Response to Anonymous Reviewer 4

The authors would like to acknowledge the reviewer, whose comments greatly allowed to clarify and improve the quality of the paper.

Point to point response:

General Comments

The paper describes four methods to monitor radar calibration and presents strategies how these methods can be combined. The methods are self-consistency check, ground clutter return observation, solar monitoring, and reflectivity inter-calibration of two radars' data in overlapping areas.

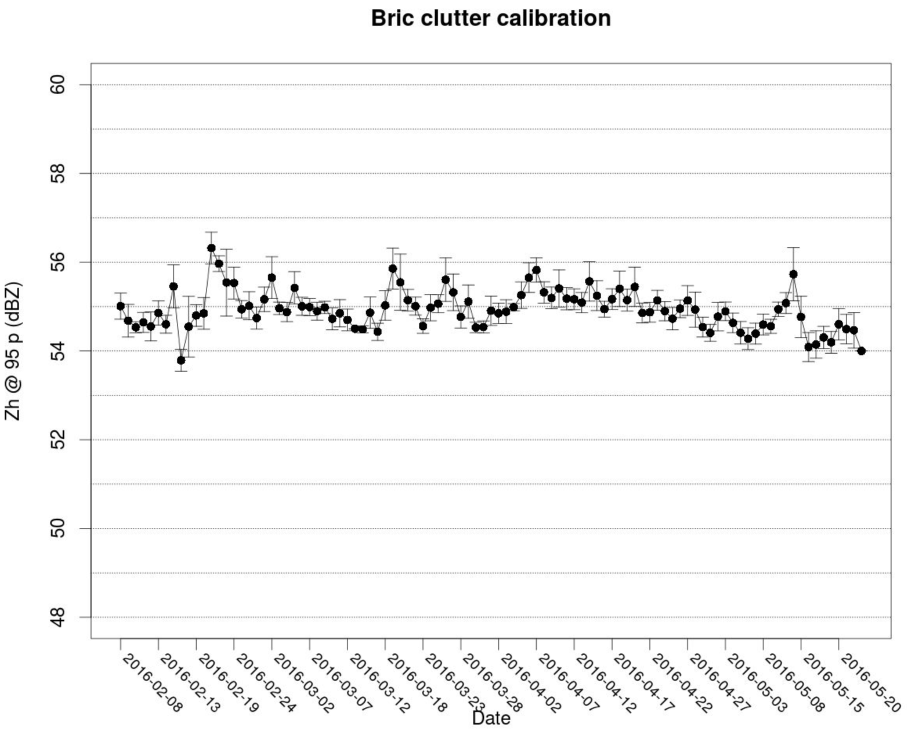
The method description and result presentation is in general quite clear. The results demonstrate that the proposed methods are useful monitoring tools. This is in particular evident from one radar having various calibrations issues and the other one not.

Parts of the algorithm description and the methods itself should be improved. This affects in particular the following subjects:

a) The intensity of ground clutter return is not only depending on weather condition and vegetation (as the authors write), but also significantly on the vertical distance between beam axis center and ground, i.e. on the effective elevation angle. The effective elevation angle is not necessarily constant; it depends on anaprop conditions and also on the limited pointing accuracy of a radar system. In particular when discussing the long-time variability of ground clutter monitoring (as shown in figures 13 and 15) one needs to know the approximate influence of elevation angle error on clutter intensity.

If for example 0.5 degree nominal elevation angle data are used for monitoring, one could provide the ECDFs (as in figure 12) once for 0.5 deg data and once for 0.6 deg data (using a sufficiently large data base, e.g. a couple of hours in clear air), and discuss the clutter differences resulting from a 0.1 degree elevation difference (which potentially amounts the typical accuracy of the effective elevation angle).

