Manuscript revision and response to the referees

We would like to thank the referees again for their valuable comments. With this document, we provide a comprehensive response that attempts to track the discussion from the original referee comments over our response in the interactive discussion to our final response including the decision on specific changes in the manuscript. To allow for this tracking, we use the following color code:

black: original referee comment
blue: our original response in the Interactive discussion
green: our final response and the specific changes made in the manuscript revision

We are confident that addressing the referee comments has substantially improved the paper, and we hope that the quality of the paper now allows for publication in AMT.

Short Comment # 1 (by Daniel Michelson)

[...]

1) How to characterize data quality? Significant effort has been devoted to systematic representation of weather radar data quality in Europe, through COST Actions (717, 731) and EUMETNET OPERA, giving framework approaches for dealing with data quality. The authors have followed Zhang et al. (2011) for representing data quality resulting from beam blockage. It would be useful to have some context in the paper acknowledging previous work and including a rationale for choosing the Zhang et al. approach.

RESPONSE: We now refer to the framework for quality representation in OPERA, and briefly compare it to the approach of Zhang et al. (2011). It should be noted, though, that our general approach is open to other definitions of overall data quality, as long as such an overall value can be used to compute weights. Furthermore, we used only one single quality variable (beam blockage fraction). Hence, the challenge to aggregate several variables to a single index is not prominent in our study.

ACTION:
The following paragraph was added to the beam blockage subsection:
An alternative function to transform partial beam blockage to a quality index has been presented in other studies (Figueras i Ventura and Tabary, 2013; Fornasiero et al., 2005; Osródka et al., 2014; Rinollo et al., 2013) where the quality is zero (0) if BBF is above a certain threshold, and then linearly increases to one (1) above that threshold. It should be noted that these approaches are equally valid and can be used in determining the quality index based on beam blockage.

The following sentence has also been added to the conclusions:
In addition, with the significant effort devoted to weather radar data quality characterization in Europe (Michelson et al., 2005), and the number of approaches in determining an overall quality index based on

different quality factors (Einfalt et al., 2010), it is straightforward to extend the approach beyond beam blockage fraction.

2) What advantages does this work offer when it comes to addressing topographical beam blockage with GR compared to previous work? The paper references Bech et al. (2003) which is a benchmark paper. There are other implementations of the same approach that use other DEM data, e.g. GTOPO30. The authors' use of high-resolution SRTM data is interesting, but are the results better than using ~1 km GTOPO30 data?

RESPONSE: Generally, any increase in DEM resolution should be expected to increase the accuracy of our estimate of the beam blockage fraction (see e.g. Kucera et al. 2004). That effect could be particularly prominent in the near range of the radar as has been shown e.g. for high-resolution airborne laser DEMs by Cremonini et al. (2016). Yet, there is no reference (truth) that could be used to actually verify beam blockage estimates for any underlying DEM. What could be done, however, is to repeat the analysis with GTOPO30 data (i.e. 1 km resolution), in order to investigate the sensitivity of results, or whether our estimate of calibration bias becomes more or less consistent using GTOPO30. Yet, we already consider the paper as quite long, especially considering the requested changes in the course of this review process. We are thus hesitant to include such additional analysis that is not expected to add substantial new insight. Instead, we will add a brief discussion on the potential effects of DEM resolution on the quantification of partial beam blockage.

ACTION: We added the following sentence in section 3.1:
"*While Bech et al. (2003) used the GTOPO30 DEM at a resolution of around on kilometer, higher DEM resolutions are expected to increase the accuracy of estimates of beam blockage fraction, as shown by e.g. Kucera et al. (2004), in particular the near range of the radar (Cremonini et al., 2016).*"

3) GR calibration. There is one sentence in 3.2(2) indirectly indicating that the Subic radar may have been calibrated during the time period covered by the study. This needs to be clarified. Was the radar calibrated during this time? More than once? Are the results available? Any other maintenance that could have impacted calibration levels? This is very important to understand results like those presented in Figure 8. Also, the methods presented in this paper have been applied to data from one GR, yet they would be much more valuable if also applied to data from a second GR. Doing so would reveal which radar is "hot" and which is "cold" and whether there are any other systematic differences that are unique to each GR.

RESPONSE: We entirely agree that it would be interesting to apply the methodology to another or even several other radars in order to investigate specific characteristics of individual radars, or in other words, differences and similarities. Yet, we consider this study as a proof-of-concept in which we present the underlying methodology, and show that it adds value in estimating calibration bias for a single radar. An inter-comparison with several radars would be an additional study. Such a study should not only compare the behaviour of different radars independent of each other. It should also investigate the effect of "recalibration" on the consistency between two or more radars in regions of overlap (c.f. Warren et al. 2018). However, we consider such an analysis beyond the scope of the present paper.

Unfortunately, we were unable, despite repeated attempts, to retrieve detailed information on maintenance operations from the radar operator. The only information provided by PAGASA engineers was that in 2012 and 2013, there were problems with the transmitter which caused the output power to be very low (~80 dBm as opposed to the required 89.2 dBm). The modulator was replaced in 2014. In 2015, the magnetron was replaced. In October 2016, the supplier was changed to SELEX. We also have to assume that after performing major changes in the radar hardware, the radar engineers carried out a calibration.
ACTION: We provided the following clarification as the last sentence of 4.2(3):
We have to assume that a fundamental issue with regard to calibration maintenance was addressed between 2013 and 2014 from hardware changes (i.e. replacement of magnetron). Unfortunately, we were not able to retrieve detailed information on maintenance operations that might explain the changes in bias of the radar throughout the years.

Technical "corrections"

4) Including information throughout the paper on what software calls are available and have been used is irrelevant and should be avoided. Instead, I recommend a small section following the recommendations given by Irving: https://doi.org/10.1175/BAMS-D-15-00010.1

RESPONSE: In the revised version of the paper, we will discuss the context of Irving (2016). In fact, our paper moves beyond the "minimum standard" suggested by Irving, since we not only provide the code (a doi pointing to the wradlib package version will be generated for the final version of the paper), but we combine the enhancement of an existing software package (wradlib, extensively documented) with a fully documented application context (a jupyter notebook), combined with the data, so that we do not have to provide a log file, as suggested by Irving, but directly allow the user to run the analysis, and modify it in order to adapt to different application contexts. However, we agree that it makes sense to remove the references to explicit function calls from the main text. In order to conform with Irving's suggestions, we also add a brief section with the description of the underlying software, its dependencies, and the notebook to reproduce the results.

ACTION: The subsection **2.5 Computation details** has been added in the manuscript. Explicit references to functions have been removed from the text.

5) References to Morris and Schwaller and Schwaller and Morris are inconsistent. In the list of references, both are given from 2011, but the paper references one from 2009. The Morris and Schwaller reference appears to be incomplete in the list of references.

RESPONSE: The Schwaller and Morris citations will be updated to consistently refer to Schwaller and Morris (2011). The Morris and Schwaller reference will be updated to show complete information.

ACTION: The citations and the reference have been updated.

6) 2.1.2 Version 6 of the GPM 2AKu products is stated, but is not the current version 05B? Reference(s) to product documentation are needed.

RESPONSE: The version will be corrected to 5A instead of 6. The latest version of the product used in this paper is 5A (downloaded February 2018). The references to the product documentation will be added in the paper.
ACTION: The version number has been corrected to 5A instead of 6. The reference was added in the paper.

7) 2.3 Where does the information on bright-band height and width come from? Also precipitation type and rain indicators? Please add.

RESPONSE: The parameters are extracted from the TRMM 2A23/2A25 and GPM 2AKu products themselves. In the paper we refer to Table 3 of Warren et al. (2018) for the exact list of parameters. We will also clarify that in the text as:"*Several meta-data parameters were extracted from the TRMM 2A23 and GPM 2AKu products for each SR gate, such as [...]*"

The following table describes which parameters were extracted for each product, and how they are used in the analysis. This table will be added to the documentation in the code repository, and also to a supplementary to the paper.

Table A. Parameters extracted from TRMM 2A23 and GPM 2AKu products and the derived variables used in 3D matching

2A23 (TRMM)	2AKU (GPM)	Derived variable
rainFlag	flagPrecip	Rain/no-rain indicator
rainType	typePrecip	Precipitation type
status	landSurfaceType	Surface type
HBB	heightBB	Brightband height
BBwidth	widthBB	Brightband width
dataQuality	dataQuality qualityBB qualityTypePrecip	Overall data quality
correctZFactor	zFactorCorrected	Attenuation-corrected reflectivity
sclocalZenith	localZenithAngle	Zenith angle
-	binClutterFreeBottom	Range bin number for clutter free bottom

ACTION: We have clarified the source of the brightband width and height data in Section 3.2. We have decided, however, not to introduce a supplementary section to the paper and thus we will not include the above-mentioned Table A. Instead, we hope that the reference to Warren et al. (2018) Table 3 will suffice.

8) 2.3 GR data are acquired every 9 minutes, but matched within a 5-min window. How is this done?

RESPONSE: With GR sweeps being repeated every 9 minutes, the maximum time difference between overpass and the closest GR sweep would be 4.5 minutes. With a buffer of 30 seconds, we set five minutes as the search window *before* and *after* the overpass. We will clarify that issue in the paper.

ACTION: We have clarified the issue in Section 3.2.

9) Figures 5-7. Sub-plot (e) is a great way to visualize this kind of result, but clearer colours are needed. I'm suspecting that light-gray points are covered by dark grey points. A colour table might be a better approach, perhaps combined with slightly smaller point sizes.

RESPONSE: We agree that the very light colors for small quality values are difficult to interpret. Then again, we were, after some experiments, unable to adequately convey the visual message of weighting low quality samples less than high quality samples by using two different colors at both ends of the colormap. A color lookup table could not resolve the issue. Instead, we decided to start the "left" end of the colormap with a darker color, so that low quality samples become more visible. We will also implement the referee comment to decrease the point size in order to minimize overlaps.

ACTION: The colormap for Figures 5-7 subplot d and e have been shifted so that the left/lower end has a darker color, to increase visibility of low quality samples. The point sizes were also decreased to minimize overlaps. In addition, the colormaps of subplots a and b have been updated to the more colorblind friendly Homeyer colormap.

10) Figure 5 caption: Replace ZPR with ZSR
RESPONSE: Z_{PR} will be replaced with Z_{SR} in the caption
ACTION: Proposed changes have been implemented.

11) Figure 8. Might want to clarify in the caption that data from Jan-Apr are not used because this is the dry season.

RESPONSE: The suggestion will be implemented.
ACTION: Proposed changes have been implemented.

12) Just a thought: what impact can radome wetting have on the results? Radome wetting is still an issue even if you exclude data near the radar. But is it an issue at all at S band?

RESPONSE: In their review paper on sources of uncertainty, Villarini and Krajewski (2010) quoted Austin (1987) in that *"the radome attenuation is significant only for wavelengths smaller than or equal to 5 cm and negligible for wavelengths as long as 10 cm"* (i.e. S-band). Merceret and Ward (2000) reported wet radome attenuation for S-band to remain below 1 dB for rainfall intensities up to 100 mm/h. In summary, we expect wet radome attenuation to be a negligible effect in our study, and we will reference Austin (1987), Merceret and Ward (2002), and Villarini and Krajewski (2010) in order to support that assumption.

ACTION:
The portion of the conclusion that briefly discusses C-band radars are updated to:
For example, for the other C-band radars in the Philippine radar network, considering path-integrated attenuation would be vital. While we expect the wet radome attenuation to be negligible for this S-Band radar (Austin (1987), Merceret and Ward (2002), and Villarini and Krajewski (2010)), the same cannot be said for C-band radars.

Referee Comment # 1 (by Marco Gabella)

[...]

0) I would also suggest the make the final part of the title more specific, for instance " … by using a quality filter based on beam blockage fraction"

RESPONSE: The title will be updated to "Enhancing the consistency of spaceborne and ground-based radar comparisons by using beam blockage fraction as a quality filter".

ACTION: The title has been updated to "Enhancing the consistency of spaceborne and ground-based radar comparisons by using beam blockage fraction as a quality filter"

1) I hope the authors can agree with the following three considerations:

a) Quantitative interpretation of radar measurements are based on A MODEL of the backscattering targets.
b) Such A MODEL is an approximation of a very complex reality (Nature).
c) There is never sufficient information in radar measurements to resolve such complexity.

Having said that, I think I can now recommend more emphasis in the text related to the very different wavelengths and sampling volumes (for instance, you may want to have a look at the figures in the paper by Joss et al., 2006) characterizing GR (3 GHz) vs SPR (14 GHz, attenuating frequency!). Yes, one can try to correct for attenuation (e.g., Iguchi et al., 2000), he can even try to convert Z from 10 to 2 cm, but the uncertainties affecting the retrieved quantities are large! (See a) b) c) above …)

By the way, when introducing Eq. (2), you mention Cao et al. (2013) and coefficients in Table 1 for dry snow and hail … I have just quickly opened the pdf and saw that Table 1 lists (retrieval/ simulated) BIAS and RMSE?!?

I am confident that after (re-)considering the above mentioned issues, after thinking of the (necessarily) simplifying approach for beam occultation correction (Gaussian shape for the main lobe of the antenna radiation pattern, instead of the simple and practical linear approach proposed by Bech et al., which is an unrealistic "top-hat" radiation pattern), …, the authors will feel more comfortable with what they call "short term variability" at page 15, line 7; furthermore, they will not list "short term variability" at the first place, rather … at the last one!

RESPONSE: First, we would like to thank the referee for spotting the mix-up in references: In fact, the conversion coefficients were reported in another paper of Cao et al. (2013), and we will correct the reference accordingly:

Qing Cao, Yang Hong, Youcun Qi, Yixin Wen, Jian Zhang, Jonathan J. Gourley, Liang Liao (2013): Empirical conversion of the vertical profile of reflectivity from Ku-band to S-band frequency, J. Geophys. Res. Atm., 118, 1814-1825, doi:10.1002/jgrd.50138.

Second, we appreciate very much that the referee puts our discussion of "short term variability" into perspective. In order to avoid misunderstandings, though, we would like to emphasize that our notion of "short term variability" does *not* necessarily imply short term variability of ground radar (mis-)calibration. In the paragraph on "short term variability", we reiterate the obvious result that our bias estimates vary substantially between overpasses. We then enumerate potential causes for that variability. Admittedly, putting "hardware instability" first in that enumeration might not be justified if we intended to list the causes ordered by their relevance to explain variability. However, the present study does not provide the scope to further disentangle and rank the different sources of uncertainty, or, in other words, the different sources of ΔZ-variability between overpasses. Yet, we agree with the referee that the role of different wavelengths and sampling volumes, the role of attenuation correction at Ku band, the limitations of assuming a Gaussian antenna pattern, as well as the general limitation of the underlying backscattering model have not been sufficiently highlighted in our list of potential causes of variability. We will add the corresponding discussion and related references to the paragraph.

We will also implement the referee's suggestion to revise the order of points discussed on page 15 of the original manuscript. We will also expand the label of the paragraph from "short term variability" to "Short term variability of bias estimates between overpasses" in order to avoid misunderstandings. In line with comment #15 of this referee, the order will be changed to: 1) Effect of quality weighting on bias estimation, 2) GPM and TRMM radars are consistent, 3) Change of bias over time, 4) Short term variability of bias estimates between overpasses.

ACTION: The mix-up in the references has been corrected. The order of the four discussed items at the end of section 4.2 has been revised according to the referee's suggestion. We enhanced the scope of the 4th item (Short term variability of bias estimates between overpasses) in order to provide more context with regard to various sources of uncertainty that still affect the comparability and consistency of GR and SR observations, and thus the stability of our bias estimates. That way, we also address comments 1-3 of Referee Comment #2. We also added a brief discussion of uncertainties of our beam blockage estimation (e.g. Gaussian antenna pattern, beam propagation under different refraction, DEM errors and insufficient resolution, DEM interpolation errors) under paragraph 4.2(1).

2) What happened to the GR in 2014? (lines 20-22, page 15 and Figure 8): +1.4 dB overestimation, after two year of clear and "heavy" under-estimation! (–4.1 dB and –2.5 dB, respectively). What a jump! Was it hardware related? Software? Both? From a weather service viewpoint, it is interesting that this paper bring in the important concept of GR calibration and monitoring, see e.g. the recent successful workshop (https://www.dwd.de/EN/specialusers/research_education/seminar/2017/wxrcalmon2017/wxrcalmon_en_node.html) However, if the authors provided possible explanations of what happened, the paper would become even more interesting and valuable. If you are interested in knowing more regarding monitoring and calibration of modern radar, you may find recent paper regarding: the Transmitter chain (e.g., Reimann et al., 2016); the Receiver chain (for instance, using the Sun: Gabella et al., 2016; Hubbert, 2017,) both Transmitter and Receiver chains, using a 24 GHz vertically pointing radar and disdrometers (Frech et al., 2017).

- Gabella, M.; Boscacci, M.; Sartori, M.; Germann, U. Calibration accuracy of the dual-polarization receivers of the C-band Swiss weather radar network. Atmosphere, 2016, 7, 76.
- Reimann, J.; Hagen, M. Antenna pattern measurements of weather radars using the Sun and a point source. J. Atmos. Ocean. Technol. 2016, 33, 891–898.
- Frech, M.; Hagen, M.; Mammen, T. Monitoring the absolute calibration of a polarimetric weather radar. J. Atmos. Ocean. Technol. 2017, 34, 599–615.

However, maybe, using a more robust definition for the SPR-GR reflectivity Bias, it will come out that the jump is smaller than 3.9 dB; for what concerns the assessment of the Mean Field Bias and its statistical evaluation, please have a look at following point #3)

RESPONSE: As of today, we were unable, despite repeated attempts, to retrieve detailed information on maintenance operations from the radar operator. The only information provided by PAGASA engineers was that hardware replacements happened in 2014 and 2015, and the supplier changed in 2016, as already elaborated in our response to Short Comment #1 (comment no 3). As much as we appreciate the recommendations and background information on transmitter and receiver chains provided by the referee, we hope he agrees that, based on the level of information provided by the operator, any discussion of specific causes for the jumps would remain speculatory. We agree with the referee that this lack of information - that should exist somewhere - leaves us somehow dissatisfied.

ACTION: As already pointed out in our response to Short Comment #1 (comment no 3), we added a brief statement in at the end of 4.2(3):
We have to assume that a fundamental issue with regard to calibration maintenance was addressed between 2013 and 2014 from hardware changes (i.e. replacement of magnetron). Unfortunately, we were not able to retrieve detailed information on maintenance operations that might explain the changes in bias of the radar throughout the years.

3) From my viewpoint, the study is a bit limited in the definition of Bias assessment and the corresponding statistical metrics for the evaluation. For instance, it is going to be straightforward for the authors to derive other statistical parameters and present them in a summarizing Table that can complement the nice and informative figure 8. First of all, in addition to the annual mean of {ΔZdB*} (lines 20-22, page 15) also the standard deviation of {ΔZdB*}. Then, I would suggest a more robust definition for the Bias: instead of using dBZ, you use Z values in linear units: Z=10(dBZ/10). Then you derive a weighted average for the numerator (denominator) using linear Z of the GR (SPR). Finally, you compute 10 Log of such ratio (dB). This annual Log_of_the_MFB is more resilient than the Mean_of_the_Log presented in the paper. To avoid weighted-average or in a probability matching scheme, you may want to consider only bins with QBBF larger than say, 0.9 (or larger). After having done this selection, consider the difference (BIASxx) between different quantiles (probability matching): xx=50 (median), 75, 84, 90, 95, 99 percentiles. Maybe, BIASxx is not constant, rather it depends on the percentile? Finally, for these QBBF-selected bins, you may explore the value of the average bias E{ΔZdB} as a function of the intensity of the echo of the GR (using for instance intervals of 3 or 5 dBZ; obviously, you will have less and less samples for larger values of dBZGR). Does this Mean_of_the_Log Bias remain more or less constant? Or do you see a trend? (Maybe, SPR has residual attenuation for large reflectivity values?). Interesting, is not it?

