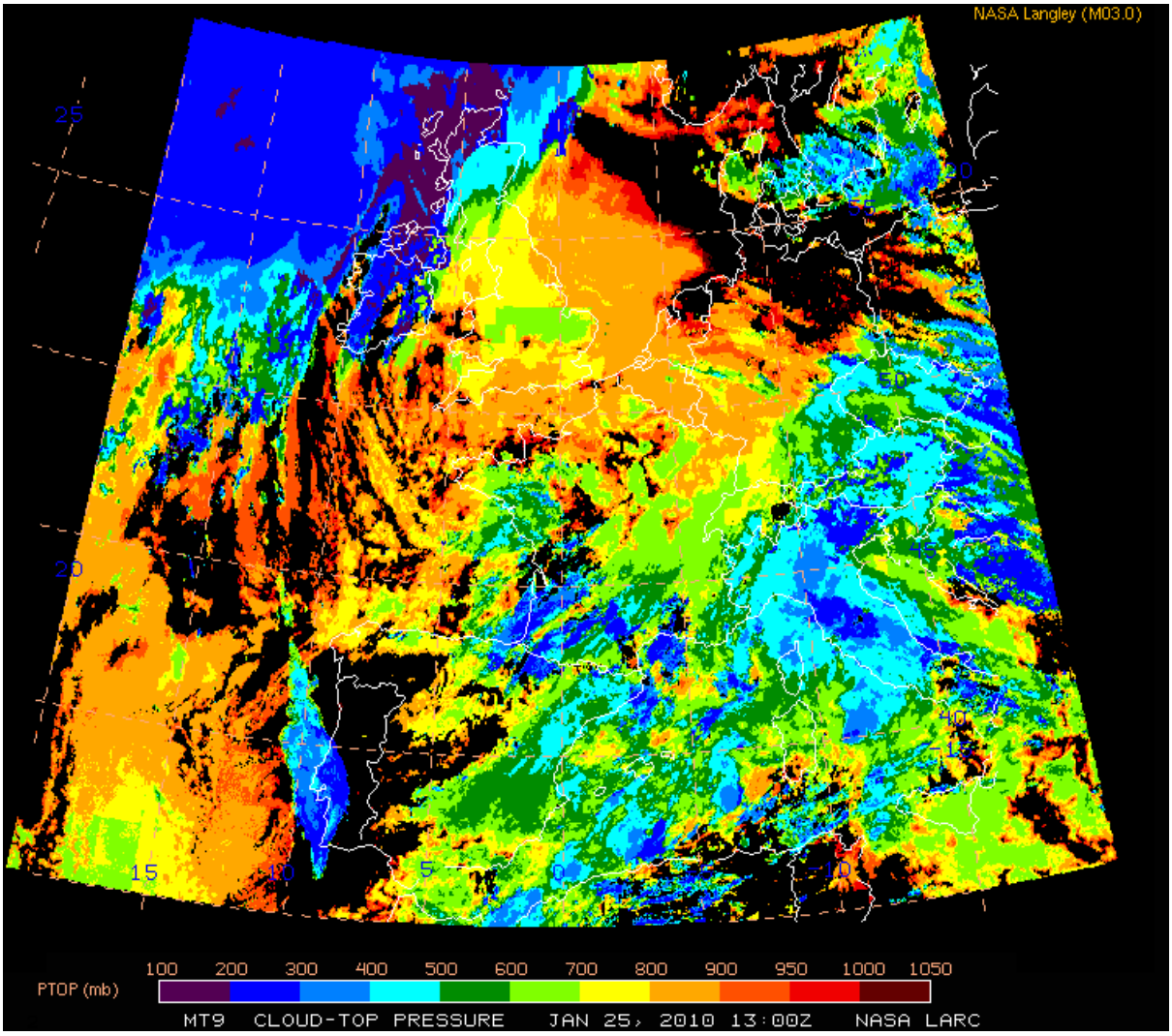

Figure 5.