This is actually a good idea to objectively quantify the influence of the effective elevation angle. Unfortunately we may not change the operational scan, but we may now show the results over a longer time period including winter and spring to show the overall stability of the clutter calibration. The following image displays the clutter calibration from February 2016 to May 2016 for Bric della Croce radar, without the threshold at 20dBZ on Zh. In this period, the variability of the 95th percentile is within 1dB.



b) It is not clear why one method, namely the self-consistency, was performed only at the beginning and at the end of the observation period, and not repeatedly during the many weeks in between. Also, the self-consistency method strongly depends on exact differential reflectivity calculation. It is somewhat questionable that the most promising method for Zdr calibration, namely vertically pointing in rain, seems to have not been performed, although the radar systems in question are able to do so.

The Zdr calibration is not performed with the vertically pointing procedure since the operational scan strategy does not include the vertical scan, and the aim of the proposed approach is to specifically exploit only the data from operational scans, without devising any additional *ad hoc* scan.
The self-consistency method has been expanded to all the significant rainy events, i.e. the same events that are used for the inter-calibration.

c) For the inter-calibration method, attenuation seems not to be considered properly. While dry-radome conditions and sufficiently high RhoHV only are considered (but unfortunately this is explained only in the results section 4.2.1 and not already in the method describing section 3.2), attenuation as evident from differential phase shift seems not to be considered, but can have significant impact on the reflectivity data of one radar only, in particular in convective situations and for the C-band frequencies used here.

The data used in the calibration techniques with precipitation targets (self consistency and intercalibration) are corrected for attenuation and differential attenuation. This is now explained better in Section 2.1.

d) The authors seem to confuse solar monitoring of differential reflectivity with differential reflectivity calibration. On page 12 line 17 (when describing figure 17) they write: "This bias is considered to correct the Zdr measurements in the radar post-processing chain." But what does this mean? A negative Zdr average of the solar monitoring is not necessarily an indication of a Zdr mis-calibration. Instead, it may just be the compensation of a difference between the calibration constants C in eq. (8) of the horizontal and vertical channels, respectively.

This point has been clarified in the text. We used the Zdr calibration in drizzle to correct the Zdr calibration and the Sun calibration to monitor the solar Zdr.

e) Results are shown only for a period when precipitation at ground level and low atmospheric levels is liquid. The radars used for this study are operated in a region where a significant amount of such echoes is from solid precipitation during the winter months. If the authors cannot provide some results for such cases, they should at least discuss on potential limitations of each particular calibration monitoring method during winter conditions.

This is definitely correct. Both the self consistency and the intercalibration approach are intended for use in the liquid phase. In addition, the winter is the driest period in our region, so the clutter and Sun calibration becomes especially relevant during this part of the year. These limitations are now discussed in the conclusions.

Specific Comments
Methods should fully be described in the corresponding sections 3.1 to 3.4. Page 9 lines 2 to 7 belong to section 3.1 and not 4.1. Page 10 lines 18 to 24 belong to section 3.2 (as a refinement of the method description) and not to section 4.2.1. More such examples follow below.

Moved to the proper sections.

Page 4 line 28 states "Rdp(Kdp, Zdr)", but in eqs. (2) and (3) it is Rdp(Kdp) only.

Corrected.

The clutter mask mentioned in section 3.3 should be better described. Is it based on reflectivity data only, or are polarimetric moments considered? How exactly is determined if an actual measurement is clutter only? By considering reflectivity only, or by considering polarimetric moments as well? If such details were described in the cited references (Silberstein, Wolf), the authors should at least outline them here.

The clutter echoes are identified by the hydroclassifaction algorithm (Bechini and Chandrasekar, 2015). Empirical thresholds are applied to the volumes in order to be used for the clutter statistics: the percentage of meteorological echoes should be less than 1% and the percentage of clutter echoes greater than 12% of the total echoes inside the volume.

Page 7 line 17: The sun’s apparent angular diameter is not constant at 0.54 degrees but varies by about 3 percent (largest in December, smallest in June). Would that have influence on the solar calibration monitoring results?

We simulated several Sun apparent angular diameter in the range (0.57-3%;0.57+3%). The difference between the computed PTOA values is less than the uncertainty of the PTOA estimate.