RESPONSE: We would like to thank the referee for sharing these ideas. We will implement the suggestions as follows:

- We will add a visual representation of the standard deviation of the annual mean {ΔZdB*} in figure 8;
- We will recompute the bias estimates based on the referee's suggestion to first convert reflectivity to linear units before computing the weighted average;
- We will analyse the sensitivity of results in case we replace the weighted average by a simple quality threshold below which the samples will be discarded in the computation of calibration bias; however, we have the feeling that the paper is already very long, so we suggest to put the results of that analysis in a supplementary and only briefly refer to that in the main paper. Of course, using only partial beam blocking as a quality variable has very specific implications as to the effect of thresholding: any additional sample that exhibits a higher degree of partial beam blockage, and that we include in our computation of average reflectivity, will lead to a lower estimate of average ground radar reflectivity in the sample. Then again, reducing sample size through excessive filtering increases the standard deviation. That problem cannot really be resolved, but using the weighted average appears to us as the least arbitrary solution;
- Finally, we will add an analysis in which we investigate the dependency of our bias estimate on the intensity of the ground radar echo. Again, we suggest to present the results of that analysis in a supplementary, and only briefly discuss them in the main text.

ACTION:
Dashed lines are added in Figure 8a to represent the standard deviation of the annual mean.

After some deliberation, we would argue that processing the data in linear units would bias the results to higher reflectivities. Especially in the case where there is less confidence in SR measurements at high reflectivity (due to attenuation), that is something we definitely would like to avoid. In addition, since it is the bias in dB that we are interested in, this is the unit that should be used for averaging. We also could not confirm the hypothesis that the computing the bias in linear units would lead to more robust bias estimates, and, as a consequence, to a lower variability of bias estimates between overpasses.

Altogether, we appreciate the suggestion of the referee to consider an alternative way to calculate the mean bias, yet we decided to keep the original approach of taking the mean bias as the average of the residuals between dBZ_{GR} and dBZ_{SR}.

As for the probability matching, we have to admit that our above (blue) response was based on a misunderstanding of the referee's comment: originally, we thought the suggestion was to test the effect of discarding samples below a certain quality threshold instead of computing a weighted average. After understanding this mishap, we carried out the suggested analysis of a probability (or quantile) matching: We related the values of GR reflectivity at varying percentiles to the same SR reflectivity percentile, and plotted the difference of the two percentiles ($P_{i,GR}$ -$P_{i,SR}$) as a function of the percentile (i) itself. The results for the different years is shown in Figure A.

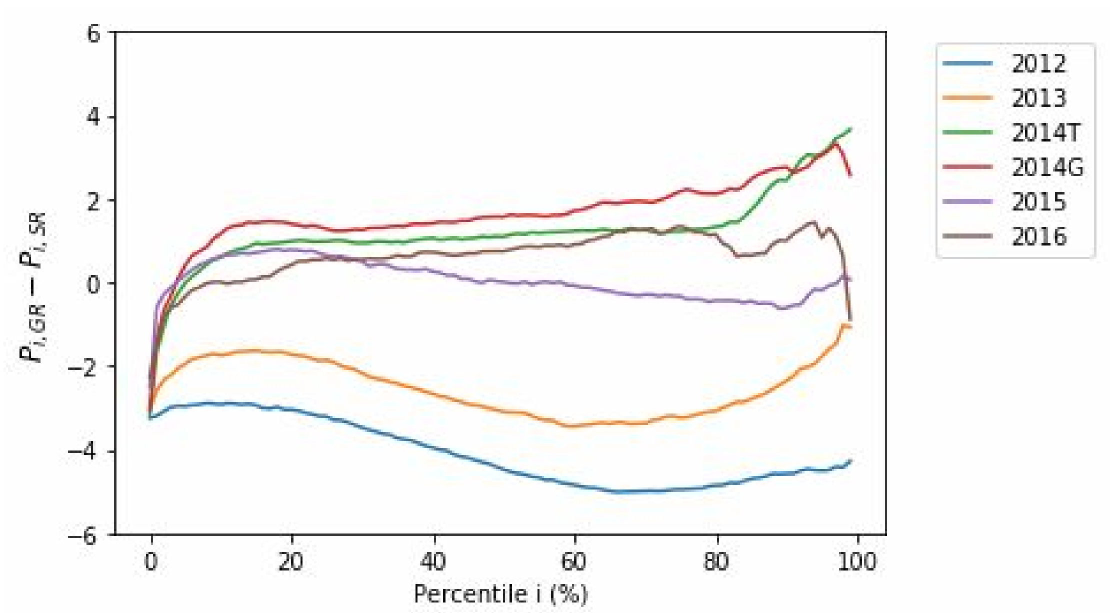

Figure A. Difference of Z_{GR} and Z_{SR} at varying percentiles

In an ideal case, the difference should not vary depending on the percentile. For all years, we can observe an increase of the differences for very low percentiles and very high percentiles. That is consistent with the findings from Warren et al. (2018): for low percentiles, the effect could be a direct consequence of the low sensitivity of the SR. For high percentiles, the increase might be related to the undercorrection of attenuation in the SR beam. Yet, for 2012, 2013, and, to a lesser extent, 2015, the bias difference decreases over a broad range of percentiles, reaches a minimum between 60 and 80 %, and then increases again. Unfortunately, we have not found an adequate explanation for that behaviour, yet.

The final analysis required by the referee relates closely to the probability matching, since it investigates how the estimated bias (or the difference between GR and SR) depends on ground radar reflectivity (hence, instead of looking at percentiles, we look at specific reflectivity classes in intervals of 1 dBZ). Instead of showing the average behaviour over a full wet season, we decided to look at specific overpasses, in order to avoid averaging over different conditions and processes over time. Figure B shows the median GR-SR reflectivity difference as a function of GR reflectivity for each overpass for the 20 overpasses with the highest number of matched samples. GR reflectivity class intervals with less than 50 matched points are shaded with a lighter color. As a consequence of this procedure, only very few overpasses provide us with a complete picture. For those overpasses, though, which have a sufficient number of samples over a wide range of reflectivity classes, the results are quite incoherent: some overpasses exhibit a similar behaviour as mentioned above (for the probability matching), and as also discussed by Warren et al. (2018): for low and high GR reflectivities, ΔZ tends to increase with reflectivity, while for intermediate reflectivities, ΔZ is rather constant. Other overpasses show a more continuous increase of ΔZ with GR reflectivity.

It might well be worth following up on these analyses: if we could understand this behaviour better, there might be reason to limit the computation of calibration bias on intermediate reflectivities, or on overpasses in which ΔZ does not substantially depend on reflectivity. At this point, however, the analysis distracts, in our opinion, too much from the focus of the paper - the effect of quality weighting. Furthermore, our understanding of these results is yet limited. As a consequence, we decided that we will not put the corresponding results neither in the main manuscript nor in a supplementary, and we hope that the referee can agree.

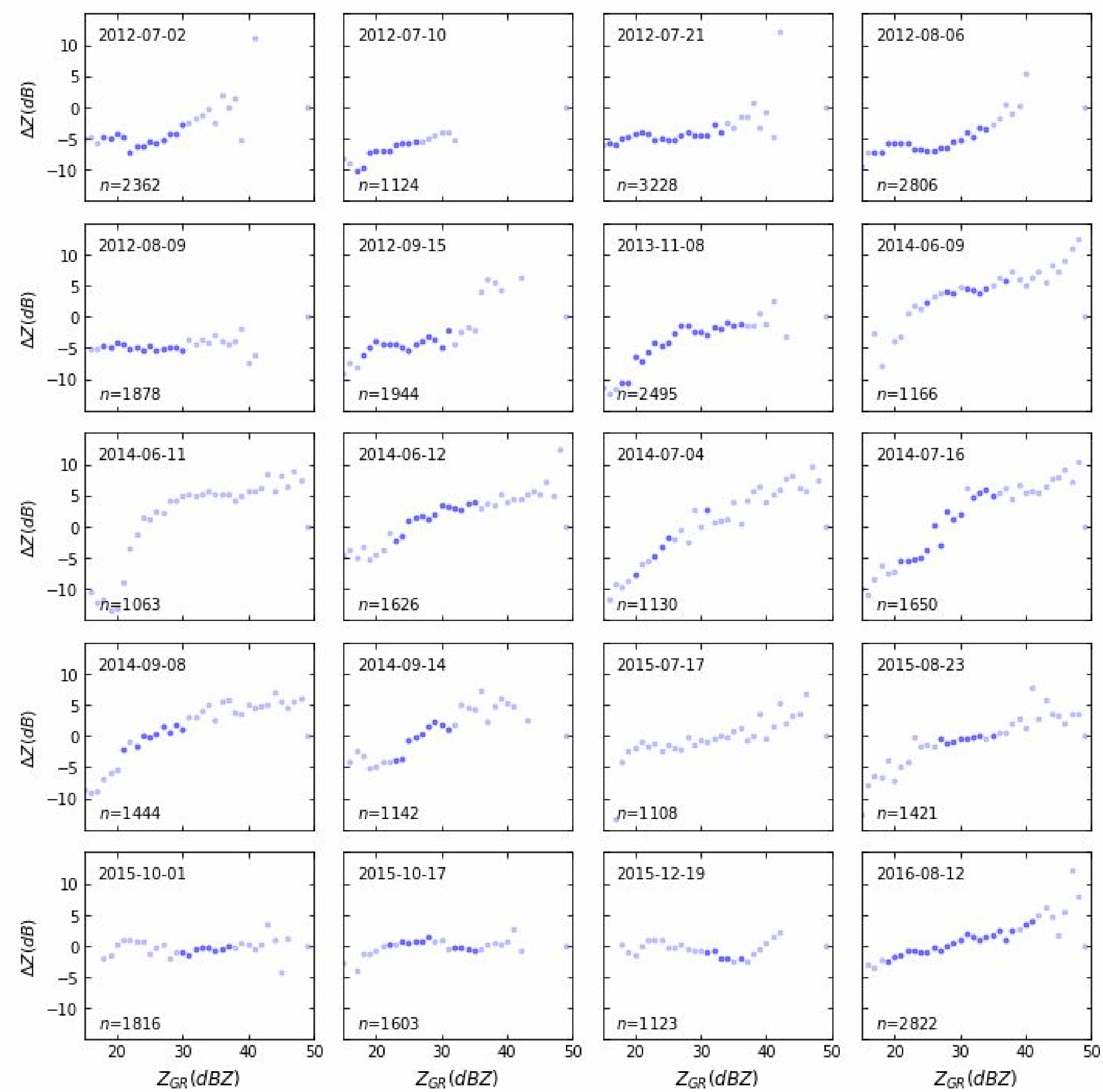

Figure B. Median GR-SR reflectivity difference as a function of GR reflectivity for each overpass for the 20 overpasses with the highest number of matched samples. GR reflectivity class intervals with less than 50 matched points are shaded with a lighter color.

4) Finally, the last issue is related to literature: while several TRMM PR vs GR papers are listed, there is a lack of DPR-related studies and DPR technical literature. regarding the latter, I have suggested at the end (GPM related references), three papers published in 2014 and 2015. Regarding the former, I have listed our recent DPR-related studies in the complex terrain of Switzerland; I am confident the authors will be able to find additional GPM papers also in other parts of the world. Furthermore, is Cao et al. double citation (at page 6) correct? Does Morris and Schwaller (2009) exist? (line 7, page 1 citation) Regarding

GPM, please, do not forget to mention that your analysis neglect Ka-band observations (please briefly discuss the reason of such a choice).

RESPONSE: Indeed the references were incomplete, missing DPR technical literature. References will be added accordingly, including intercomparisons between ground radar and the GPM DPR.
We will remove one of the references to Cao et al. (2013) on page 6, and correct the actual reference (as already pointed out above). The Morris and Schwaller (2009) citation mistake will also be corrected.

We will add a brief note that GPM Ka-band observations have not been considered in the present study, reasons being Ka band being more prone to attenuation, and limited validity of the Rayleigh scattering hypothesis in a substantial portion of rainfall cases (see e.g. Baldini et al. 2012).

ACTION: The DPR technical literature and GPM papers have been added. The Morris and Schwaller (2009) citation mistake is corrected. The redundant citation for Cao et al. (2013) has been removed.

Furthermore, we added the following sentence to section 2.1:
"*Precipitation radar data were gathered from TRMM 2A23 and 2A25 version 7 products [...] and GPM 2AKu version 5A products [...].* ***Ka band observations have not been considered due to higher susceptibility to attenuation, and a limited validity of Rayleigh scattering in a substantial portion of rainfall cases (Baldini et al. 2012).*** *From the collection of overpasses within these dates [...]*"

[...]

<u>**5)** Introduction</u>

- Line 4: … to monitor the bias of the gauge adjustment factor to be applied to precipitation estimates of the GR.
- Lines 6: … to quantify the GR reflectivity bias with respect to the reference (namely, SR reflectivity value after conversion from Ku-band to S-band).

RESPONSE: The above mentioned lines in the abstract will be revised accordingly.
ACTION: Proposed changes have been implemented.

6) In fact, I would propose the following terminology:

- Setting the Bias as close as possible to 0 dB between radar QPE and in situ measurements: *gauge-adjustment*
- Assessing the Bias between reflectivity of two radars: *relative calibration*
- Forcing to 0 dB the Bias in measured Power (dBm) between an external or internal reference Noise Source and the radar at hand: *absolute calibration*

RESPONSE: We agree with the suggestion and will revise the manuscript accordingly, introducing the labels "gauge adjustment", "relative calibration", and "absolute calibration" in the first section.

ACTION: Proposed changes have been implemented.

7) Section 2.1.2

The fact that only 283 overpasses were within the selected, reasonable 120 km range should be mentioned here. There is no reason to wait until former(see **) sec. 3.1.1. Similarly, you can at least anticipate that the number will considerably decrease upon conditional requirements such as min. # of "wet" pixels, time difference, min. # of bins above both GR and SPR sensitivity.

RESPONSE: We agree that it might be confusing to state numbers of overpasses or valid samples without already anticipating the effect of spatial limitations or additional filter requirements. We will revise the manuscript accordingly by stating these effects early in the paper.

ACTION: The paragraph in section 2.1.2 is updated as follows:
[...] from 1 June 2014 to 31 December 2016. ***From the collection of overpasses within these dates, only 183 TRMM overpasses and 103 GPM passes were within the radar coverage.*** *The data were downloaded [...]*

The paragraph in the Results and Discussion: SR-GR Matching (former section 3.1.1.) has also been updated accordingly:
From the 183 TRMM and 103 GPM overpasses that intersected with the 120 km Subic radar range, only 74 TRMM and 40 GPM overpasses were considered valid after applying the selection criteria listed in Table 2.

8) Question: have you only used only months from June to November? Not clear from the text. Please rephrase. In fig. 8, I see two overpasses in December (2012 and 2014). By the way, in Dec. 2012 E{ΔZdB*} is almost 5! dB, while the annual average is -4.1 dB?!? (see my previous points 2) and 3))

RESPONSE: We used only the months from June to December, which coincides with the rainy season in the area. We will clarify that in the text. As for the case of December 2012, upon checking the particular GR-SR match (December 5, 2012), the value of ΔZ^* is indeed very high (4.2 dB) compared to the average. Looking at the GR and SR data, the number of samples seems sufficient (n=382), and the GR overestimation is consistent for the different elevation angles. As a result, we cannot provide a consistent explanation for this outlier.

ACTION: We clarified the use of only the months of June to December in the text.

9) By the way, I would propose the following structure for Sections and Subsections

2. Data
2.1 Spaceborne Precipitation Radar (SPR)
2.2 GR

3. Method
3.1 Partial beam shielding and quality index based on beam blockage fraction
3.2 SPR-GR volume matching

3.3 Assessment of the average reflectivity Bias

<u>Section 3.1: </u>I would move it (including former fig. 4) inside the new Section 3.1 (former Sec. 2.2)

4. Results and Discussion
4.1 Single event comparison
4.1.1 Case1
4.1.2 Case2
4.2 Overall June-November comparison during the 5-year observation period

RESPONSE: We would like to thank the referee for this suggestion. We agree with the proposed structure, and will update the paper accordingly.

ACTION: The structure has been revised according to the referee's suggestion.

10) Page 5, Line 4-5: Please delete the sentence, the reader is able to read the simple algebra in eq. (1).

RESPONSE: The sentence will be deleted.

ACTION: The sentence has been deleted.

11) Page 10, Lines 19-25: misleading. I cannot possibly agree. On the contrary, my interpretation is that partial beam blockage plays approximately the same role (0.7 dB difference between the silly estimate that include blockage and the conservative one that exclude all cases where BBF > 0.5). Please rephrase.

RESPONSE: We agree that our interpretation was hard to follow, and the corresponding paragraph kind of confusing. That was also pointed out in comment #11 of referee #2. We will rephrase that part of the paper, and hope that our point will become clearer - because we still think it is quite an important one! At a low elevation angle, substantial parts of the sweep are affected by **total** beam blockage. The affected bins are either below the detection limit, or they do not exceed the GR threshold specified in Table 2 of the manuscript. As a consequence, these bins will not be considered in our matched samples, and will thus not influence our bias estimate - irrespective of using partial beam blockage as a quality filter. At a higher elevation angle, though, the same bins might not be affected by **total** beam blockage, but by **partial** beam blockage, as also becomes obvious from Fig. 4 of the manuscript. If we consider these bins in our matched samples, they will cause a systematic error in our estimate of calibration bias, unless we use the partial beam blockage fraction as a quality filter by computing a quality-weighted average of reflectivity. As a consequence, the effect of quality-weighted averaging (with partial beam blockage fraction as a quality variable) can be most pronounced at "intermediate" elevation angles, depending of course on the specific topography and the relative position of the ground radar. We had referred to that effect as "counterintuitive" since one might naively expect that the detrimental effects of beam blockage on our estimate of calibration bias would *generally* decrease with increasing elevation angle.

ACTION: The part in the paper is rephrased as follows:
*This case demonstrates how partial beam blockage affects the estimation of GR calibration bias. At a low elevation angle, substantial parts of the sweep are affected by **total** beam blockage. The affected bins are either below the detection limit, or they do not exceed the GR threshold specified in Table 2. As a consequence, these bins will not be considered in the matched samples, and will thus not influence the bias estimate - irrespective of using partial beam blockage as a quality filter. At a higher elevation angle, though, the same bins might not be affected by **total** beam blockage, but by **partial** beam blockage, as also becomes obvious from Figure 4. Considering these bins in the matched samples will cause a systematic error in the estimate of calibration bias, unless we use the partial beam blockage fraction as a quality filter by computing a quality-weighted average of reflectivity. As a consequence, the effect of quality-weighted averaging (with partial beam blockage fraction as a quality variable) can be most pronounced at "intermediate" elevation angles, depending on the specific topography and the relative position of the ground radar.*

12) Page 13: Would you please add a complementary figure at ELEV= 1.5 for the 1.10.2015 overpass? Just like you did for the 8.11.2013 overpass.

RESPONSE: We thank the referee for the suggestion, however, we are hesitant to add the additional figure as it does not provide additional insight as compared to the comparison of two sweeps for 2013-11-08, while adding to the length of the manuscript. As a compromise, we suggest to add the additional figure to the supplementary material.

