Dear authors,

Thanks a lot for revising the manuscript. I still have major concerns, which have to be solved before the manuscript can be published. After modifications, I would like to look through the revised version.

Below I address authors' comments. The authors responses are in blue, my new comments/replies are in black.

1. The main objective of the paper is to present a calibration methodology. The methodology itself is not affected by the use of a real or theoretical target RCS. Actually, once the real target RCS is retrieved, any possible bias in the results can be corrected without changing the calibration method. We now state this more clearly in lines 190-196. The company that manufactures the targets declares having a cutting accuracy better than 0.1 mm and an alignment precision better than 0.1°, therefore we can expect a bias but it should be on the order of 1-2 dBsm. We also include now how to account for the uncertainty of an eventual target characterization (eq. 6a and lines 231-234), and indicate that the uncertainty of the target calibration may increase the uncertainty in the results (lines 530-535). Finally, as future work we now include the need of a target characterization in an anechoic chamber to correct any bias introduced by the use of the theoretical model (lines 598-600).

Citation: "However, since at the writing time we do not have an experimental characterization for our targets, we rely on the: theoretical model. This is not a major issue because, once an experimental characterization of the target becomes available, it can be used to correct any calibration bias by rectifying the value of Gamma used in the calculations"

I do not agree with the authors. What is described in the manuscript is a method (i.e. description of steps to get knowledge), not a methodology (analysis of a set of methods). And in my opinion a calibration method is worth nothing without a proper characterization of a calibration target. I think this is the first thing one should do for the radar calibration – characterize the reference target. Currently it sounds to me, that after the proposed calibration procedure another calibration steps would be required (characterization of the target and application of another bias correction) when the target is measured in a chamber. The authors claim, "A detailed analysis enabled the design of a calibration methodology which can reach a cloud radar calibration uncertainty of **0.3 dB based on the equipment used in the experiment**". This can be misleading for a reader. The authors do not reach the claimed value (0.3 dB) in the current work. As authors estimate, the real uncertainty is not known at the moment and may be in the order of 2 dB (dBsm are not proper units here, since this value is unitless in linear scale). I suggest two ways to solve this problem:
- Authors characterize the target in a chamber and add these results (cross section and its uncertainties) in the manuscript.
- Authors use sigma_rcs = 2 dB in Eq. 6a, reevaluate the results, and write explicitly in the abstract, main text, and conclusions that the uncertainty of the proposed method **at the current stage** is not better than ... dB due to uncharacterized reference target. Otherwise, it is not honest to neglect a large uncertainty source just because it is not characterized.

I would strongly recommend the authors to follow one of these ways.

2. The problem with the power units arises because power output in the BASTA radar is in an arbitrary power unit. We define this power unit as $dB(AU) = 10log_{10}(AU)$. The arbitrary unit defined as AU is proportional to watts multiplied by a unitless digital gain k_d, which depends on

The introduced changes are even more confusing. The calibration terms characterize a ratio of a real measure over the calculated one. Therefore, calibration terms must be unitless in linear scale and in dB in the logarithmic scale. I do not understand what a calibration term in dB(AU^-1m^-2) means.

The lines 72 – 76 are confusing. It is stated that kd is included to account for the units of the measured power which is in 10*log10(AU). One sentence later it is stated that kd is unitless. If it is unitless then the equation 1a has problems with units again. The nominator is unitless, the denominator has units of m^2*W as it was in the original version. I kindly ask the authors to carefully reconsider the units again.

In fact, the previous version was better, the only problem was with units notation, i.e. dB was used instead of dBm and dBsm (please see my previous comments). Please modify the units in such a way that the calibration factors are given in dB (unitless in linear scale). And please modify the units throughout the manuscript accordingly.

3. During calibration we used a Hann time window, which is the default for the BASTA radar. This is now mentioned in line 178. Additionally, we include a new figure (Figure 3) to show which gates are used to estimate the target signal. The integration of additional gates increases the signal power by less than 0.01 dB, as indicated in lines 183-185.

Thanks. It is clear now.

4. This change is included in every mention of receiver losses as L r (T, r), and is therefore propagated to the RCS and reflectivity calibration terms as well, which now depend on temperature and range (C Γ (T, r), C Z (T, r)).

I would recommend to use IF instead of r because for a different chirp configuration (slope) the relations between IF bins and range gates may change.

5. Section 5.5
In this newly added section the authors, as far as I understand, assume that during the 'passive' observations the power variability along IF depends only on gain changes. In general case this is not true:

$$Pr(IF) \sim G(IF)*(Tsys(IF) + Tamb)$$

Here Pr(IF) is the received power at IF in W, G(IF) is the linear gain of the receiver chain at IF (unitless), Tsys(IF) is system noise temperature at IF in K, Tamb is brightness temperature of the sky (or an object the radar was pointed to) in K, ~ is the proportionality sign. From this equation one can see that the received power depends on two parameters, namely the gain and the system noise temperature. If I understand right, the authors did so-called single point calibration. Using the single point calibration is it

not possible to separate the gain and the system noise temperature. Therefore, typically two-point calibrations are used in radars and radiometers. Also the authors need to know Tamb (at least with respect to Tsys(IF)). I kindly ask the authors to clarify how they took these aspects into account to calibrate all the IF bins of the radar receiver.

