The authors thank the reviewer for the constructive and useful comments. In the table below we have addressed the reviewer's comments in detail. We believe the reviewer's comments refer to the initially submitted version of the paper. In response to the reviewer's original comments we had already made changes to the manuscript addressing the reviewer's comments. Those changes were already included in the discussion paper. As a result, the line numbers referred to by the reviewer cannot be tracked in the discussion paper.

We included other minor changes as specified in the table below.

The structure needs to adjust. The introduction section failed to review the relevant literature and properly link this study to the existing literature. After reading the introduction, I hardly understand the necessity and goals of the study. The authors did not mention the aims of the study until the third section (L87, L99). After that, L90-97 abruptly summarized the outcome of the study. This part should be moved to the conclusion section.	The introduction provides the variable definition and explains its relevance in Climate Change studies. A number of relevant references had already been indicated in the previous version of the manuscript; these have been further completed in the revised version. The authors are open and would be glad to accept suggestions for other works to be referred to. The study background is defined at Pag. 7537, lines 21-24; The goal of the study is described on Pag. 7537, line 25 to Pag. 7538, line 2; The outcome of the study has been moved in the conclusions as suggested by the reviewer (Pag. 7549, lines 1-9);
As an assessment study, this manuscript used only limited validation datasets. In terms of in-situ measurements, only one site from SAFARI was selected. Have the authors checked other flux networks, i.e., BSRN, FLuxnet? With regard to peer satellite products, albedo data from another geostationary satellite were used. I think comparison with MODIS and/or CLARA-SAL will be of greater interest to the readership. The datasets are well validated and used extensively by many users	This paper, as explained on Pag. 7537 line 28 to Pag. 7538 line 2, only focuses on the most relevant positive outcome, i.e. the high temporal stability and presents a strategy to address the most relevant weakness, i.e. the degradation in quality due to undetected clouds found in the ALBEDOVAL study. Other datasets (both in-situ and satellite based), including those mentioned by the reviewer and not listed in this paper, have been used for the validation study (ALBEDOVAL). The final report, referenced in the discussion paper, is publically available. One should also consider that, as mentioned in the current version of the

	paper (Pag. 6 lines 4-10), a systematic comparison of the MSA with in-situ observations is problematic due to the lack of direct albedo measurements representative of the large areas covered by individual MSA pixels. For example, most of the FluxNet or BSRN sites providing surface albedo observations are located in rather heterogeneous landscapes and are thus not well suited for MSA validation. In the context of the ALBEDOVAL study, a very limited number of in situ sites have been identified as potential candidates for a direct case-based comparison between in situ observations and MSA values.
The presentation of the methodology and results needs improvement. Let us still use the comparison with SAFARI data as example. Many details about the data and processing are missing: how is albedo measured and calculated? What is the quality of in-situ measurements and is there any quality assurance applied? Are measurements of the entire day used to calculate albedo for the day? The authors also ignored the difference between blue-sky albedo (SAFARI) and white/black-sky albedo (MSA ACP). The blue-sky albedo can be easily calculated from DHR and BHRiso of MSA ACP, because aerosol load is jointly retrieved with surface albedo. Comparison results were plotted in Figure 8 and 9, but little discussion was given in the text. More importantly, the comparison results were not quantified (i.e., bias, standard deviation).	This paper does not focus on the SAFARI measurement campaign. The details of such a campaign are listed in the SAFARI reference. On Pag. 7544, lines 8-9, the authors clearly define the quantity measured by the SAFARI campaign instrument. It is true that there is no quantitative analysis but this is explained on Pag. 7544, lines 4-13. The comparison with the SAFARI campaign has been mentioned in the cloud contamination section in order to show that in presence of clouds the retrieval of MSA has lower quality.
Results of consistency assessment were documented in Table 3 and Figures 1-4. The analysis and discussion about the results are insufficient. The signs of the regression slopes are different across location. What are the possible reasons behind positive/negative slopes? It is true the overall intra-annual variations	The authors added the following sentence in the text: *"If the true change in land surface albedo, represented by the time series slope, is exactly zero, one can expect the retrieved slopes to be distributed around the zero and the scatter be associated with the measurement*

are generally smaller than 0.01/decade. However, the intra-daily variations from one retrieval to another are significant. Variations of BHR_iso are greater than those of DHR. These issues were not touched in the discussion.	*uncertainty."* A more detailed explanation is as follows: We calculate the mean slope and standard deviation over all slopes reported in Table 3 and we get: BHR: -0.00186 +/- 0.019 DHR: 0.00005 +/- 0.014 We now test the hypothesis that the mean slope is significantly different from zero, i.e. our null hypothesis is that the slope is actually zero. We use a two-sided t-test with the numbers referred to above for BHR (N=8, DOF=N-1, M = 0.0019, S=0.019), yielding at t-value of t = M/sqrt(N*S^2) = 0.03536. The resulting two-sided p-value (for t=0.03536 and DOF=7) is 0.9728. Thus there is 97% likelihood that the null-hypothesis is true, i.e. the data support the assumption the actual slope is zero. If we use the DHR values for the test, we end up with an even higher likelihood of 99.9%. Note, these are low-number statistics so we should not over-emphasize the value of statistical significance testing. However, the variations we are seeing are definitely not out of scope with what one would expect. The reason behind the difference between DHR and BHRis is not fully understood yet and it still needs to be addressed. One sentence has been added: *"The variations in BHRiso appear greater than the corresponding one in DHR. The reason behind this difference is not clear. One possible explanation could be the higher sensitivity of the BHRiso to aerosol retrieval and quantification."*
T in EQ(3) and 3 in EQ(1) are two	The 3σ threshold is the standard one for

important thresholds for the new cloud removal strategy. The authors need to explain how the parameters were set and how they could affect the cloud removal.	the “near certainty” condition in all statistical acceptance tests. The following sentence has been added: *“This is a standard threshold value for a ``near certainty'' condition.”* The reason for T=0.4 as value for the threshold is explained on Pag. 7548, line 1-6: the *“this value corresponds to an increase of 40 % with respect to the background value”.*