ACTION: We decided not to include a supplementary section and thus also not the additional figure for the higher sweep for the 1.10.2015 overpass. While it visually reaffirms the discussion made with respect to the effect of higher elevation angles on the partial beam blockage, it does not provide an additional insight. We hope that the referee can agree.

13) Page 14 and Figure 8. Some journals ask for a graphical abstract as a self-explanatory image to appear alongside with the abstract. I think Fig. 8 would be perfect for such scope. It is nice and rich of information. Suggestion: could you please use color. For instance, the 1.10.2015 and 8.11.2013 overpasses could be in color. By the way, the 8.11 circle in picture a) seems to be very close to 0 dB, while in Fig. 7 it is written that $E\{\Delta ZdB^*\}$ is -1.1 dB. Am I missing something? Is it related to what you wrote in lines 3-6? These sentences are not clear to me, could you rephrase, please? Furthermore, regarding picture b), do not forget to emphasize that if the QBFF works properly then: $E\{\Delta ZdB^*\}$– $E\{\Delta ZdB\}$ should be negative in 2012 and 2013 (almost all the point in a) are below the 0 dB dotted line), positive in 2014 (almost all the point in a) above the 0 dB dotted line …).

RESPONSE: We thank the referee for the suggestion to highlight the two case studies in Figure 8 by color, and we will implement the suggestion accordingly. We are also grateful for suggesting a potential error, however, in this case, we do not agree: the triangle for Nov 8, 2013, represents correctly the bias estimate on that date, as an average over samples from all sweeps (-3.7dB). Apart from that, Fig. 7 refers to the overpass on October 1, 2015.

We also thank the referee for pointing out the issue of negative differences E{ΔZ*} - E{ΔZ} in Fig. 8b which we missed to discuss sufficiently in the manuscript. First, we would like to clarify that if the QBBF works properly, the difference E{ΔZ*} - E{ΔZ} should be positive - the areas suffering from partial beam blockage registers weaker signals (i.e. lower reflectivity) than expected producing the "old" lower mean bias, and giving them low weights in the calculation of mean bias brings the "new" (quality-weighted) mean bias up. In the same vein, the difference in standard deviation should be negative - the "new" standard deviation that considers quality is lower than the "old" standard deviation that does not consider quality, so that the difference between "new" and "old" standard deviation is negative. The negative differences E{ΔZ*} - E{ΔZ} are therefore inconsistencies, caused by the effect of filtering in the case of very small sample sizes. We will include this clarification in the revision.

ACTION: In Figure 8, we have added green vertical lines at the dates of the case studies so as not to interfere with the color symbolism of the markers and bars.

The paragraph in Section 4.2(1) on **Effect of quality weighting on bias estimation** has been updated to:
Figure 8b and c together illustrate the benefit of taking into account GR data quality (i.e. beam blockage) when we estimate GR calibration bias. It does not come as a surprise that the difference between ΔZ and ΔZ is mostly positive because the areas suffering from partial beam blockage register weaker signals (i.e. lower reflectivity) than expected, producing a lower mean bias. Giving the associated volume-matched samples low weights in the calculation of the mean bias brings the quality-weighted bias up. In the same vein, the beam-blocked bins introduce scatter, and assigning them low weights decreases the standard deviation. Figure 8c shows, as a consequence, that the quality weighted bias estimates are consistently more precise: in the vast majority of overpasses, the quality weighted standard deviation is substantially smaller than the simple standard deviation. That result is also consistent with the case study result shown above. It should be noted, though, that for some overpasses, the quality weighting procedure (which is in effect a filtering) can cause an increase in the bias estimate and/or the standard deviation of that estimate. That effect occurs for overpasses with particularly low numbers of matched samples, and, presumably, with rainfall in regions in which our estimated beam blockage fraction is subject to higher errors (caused by e.g. the inadequateness of the assumed Gaussian antenna pattern, variability of atmospheric refractivity, or errors related to the DEM, its resolution and its interpolation to ground radar bins). In total, however, the effect of decreasing standard deviation vastly dominates.*

14) Page 15. I would change the order of your points and list your point (1) at the end, as # (4) [see my comment 1) at page 1)]. I would start from (3), which is the scope of this paper: indeed an intelligent weighted-average based on QBFF shows a better standard deviation of ΔZdB*. By the way, I recommend adding a table and/or a figure (histogram) that summarizes the statistical properties of σ*{ ΔZdB*} and σ {ΔZdB}. Then, I would introduce the important result regarding the consistency of GPM and TRMM radars, followed by the changes of the bias in time

RESPONSE: As already pointed out in our response to comment #1 of the referee, we will change the order of points as suggested. However, we decided not to introduce additional figures in terms of histograms of bias, differences in bias, or standard deviations. These histograms would have to be provided separately for each year, because it is obvious from the time series that they would represent different populations. Apart from avoiding to introduce many new figures, the informative value of these histograms is not too high due to the limited number of samples. Instead, we will implement the referee's suggestion from comment #3 by including the standard deviation of the annual mean {ΔZdB*} in Fig. 8a.

ACTION: The order of discussed points has been changed (see also our ACTION to comment 1 of this referee). We modified FIg. 8 by showing the standard deviation of the mean annual bias by dashed lines. Please note that mean and standard deviation of the annual bias are not computed from the mean bias of each overpass, but from the total set of matched samples of an entire wet season. That is why the standard deviation is wider than the actual variability of the bias between overpasses.

15) Page 16.
Line 5, delete coherent.
RESPONSE: The word "coherent" will be deleted.
ACTION: The word "coherent" has been deleted.

16) Line 14-16. Sorry, you cannot summarize the (mis-) calibration of the GR by simply going from 2012 (–4.1 dB) to 2016 (+0.6 dB) and omit, for instance, the +1.4 jump in 2014. [see my comment 2) at page 2)].

RESPONSE: We will revise the manuscript accordingly by providing a more complete and coherent summary of the temporal changes of our bias estimate.

ACTION: The sentence has been revised to provide a more complete summary of the temporal changes of the bias estimates, and now reads as:
Analyzing five years of archived data from the Subic S-band radar (2012-2016), we also demonstrated that the calibration standard of the Subic radar substantially improved over the years, from bias levels around -4.1 dB in 2012 to bias levels of around 1.4 dB in 2014 and settling down to a bias of 0.6 dB in 2016. Of course, further studies looking at more recent years are necessary to evaluate the steadiness of the bias.

17) Line 17-19. Pleonastic. I would delete it.

RESPONSE: We would like to refer to our response to the referee's comment #11: we hope that we were able to clarify a misunderstanding there. Given that the referee agrees with our clarification, we think that lines 17-19 on page 16 are not pleonastic, but rather an important note to emphasize that moving to higher elevation angles does not necessarily help to avoid the problems introduced by beam blockage in the specific case of comparing GR and SR observations. Nevertheless, we will also revise the corresponding paragraph in the conclusions section in order to make it more comprehensible.

ACTION: We hope that the revision of the paragraph in the discussion of Case Study #1 has clarified the misunderstanding. We wanted to emphasize that moving to higher elevation angles does not necessarily help to avoid the problems introduced by beam blockage in the specific case of comparing GR and SR observations. With the clarification, we think that the corresponding lines in the conclusion is now comprehensible as it is.

18) Line 26. Why do you discuss C-band radar technology ?

RESPONSE: Lines 24-28 on page 16 of the original manuscript were intended to provide a brief perspective for future studies, in which we mention that for C-band radars, it would be important to include path-integrated attenuation as a quality variable. In the revised version, we will clarify that point.

ACTION: In the conclusion (section 5), we changed the corresponding sentence as follows:
For example, if we consider C-band instead of S-band radars, path-integrated attenuation needs to be taken into account for the ground radar, and wet radome attenuation probably as well (Austin, 1987; Merceret, 2000; Villarini and Krajewski, 2010).

19) Minor points
My proposal for radar acronyms: 2-character for ground, namely GR; 3-character for satellite radar. Would you please use TPR for TRMM, DPR for GPM and SPR in those cases where you refer to both, independently of the platform

RESPONSE: We appreciate the suggestion. Yet, we think that distinguishing the different spaceborne platforms via acronyms might cause more confusion than clarification, in particular since we rarely address the different platforms separately in the main text. We would thus prefer to stick with GR vs. SR in general.

ACTION: Based on our above (blue) response, we decided to keep the acronyms as is.

Referee Comment #2 (Anonymous)

[...]

Comments on other sources of uncertainty in calibration assessment:

1) Attenuation at Ku-band:
The authors should address the uncertainties with attenuation correction at Ku-band. The

attenuation correction tech. used for just Ku-band is the HB-SRT method (Seto and Iguchi 2015). It is known that using the HB method alone does not work well in higher rain rates (> 20 mm hr-1, Seto and Iguci 2011, but as low as 12 mm hr-1 Rose and Chandrasekar 2005). Furthermore, the SRT method is more uncertain over land (larger standard deviation of the surface backscatter cross-section, Meneghini et al. 2000). It is anticipated that since the radar is located in the tropics both of the issues above could occur (more likely in convective precipitation). Please discuss these uncertainties and how they could impact your results of the bias correction. It is mentioned in the conclusions that for C-band attenuation correction is vital, but GPM and TRMM are Ku-band, thus isn't it vital as well?

RESPONSE: We agree that attenuation correction is vital for both GPM and TRMM at Ku-band, and there is certainly a large body of literature concerned with the related effects, including the effects of nonuniform beam filling (NUBF) on the attenuation correction procedure. In the present study, we have only used the attenuation-corrected reflectivity values without considering the uncertainty associated with the correction procedure. In the revised manuscript, we will explicitly refer to the uncertainty introduced by attenuation correction. We will also, in the conclusions, provide an outlook on including the spaceborne reflectivity observations in the framework of quality-weighted averaging, just as we suggested for the ground radar observations. That would imply to use the estimates of PIA which are provided through the SR meta-data as a quality variable and thus to consider it in the quality-weighted average of SR reflectivity in the matched samples.

ACTION: We have addressed attenuation (correction) at Ku band explicitly in our discussion of uncertainties in section 4.2(4). In the conclusions (section 5), we provided an outlook on including SR observations in the framework of quality-weighted averaging, with the addition of this sentence:
*The framework could also be extended by explicitly assigning a quality index to SR observations, too. In the context of this study, that was implicitly implemented by filtering the SR data e.g. based on bright band membership. An alternative approach to filtering could be weighting the samples based on their proximity to the bright band, the level of path-integrated attenuation (as e.g. indicated by the GPM 2AKu variables **pathAtten** and the associated reliability flag (**reliabFlag**)) or the prominence of non-uniform beam filling (which could e.g. be estimated based on the variability of GR reflectivity within the SR footprint, see e.g. (Han et al., 2018)).*

2) Ground Clutter for the SR:
In radar gates near the surface, with respect to the SR, ground clutter is a problem. How are the authors dealing with ground clutter from the SR? Are they using gates below the lowest clutter free bin estimate (included in the GPM file)? If so, is the lowest clutter free gate being assigned to all the gate below it? If you plot it out, a lot of times that's what is done. Essentially the data looks smeared from the lowest clutter free bin to the surface, which isn't to realistic and it is suggested to just not consider these gates. Please comment on this, potentially in Section 2.3. If you are including these interpolations, you may wish to not (it will introduce error).

RESPONSE: Thanks for pointing out this issue which has not yet been sufficiently clarified in the original manuscript. While TRMM 2A25 contains a clutter flag for the variable "Corrected Z-factor" (-8888 indicates ground clutter), the GPM 2AKu product contains a variable "binClutterFreeBottom" to indicate the lowest clutter free bin in a ray. In both cases, TRMM and GPM, we use the SR clutter information to discard the affected bins. We will clarify that point in the revised manuscript, using both table 2 (filtering criteria), and the new table with metadata variables that we introduced as a response to comment #7 of SC1 (as part of the the supplementary).

ACTION: First, we would like to apologize for a mistake we made in our initial response. We had in fact not considered the lowest clutter free bin for the GPM data in the submitted manuscript. We discarded clutter affected bins based on TRMM clutter flags, but did not explicitly consider the variable "binClutterFreeBottom" for GPM. We would like to thank the referee for bringing up this point. Upon investigating the binClutterFreeBottom variable (i.e. the lowest clutter-free bin), we have noticed that the average altitude of that bin is at about 3000 meters (with a range of 1300m-7200m) in the study area. Given that the highest mountain peaks are around 2000 meters, we do not have much confidence in that variable. Assuming clutter-contaminated SR bins at such altitudes and thus discarding the corresponding SR bins would exclude the majority of the volume covered by our ground radar. We thus decided to use the GPM data as is, although the poor clutter identification surely merits further attention in the future, and has also been confirmed by other studies (e.g. Watters et al., ERAD 2018 book of abstracts)

3) NUBF:
Please also include some discussion of the potential impacts of non-uniform beam filling (NUBF) on your analysis. Edges of large systems, individual cumulus showers could result in NUBF in SR because of the quasi-large footprint. Lowering the reflectivity value in the gate.

RESPONSE: We agree that non-uniform beam filling can cause errors in particular for the SR platform which might become more pronounced in case of path-integrated attenuation is present and being corrected for. Durden et al. (1998) provided an excellent discussion of potential effects. Han et al. (2018) attempted to consider the effect in case GR and SR observations are matched, by using the - comparatively highly resolved - GR observations in order to compute the standard deviation of reflectivity in an SR footprint as a measure of NUBF. From the literature, it is hard to tell how much systematic error is introduced in SR measurements by the effects of NUBF. However, the three comments of this referee (reg. attenuation, clutter, NUBF) were very helpful for us to understand the necessity of extending the framework of quality-weighted averaging to the SR, too. So while we consider our present manuscript as a proof-of-concept in the consideration of quality, follow up studies should attempt to achieve a more general implementation that not only includes additional quality variables for the GR data, but that also applies these to the SR observation which already come with extremely rich and helpful meta-data to support such attempts. While our study tries to minimize the effects of NUBF (by setting a minimum fraction of GR bins within the SR footprint to exceed a minimum reflectivity threshold, see table 2 of the original manuscript), a future framework for SR quality might rather consider the variability of GR bins in the SR footprint, as suggested by Han et al. (2018).

ACTION: We explicitly included NUBF in our list of uncertainties in section 4.2(4). Furthermore, the following sentence was added to the conclusions (section 5):
[...] proximity to the bright band, **the level of path-integrated attenuation (as e.g. indicated by the GPM 2AKu variables "pathAtten" and the associated reliability flag ("reliabFlag"),) or the prominence of non-uniform beam filling (which could e.g. be estimated based on the variability of GR reflectivity within the SR footprint, see e.g. Han et al. (2018))**. *In addition, [...]"*

Specific Comments:

4) Page 2, line 5: Please add the Kummerow et al. (1998) paper for TRMM, and the Hou et al. (2014) for the GPM reference (page 2, line 6). This will help readers who are not entirely familiar with both platforms.

RESPONSE: The Kummerow et al. (1998) and Hou et al. (2014) citations and references will be added.
ACTION: The references have been added.

5) Page 6, line 3: "The gates below and above the brightband were considered in the comparison". Please provide a brief reason why this is done. I do not want to assume the author's reasoning.

RESPONSE: According to Warren et al. (2018), the frequency-corrected reflectivities within the melting layer (bright band) appear underestimated compared to the ones below and above the melting layer. In addition, while usually the samples above the brightband are used in GPM validation, there are significantly more samples below the melting layer, especially in a tropical environment such as the Philippines.

ACTION: The following explanation is added in section 3.2:
...Only gates below and above the brightband were considered in the comparison. *Warren et al. (2018) found a positive bias in GR-SR reflectivity difference for volume-matched samples within the melting layer, compared to those above and below the melting layer. They speculated that this was due to underestimation of the Ku- to S-band frequency correction for melting snow. In addition, while usually the samples above the brightband are used in GPM validation, there are significantly more samples below the melting layer, especially in a tropical environment such as the Philippines.* To ensure that there are sufficient bins...

6) Figure 4 & Section 3.1: It is not clear what you are plotting. The figure titles state the quality index but the figure caption and text states beam blockage fraction. Please clarify.

RESPONSE: The caption has been updated to match the figures: *Quality index map of the beam blockage fraction for the Subic radar at (a) 0.0° (b) 0.5° (c) 1.0° and (d) 1.5° elevation angles.*

ACTION: The caption has been updated.

7) Section 3.1.1: Why are the number of overpasses here different than when they were listed earlier (section 2.1.2)? I am referring to the numbers before applying the criteria in Table 2.

RESPONSE: Applying the criterion of "Minimum number of pixels tagged as rain = 100" eliminates several overpasses. Only this criteria affects the number of overpasses, not the others listed in Table 2. We will clarify this in the paper.

ACTION: The number of overpasses stated in section 2.1 has been updated to reflect the number of overpasses that intersected with the radar coverage. This is now consistent with the numbers mentioned in the results section.

8) Case studies (Section 3.1.2 and 3.1.3): Could you include the mean BB level height? You can add it to the bottom right with the other statistics. Also comment on fraction of stratiform vs convective. These two will help readers assess the amount of attenuation and NUBF that could be involved (e.g. uncertainty in the SR measurements).

RESPONSE: The mean BB level height will be added to the figure as suggested. While stratiform rain dominates the precipitation type for most cases, convective rain is significantly represented, hence we decided to keep both rain types in the analysis.

ACTION: We added the mean BB level height to the caption of Figures 5 and 7.
Figure 5. [...] The mean bright band level is at a height of 4685 meters.
Figure 7. [...] The mean bright band level is found at 4719 meters for this case.

9) Figure 5 + 6 + 7 a and b: Suggestion. Consider changing the colorscale to one that is perceptually uniform and color-deficient friendly. For example, try the HomeyerRainbow or the LangRainbow included in Pyart (https://github.com/ARM-DOE/pyart)

RESPONSE: We thank the referee for the suggestion. Upon trying the different colormaps proposed, we decided that we will go with the HomeyerRainbow colormap. The figures will be updated to reflect the new colormap.

ACTION: The colormaps for Figure 5+6+7 a and b have been updated to follow the HomeyerRainbow colormap in PyART. The scale of the colormap for subplots c and d have also been shifted such that the lowest value is darker than the previous version, for better visibility, following the suggestion in comment (9) of Short Comment #1. The point size of the scatter plot has also been reduced to minimize overlaps.

10) Page 10, Line 12: "Major parts of that sector did not receive any signal due to total beam blockage". Where is this occurring? The reader can refer back to Figure 4, but it might be helpful to outline the circles with a thin black line in Figure 5d where there is SR data, but no GR data. That way the readers would see where there is 100% beam blockage and thus no signal from the GR, but also gain insight of size of the precipitating system.

RESPONSE: The figures for the case studies show only the matched bins, but the referee is right, information such as location of bins where there is SR signal but no GR signal and the size of precipitating system are not conveyed. We will address this by showing all the available SR bins for the first panel and outlining the circles with SR data but no GR data in black, as suggested.

ACTION: In Figures 5-7, the SR bins where SR data is present but not GR data is encircled in black, as suggested. Correspondingly, the text in Section 4.1.1 now reads:
*Major parts of that sector did not receive any signal due to total beam blockage**, highlighted in Figure 5a with black circles showing the bins where the GR did not have valid observations.***

11) Page 10, Line 24-25: "That might be considered counterintuitive, as one might expect the blockage to disappear with higher elevations". Please provide some discussion explaining why this is the case.

RESPONSE: We thank the referee for pointing out the lack of adequate explanation. As can be seen also from the comments of referee #1, this paragraph appears to be confusing in the original manuscript. We will revise the paragraph accordingly in order to make our point clearer. Please also refer to our response to the comment #11 of referee #1.

ACTION: The confusing paragraph has been updated and a better discussion of the effect of "intermediate" elevation angles on the partial beam blocking has been included. Please also refer to our response to the comment #11 of referee #1.

