A review of "Enhancing the consistency of spaceborne and ground-based radar comparisons by using quality filters" by Irene Crisologo et al., submitted to AMT

The paper presents a large data set (5 years of data, 283 overpasses downloaded) and interesting comparison between a single ground-based radar (GR) and two satellite precipitation radars (SPR):

- the first ever spaceborne precipitation radar (PR) on-board the TRMM satellite (183 overpasses from June 1, 2012 to September 30, 2014).
- the follow-on of TRMM PR on board the GPM core satellite (103 overpasses, from June 1, 2014 to November 2016). GPM carries a dual-frequency Ku- Ka- band radar (DPR) rather than the single-frequency Ku-band PR on board TRMM.

Many efforts have been made to provide spaceborne radars with long-term electronic stability. The calibration factor has assumed to have an accuracy better than 1 dB. As a consequence, the authors have used (quality-weighted) average of geo-located TRMM PR and GPM DPR reflectivity values as reference to check the quality of the reflectivity bias of the S-band GR (Note that the number of overpasses with "enough" reflectivity values reduces to "only" 74 for TRMM and 40 for GPM)

The authors have done a thorough job of comparing (quality-)averaged reflectivity of the GR versus the satellite one. Hence, I find the paper valuable and interesting: it is another example of weather radar monitoring and reflectivity bias assessment using the SPR as reference. It shows once more the importance of monitoring (calibrate and adjust) GRs reflectivity before a quantitative use for precipitation estimation. I think it is particularly valuable and illustrative for National Weather Services worldwide. However, I feel that there are some issues that need to be specifically addressed in the text/Tables before the paper becomes in publishable form. Schematically, I have identified four points, which will be described below

1) Omission (or insufficient description) in the text of limiting factors in the quantitative interpretation of radar reflectivity.
2) GR monitoring based on SPR: possible causes of the discontinuity affecting the GR?
3) (Radar reflectivity) Mean Field Bias definition and assessment.
4) Literature

I also propose a different structure for the paper: Sec. 2 Data; Sec. 3 Methods; Sec. 4 Results and Discussion (see page 4 for details)

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"

Yours Sincerely,
Marco Gabella

Locarno-Monti, 22.5.2018

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[1] (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!

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)

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 $\{\Delta Z_{dB}*\}$ (lines 20-22, page 15) also the standard deviation of $\{\Delta Z_{dB}*\}$. 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 Q_{BBF} larger than say, 0.9 (or larger). After having done this selection, consider the difference ($BIAS_{xx}$) between different quantiles (probability matching): xx=50 (median), 75, 84, 90, 95, 99 percentiles. Maybe, $BIAS_{xx}$ is not constant, rather it depends on the percentile? Finally, for

[1] Bech et al. wrote " … less idealized, but more realistic, is the case of a Gaussian beam."

these Q_{BBF}-selected bins, you may explore the value of the average bias $E\{\Delta Z_{dB}\}$ 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 dBZ_{GR}). 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?

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).

ADDITIONAL SUGGESTED REFERENCES

Iguchi, T., T.Kozu, R. Meneghini, J.Awaka, and K. Okamoto, “Rain-profiling algorithm for the TRMM Precipitation Radar,” J. Appl. Meteorol., vol. 39, pp. 2038–2052, 2000.

Keenan, T., E. Ebert, V. Chandrasekar, V. Bringi, and M. Whimpey (2003), Comparison of TRMM satellite-based rainfall with surface radar and gauge information, paper presented at 31st Conference on Radar Meteorology, Am. Meteorol. Soc., Seattle, Wash., 6–12 Aug.

Joss, J., et al, 2006: Variation of weather radar sensitivity at ground level and from space: Case studies and possible causes, Meteorol. Z., 15, 485–496, doi:10.1127/0941-2948/2006/0150.

GPM related references

Kubota, T., and Coauthors, 2014: Evaluation of precipitation estimates by at-launch codes of GPM/DPR algorithms using synthetic data fromTRMM/PRobservations. IEEEJ. Sel. Top.Appl.Earth Obs. Remote Sens., 7, 3931–3944, doi:10.1109/JSTARS.2014.2320960.

Furukawa, K., T. Nio, T. Konishi, R. Oki, T. Masaki, T. Kubota, T. Iguchi, and H. Hanado, 2015: Current status of the dual-frequency precipitation radar on the global precipitation measurement core spacecraft. Proc. SPIE, 9639, 96 390

Toyoshima, K., H. Masunaga, and F. A. Furuzawa, 2015: Early evaluation of Ku- and Ka-band sensitivities for the Global Precipitation Measurement (GPM) Dual-Frequency Precipitation Radar (DPR). SOLA, 11, 14–17, doi:10.2151/sola.2015-004.

Speirs et al., 2017: A Comparison between the GPM Dual-Frequency Precipitation Radar and Ground-Based Radar Precipitation Rate Estimates in the Swiss Alps and Plateau, *J. Hydrometeorol.*, DOI: 10.1175/JHM-D-16-0085.1

Gabella et al., 2017: 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; 19 pp.; doi:10.3390/rs9111147

Specific Suggestions

Introduction

-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).
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

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.
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 $E\{\Delta Z_{dB}*\}$ is almost 5! dB, while the annual average is -4.1 dB?!? (see my previous points 2) and 3))

**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

Section 3.1: 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

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

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.

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.

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 $E\{\Delta Z_{dB}*\}$ 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 Q_{BFF} works properly then: $E\{\Delta Z_{dB}^*\} - E\{\Delta Z_{dB}\}$ 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 …).

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 Q_{BFF} shows a better standard deviation of ΔZ_{dB}^*. By the way, I recommend adding a table and/or a figure (histogram) that summarizes the statistical properties of $\sigma^*\{\Delta Z_{dB}^*\}$ and $\sigma\{\Delta Z_{dB}\}$.
Then, I would introduce the important result regarding the consistency of GPM and TRMM radars, followed by the changes of the bias in time

Page 16.
Line 5, delete coherent.
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)].
Line 17-19. Pleonastic. I would delete it.
Line 26. Why do you discuss C-band radar technology ?

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?