The retrieval of single layer cloud is at pressures between 100-200 mb, around the boxed location shown in Figure 4 above, at an altitude associated with cirrus. Again, these results indicate that there is unlikely to be underlying water cloud beneath the cirrus at the time of the PARASOL overpass. If there were underlying water cloud, then, the retrieved cloud-top pressure would be biased towards significantly higher pressures. This statement is consistent with the lidar image shown in Figure 2 above. The reviewer asks for "100% certainty" that there is no underlying water cloud beneath the 12 contested PARASOL pixels. Of course, in science, there is no such thing as "100% certainty," as the above retrievals shown in Figures 4 and 5 are themselves subject to retrieval error. However, it is our opinion, that taken the results above, we can say that the evidence for there being underlying water cloud beneath the cirrus covering the 12 pixels is unlikely. Of course, we cannot re-produce these images in the revised manuscript but we state the result above and provide the link to the relevant NASA cloud product webpage. For reasons of brevity we do not include the above cloud product images in our re-revision, we merely state the findings, and the discussion about lower level water cloud can be found on pages 31-32, lines 1-25 and 1-11, respectively in the re-revised manuscript.

Given that it is unlikely that there was underlying water cloud beneath the 12 cirrus pixels, how can we improve the assignment of particular shapes of the phase function to those

pixels? This too has proven to be possible. We described previously in our reply to reviewer 1, the reasons as to why we do not have more than one cirrus case, and these have not changed. However, the reviewer is correct to question the robustness of just applying the rmse analysis to assign particular shapes of the scattering phase function to those pixels. To this end, we sought, in addition to rmse, a quantitative statistic that could measure the difference in variance between at least two samples. In the revised manuscript, we apply the Levene (1960) [Levene, H. (1960). In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin et al. eds., Stanford University Press, pp. 278-292] test statistic. This particular statistic is chosen because other alternative methods are very sensitive as to whether the test distributions are strictly normal. We test whether the variances between samples are different at the 5% significance level, as this level of statistical confidence is generally applied, since 10% and 1% might be considered too easy on the one hand but too onerous on the other, respectively.

Please note that the discussion of Figure 10 (a) and (b) has also now been revised on page 29 of the re-revised manuscript. In the figure, the spherical albedo differences obtained using the fully randomized phase function is now Fig. 10 (a) and the pristine result is shown as Fig. 10 (b). This change naturally accommodates discussion of the Levene results which immediately follow. Essentially, in this revision, in the case of Fig. 10 (b), we agree with the reviewer that it is difficult to associate one model using rmse alone, and so the discussion has been revised accordingly.

In the re-revised manuscript, pages 22-23 lines 20-22, 1-16, respectively, we describe the Levene test statistic as follows:

"The rmse is one general measure for choosing the best-fit model to the observations. However, once a set of rmse-minimisers has been identified, one should assess where the identification is statistically significant. This is needed to rule-out the possibility that the differences in rmse for the different distortions could have resulted from chance. To test this, we apply the Levene (1960) test statistic, as it is less sensitive to the condition that the data must be normally distributed than the usual F statistic, which is generally used to test whether the variances between two samples are equal, provided the data follow a normal distribution. In the Levene test, the samples, k, are tested for homogeneity of variances between the k samples. The total number of data points contained in all samples is given by N. The Levene null hypothesis is that variances between k samples are equal. The Levene null hypothesis is rejected, at some level of significance, α, if the Levene test statistic, W, is greater than $F_\alpha(k-1, N-k)$, where $F_\alpha(k-1, N-k)$ is the upper critical value of some F distribution with k-1 and N-k degrees of freedom. We consider pixels for which the derived rmse values do not exceed 100% to require testing for significance using the W-test. For these pixels, the W-test statistic is applied, to test whether the model variances in ΔS_j are different at the 5% significance level. If the null hypothesis is rejected, then that pixel is assigned a particular model phase function. The 5% or α=0.05 significance level is chosen, as this is simply between α=0.1 (10%) and α=0.01 (1%) significance levels, so that the model test is neither too easy or unrealistically too hard, respectively."