12) Page 16, Lines 13 – 16. 'We could' and 'we could also' imply that you did not conduct this analysis when it seems you have. I suggest to change these phrases to be definitive. 'We showed that…' 'we also demonstrated that…'
RESPONSE: The sentences will be updated as suggested.
ACTION: The sentences have been updated as suggested.

Technical corrections:

13) Page 3, line 20: The most current GPM version is version 5, version 6 is not released yet.
RESPONSE: The version will be corrected (version 5A instead of 6).
ACTION: The version has been corrected (version 5A instead of 6).

14) Page 18, line 18: Reference Cao et al. 2013 is incorrect. It should be:
Empirical conversion of the vertical profile of reflectivity from Ku-band to S-band frequency

RESPONSE: We apologize for the mixup. The citation and reference will be corrected to refer to

Cao, Qing, Yang Hong, Youcun Qi, Yixin Wen, Jian Zhang, Jonathan J. Gourley, and Liang Liao. 2013. "Empirical Conversion of the Vertical Profile of Reflectivity from Ku-Band to S-Band Frequency." *Journal of Geophysical Research: Atmospheres* 118 (4): 1814–25. https://doi.org/10.1002/jgrd.50138.

ACTION: The citation and reference has been corrected.

15) The reference Warren et al. should be 2018, published Feb 2018 in J. Atmo. + Ocean. Tech.. Page 2, line 8;Page 3,line 25;Page 5,line 11;Page 15,line 14
RESPONSE: The citations and reference will be corrected.
ACTION: The citations and references have been corrected.

16) Figure 4: Missing y-ticks and tick labels on bottom left subplot
RESPONSE: Axis labels will be restored in Figure 4. The color scheme has been changed so that the lightest color is made a bit darker for better visibility in Figures 5-7 subplots d and e, following the suggestion of another reviewer.
ACTION: The axis labels have been restored.

17) Page 8, line 5-6. No need for new paragraph. You can combine the two.
RESPONSE: The paragraphs will be combined as suggested.
ACTION: The paragraphs have been combined.

18) Figure 5: Figure caption has Zpr instead of Zsr
RESPONSE: Z_{PR} will be replaced with Z_{SR} in the caption
ACTION: The caption has been updated.

References

Austin, P.M. (1987): Relation between measured radar reflectivity and surface rainfall. Mon Weather Rev., 115, 1053-1071.

Baldini, L., V. Chandrasekar, D. Moisseev (2012): Microwave radar signatures of precipitation from S band to Ka band: application to GPM mission, European Journal of Remote Sensing, 45:1, 75-88, DOI: 10.5721/EuJRS20124508.

Biswas, S. K. (2017): Cross Validation of Observations from GPM Dual-Frequency Precipitation Radar with S-Band Ground Radar Measurements over the Dallas — Fort Worth Region. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX, USA: IEEE. https://doi.org/10.1109/IGARSS.2017.8127393.

Cremonini, R., Moisseev, D., and Chandrasekar, V. (2016): Airborne laser scan data: a valuable tool with which to infer weather radar partial beam blockage in urban environments, Atmos. Meas. Tech., 9, 5063-5075.

Durden SL, Haddad ZS, Kitiyakara A, Li FK (1998): Effects of nonuniform beam filling on rainfall retrieval for the TRMM precipitation radar. J. Atmos. Oceanic Technol. 15: 635.

Han, J., Z. Chu, Z. Wang, D. Xu, N. Li, L. Kou, F. Xu, Y. Zhu (2018): The establishment of optimal ground-based radar datasets by comparison and correlation analyses with space-borne radar data, Meteorol. Appl. 25, 161-170.

Irving, Damien. 2016. "A Minimum Standard for Publishing Computational Results in the Weather and Climate Sciences." Bulletin of the American Meteorological Society 97 (7): 1149–58. https://doi.org/10.1175/BAMS-D-15-00010.1.

Kucera, P. A., W. F. Krajewski, and C. B. Young (2004): Radar Beam Occultation Studies Using GIS and DEM Technology: An Example Study of Guam. Journal of Atmospheric and Oceanic Technology 21 (7): 995–1006.

Merceret, F. J., J. G. Ward (2002): Attenuation of Weather Radar Signals Due to Wetting of the Radome by Rainwater or Incomplete Filling of the Beam Volume, Technical Report NASA/TM-2002-211171, NAS 1.15:211171, 20 p., URL: https://ntrs.nasa.gov/search.jsp?R=20020043890

Villarini, G., W. F. Krajewski (2010): Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surv Geophys (2010) 31:107–129

NASA. 2017. "Release Notes for the PR Level 1 Products." Available online at https://pps.gsfc.nasa.gov/Documents/ReleaseNote_PU1_productV05.pdf.

Warren, Robert A., Alain Protat, Steven T. Siems, Hamish A. Ramsay, Valentin Louf, Michael J. Manton, and Thomas A. Kane. 2018. "Calibrating Ground-Based Radars against TRMM and GPM." Journal of Atmospheric and Oceanic Technology, February. https://doi.org/10.1175/JTECH-D-17-0128.1.

Watters, Daniel, Alessandro Battaglia, Kamil Mroz, and Frederic Tridon. 2018. "On the Capabilities of the GPM Core Observatory over Great Britain and Ireland." In 10th European Conference on Radar in Meteorology and Hydrology. Ede, Netherlands.

Zhang, JIAN, YOUCUN Qi, K. Howard, C. Langston, and B. Kaney. 2011. "Radar Quality Index (RQI)—A Combined Measure of Beam Blockage and VPR Effects in a National Network." In Proceedings, International Symposium on Weather Radar and Hydrology.

Other changes made in the manuscript

1. The first two sentences in the abstract were deleted.

2. Slight changes in Figures:
    a. Figure 1a - modified axis labels
    b. Figure 1b - modified background color

c. Figure 4 - removed boxes
d. Figure 8b and c - used broken y-axis

3. The date of download for TRMM and GPM was included in Section 2.1.
"The data were downloaded from NASA's Precipitation Processing System (PPS) through the STORM web interface (https://storm.pps.eosdis.nasa.gov/storm/) on 15 February 2018 for TRMM and 14 June 2018 for GPM."

4. TRMM v7 and GPM v5 PR products differ by +1.1 dB (NASA 2017). An additional sentence to mention this was added in Section 2.1:

   > *"It is important to note that, at the time of writing, changes in calibration parameters applied in the GPM Version 5 products resulted in an increase of +1.1 dB from the corresponding TRMM version 7 products (NASA 2017)."*

   And in Section 4.2(2):

   > *"The difference between TRMM version 7 and GPM version 5 reflectivities mentioned in Section 2.1 falls within the uncertainties in the annual estimated mean bias, which makes us confident that the substantial year-to-year changes of our bias estimates are based on changes in GR calibration."*

Enhancing the consistency of spaceborne and ground-based radar comparisons by using beam blockage fraction as a quality ~~filters~~filter

Irene Crisologo[1], Robert A. Warren[2], Kai Mühlbauer[3], and Maik Heistermann[1]

[1]Institute of Earth and Environmental Sciences, University of Potsdam, Germany
[2]School of Earth, Atmosphere and Environment, Monash University, Australia
[3]Meteorological Institute, University of Bonn, Germany

Correspondence: Irene Crisologo (crisologo@uni-potsdam.de)

Abstract.

~~Coinciding monsoon and typhoon seasons in the Philippines cause torrential rainfall, and associated hazards such as flooding and landslides. While early warning systems require accurate radar-based rainfall estimates, low-density rain gauge networks in the Philippines make it challenging to monitor the calibration of the ground-based radars (GRs). As an alternative, we~~ We explore the potential of spaceborne radar (SR) observations from the Ku-band precipitation radars on board the TRMM and GPM satellites as a reference to quantify the ~~calibration bias of an S-band GRin the Philippines~~ground radar (GR) reflectivity bias. To this end, the 3D volume-matching algorithm proposed by ~~Schwaller and Morris (2009)~~ Schwaller and Morris (2011) is implemented and applied to five years (2012-2016) of observations. We further extend the procedure by a framework to take into account the data quality of each ground radar bin. Through these methods, we are able to assign a quality index to each matching ~~SR-GR~~ SR–GR volume, and thus compute the GR calibration bias as a quality-weighted average of reflectivity differences in any sample of matching ~~GR-SR~~ GR–SR volumes. We exemplify the idea of quality-weighted averaging by using the beam blockage fraction as ~~a~~ the basis of a quality index. As a result, we can increase the consistency of SR and GR observations, and thus the precision of calibration bias estimates. The remaining scatter between GR and SR reflectivity, as well as the variability of bias estimates between overpass events indicate, however, that other error sources are not yet fully addressed. Still, our study provides a framework to introduce any other quality variables that are considered relevant in a specific context. The code that implements our analysis is based on the open source software library *wradlib*, and is, together with the data, publicly available to monitor radar calibration, or to scrutinize long series of archived radar data back to December 1997, when TRMM became operational.

1 Introduction

Weather radars are essential tools in providing high quality information about precipitation with high spatial and temporal resolution in three dimensions. However, several uncertainties deteriorate the accuracy of rainfall products, with calibration contributing the most amount (Houze Jr et al., 2004), while also varying in time (Wang and Wolff, 2009). While adjusting ground radars (GR) by comparison with a network of rain gauges (also know as *gauge adjustment*) is a widely used method, it suffers from representativeness issues. Furthermore, comparison with rain gauge observations gauge-adjustment accumulates uncertainties along the entire rainfall estimation chain (e.g. including the uncertain transformation from reflectivity to rainfall rate), and thus does not provide a direct reference for the measurement of reflectivity. The Relative calibration (defined as the assessment of bias between the reflectivity of two radars) has been steadily gaining popularity, in particular the comparison with space-borne precipitation radars (SR) (such as those the precipitation radar on-board the Tropical Rainfall Measuring Mission (TRMM; 1997-20141997–2014; Kummerow et al. (1998)) and the dual-frequency Precipitation Radar on the subsequent Global Precipitation Measurement mission (GPM; 2014-present))has been steadily gaining popularity, since 2014–present; Hou et al. (2013))). Several studies have shown that surface precipitation estimates from GRs can be reliably compared to precipitation estimates from SRs for both TRMM (Amitai et al., 2009; Joss et al., 2006; Kirstetter et al., 2012) and GPM (Gabella et al., 20 In addition, a major advantage of relative calibration and gauge adjustment in contrast to the absolute calibration (i.e. minimizing the bias in measured power between an external or internal reference noise source and the radar at hand) is that they can be carried out a posteriori, and thus be applied to historical data.

Since both ground radars and space-borne precipitation radars provide a volume-integrated measurement of reflectivity, thus allowing for a direct comparison of the observations can be done in three dimensions (Anagnostou et al., 2001; Gabella et al., 2006, 2011; ? In addition(Anagnostou et al., 2001; Gabella et al., 2006, 2011; Keenan et al., 2003; Warren et al., 2018). Moreover, as the space-borne radars are and have been constantly monitored and validated (with their calibration accuracy proven to be consistently within 1 dB; Kawanishi et al. (2000); Takahashi et al. (2003)) (TRMM: Kawanishi et al. (2000); Takahashi et al. (2003); GPM: Furukawa et al. (2015); Kubota et al. (2014); Toyoshima et al. (2015)), they have been suggested as a suitable reference for ground radar calibration relative calibration of ground radars (Anagnostou et al., 2001; Islam et al., 2012; Liao et al., 2001; Schumacher and Houze Jr, 2003).

Comparing measurements of space-borne radars with ground radars Relative calibration between SRs and GRs was originally suggested by Schumacher and Houze (2000), but the first method to match SR and GR reflectivity measurements was developed by Anagnostou et al. (2001). In their method, SR and GR measurements are resampled to a common three-dimensional grid. Liao et al. (2001) developed a similar resampling method. Such 3D-resampling methods have been used in comparing SR and GR for both SR validation and GR bias determination (Bringi et al., 2012; Gabella et al., 2006, 2011; Park et al., 2015; Zhong et al., 20 (Bringi et al., 2012; Gabella et al., 2006, 2011; Park et al., 2015; Wang and Wolff, 2009; Zhang et al., 2018; Zhong et al., 2017) Another method was suggested by Bolen and Chandrasekar (2003) and later on further developed by Schwaller and Morris (2011), where the SR-GR SR–GR matching is based on the geometric intersection of SR and GR beams. This geometry matching algorithm confines the comparison to those locations where both instruments have actual observations, without interpolation or extrapolation. The method has also been used in a number of studies comparing SR and GR reflectivities (Chandrasekar et al., 2003; Chen and Chandrasekar, 2016; Islam et al., 2012; Kim et al., 2014; Wen et al., 2011). A sensitivity study by Morris and Schwaller (2011) found that method to give more precise estimates of relative calibration bias as compared to grid-based methods.

Due to different viewing geometries, ground radars - in contrast to space-borne precipitation radars - and spaceborne radars are affected by processes that can deteriorate reflectivity measurementsdifferent sources of uncertainty and error. Observational errors with regard to atmospheric properties such as reflectivity are, for example, caused by ground clutter or partial beam blocking. Persistent systematic errors in the observation of reflectivity by ground radars are particularly problematic: the intrinsic assumption of the bias estimation is that the only systematic source of error is radar calibration. It is therefore particularly important to address such systematic observation errors.

In this study, we demonstrate that requirement with the example of partial beam blocking. The analysis is entirely based on algorithms implemented in the open source software library *wradlib* (Heistermann et al., 2013b), including a technique to infer partial beam blocking by simulating the interference of the radar beam with terrain surface based on a digital elevation model. Together, that approach might become a reference for weather services around the world who are struggling to create unbiased radar observations from many years of archived single-polarized radar data, or to consistently monitor the bias of their radar observations. We demonstrate the approach in a case study with five years of data from the single-polarized S-band radar near the city of Subic, Philippines, which had been shown in previous studies to suffer from substantial miscalibration (Abon et al., 2016; Heistermann et al., 2013a). The approach, however, is limited to within ±35 degrees latitude

2 Data

2.1 Space-Borne Precipitation Radar

Precipitation radar data were gathered from TRMM 2A23 and 2A25 version 7 products (NASA, 2017) for overpass events intersecting with the Subic ground radar coverage between 1 June 2012 to 30 September 2014, and GPM 2AKu version 5A products (Iguchi et al., 2010) from 1 June 2014 to 31 December 2016. Ka band observations have not been considered due to higher susceptibility to attenuation, and a limited validity of Rayleigh scattering in a substantial portion of rainfall cases (Baldini et al., 2012). From the collection of overpasses within these dates, only 183 TRMM overpasses and 103 GPM passes were within the radar coverage. The data were downloaded from NASA's Precipitation Processing System (PPS) through the STORM web interface (https://storm.pps.eosdis.nasa.gov/storm/) on 15 February 2018 for TRMM and ±65 degrees latitude 14 June 2018 for GPM. The parameters of TRMM/GPM extracted for the analysis are the same as Warren et al. (2018; their Table 3).

3 Methods

Table 1. Characteristics of the Subic radar and its volume scan strategy. The numbers in parentheses correspond to scans in 2015, where the scanning strategy was different due to hardware issues.

	Subic Radar
Polarization	Single-Pol
Position (lat/lon)	14.82° N 120.36 ° E
Altitude	532 ~~(~~m.a.s.l. ~~)~~
Maximum Range	120 km (150 km)
Azimuth resolution	1 °
Beam width	0.95 °
Gate length	500 m (250 m)
Number of elevation angles	14 (3)
Elevation angles ~~(°)~~	0.5, 1.5, 2.4, 3.4, 4.3, 5.3, 6.2, 7.5, 8.7, 10, 12, 14, 16.7, 19.5 (°) (0.0, 1.0, 2.0)
Volume cycle interval	9 minutes
Data available since	April 2012
Peak power	850 kW
Wavelength	10.7 cm

It is important to note that, at the time of writing, changes in calibration parameters applied in the GPM Version 5 products resulted in an increase of +1.1 dB from the corresponding TRMM version 7 products (NASA, 2017).

2.1 ~~Data~~Ground Radar

2.1.1 ~~Ground Radar~~

The Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) maintains a nationwide network of ten weather radars, eight of which are single-polarization S-band radars and two are dual-polarization C-band radars. Subic radar, which covers the greater Metropolitan Manila area, has the most extensive set of archived data. The radar coverage includes areas that receive some of the highest mean annual rainfall in the country.

The Subic radar sits on top of a hill at 532 m.a.s.l. in the municipality of Bataan, near the border to Zambales (location: 14.82 °N, 120.36 °E) (see Figure 1). To its south stands Mt. Natib (1253 m.a.s.l.) and to its north runs the Zambales Mountain Range (highest peak stands at 2037 m.a.s.l.). To the west is the Redondo Peninsula in the southern part of the Zambales province, where some mountains are also situated. Almost half of the coverage of the Subic radar is water, with Manila Bay to its southeast and the West Philippine Sea to the west. Technical specifications of the radar are summarized in Table 1. Data from April 2012 to December 2016 were obtained from PAGASA. Throughout the five years the scan strategy remained the same, except for 2015 when it was limited to only three elevation angles per volume due to hardware issues. The standard scanning strategy was re-implemented in 2016.

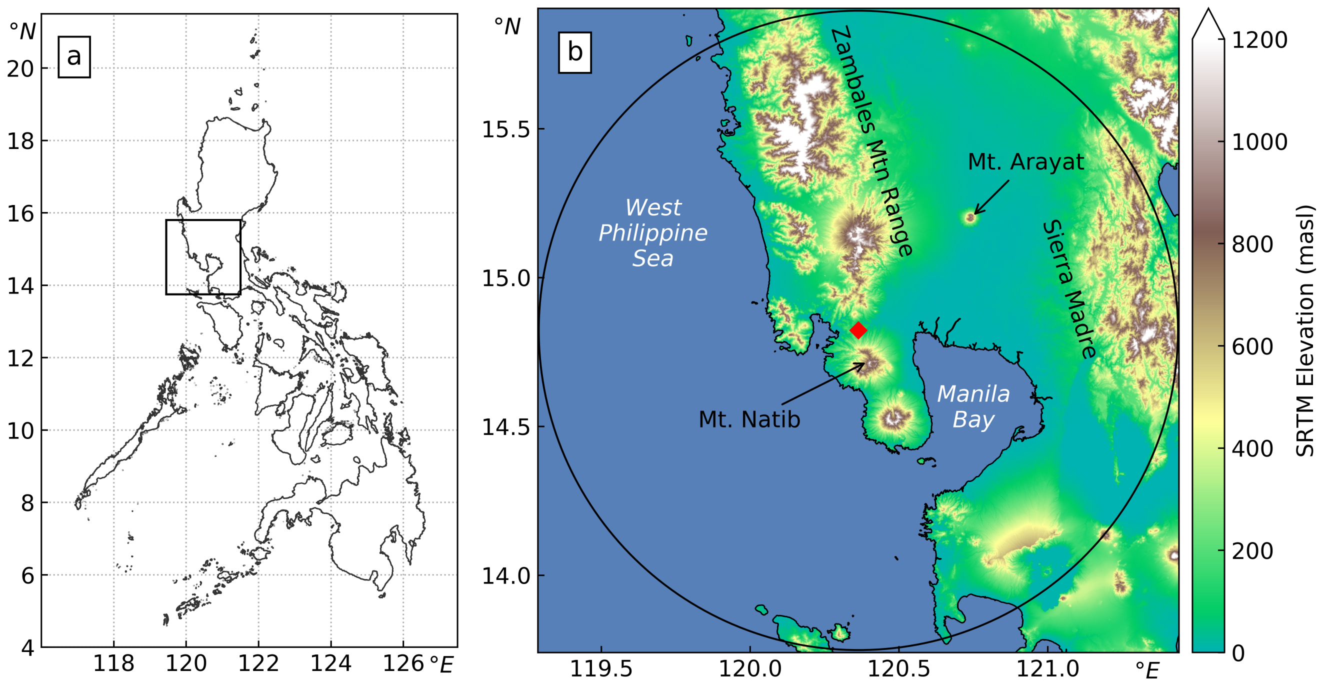

Figure 1. (a) ~~A map~~ Map of the Philippines showing the region of study and (b) the 120km coverage of the Subic radar (location marked with red diamond) with the SRTM Digital Elevation Model of the surrounding area.