6. However, we did another revision of the scanning data and concluded that, at present, it is not possible to retrieve alignment information with an accuracy comparable to the antenna beam-width. This is now stated in lines 294-295. The reason is that the repeatability of the scanner positioning is not sufficient to allow a reliable retrieval under our current procedure. Additionally, we now include a discussion on how parallax errors can influence the measurements (286-290), and indicate that calibration results are compatible with parallax errors smaller than the radar beamwidth (296-298). Since we don't have information on the exact alignment, we now mention the parallel antennas only as an hypothesis (245, 299-300). Finally, we improved the calibration methodology by indicating how parallax errors can be taken into account, suggesting the addition of an additional range dependent correction function (300-301), and by introducing an uncertainty term representing the error in the antennas alignment estimation (eq. 6b, lines 243-245 and 301-302).

The assumption on parallel antennas can lead to large uncertainties. The problem with two antennas is that it is possible to measure the pattern of the receiving antenna with an external transmitter but it is often not possible to measure the transmitting antenna. Basically, with the proposed method only two points of the possible range dependent bias are characterized.

Instead of leaving this large uncertainty source untouched, I would encourage the authors to make a relatively simple estimation of possible impacts (just theoretical calculations, taking into account different divergences (magnitude and direction) of the two antennas and bias measurements at two distances). This would definitely improve the quality of the manuscript. The result of this theoretical estimation would give a proxy for sigma_a in Eq. 6a which is currently, if I understand it right, completely neglected.

Just to better understanding I give a couple of figures:
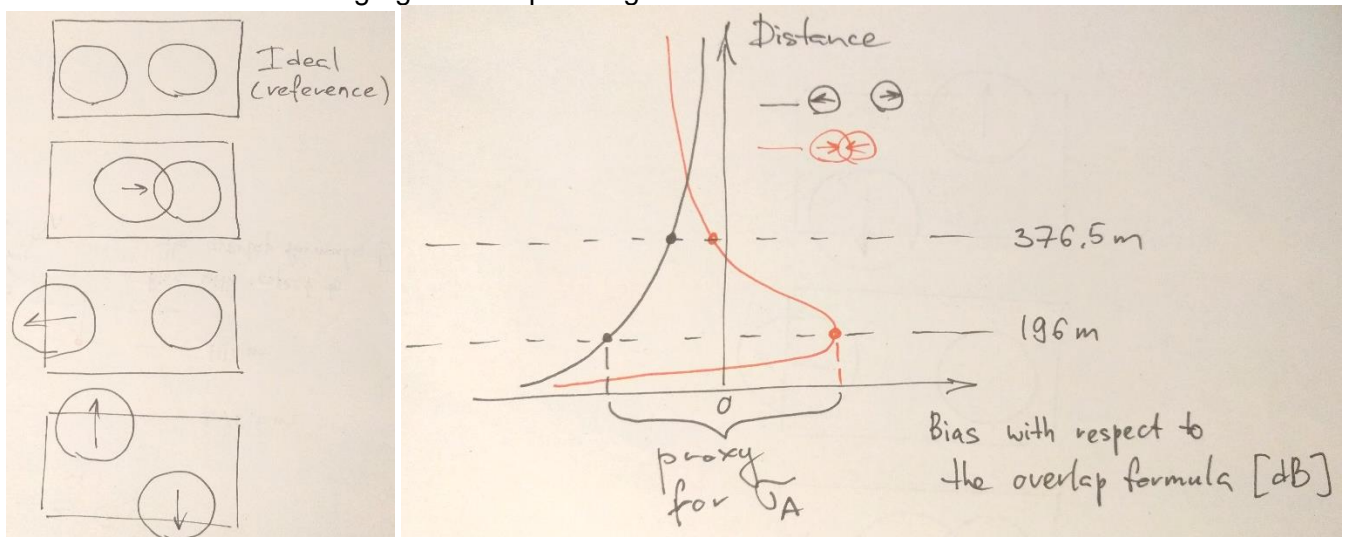
On the left figure you can see different divergence directions. On the right figure I illustrate the impact (qualitatively). The authors could perform such calculations and give an estimate for sigma_a (maximum divergence from 0 dB line).

7. To verify if data did follow a linear relationship, we did a new plot with the point density of all samples together. This figure has been added to the paper (Figure 7). In this figure it is easier to observe that deviated points are rather exceptional, with most points close to the regression. From this figure we think the 0.13 dB RMSE value is representative for most samples. We also modified Figure 6 (D). Now it is only used to introduce the data set, with the linear fit shown in new Figure 7. This produced text changes in lines 351-355, and 360-371.

In Fig. 7 the authors just masked the problem I am talking about. I agree that a majority of samples follow the linear model. But some complete iterations (like green points in Fig 6d) are off by more than 0.5 dB.