Note also, the word "null" used to describe the no-retrievals in previous versions of the manuscript has been changed to "indeterminate" to avoid confusion with the term "null hypothesis". To apply the above test, for those 12 pixels, we compared the minimised rmse selected phase function model to all other models for those same pixels. We compared the variances in ΔS_j in groups of 2, so k=2, against the variances predicted by the other models, then, evaluated whether the null hypothesis was accepted or rejected at the 5% level of

significance for each of the model comparisons. Previously, for the 12 pixels we showed that, using minimised rmse values only, 5 of the pixels were assigned phase function shapes of zero distortion, 3 pixels were assigned distortion values of 0.25 and 4 were assigned distortion values of 0.15. The variances obtained for the 5, 3, and 4 pixels were combined for each of the rmse determined best-fit scattering phase function. So, the first test has 5 PARASOL pixels etc. The results are described as follows in the re-revised manuscript, on pages 29-30, lines 17-25, 1-25, respectively.

"From Fig. 9 (c) it can be seen that using minimised rmse test, five of the twelve pixels are associated with pristine model phase functions (distortion=0), whilst four pixels are associated with slightly distorted phase functions (distortion=0.15), and the other three pixels are associated with moderately distorted phase functions (distortion=0.25). All pixels associated with each of the above 3 distortion values were combined together. For each of the distortions, the W statistic was obtained in groups of 2, so that k=2. The variances in the spherical albedo differences obtained with the rmse determined best-fit phase function were compared against the variances obtained assuming all other model phase functions. For each group of 2, the test W statistic was computed and then compared against the tabulated upper critical value of the $F_{\alpha}(k-1,N-k)$ distribution, to accept or reject the null hypothesis at the 5% significance level. The results of this analysis are presented in Table 2.

Table 2. The Levene test statistic, W, applied to test homogeneity of variances in spherical albedo differences between two groups of scattering phase function models for each set of pixels. Where in the table the two phase function models are represented for each set of pixels by their assumed distortion values called model pair, the total number of pixels used in each test is n. The null hypothesis is given by H0, which is either rejected or accepted, k is the number of samples, N is the total number of observations in the two samples, and F0.05(k, N-k) is the value of the tabulated upper critical value at the 5% significance level composed of k and N-k degrees of freedom.

Model Pair	n	k	N	W	$F_{0.05}(k, N-k)$	H_0
Full/0	5	2	70	1.61	3.93	accept
Full/0.25	3	2	42	<1	4.1	accept
0/0.25	3	2	42	<1	4.1	accept
Full/0.15	4	2	56	4.022	4.020	reject
0.25/0.15	4	2	56	<1	4.020	accept
0/0.15	4	2	56	<1	4.020	accept

In the case of the five pixels associated with pristine phase functions, it can be seen from Table 2 that the Levene null hypothesis must be accepted. Therefore, the variances in the spherical albedo differences determined using the rmse best-fit model are not sufficiently different from the variances obtained using all other phase function models. A similar result to the above was found for the three pixels, which were associated with the moderately distorted phase function (distortion=0.25). For the four pixels associated with the slightly distorted model phase function (distortion=0.15), Table 2 shows that the null hypothesis can

be rejected, when its variances are compared against the variances obtained assuming the fully distorted model phase function (distortion=0.4 with spherical air bubble inclusions). However, for all other assumed models, for this group of four pixels, the null hypothesis must be accepted. The results contained in Table 2 shows that using minimised rmse values alone may not be sufficient to select model phase functions on a pixel-by-pixel basis and that some other test statistic is required to compliment the rmse method.

The Levene test statistic was also applied to some pixels associated with the most randomized phase function, to test if W>>F for these pixels. The results of this test are presented in Table 3.

Table 3. Same definitions as Table 2 but the Levene test statistic is applied to a group of 7 pixels, where the fully randomized model phase function was found to best fit spherical albedo differences using minimised rmse values. The model pair tests are between all other scattering phase function models and the fully randomized scattering phase function model.