2.1.1 ~~Space-Borne Precipitation Radar~~

3 Method

~~Precipitation radar data were gathered from TRMM 2A23 and 2A25 version 7 products for overpass events intersecting with the Subic ground radar coverage between 1 June 2012 to 30 September 2014, and GPM 2AKu version 6 products from 1 June 2014 to 31 December 2016, covering the monsoon season of June to August which overlaps with the typhoon season of June to November. On that basis, we collected 225, 378, and 295 TRMM overpasses for 2012, 2013, and 2014, respectively. For GPM, there were 178, 224, and 211 for 2014, 2015, and 2016. The data were downloaded from NASA's Precipitation Processing System (PPS) through the STORM web interface (https://storm.pps.eosdis.nasa.gov/storm/). The parameters of TRMM/GPM extracted for the analysis are the same as Warren et al. (2017; their Table 3).~~

3.1 Partial ~~Beam Blocking~~beam shielding and quality index based on beam blockage fraction

In an ideal situation, SR and GR should have the same measurements for the same volume of the atmosphere, as they are measuring the same target. However, observational differences may arise due to different view geometries, different operating frequencies, different environmental conditions of each instrument, and different processes along the propagation path of the beam. As pointed out before, we focus on beam blockage as an index of GR data quality.

In regions of complex topography, ground radars are typically affected by the effects of beam blockage, induced by the interaction of the beam with the terrain surface ~~or other solid obstacles along the beam propagation path,~~ resulting into a weakening or even loss of the signal.

To quantify that process within the Subic radar coverage, a beam blockage map is generated following the algorithm proposed by Bech et al. (2003). It assesses the extent of occultation using a digital elevation model (DEM). While Bech et al. (2003) used the GTOPO30 DEM at a resolution of around on kilometer, higher DEM resolutions are expected to increase the accuracy of estimates of beam blockage fraction, as shown by Kucera et al. (2004), in particular the near range of the radar (Cremonini et al., 2016). The DEM used in this study is from the Shuttle Radar Topography Mission (SRTM) data, with 1-arc-second (approximately 30-meter) resolution. The DEM was resampled to the coordinates of the radar bin centroids, using spline interpolation, in order to match the polar resolution of the radar data (500 m in range and 1° in azimuth, extending to a maximum range of 120 km from the radar site; see Figure 1). A beam blockage map is generated for all available elevation angles.

~~The function `wradlib.qual.beam_block_frac()` which is an implementation of this algorithm was used to calculate the~~

The beam blockage fraction was calculated for each bin and each antenna pointing angle. The ~~function `wradlib.qual.cum_beam_b` was used to compute the~~ cumulative beam blockage was then calculated along each ray. A cumulative beam blockage fraction (BBF) of 1.0 corresponds to full occlusion, and a value of 0.0 to perfect visibility.

The quality index based on beam blockage fraction is then computed ~~as follows, following the equation from Zhang et al. (2011)~~ followin Zhang et al. (2011) as:

$$Q_{BBF} = \begin{cases} 1 & BBF \leq 0.1 \\ 1 - \frac{BBF-0.1}{0.4} & 0.1 < BBF \leq 0.5 \\ 0 & BBF > 0.5 \end{cases} \quad (1)$$

~~With this index , all the values with near perfect visibility have the highest quality , while those below 50% visibility have zero quality~~

A slightly different formulation to transform partial beam blockage to a quality index has been presented in other studies (Figueras i Ventura and Tabary, 2013; Fornasiero et al., 2005; Ośródka et al., 2014; Rinollo et al., 2013) where the quality is zero (0) if BBF is above a certain threshold, and then linearly increases to one (1) above that threshold. It should be noted that these approaches are equally valid and can be used in determining the quality index based on beam blockage.

Figure 2 shows the beam blockage map for the two lowest elevation angles of each scanning strategy. Figure 2a and c are for 0.0° and 1.0°, which are the two lowest elevation angles in 2015, while Figure 2b and d are for 0.5° and 1.5°, which are the two lowest elevation angles for the rest of the dataset.

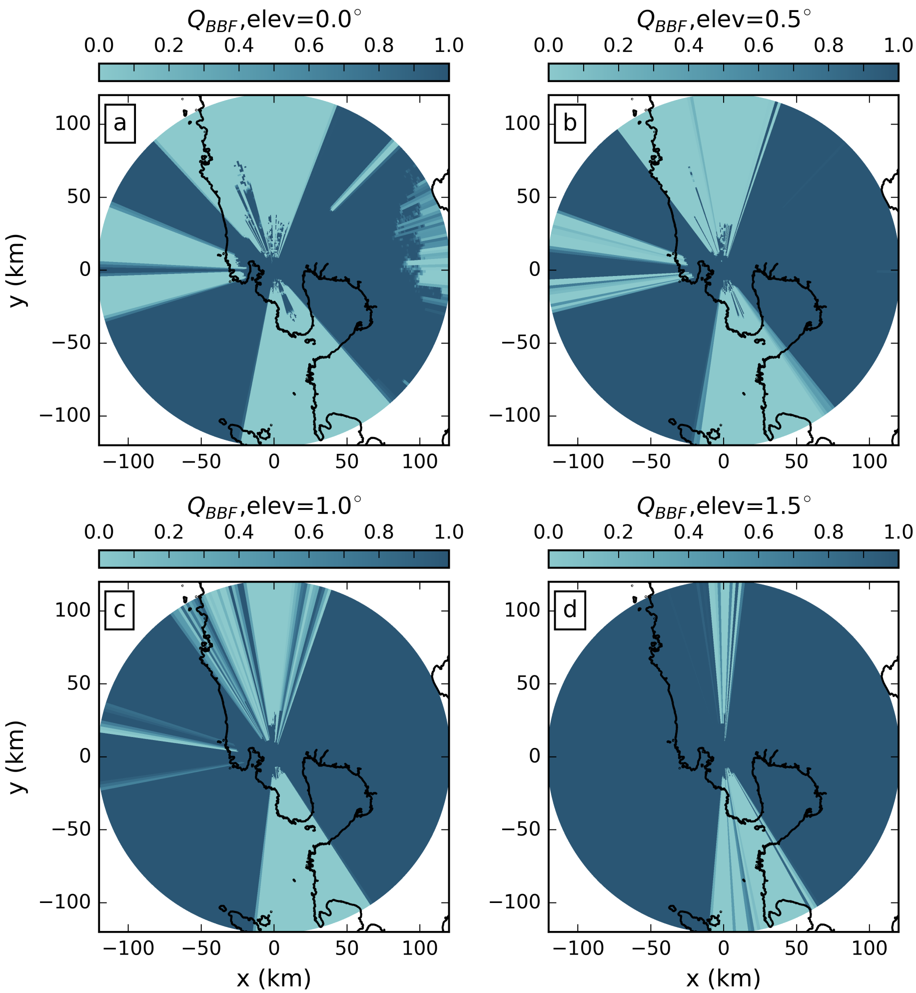

Figure 2. Quality index map of the beam blockage fraction for the Subic radar at (a) 0.0° (b) 0.5° (c) 1.0° and (d) 1.5° elevation angles.

3.2 ~~SR-GR Volume Matching~~

~~For each SR gate, several~~ As expected, the degree of beam blockage decreases with increasing antenna elevation, yielding the most pronounced beam blockage at 0.0°. Each blocked sector can be explained by the topography (see Figure 1), with the Zambales Mountain Range causing blockage in the northern sector, Mt. Natib in the southern sector, and the Redondo peninsula mountains in the western sector. The Sierra Madre mountains also cause some partial beam blocking at the far east, and a narrow partial blocking northeast of the station where Mount Arayat is located. As the elevation angle increases, the beam blockage becomes less pronounced or even disappears. Substantial blockage persists, however, for the higher elevation angles in the northern and southern sectors.

3.2 SR–GR Volume Matching

SR and GR data were matched only for the wet period within each year, which is from June to December. Several meta-data parameters were extracted from the TRMM 2A23 and GPM 2AKu products for each SR gate, such as the corresponding ray's bright band height and width, gate coordinates in three dimensions (longitude and latitude of each ray's Earth intercept and range gate index), time of overpass, precipitation type (*stratiform*, *convective*, or *other*), and rain indicators (*rain certain* or *no-rain*). The parallax-corrected altitude (above mean sea level) and horizontal location (with respect to the GR) of each gate were determined as outlined in the appendix of ~~?~~Warren et al. (2018). From the bright band height/width and the altitude of each SR gate, the bright band membership of each gate was calculated by grouping all rays in an overpass and computing the mean brightband height and width~~, using the `wradlib.qual.get_bb_ratio` function~~. A ratio value of less than zero indicates that the gate is below bright band, greater than one indicates that the gate is above the bright band, and a value between zero and one means that the gate is within the bright band. ~~The~~ Only gates below and above the brightband were considered in the comparison. Warren et al. (2018) found a positive bias in GR–SR reflectivity difference for volume-matched samples within the melting layer, compared to those above and below the melting layer. They speculated that this was due to underestimation of the Ku- to S-band frequency correction for melting snow. In addition, while usually the samples above the brightband are used in GPM validation, there are significantly more samples below the melting layer, especially in a tropical environment such as the Philippines. To ensure that there are sufficient bins with actual rain included in the comparison, overpasses with less than 100 gates flagged as rain certain were discarded.

For each SR overpass, the GR sweep with the scan time closest to the overpass time within a ~~5-min window~~ 10-min window (±5-min from overpass time) was selected. Both the SR and GR data were then geo-referenced into a common azimuthal equidistant projection centered on the location of the ground radar.

In order to minimize systematic differences in comparing the SR and GR reflectivites caused by the different measuring frequencies, the SR reflectivities ~~are~~ were converted from Ku to S Band ~~(using the function `wrl.trafo.KuBandToS`)~~ following the formula~~from ?~~:

$$Z(S) = Z(Ku) + \sum_{i=0}^{4} a_i [Z(Ku)]^i \quad (2)$$

where the a_i are the coefficients for dry snow and dry hail, rain, and in between at varying melting stages (Table 1 of ~~?).~~ Cao et al. (2013)). We used the coefficients for snow in the reflectivity conversion above the brightband, following Warren et al. (2018).

The actual volume matching algorithm closely follows the work of Schwaller and Morris (2011), where SR reflectivity is spatially and temporally matched with GR reflectivity without interpolation. The general concept is highlighted by Figure 3: Each matching sample consists of bins from only *one SR ray* and *one GR sweep*. From the SR ray, those bins were selected that intersect with the vertical extent of a specific GR sweep at the SR ray location~~(as a result of GR beam width and distance from radar)~~. From each GR sweep, those bins were selected that intersect with the horizontal footprint of the SR ray at the corresponding

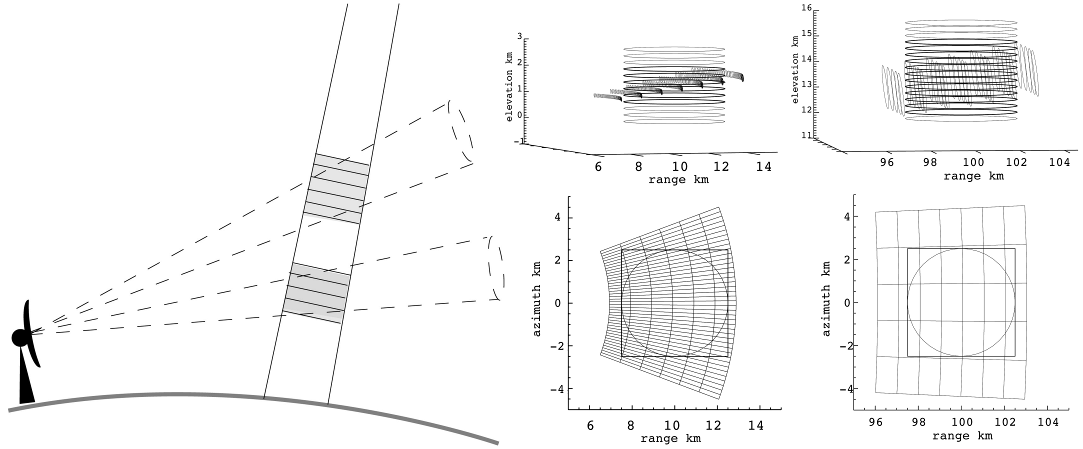

Figure 3. Diagram illustrating the geometric intersection. Left panel shows a single SR beam intersection ground radar beams intersecting GR sweeps of two different elevation angles. The two top right panels illustrate the intersection of SR-GR sample volumes in the near and far ranges and the two bottom right panels show the projection of these intersections along an SR ray. From Schwaller and Morris (2011) ©American Meteorological Society. Used with permission.

Table 2. Filtering criteria for the matching workflow.

Criteria	Condition
Minimum number of pixels in overpass tagged as 'rain'	100
Bright band membership	below and or above
Range limits (min - maxGR range limits (min–max)	15 km - – 115 km
Minimum fraction of bins above minimum SR sensitivity	0.7
Minimum fraction of bins above minimum GR sensitivity	0.7
Maximum time difference between SR and GR	5 min
Minimum PR reflectivity considered	18 dBZ
Minimum GR reflectivity considered 0 dBZ	

altitude. The SR and GR reflectivity of each matched volume was computed as the average reflectivity of the intersecting SR and GR bins, where the averaging is done in linear units (mm^6/m^3).The entire workflow of this volume matching procedure is based on wradlib functions and available at `https://github.com/wradlib/radargpm-beamblockage`.

The nominal minimum sensitivity of both TRMM PR and GPM KuPR is 18 dBZ, so only values above this level were considered in the calculation of average SR reflectivity in the matched volume. In addition, the fraction of SR gates within a matched volume above that threshold was also recorded. On the other hand, all GR bins are included in the calculation of average GR reflectivity, after setting the bins with reflectivities below 0.0 dBZ to 0.0 dBZ, as also suggested by Schwaller and Morris (2011)suggested by Morris and Schwaller (2011). The filtering criteria applied in the workflow are summarized in Table 2.

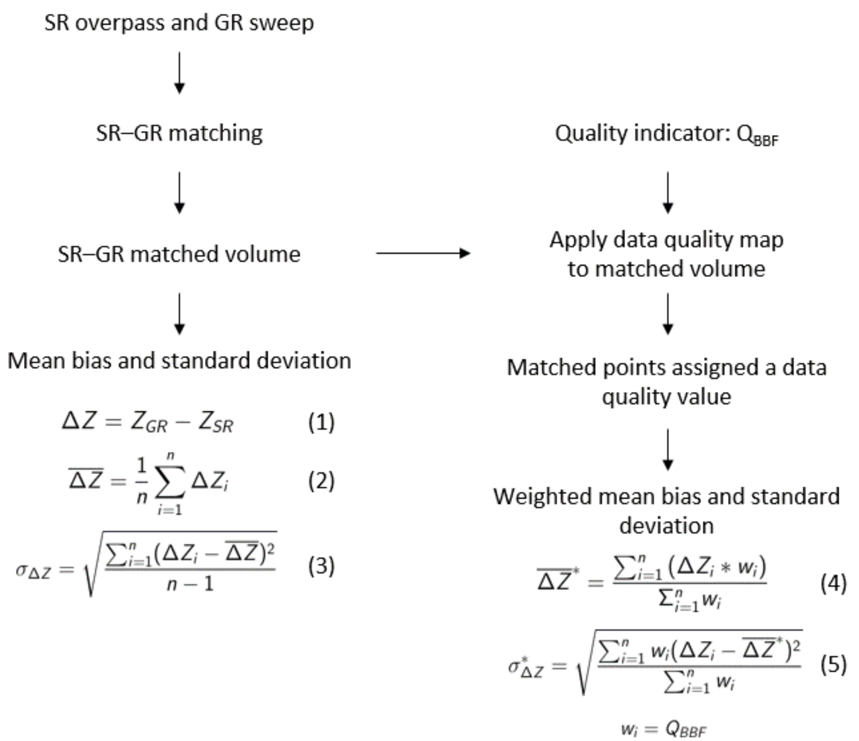

Figure 4. Flowchart describing the processing steps to calculate the mean bias and the weighted mean bias between ground radar data and satellite radar data. The results of each step are shown in Section 4.

3.3 ~~Estimating GR calibration~~ Assessment of the average reflectivity bias

Beam blockage and the corresponding GR quality maps were computed for each GR bin (cf. Section 3.1.) For each matched ~~SR-GR~~ SR–GR volume, the data quality was then based on the minimum quality of the GR bins in that volume.

To analyze the effect of data quality on the estimation of GR calibration bias, we compared two estimation approaches: a simple mean bias that does not take into account beam blockage, and a weighted mean bias that considers the quality value of each sample as weights. The corresponding standard deviation and weighted standard deviation were calculated as well. The overall process is summarized in Figure 4. This way, we provide an overview of the variability of our bias estimates over time.

4 ~~Results and Discussion~~

3.1 Computational details

3.2 ~~Beam Blockage Map~~

In order to promote transparency and reproducibility of this study, we mostly followed the guidelines provided by Irving (2016) which have also been implemented by a number of recent studies (Blumberg et al., 2017; Irving and Simmonds, 2016; Rasp et al., 2018).

~~Figure 2 shows the beam blockage map for the two lowest elevation angles of each scanning strategy. Figure 2a and c are for 0.0° and~~ The entire processing workflow is based on *wradlib* (Heistermann et al., 2013b), an extensively documented open-source software library for processing weather radar data. At the time of writing, we used version 1.0~~°, which are the two lowest elevation angles in 2015, while Figure 2b and d are for 0.5° and 1.5°, which are the two lowest elevation angles for the rest of the dataset.~~ .0 released on 01 April 2018, based on Python 3.6. The main dependencies include Numerical Python (NumPy; Oliphant (2015)), Matplotlib (Hunter, 2007), Scientific Python (SciPy; Jones et al. (2014)), h5py (Collette, 2013), netCDF4 (Rew et al., 1989), and gdal (GDAL Development Team, 2017).

~~Beam blockage fraction for the Subic radar at (a) 0.0° (b) 0.5° (c) 1.0° and (d) 1.5° elevation angles.~~

~~As expected, the degree of beam blockage decreases with increasing antenna elevation, yielding the most pronounced beam blockage at 0.0°. Each blocked sector can be explained by the topography (see Figure 1), with the Zambales Mountain Range causing blockage in the northern sector, Mt. Natib in the southern sector,~~ Reading the TRMM 2A23 and ~~the Redondo peninsula mountains in the western sector. The Sierra Madre mountains also cause some partial beam blocking at the far east, and a narrow partial blocking northeast of the station where Mount Arayat is located.~~ 2A25 version 7 data, GPM 2AKu version 5A data, and the Subic ground radar data in the netCDF format converted through the EDGE software of EEC radars was done through the input/output module of wradlib. The beam blockage modeling is based on the Bech et al. (2003) method implemented as a function in wradlib's data quality module. The volume-matching procedure is built upon the georeferencing and zonal statistics modules, accompanied by Pandas (McKinney, 2010) for organizing and analyzing the resulting database of matched bins. Visualization was carried out with the help of matplotlib (Hunter, 2007) and Py-ART (Helmus and Collis, 2016).