Page 8 line 24 and figure 15 caption: In the text, "Fit residual standard error" is not clear. The caption mentions "square root of the differences between the measured solar power and the theoretical model". Is both the same? Does "theoretical model" in the caption refer to the "Nonlinear Least Square method" of the text? (On a side note, the text is section 3.4 and figure 15 belongs to section 4.2.3.)

The fit residual standard error is an output of the “Nonlinear Least Square” method, which is used to implement the comparison between theoretical model (Altube et al, 2015) and the measurements. Actually, the uncertainty of the fit, for each day, is evaluated as the square root of the differences between all the measured solar powers and the corresponding values computed by the theoretical model.

Page 9 line 13: "Zdr < 0" is not a good indicator of attenuation. Such may happen either if the system is not properly calibrated, or be due to random measurement accuracy. Instead, differential phase shift should be used as a measure of total path attenuation. Note again that such details should be mentioned with the method description and not with the results only.

The self-consistency technique is applied to attenuation-corrected data. Thus, the differential phase shift is not considered and differential reflectivity values less than 0dB are removed since they are unphysical in rain.

Page 9, around line 10: How is "data collected in rain" determined? Manually? Using a hydrometeor-classification? Also, this belongs to the method description in chapter 3.

The Bechini and Chandrasekar (2015) hydroclassification scheme is adopted to select echoes in rain medium. Moved to Chapter 3.

Figures 6 and 7 (and text page 9 around line 30): What means the "dBR > 11" threshold: both Rdr and Rdp above threshold, or only one (which one)?

Both rain rates must be greater than 11 dBR (logarithmic rainfall rate).

Bottom of page 10: Instead of describing "warm" and other colors, authors should give a color scale to figures 6, 7, 9, 10 and 11.

Added colorbar in Fig. 6, 7, 9, 10, 11.

Page 11, around line 15: removing all data below 20 dBZ significantly (?) alters the ECDF and thus potentially the monitoring stability. The authors should comment on that. And again, this belongs to chapter 3 and not 4.

Removing the lowest echoes the ECDF slope around the 95th percentile increases and, as consequence, the daily uncertainty decreases. Moved to Chapter 3.

Page 11 line 32: The sun’s "received power in dBm" is from both the sun’s emission and the clear air thermal noise. The latter is somewhere between -120 and -110. How is the thermal noise determined and subtracted? Figure 14 shows contour lines for the sun’s emission only, but are the radar measurements also sun’s emission only, or measurements including the thermal noise?

The thermal noise, as the reviewer suggested, is about -110dBm, while the solar power is about -98dBm. Removing this estimate of the thermal noise from the received solar power, the differences on the fit estimates are less than their uncertainty. Since we do not have at the moment a real-time accurate noise estimation, we preferred to simply use the measured power.

Page 12 line 11: "The daily PTOA value of the received solar power is generally comparable with the DRAO reference". This is no good statement (comparisons can almost always be made). Instead, the authors should write e.g. "the mean difference is X dB, and the correlation is Y".

Modified as suggested.

Page 12 lines 25 to 32: this describes Figure 18, but the self-consistency results are not included in Figure 18.

Self-consistency results are now included in Fig.18

Page 13 lines 5 to 7: If the corruption of solar signal by radio interference was observed, why was it not corrected? At least the "solar" measurements in question should have been removed before calculating the results of Fig. 18. And why are these results for the Monte Settepani radar so much worse than for the Bric della Croce radar?

We didn’t remove the radio interference in this work also to show their effect on the PTOA uncertainty. We computed the RMSE for both radar and we found 0.96dB for Monte Settepani and 0.64dB for Bric della Croce. The difference of 0.32dB is likely related to the different scan strategies for the two radars, which allows to collect different amounts of solar interferences.

Figure 12 (ECDFs): What is the meaning of the many lines with different colors?

Each line represents the ECDF for a single PPI.

Figure 13 (time series of ECDF): Instead of "Mean values of the daily values of the 95th quantile" (which probably means "daily mean values of the 95th quantiles of all day’s scans"), one could also have derived one 95th quantile value using all data of one day together. The proper English term here is "95th percentile", not "95th quantile".

Corrected.