Model Pair	n	k	N	W	$F_{0.05}(k, N-k)$	H_0
0/Full	7	2	98	18.289	3.93	reject
0.15/Full	7	2	98	19.436	4.1	reject
0.25/Full	7	2	98	12.918	4.1	reject

In this case, seven pixels were selected between latitudes 57.92° to 58.92°, and between longitudes -3.42° to -3.71°. As before, the seven pixels were combined, and the resulting variances in spherical albedo differences obtained assuming the most randomized phase function were compared against the variances obtained assuming model distortion values of 0.15, 0.25 and 0, respectively. The results from Table 3 shows that the null hypothesis can be very strongly rejected at α=0.05 (5% significance level). The results of this analysis strongly suggest that the selection of the most randomized phase function using minimized rmse values is acceptable, as illustrated by the example case in Figure 10 (a)."

Clearly, the above results indicate, that if we include another test, other than rmse, then we cannot conclude that each of the 12 pixels can be assigned a particular shape of the scattering phase function. What we can, however say, with some confidence, from these results, is that when the NWP model predicted RH_i>1.0, these pixels are more likely to be associated with randomized phase functions, such that these phase functions, at least between scattering angles of 80°-130°, are featureless. We cannot say anything about RH_i<1, except none of the presented models are sufficiently different in their variances to obtain a unique result. The reason for this could have something to do with there being some structure in the phase function not explained by the tested phase function models. Alternatively, the more simple explanation is that there is insufficient scattering angle information available to discriminate between models. These new results mean that Figure 12, showing correlation between distortion value and RHi in the original manuscript has been removed and Figure 11a has been replaced with the following Figure 6a below. Note Figure 11b, shown below as Figure 6b, has been improved to show a stronger variation in RH_i.

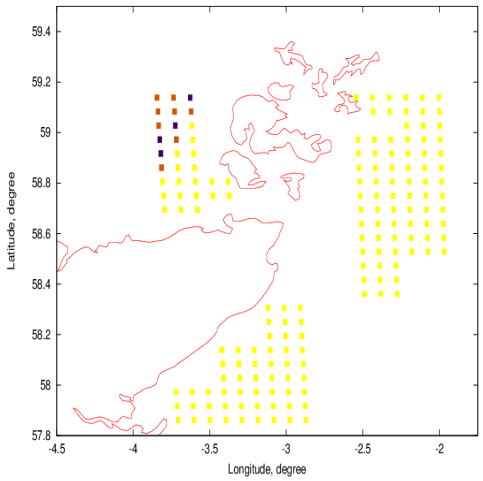

Figure 6a.

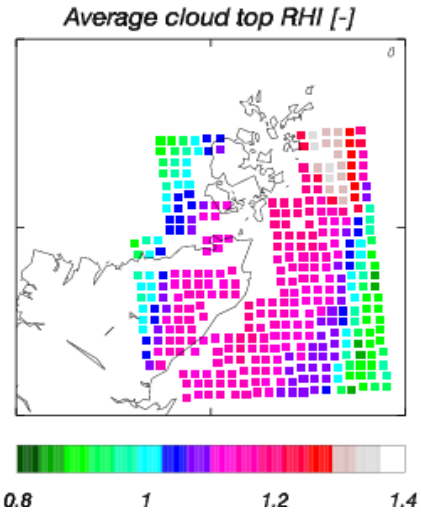

Figure 6b.

In Figure 6 the brown pixels show the locations where none of the pixels could be assigned a shape of the scattering phase function, and the blue pixels indicate the pixel locations where phase functions having distortion values of between 0 and 0.25 could equally be assigned.

The discussion of Figure 6 (Figure 11a in the re-revised manuscript) has been revised as follows on pages 33-34, lines 11-25, and 1-5, respectively.