~~As the elevation angle increases, the beam blockage becomes less pronounced or even disappears. Substantial blockage persists, however, for the higher elevation angles in the northern and southern sectors~~An accompanying GitHub repository that hosts the Jupyter notebooks of the workflow and sample data is made available at `https://github.com/wradlib/radargpm-bea`

3.1.1 ~~SR-GR Matching~~

4 Results and Discussion

~~There were~~

4.1 Single event comparison

From the 183 TRMM and 103 GPM overpasses that ~~overlapped~~ intersected with the 120 km Subic radar range~~from 2012 to 2016. However these numbers decreased to~~ , only 74 ~~(for TRMM)~~ TRMM and 40 ~~(for GPM)~~ GPM overpasses were considered valid after applying the selection criteria listed in Table 2. In order to get a better idea about the overall workflow, we first exemplify the results for two specific overpass ~~events - one~~ events—one for TRMM, and one for GPM.

4.1.1 Case 1: 08 November 2013

For the TRMM overpass event on November 8, 2013, the top row of Figure 5 shows SR (a) and GR (b) reflectivity as well as the resulting differences (c) for matching samples at an elevation angle of 0.5°. Each circle in the plots represents a matched volume. A corresponding map of Q_{BBF} is shown in (d) while (e) shows a scatter plot of GR versus SR reflectivities, with points coloured according to their Q_{BBF}. The reflectivity difference map and scatter plot indicate significant variability with absolute differences of up to ~~-10~~ and exceeding 10 dB. Large differences can be observed at the edges of the southern sector affected by beam blockage (cf. also Figure 2). Major parts of that sector did not receive any signal due to total beam blockage, highlighted in Figure 5a with black circles showing the bins where the GR did not obtain valid observations. At the edges, however, partial beam blockage caused substantially lower GR reflectivity values. As expected, large negative differences of Z_{GR} ~~-~~ – Z_{SR} are characterized by low quality.

Consequently, the estimate of the calibration bias substantially depends on the consideration of partial beam blockage (or quality). Ignoring quality (simple mean) yields a bias estimate of -1.9 dB while the quality-weighted average yields a bias estimate of -1.2 dB. Accordingly, the standard deviation is reduced from 3.4 to 2.6 dB, indicating ~~an increased precision of the~~ a more precise bias estimate.

This case demonstrates how partial beam blockage affects the estimation of GR calibration bias. ~~It is important to note that beam blockageonly becomes an issue once the reflectivity observed by the ground radar exceeds the detection limit~~At a low elevation angle, substantial parts of the sweep are affected by *total* beam blockage. The affected bins are either below the detection limit, or they do not exceed the GR threshold specified in Table 2. As a consequence, ~~total beam blockage does not interfere with the bias estimation.~~ these bins will not be considered in the matched samples and will thus not influence the bias estimate, irrespective of using partial beam blockage as a quality filter. At a higher elevation angle, though, the same bins might not be affected by *total* beam blockage, but by *partial* beam blockage, as also becomes obvious from Figure 2. Considering these bins in the matched samples will cause a systematic error in the estimate of calibration bias, unless we use the partial beam blockage fraction as a quality filter by computing a quality-weighted average of reflectivity. As a consequence, the effect of quality-weighted averaging (with partial beam blockage fraction as a quality variable) can be most pronounced at "intermediate" elevation angles, depending on the specific topography and its location with respect to the ground.

The effect becomes ~~even more~~ obvious for the next elevation angle. Figure 6 is equivalent to Figure 5, but for an elevation angle of 1.5°: As the sector of total beam blockage shrinks at that elevation, the impact of partial beam blockage on the estimation of GR calibration bias increases. ~~That might be considered counterintuitive, as one might expect the effect of beam~~

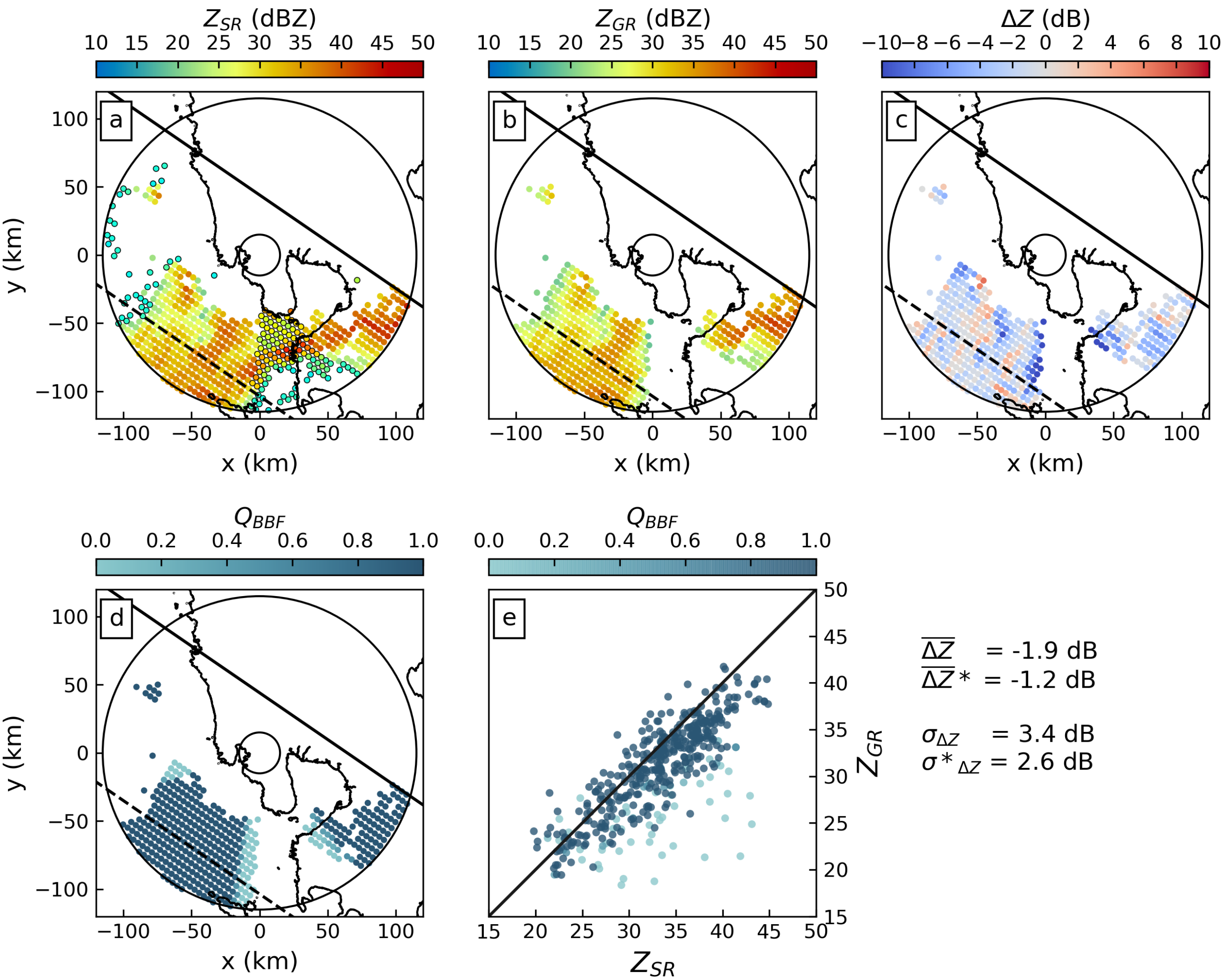

Figure 5. ~~PPI~~ GR-centered maps of volume-matched samples from 8 November 2013 at 0.5° elevation angle of (a) SR reflectivity, (b) GR Reflectivity, (c) difference between GR and SR reflectivities, and (d) Q_{BBF}. (e) Scatter plot of Z_{GR} vs ~~Z_{PR}~~ Z_{SR} where each point is colored based on the data quality (Q_{BBF}). The solid line in (a)~~-~~–(d) is the edge of the SR swath, the other edge lies outside the figure. The ~~dotted~~ dashed line denotes the central axis of the swath. The solid concentric circles demarcate the 15 km and 115 km ranges from the radar. In (a) observations that are present in the SR data but not detected by the GR are encircled in black. The mean bright band is at a height of 4685 meters.

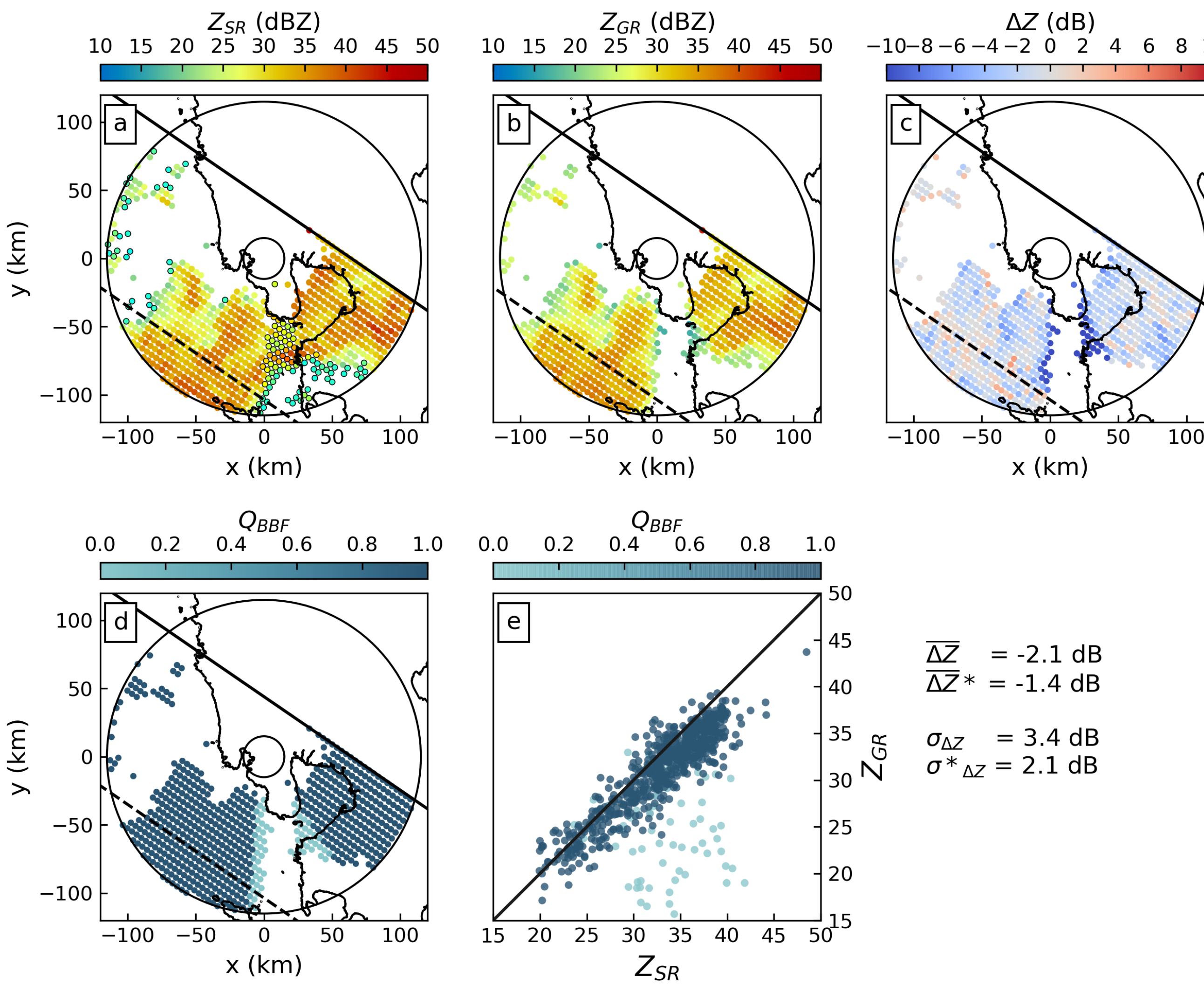

Figure 6. Same as in Figure 5 but for 1.5° elevation angle

~~blockage to disappear at higher elevations.~~ For an antenna elevation of 1.5°, ~~however, ignoring beam blockage~~ some bins in areas of partial beam blockage have very large negative biases (over 20 dB). Ignoring beam blockage for this elevation angle yields a bias estimate of -2.1 dB (simple mean), while the quality-weighted average yields a bias of -1.4 dB. At the same time, considering quality substantially reduces the standard deviation from 3.4 dB to 2.1 dB.

4.1.2 Case 2: 01 October 2015

The second case confirms the findings in the previous section for a GPM overpass on October 1, 2015. That overpass captured an event in the northern and eastern part of the radar coverage where partial beam blockage is dominant, as well as a small part of the southern sector with partial and total beam blockage. Figure 7 shows the results of that overpass in analogy to the previous figures, for an antenna elevation of 0.0 degree. The figure shows a dramatic impact of partial beam blockage, with a dominant contribution from the northern part, but also clear effects from the eastern and southern sectors. The scatter plot of Z_{GR} over Z_{SR} (e) demonstrates how the consideration of partial beam blockage increases the consistency between GR and SR observations and allows for a more reliable estimation of the GR calibration bias: Ignoring partial beam blockage (simple mean) yields a bias of -2.7 dB, while the quality-weighted average bias is -1.1 dB. Taking into account quality decreases the standard deviation from 3.8 dB to 2.7 dB.

4.2 ~~Calibration bias over time~~Overall June-November comparison during the 5-year observation period

Finally, we applied both the simple and the quality-weighted mean bias estimation to each of the TRMM and GPM overpasses from 2012 to 2016 that met the criteria specified in Section ~~??~~ 3.2, Table 2. As pointed out in Section ~~??~~3.2, the matching procedure itself is carried out per GR sweep, i.e. separately for each antenna elevation angle.

As a result, we obtain a time series of bias estimates for GR calibration, as shown in Figure 8. In this figure, the calibration bias for each overpass is computed from the full GR volume, i.e. including matched samples from all available antenna elevations. In the upper panel (a), each marker represents the quality-weighted mean bias for a specific SR overpass (circles for GPM, triangles for TRMM). The center panel (b) highlights the differences between the quality-weighted and the simple mean approach, ~~.e. it quantifies~~ by quantifying the effect of taking into account GR data quality (~~.e.~~ in this case, partial beam blockage). The bottom panel (c) shows the differences between the quality-weighted standard deviation and the simple standard deviation of differences, illustrating how taking into account GR quality affects the precision of the bias estimates.

The time series provide several important insights:

(1) ~~**Short term variability**~~**Effect of quality weighting on bias estimation**: ~~There is a strong variability of the estimated calibration bias between overpasses (Figure 8a) . That variability appears to have a strong random component, and it is clearly not a desirable property: typically, we would not expect changes in calibration bias to occur at the observed frequency and amplitude. Yet, it is beyond the scope of this study to disentangle the sources of this variability. We have to assume that the fluctuations are a cumulative result of various effects that might include~~ Figure 8b and c together illustrate the benefit of taking into account GR data quality (i.e. beam blockage) when we estimate GR calibration bias. It does not come as a surprise that the difference between ΔZ^* and ΔZ is mostly positive because the areas suffering from partial beam blockage register weaker

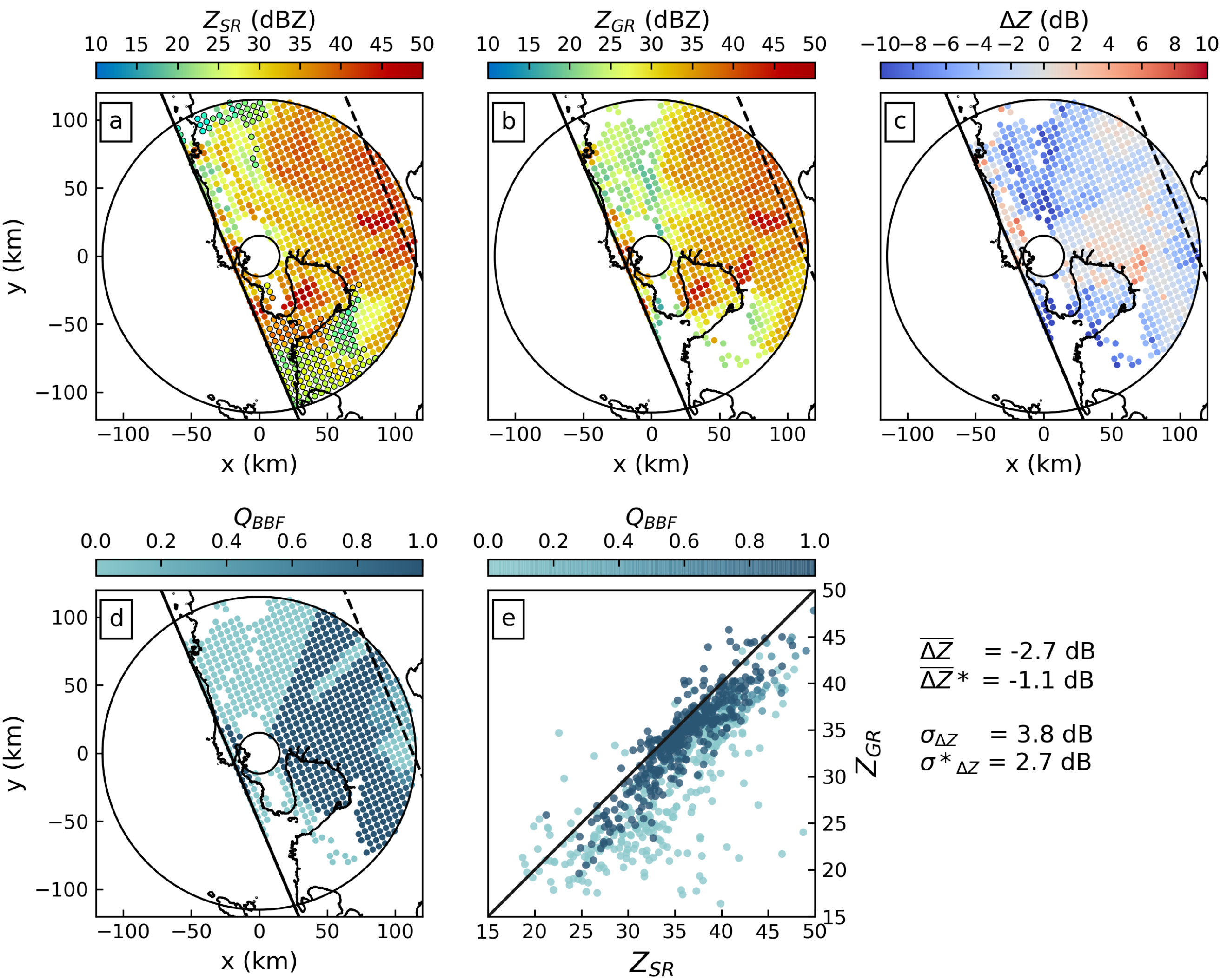

Figure 7. As in Figure 5 but for the overpass 01 October ~~2015~~2015. The mean bright band level is found at 4719 meters for this case

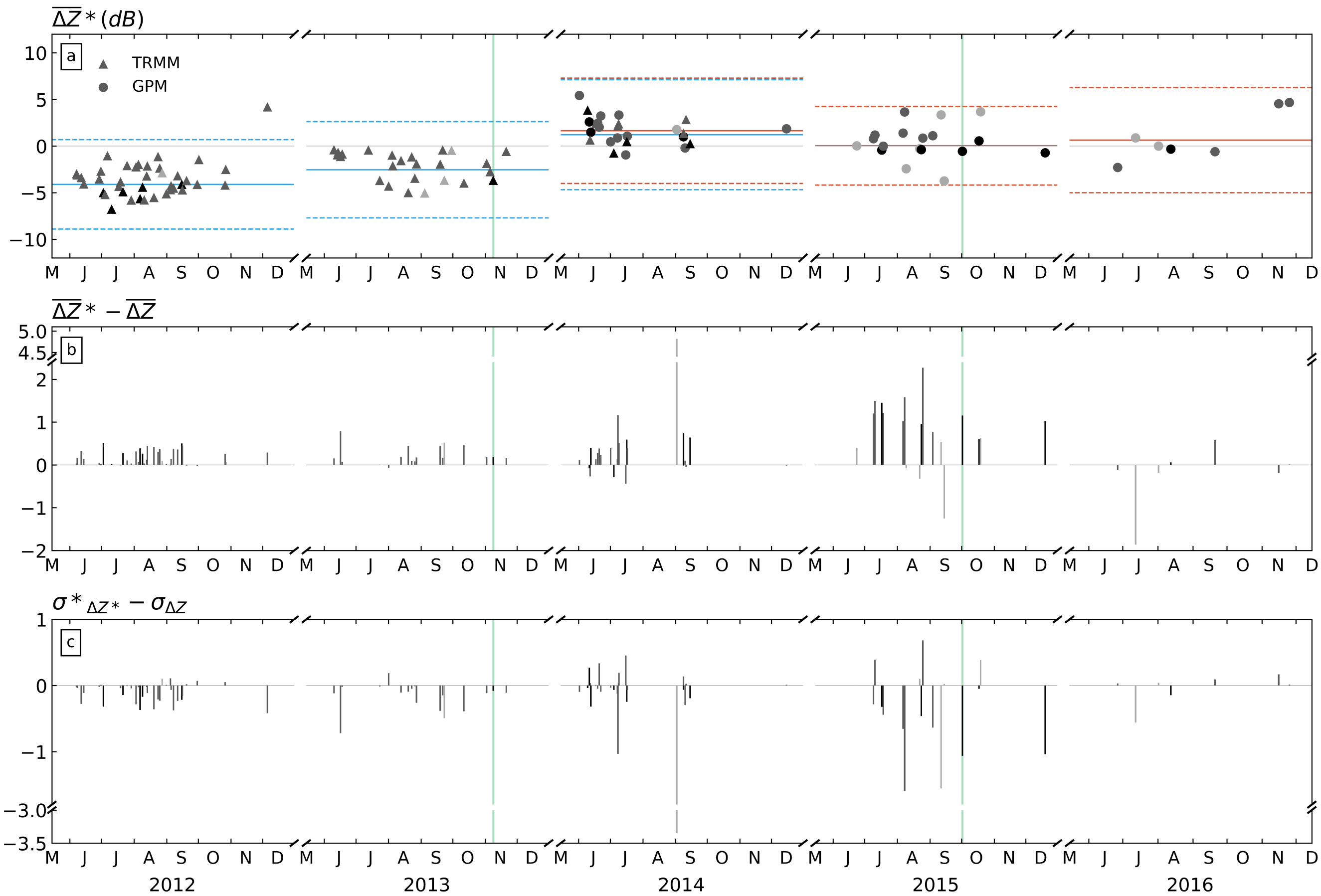

Figure 8. (a) Time series of the weighted mean bias (ΔZ^*) from 2012 to 2016. Analysis covers only the wet season from June to December. Triangle markers represent ~~comparison with~~ TRMM overpasses while circle markers are ~~comparisons with~~ GPM overpasses. Symbols are colored according to the number of volume-matched ~~sample pairs~~ samples on a logarithmic scale: light gray = ~~10-99~~10–99, medium gray = ~~100-999~~100–999, and black = 1000+. ~~Solid~~ Blue and orange solid (dashed) horizontal lines represent the weighted average (standard deviation) of ~~all~~ *all* individual matched ~~points~~ samples within the year for TRMM and GPM, respectively. (b) The difference between the weighted mean biases (ΔZ^*) and the simple mean biases (ΔZ). (c) The standard deviation of the weighted mean bias ~~values~~ minus the standard deviation of the simple mean bias values. The green vertical lines indicate the date of the two case studies.

signals (i.e. lower reflectivity) than expected, producing a lower mean bias. Giving the associated volume-matched samples low weights in the calculation of the mean bias brings the quality-weighted bias up. In the same vein, the beam-blocked bins introduce scatter, and assigning them low weights decreases the standard deviation. Figure 8c shows, as a consequence, that the quality weighted bias estimates are consistently more precise: in the vast majority of overpasses, the quality weighted standard deviation is substantially smaller than the simple standard deviation. That result is also consistent with the case study result shown above. It should be noted, though, that for some overpasses, the quality weighting procedure (which is in effect a filtering) can cause an increase in the bias estimate and/or the standard deviation of that estimate. That effect occurs for overpasses with particularly low numbers of matched samples, and, presumably, with rainfall in regions in which our estimated beam blockage fraction is subject to higher errors (caused by e.g. hardware instability, fluctuations in the refractive index of the atmosphere, dynamic changes in the atmosphere in the time interval between SR overpass and GR sweep, non-meteorological echoes, residual errors in the geometric intersection of the volume samples, and other random error components, as noted in other studies (Anagnostou et al., 2001; Kim et al., 2014; Schwaller and Morris, 2011; Wang and Wolff, 2009; **?**). inadequateness of the assumed Gaussian antenna pattern, variability of atmospheric refractivity, or errors related to the DEM, its resolution and its interpolation to ground radar bins). In total, however, the effect of decreasing standard deviation vastly dominates.

(2) **GPM and TRMM radars are consistent**: In 2014, both TRMM and GPM overpasses are available. That period of overlap shows that the GR calibration bias estimates that are based on both TRMM and GPM observations can be considered as homogeneous. Using TRMM data, the average calibration bias for all 2014 overpasses amounts to 1.6 ± 1.3 dB, while using the GPM overpasses yields a bias of 1.8 ± 1.5 dB. The difference between TRMM version 7 and GPM version 5 reflectivities mentioned in Section 2.1 falls within the uncertainties in the annual estimated mean bias, which makes us confident that the substantial year-to-year changes of our bias estimates are based on changes in GR calibration.

(3) **Change of bias over time**: Despite the variability of bias estimates between the individual overpass events, the time series still provides us with a clear signal: The bias estimates appear to fluctuate around an average value that appears to be quite persistent over the duration of the corresponding wet seasons of the different years, so over intervals of several months. Considering the average calibration bias over the different wet seasons (horizontal lines in Figure 8a), we can clearly observe changes in calibration bias over time. The bias was most pronounced in 2012 and 2013, with average bias estimates around -4.1 dB for 2012 and -2.5 dB for 2013. For 2014, the absolute calibration bias was much smaller, at a level of 1.4 dB while for 2015 and 2016, the situation improved further, with an average bias of 0.03 0.0 dB in 2015 and 0.6 dB in 2016. It is important to note that these values were computed as the average bias and its standard deviation across all matched volumes and not as the average of biases across overpasses, as the overpasses have different sample sizesbias estimates across overpasses. Accordingly, the standard deviation (as indicated by the dashed lines) is quite high since it includes all the scatter from the individual overpasses. We have to assume that a fundamental issue with regard to calibration maintenance was addressed between 2013 and 2014 , although we could not confirm that with the radar operator.

in the context of hardware changes (3) **Effect of quality weighting on bias estimation**: Figure 8b and c together illustrate the benefit of taking into account GR data quality (i.e. beam blockage)when we estimate GR calibration bias. It is important to understand that only the center and the bottom panel together prove that benefit: It does not come as a surprise that the quality-weighted mean bias is consistently higher than the simple mean bias replacement of magnetron). Unfortunately, we were not able to retrieve detailed information on maintenance operations that might explain the changes in bias of the radar throughout the years.

Short term variability of bias estimates between overpasses: There is a strong variability of the estimated calibration bias between overpasses (Figure 8b) since we give less weight to those GR bins affected by partial beam blockage. But only Figure 8c shows that the quality weighted bias estimates are also consistently more precise: in the vast majority of overpasses the quality weighted standard deviation is substantially smaller than the simple standard deviation. That result is also consistent with the case study result shown above. However, when the matched sample size is already small, further filtering of samples increases the standard deviation.

(4) **GPM and TRMM radars are consistent**: In 2014, both TRMM and GPM overpasses are available. That period of overlap shows that the GR calibration bias estimates that are based on both TRMM and GPM observations can be considered as homogeneous. Using TRMM data, a) and spatially within each overpass (Figures 5 to 7). That variability is clearly not a desirable property, as we would not expect changes in calibration bias to occur at the observed frequency, amplitude, and apparent randomness. As a consequence, we have to assume that the variability is a cumulative result of various and dynamic sources of uncertainty along the entire process of observation, product generation, matching, and filtering. That assumption is well in line with many other studies (such as Anagnostou et al. (2001); Durden et al. (1998); Joss et al. (2006); Kim et al. (2014); Meneghin to name only a few) which discuss e.g. fundamental issues with the backscattering model for different wavelengths and sampling volumes; the uncertainty of beam propagation subject to fluctuations in atmospheric refractivity; residual errors in the geometric intersection of the volume samples; uncertainties in SR reflectivity subject to the effects of attenuation correction at Ku band, non-uniform beam filling and undesirable synergies between the two; rapid dynamics in backscattering target during the time interval between SR overpass and GR sweep; effects non-meteorological echoes for both SR and GR; and, presumably, also short-term hardware instabilities. Considerung these uncertainties, together with the fact that the quality weighting in our case study explicitely accounts for beam bloackage only, the average calibration bias for all 2014 overpasses amounts to 1.6 ± 1.3 dB, while using the GPM overpasses yields a bias of 1.8 ± 1.5 dB. That makes us confident that the substantial year-to-year changes of our bias estimates are based on changes in GR calibration, not based on differences between TRMM and GPM sensorsshort term variability becomes plausible. Yet, it is beyond the scope of this study to disentangle the sources of this variability.

5 Conclusions

In 2011, Schwaller and Morris presented a new technique to match spaceborne radar (SR) and ground-based radar (GR) reflectivity observations, with the aim to determine the GR calibration bias. Our study extends that technique by a coherent approach to take an approach that takes into account the quality of the ground radar observations: . Each GR bin was assigned a quality index between 0 and 1, which was again used to assign a quality value to each matched volume of SR and GR observations. For any sample of matched volumes (e.g. all matched volumes of one overpass, or a combination of multiple overpasses), the calibration bias can then be computed as a quality-weighted average of the differences between GR and SR reflectivity in all samples.

We exemplified that approach by ~~inferring~~ applying a GR data quality index based on the beam blockage fraction, and we demonstrated the added value for both TRMM and GPM overpasses over the ~~120~~ 115 km range of the Subic S-band radar in the Philippines for a five year period.

Although the variability of the calibration bias estimates between overpasses is high, we ~~could show~~ showed that taking into account partial beam blockage leads to more consistent and more precise estimates of GR calibration bias. Analyzing five years of archived data from the Subic S-band radar (2012-2016), we ~~could also demonstrate~~ also demonstrated that the calibration standard of the Subic radar substantially improved over the years, from bias levels around -4.1 dB in 2012 to bias levels ~~around~~ of around 1.4 dB in 2014 and settling down to a bias of 0.6 dB ~~.~~

in 2016. Of course, more recent comparisons with GPM are needed to verify that this level of accuracy has been maintained. Case studies for specific overpass events also showed that the necessity to account for partial beam blockage might even increase for higher antenna elevations. That applies when sectors with total beam blockage (in which no valid matched volumes are retrieved at all) turn into sectors with partial beam blockage at higher elevation angles.

Considering the scatter between SR and GR reflectivity in the matched volumes of one overpass (see case studies), as well as the variability of bias estimates between satellite overpasses (see time series), it is obvious that we do not yet account for various sources of ~~systematic and random sources of observational errors~~uncertainties. Also the simulation of beam blockage itself might still be prone to ~~various errors(e. g. from geometric inaccuracies, variability of atmospheric refractivity, interpolation effects, errors in the DEM, or limited DEM resolution).~~ errors. Nevertheless, the idea of the quality-weighted estimation of calibration bias presents a consistent framework that allows for the integration of any quality variables that are considered important in a specific environment or setting. For example, if we consider C-band ~~radars, e.g. considering~~ instead of S-band radars, path-integrated attenuation ~~would be vital~~needs to be taken into account for the ground radar, and wet radome attenuation probably as well (Austin, 1987; Merceret, 2000; Villarini and Krajewski, 2010). The framework could also be extended by explicitly assigning a quality index to SR observations, too. In the context of this study, that was implicitly implemented by filtering the SR data e.g. based on bright band membership. An alternative approach to filtering could be weighting the samples based on their proximity to the bright band, the level of path-integrated attenuation (as e.g. indicated by the GPM 2AKu variables *pathAtten* and the associated reliability flag (*reliabFlag*)) or the prominence of non-uniform beam filling (which could e.g. be estimated based on the variability of GR reflectivity within the SR footprint, see e.g. (Han et al., 2018)).

In addition, with the significant effort devoted to weather radar data quality characterization in Europe (Michelson et al., 2005), and the number of approaches in determining an overall quality index based on different quality factors (Einfalt et al., 2010), it is straightforward to extend the approach beyond beam blockage fraction.

Despite the fact that there is still ample room for improvement, our tool that combines ~~SR-GR~~ SR–GR volume matching and quality-weighted bias estimation is readily available for application or further scrutiny. In fact, our analysis is the first of its kind that is entirely based on open source software, and thus fully transparent, reproducible and adjustable (see also ~~?~~Heistermann et al. (2014)). Therefore this study, for the first time, demonstrates the utilization of wradlib functions that have just recently been implemented to support the volume matching procedure and the simulation of partial beam blockage. We also make ~~available~~ the complete workflow ~~that put together the different steps of the processing chain, together~~ available together with the underlying ground and space-borne radar data. Both code and results can be accessed at the following repository `https://github.com/wradlib/radargpm-beamblockage` upon the publication of this manuscript.

Through these open-source resources, our methodology provides both research institutions and weather services with a valuable tool that can be applied to monitor ~~the~~ radar calibration, ~~and - perhaps more importantly - to~~ and—perhaps more importantly—to quantify the calibration bias for long time series of archived radar observations, basically beginning with the availability of TRMM radar observations in December 1997.

Code and data availability. Code and sample data can be accessed at `https://github.com/wradlib/radargpm-beamblockage`

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The radar data for this analysis were provided by the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA, http://pagasa.dost.gov.ph). The study was also funded by the German government through the German Academic Exchange Service (https://www.daad.de/en/).

References

Abon, C. C., Kneis, D., Crisologo, I., Bronstert, A., David, C. P. C., and Heistermann, M.: Evaluating the potential of radar-based rainfall estimates for streamflow and flood simulations in the Philippines, Geomatics, Natural Hazards and Risk, 7, 1390–1405, https://doi.org/10.1080/19475705.2015.1058862, http://www.tandfonline.com/doi/full/10.1080/19475705.2015.1058862, 2016.

Amitai, E., Llort, X., and Sempere-Torres, D.: Comparison of TRMM radar rainfall estimates with NOAA next-generation QPE, Journal of the Meteorological Society of Japan, 87, 109–118, https://doi.org/10.2151/jmsj.87A.109, https://www.researchgate.net/publication/250140241_Comparison_of_TRMM_radar_rainfall_estimates_with_NOAA_next-generation_QPE, 2009.

Anagnostou, E. N., Morales, C. A., and Dinku, T.: The use of TRMM precipitation radar observations in determining ground radar calibration biases, Journal of Atmospheric and Oceanic Technology, 18, 616–628, http://journals.ametsoc.org/doi/abs/10.1175/1520-0426(2001)018%3C0616%3ATUOTPR%3E2.0.CO%3B2, 2001.

Austin, P. M.: Relation between Measured Radar Reflectivity and Surface Rainfall, Monthly Weather Review, 115, 1053–1070, https://doi.org/10.1175/1520-0493(1987)115<1053:RBMRRA>2.0.CO;2, https://journals.ametsoc.org/doi/10.1175/1520-0493%281987%29115%3C1053%3ARBMRRA%3E2.0.CO%3B2, 1987.

Baldini, L., Chandrasekar, V., and Moisseev, D.: Microwave radar signatures of precipitation from S band to K_a band: application to GPM mission, European Journal of Remote Sensing, 45, 75–88, https://doi.org/10.5721/EuJRS20124508, https://www.tandfonline.com/doi/full/10.5721/EuJRS20124508, 2012.

Bech, J., Codina, B., Lorente, J., and Bebbington, D.: The sensitivity of single polarization weather radar beam blockage correction to variability in the vertical refractivity gradient, Journal of Atmospheric and Oceanic Technology, 20, 845–855, http://journals.ametsoc.org/doi/abs/10.1175/1520-0426(2003)020%3C0845:TSOSPW%3E2.0.CO;2, 2003.

Blumberg, W. G., Halbert, K. T., Supinie, T. A., Marsh, P. T., Thompson, R. L., and Hart, J. A.: SHARPpy: An Open-Source Sounding Analysis Toolkit for the Atmospheric Sciences, Bulletin of the American Meteorological Society, 98, 1625–1636, https://doi.org/10.1175/BAMS-D-15-00309.1, http://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00309.1, 2017.

Bolen, S. M. and Chandrasekar, V.: Methodology for aligning and comparing spaceborne radar and ground-based radar observations, Journal of Atmospheric and Oceanic Technology, 20, 647–659, http://journals.ametsoc.org/doi/abs/10.1175/1520-0426(2003)20%3C647%3AMFAACS%3E2.0.CO%3B2, 2003.

Bringi, V. N., Huang, G.-J., Munchak, S. J., Kummerow, C. D., Marks, D. A., and Wolff, D. B.: Comparison of Drop Size Distribution Parameter (D_0) and Rain Rate from S-Band Dual-Polarized Ground Radar, TRMM Precipitation Radar (PR), and Combined PR–TMI: Two Events from Kwajalein Atoll, Journal of Atmospheric and Oceanic Technology, 29, 1603–1616, https://doi.org/10.1175/JTECH-D-11-00153.1, http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-11-00153.1, 2012.

Cao, Q., Hong, Y., Qi, Y., Wen, Y., Zhang, J., Gourley, J. J., and Liao, L.: Empirical conversion of the vertical profile of reflectivity from Ku-band to S-band frequency, Journal of Geophysical Research: Atmospheres, 118, 1814–1825, https://doi.org/10.1002/jgrd.50138, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50138, 2013.

Chandrasekar, V., Bolen, S., and Gorgucci, E.: Microphysical cross validation of spaceborne radar and ground polarimetric radar, IEEE Transactions on Geoscience and Remote Sensing, 41, 2153–2165, https://doi.org/10.1109/TGRS.2003.817186, http://ieeexplore.ieee.org/document/1237374/, 2003.

Chen, H. and Chandrasekar, V.: Validation of NASA's Global Precipitation Measurement mission with a high-resolution ground radar network, in: URSI Asia-Pacific Radio Science Conference (URSI AP-RASC), pp. 836–839, IEEE, http://ieeexplore.ieee.org/abstract/document/7601343/, 2016.

Collette, A.: Python and HDF5, O'Reilly, 2013.

Cremonini, R., Moisseev, D., and Chandrasekar, V.: Airborne laser scan data: a valuable tool with which to infer weather radar partial beam blockage in urban environments, Atmospheric Measurement Techniques, 9, 5063–5075, https://doi.org/10.5194/amt-9-5063-2016, https://www.atmos-meas-tech.net/9/5063/2016/, 2016.