"The PARASOL estimations of the shape of the scattering phase function, based on applying the minimised rmse and the Levene tests are shown in Fig. 11 (a). In the figure, the yellow pixels were assigned the most randomized phase function, the brown pixels are the locations where no one model phase function could be uniquely assigned. The blue pixels show the locations where either phase function model, apart from the most randomized phase function, could be assigned. The results shown in Fig. 11 (a) are now directly compared against the NWP model predicted RH_i field at the cloud-top, which is shown in Fig. 11 (b). The NWP model results are shown at a cloud-top altitude of 10 km. On comparison with Fig. 11 (a), it can be seen from Fig. 11 (b), that the most randomized phase functions (i.e, yellow squares) generally correspond to model pixels with $RH_i > 1.0$. Conversely, the 12 pixels, where no one model phase function could be assigned, generally correspond to NWP model pixels with $RH_i < 1.0$. The results for $RH_i > 1$ are broadly consistent with the findings of Gayet et al. (2011) and Ulanowski et al. (2013). The results of the former paper suggested that featureless phase functions were generally associated with $RH_i > 1.0$. Whilst the laboratory studies of Ulanowski et al. (2013), on ice crystal analogues, indicate that at higher levels of ice supersaturation, surface roughness on the ice crystal increased. This increase in surface roughness would naturally lead to featureless phase functions (Yang and Liou, 1998; Ulanowski et al. 2006; Baran 2012, and references contained therein)."

As a result of this finding, therefore, the abstract and conclusions in the re-revision have been substantially changed to reflect these new results. The changes to abstract and conclusions are as follows. In the abstract we now state the following:

"This paper reports, for this one case study, an association between the most featureless phase function model and the highest values of NWP predicted RHi (i.e., when RHi > 1.0). For pixels associated with NWP model predicted RHi < 1, it was impossible to generally discriminate between phase function models at the 5% significance level."

In the conclusions the following is now stated on pages 36-37 lines 18-25, and 1-12, respectively in the re-revised manuscript:

"For this one cirrus case, it is found that featureless phase function models, representing highly randomized ice crystals, were shown to be generally associated with NWP model RH_i values greater than unity. In the cases where the NWP model RH_i values were found to be generally less than unity, no one single scattering phase function model could be assigned to the PARASOL pixel using a quantitative statistical measure. The possibility of these pixels being affected by the issue of underlying water cloud below the cirrus was also investigated. Using high-resolution lidar images, retrievals of cirrus optical depth obtained from PARASOL and the aircraft mounted lidar, as well as generally available space-based cloud products. It was found that it is unlikely that these pixels were affected by underlying water cloud. Given this finding, the model phase functions did not have the correct structure in the backscattering part of the phase function, or more simply, there was not enough scattering angle information to be able to discriminate more clearly between the different phase function models. Given the latter reason, it would clearly be more desirable if future space-based instrumentation could resolve more clearly, and over a greater scattering angle range, the backscattering part of the cirrus phase function. This paper has also demonstrated that high-resolution interferometer data can be used, in the presence of optically thin cirrus, to retrieve the vertical profile of RH_i. This Interferometric capability, which already exists in space through IASI, could be combined with improved resolution of multiple viewing satellites to explore the relationship between atmospheric state parameters and shape of the scattering phase function on a global scale. This paper has demonstrated the potential for obtaining such global space-based measurements."

On page 37 lines 17-19, in the conclusions, we state the final remark about climate models:

"Currently, the ice radiation scheme in climate models does not take into account the changing atmospheric state as a function of ice crystal complexity. Further research in this area will prove or disprove whether this climate model assumption needs to change."

In our opinion, we have shown that the 12 pixels considered here are unlikely to have been affected by underlying water cloud. However, application of the quantitative Levene test statistic shows that, in the case of the 12 pixels, we are unable to assign a unique shape of the scattering phase function to each pixel. These results are worthy of publication.

In this paper, we also show, for the first time, that from high-resolution upwelling radiance measurements, the vertical profile of RH_i can be retrieved in the presence of optically thin cirrus. This means that high-resolution radiance measurements from current space-based interferometers such as IASI can be used to globally retrieve the vertical profile of RH_i. This is an important prognostic variable in climate models, especially at the TTL, where such measurements are difficult to assess or indeed obtain. Moreover, we also demonstrate that with well chosen microphysics a high-resolution NWP model can predict the altitude and depth of cirrus as well as the RH_i field, which were all consistent with independent measurements presented in the re-revised paper. These results will be of interest to the atmospheric physics community and will promote further research in these areas of importance to climate models.

We thank the reviewers again for their careful attention, which has led to a significant improvement to the manuscript.