Durden, S. L., Haddad, Z. S., Kitiyakara, A., and Li, F. K.: Effects of Nonuniform Beam Filling on Rainfall Retrieval for the TRMM Precipitation Radar, Journal of Atmospheric and Oceanic Technology, 15, 635–646, https://doi.org/10.1175/1520-0426(1998)015<0635:EONBFO>2.0.CO;2, https://journals.ametsoc.org/doi/10.1175/1520-0426%281998%29015%3C0635%3AEONBFO%3E2.0.CO%3B2, 1998.

Einfalt, T., Szturc, J., and Ośródka, K.: The quality index for radar precipitation data: a tower of Babel?, Atmospheric Science Letters, 11, 139–144, https://doi.org/10.1002/asl.271, http://onlinelibrary.wiley.com/doi/10.1002/asl.271/abstract, 2010.

Figueras i Ventura, J. and Tabary, P.: The New French Operational Polarimetric Radar Rainfall Rate Product, Journal of Applied Meteorology and Climatology, 52, 1817–1835, https://doi.org/10.1175/JAMC-D-12-0179.1, http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-12-0179.1, 2013.

Fornasiero, A., Alberoni, P. P., Amorati, R., Ferraris, L., and Taramasso, A. C.: Effects of propagation conditions on radar beam-ground interaction: impact on data quality, Advances in Geosciences, 2, 201–208, https://hal.archives-ouvertes.fr/hal-00297387, 2005.

Furukawa, K., Nio, T., Konishi, T., Oki, R., Masaki, T., Kubota, T., Iguchi, T., and Hanado, H.: Current status of the dual-frequency precipitation radar on the global precipitation measurement core spacecraft, in: Sensors, Systems, and Next-Generation Satellites XIX, vol. 9639, p. 96390G, International Society for Optics and Photonics, https://doi.org/10.1117/12.2193868, https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9639/96390G/Current-status-of-the-dual-frequency-precipitation-radar-on-the/10.1117/12.2193868.short, 2015.

Gabella, M., Joss, J., Perona, G., and Michaelides, S.: Range adjustment for ground-based radar, derived with the spaceborne TRMM precipitation radar, IEEE Transactions on Geoscience and Remote Sensing, 44, 126–133, https://doi.org/10.1109/TGRS.2005.858436, http://ieeexplore.ieee.org/document/1564402/, 2006.

Gabella, M., Morin, E., and Notarpietro, R.: Using TRMM spaceborne radar as a reference for compensating ground-based radar range degradation: Methodology verification based on rain gauges in Israel, Journal of Geophysical Research, 116, https://doi.org/10.1029/2010JD014496, http://doi.wiley.com/10.1029/2010JD014496, 2011.

Gabella, M., Speirs, P., Hamann, U., Germann, U., and Berne, A.: Measurement of Precipitation in the Alps Using Dual-Polarization C-Band Ground-Based Radars, the GPM Spaceborne Ku-Band Radar, and Rain Gauges, Remote Sensing, 9, 1147, https://doi.org/10.3390/rs9111147, http://www.mdpi.com/2072-4292/9/11/1147, 2017.

GDAL Development Team: GDAL - Geospatial Data Abstraction Library, Version 2.2.3, http://www.gdal.org/, 2017.

Han, J., Chu, Z., Wang, Z., Xu, D., Li, N., Kou, L., Xu, F., and Zhu, Y.: The establishment of optimal ground-based radar datasets by comparison and correlation analyses with space-borne radar data, Meteorological Applications, 25, 161–170, https://doi.org/10.1002/met.1682, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1682, 2018.

Heistermann, M., Crisologo, I., Abon, C. C., Racoma, B. A., Jacobi, S., Servando, N. T., David, C. P. C., and Bronstert, A.: Brief communication "Using the new Philippine radar network to reconstruct the Habagat of August 2012 monsoon event around Metropolitan

Manila", Nat. Hazards Earth Syst. Sci., 13, 653–657, https://doi.org/10.5194/nhess-13-653-2013, http://www.nat-hazards-earth-syst-sci.net/13/653/2013/, 2013a.

Heistermann, M., Jacobi, S., and Pfaff, T.: Technical Note: An open source library for processing weather radar data (<i>wradlib</i>), Hydrology and Earth System Sciences, 17, 863–871, https://doi.org/10.5194/hess-17-863-2013, http://www.hydrol-earth-syst-sci.net/17/863/2013/, 2013b.

Heistermann, M., Collis, S., Dixon, M. J., Giangrande, S., Helmus, J. J., Kelley, B., Koistinen, J., Michelson, D. B., Peura, M., Pfaff, T., and Wolff, D. B.: The Emergence of Open-Source Software for the Weather Radar Community, Bulletin of the American Meteorological Society, 96, 117–128, https://doi.org/10.1175/BAMS-D-13-00240.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00240.1, 2014.

Helmus, J. and Collis, S.: The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language, Journal of Open Research Software, 4, https://doi.org/10.5334/jors.119, http://openresearchsoftware.metajnl.com/articles/10.5334/jors.119/, 2016.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, Bulletin of the American Meteorological Society, 95, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1, https://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00164.1, 2013.

Houze Jr, R. A., Brodzik, S., Schumacher, C., Yuter, S. E., and Williams, C. R.: Uncertainties in oceanic radar rain maps at Kwajalein and implications for satellite validation, Journal of Applied Meteorology, 43, 1114–1132, http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(2004)043%3C1114:UIORRM%3E2.0.CO%3B2, 2004.

Hunter, J. D.: Matplotlib: A 2D Graphics Environment, Computing in Science Engineering, 9, 90–95, https://doi.org/10.1109/MCSE.2007.55, 2007.

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., and Kubota, T.: GPM/DPR level-2 algorithm theoretical basis document, NASA Goddard Space Flight Center, Greenbelt, MD, USA, Tech. Rep, 2010.

Irving, D.: A Minimum Standard for Publishing Computational Results in the Weather and Climate Sciences, Bulletin of the American Meteorological Society, 97, 1149–1158, https://doi.org/10.1175/BAMS-D-15-00010.1, http://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00010.1, 2016.

Irving, D. and Simmonds, I.: A New Method for Identifying the Pacific–South American Pattern and Its Influence on Regional Climate Variability, Journal of Climate, 29, 6109–6125, https://doi.org/10.1175/JCLI-D-15-0843.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0843.1, 2016.

Islam, T., Rico-Ramirez, M. A., Han, D., Srivastava, P. K., and Ishak, A. M.: Performance evaluation of the TRMM precipitation estimation using ground-based radars from the GPM validation network, Journal of Atmospheric and Solar-Terrestrial Physics, 77, 194–208, https://doi.org/10.1016/j.jastp.2012.01.001, http://www.sciencedirect.com/science/article/pii/S1364682612000028, 2012.

Jones, E., Oliphant, T. E., and Peterson, P.: SciPy: open source scientific tools for Python, 2014.

Joss, J., Gabella, M., Michaelides, S. C. h. r., and Perona, G.: Variation of weather radar sensitivity at ground level and from space: case studies and possible causes, Meteorologische Zeitschrift, pp. 485–496, https://doi.org/10.1127/0941-2948/2006/0150, http://www.schweizerbart.de/papers/metz/detail/15/55160/Variation_of_weather_radar_sensitivity_at_ground_l?af=crossref, 2006.

Kawanishi, T., Kuroiwa, H., Kojima, M., Oikawa, K., Kozu, T., Kumagai, H., Okamoto, K., Okumura, M., Nakatsuka, H., and Nishikawa, K.: TRMM Precipitation Radar, Advances in Space Research, 25, 969–972, https://doi.org/10.1016/S0273-1177(99)00932-1, http://www.sciencedirect.com/science/article/pii/S0273117799009321, 2000.

Keenan, T. D., Ebert, E., Chandrasekar, V., Bringi, V. N., and Whimpey, M.: Comparison of TRMM Satellite-Based Rainfall with Surface Radar and Gauge Information, in: 31st Conference on Radar Meteorology, American Meteorological Society, https://ams.confex.com/ams/pdfpapers/64179.pdf, 2003.

Kim, J.-H., Ou, M.-L., Park, J.-D., Morris, K. R., Schwaller, M. R., and Wolff, D. B.: Global Precipitation Measurement (GPM) Ground Validation (GV) Prototype in the Korean Peninsula, Journal of Atmospheric and Oceanic Technology, 31, 1902–1921, https://doi.org/10.1175/JTECH-D-13-00193.1, http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-13-00193.1, 2014.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Schwaller, M., Petersen, W., and Zhang, J.: Comparison of TRMM 2A25 Products, Version 6 and Version 7, with NOAA/NSSL Ground Radar–Based National Mosaic QPE, Journal of Hydrometeorology, 14, 661–669, https://doi.org/10.1175/JHM-D-12-030.1, http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-12-030.1, 2012.

Kubota, T., Yoshida, N., Urita, S., Iguchi, T., Seto, S., Meneghini, R., Awaka, J., Hanado, H., Kida, S., and Oki, R.: Evaluation of Precipitation Estimates by at-Launch Codes of GPM/DPR Algorithms Using Synthetic Data from TRMM/PR Observations, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 3931–3944, https://doi.org/10.1109/JSTARS.2014.2320960, 2014.

Kucera, P. A., Krajewski, W. F., and Young, C. B.: Radar beam occultation studies using GIS and DEM technology: An example study of Guam, Journal of Atmospheric and Oceanic Technology, 21, 995–1006, http://journals.ametsoc.org/doi/full/10.1175/1520-0426(2004)021%3C0995:RBOSUG%3E2.0.CO%3B2, 2004.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The Tropical Rainfall Measuring Mission (TRMM) Sensor Package, Journal of Atmospheric and Oceanic Technology, 15, 809–817, https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2, https://journals.ametsoc.org/doi/full/10.1175/1520-0426%281998%29015%3C0809%3ATTRMMT%3E2.0.CO%3B2, 1998.

Liao, L., Meneghini, R., and Iguchi, T.: Comparisons of rain rate and reflectivity factor derived from the TRMM precipitation radar and the WSR-88D over the Melbourne, Florida, site, Journal of Atmospheric and Oceanic Technology, 18, 1959–1974, http://journals.ametsoc.org/doi/abs/10.1175/1520-0426(2001)018%3C1959:CORRAR%3E2.0.CO%3B2, 2001.

McKinney, W.: Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, vol. 445, pp. 51–56, Austin, TX, 2010.

Meneghini, R., Iguchi, T., Kozu, T., Liao, L., Okamoto, K., Jones, J. A., and Kwiatkowski, J.: Use of the Surface Reference Technique for Path Attenuation Estimates from the TRMM Precipitation Radar, Journal of Applied Meteorology, 39, 2053–2070, https://doi.org/10.1175/1520-0450(2001)040<2053:UOTSRT>2.0.CO;2, https://journals.ametsoc.org/doi/abs/10.1175/1520-0450(2001)040%3C2053:UOTSRT%3E2.0.CO%3B2, 2000.

Merceret, F. J. W.: Attenuation of Weather Radar Signals Due to Wetting of the Radome by Rainwater or Incomplete Filling of the Beam Volume, Tech. rep., https://ntrs.nasa.gov/search.jsp?R=20020043890, 2000.

Michelson, D., Einfalt, T., Holleman, I., Gjertsen, U., Friedrich, Katja, Haase, G., Lindskog, M., and Jurczyk, A.: Weather radar data quality in Europe quality control and characterisation., Publications Office, Luxembourg, oCLC: 1044365355, 2005.

Morris, K. R. and Schwaller, M.: Sensitivity of spaceborne and ground radar comparison results to data analysis methods and constraints, in: 35th Conference on Radar Meteorology, American Meteorological Society, Pittsburgh, PA, https://ams.confex.com/ams/35Radar/webprogram/Manuscript/Paper191729/AMS_35th_Radar_paper_final.pdf, 2011.

NASA: Precipitation Processing System Tropical Rainfall Measuring Mission File Specification for TRMM Products, https://pps.gsfc.nasa.gov/Documents/filespec.TRMM.V7.pdf, 2017.

NASA: Release Notes for the PR Level 1 products, https://pps.gsfc.nasa.gov/Documents/ReleaseNote_PU1_productV05.pdf, 2017.

Oliphant, T. E.: Guide to NumPy, CreateSpace Independent Publishing Platform, USA, 2nd edn., 2015.

Ośródka, K., Szturc, J., and Jurczyk, A.: Chain of data quality algorithms for 3-D single-polarization radar reflectivity (RADVOL-QC system), Meteorological Applications, 21, 256–270, https://doi.org/10.1002/met.1323, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1323, 2014.

Park, S., Jung, S.-H., and Lee, G.: Cross Validation of TRMM PR Reflectivity Profiles Using 3D Reflectivity Composite from the Ground-Based Radar Network over the Korean Peninsula, Journal of Hydrometeorology, 16, 668–687, https://doi.org/10.1175/JHM-D-14-0092.1, wOS:000352735100013, 2015.

Petracca, M., D'Adderio, L. P., Porcù, F., Vulpiani, G., Sebastianelli, S., and Puca, S.: Validation of GPM Dual-Frequency Precipitation Radar (DPR) Rainfall Products over Italy, Journal of Hydrometeorology, 19, 907–925, https://doi.org/10.1175/JHM-D-17-0144.1, http://journals.ametsoc.org/doi/10.1175/JHM-D-17-0144.1, 2018.

Rasp, S., Selz, T., and Craig, G. C.: Variability and Clustering of Midlatitude Summertime Convection: Testing the Craig and Cohen Theory in a Convection-Permitting Ensemble with Stochastic Boundary Layer Perturbations, Journal of the Atmospheric Sciences, 75, 691–706, 2018.

Rew, R., Davis, G., Emmerson, S., Cormack, C., Caron, J., Pincus, R., Hartnett, E., Heimbigner, D., Lynton Appel, and Fisher, W.: Unidata NetCDF, https://doi.org/10.5065/D6H70CW6, http://www.unidata.ucar.edu/software/netcdf/, 1989.

Rinollo, A., Vulpiani, G., Puca, S., Pagliara, P., Kaňák, J., Lábó, E., Okon, L., Roulin, E., Baguis, P., Cattani, E., Laviola, S., and Levizzani, V.: Definition and impact of a quality index for radar-based reference measurements in the H-SAF precipitation product validation, Natural Hazards and Earth System Science, 13, 2695–2705, https://doi.org/10.5194/nhess-13-2695-2013, http://www.nat-hazards-earth-syst-sci.net/13/2695/2013/, 2013.

Rose, C. R. and Chandrasekar, V.: A systems approach to GPM dual-frequency retrieval, IEEE Transactions on Geoscience and Remote Sensing, 43, 1816–1826, https://doi.org/10.1109/TGRS.2005.851165, 2005.

Schumacher, C. and Houze, Jr., R. A.: Comparison of Radar Data from the TRMM Satellite and Kwajalein Oceanic Validation Site., Journal of Applied Meteorology, 39, 2151–2164, https://doi.org/10.1175/1520-0450(2001)040<2151:CORDFT>2.0.CO;2, http://adsabs.harvard.edu/abs/2000JApMe..39.2151S, 2000.

Schumacher, C. and Houze Jr, R. A.: Stratiform rain in the tropics as seen by the TRMM precipitation radar, Journal of Climate, 16, 1739–1756, http://journals.ametsoc.org/doi/full/10.1175/1520-0442(2003)016%3C1739:SRITTA%3E2.0.CO%3B2, 2003.

Schwaller, M. R. and Morris, K. R.: A Ground Validation Network for the Global Precipitation Measurement Mission, Journal of Atmospheric and Oceanic Technology, 28, 301–319, https://doi.org/10.1175/2010JTECHA1403.1, http://journals.ametsoc.org/doi/citedby/10.1175/2010JTECHA1403.1, 2011.

Seto, S. and Iguchi, T.: Intercomparison of Attenuation Correction Methods for the GPM Dual-Frequency Precipitation Radar, Journal of Atmospheric and Oceanic Technology, 32, 915–926, https://doi.org/10.1175/JTECH-D-14-00065.1, https://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-14-00065.1, 2015.

Speirs, P., Gabella, M., and Berne, A.: A comparison between the GPM dual-frequency precipitation radar and ground-based radar precipitation rate estimates in the Swiss Alps and Plateau, Journal of Hydrometeorology, https://doi.org/10.1175/JHM-D-16-0085.1, http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0085.1, 2017.

Takahashi, N., Kuroiwa, H., and Kawanishi, T.: Four-year result of external calibration for Precipitation Radar (PR) of the Tropical Rainfall Measuring Mission (TRMM) satellite, IEEE Transactions on Geoscience and Remote Sensing, 41, 2398–2403, https://doi.org/10.1109/TGRS.2003.817180, 2003.

Toyoshima, K., Masunaga, H., and Furuzawa, F. A.: Early Evaluation of Ku- and Ka-Band Sensitivities for the Global Precipitation Measurement (GPM) Dual-Frequency Precipitation Radar (DPR), SOLA, 11, 14–17, https://doi.org/10.2151/sola.2015-004, https://www.jstage.jst.go.jp/article/sola/11/0/11_2015-004/_article, 2015.

Villarini, G. and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surveys in Geophysics, 31, 107–129, https://doi.org/10.1007/s10712-009-9079-x, https://link.springer.com/article/10.1007/s10712-009-9079-x, 2010.

Wang, J. and Wolff, D. B.: Comparisons of Reflectivities from the TRMM Precipitation Radar and Ground-Based Radars, Journal of Atmospheric and Oceanic Technology, 26, 857–875, https://doi.org/10.1175/2008JTECHA1175.1, http://journals.ametsoc.org/doi/abs/10.1175/2008JTECHA1175.1, 2009.

Warren, R. A., Protat, A., Siems, S. T., Ramsay, H. A., Louf, V., Manton, M. J., and Kane, T. A.: Calibrating ground-based radars against TRMM and GPM, Journal of Atmospheric and Oceanic Technology, https://doi.org/10.1175/JTECH-D-17-0128.1, http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-17-0128.1, 2018.

Wen, Y., Hong, Y., Zhang, G., Schuur, T. J., Gourley, J. J., Flamig, Z., Morris, K. R., and Cao, Q.: Cross Validation of Spaceborne Radar and Ground Polarimetric Radar Aided by Polarimetric Echo Classification of Hydrometeor Types, Journal of Applied Meteorology and Climatology, 50, 1389–1402, https://doi.org/10.1175/2011JAMC2622.1, http://journals.ametsoc.org/doi/abs/10.1175/2011JAMC2622.1, 2011.

Zhang, J., Qi, Y., Howard, K., Langston, C., and Kaney, B.: Radar quality index (RQI)—A combined measure of beam blockage and VPR effects in a national network, in: Proceedings, International Symposium on Weather Radar and Hydrology, http://www.nssl.noaa.gov/projects/q2/tutorial/images/mosaic/WRaH_Proceedings_Zhang-et-al_v3.pdf, 2011.

Zhang, S., Zhu, Y., Wang, Z., and Wang, Y.: Consistency analysis and correction of ground-based radar observations using space-borne radar, Journal of Atmospheric and Solar-Terrestrial Physics, 169, 114–121, https://doi.org/10.1016/j.jastp.2018.01.030, http://linkinghub.elsevier.com/retrieve/pii/S1364682617305102, 2018.

Zhong, L., Yang, R., Wen, Y., Chen, L., Gou, Y., Li, R., Zhou, Q., and Hong, Y.: Cross-evaluation of reflectivity from the space-borne precipitation radar and multi-type ground-based weather radar network in China, Atmospheric Research, 196, 200–210, https://doi.org/10.1016/j.atmosres.2017.06.016, http://linkinghub.elsevier.com/retrieve/pii/S0169809516306913, 2017.