Responses to comments from Referee #2

On „Use of dual-polarization weather radar quantitative precipitation estimation for climatology“ by Tanel Voormansik et al.(HESS-2019-624)

Referee's comment

GENERAL COMMENTS

In this paper long-term datasets from two different regions (Estonia and Northern Italy) are used to evaluate the performance of polarimetric weather radar quantitative precipitation estimates. Several years of radar and gauge data are used for this. This is a very interesting topic that is highly relevant and timely as long-term high-quality operational polarimetric datasets are becoming more and more available. The paper has a clear focus, which makes it pleasant to read. Some of the English used in the paper could be improved, but it certainly does not prohibit full understanding of the paper. I do have some questions that I would like to see clarified and some suggestions for improvements. In particular, I think there may be an error in at least one of the figures that I think the authors should look at. I think the paper should be published after major revisions. Specific comments are given below.

Authors' response

Authors would like to sincerely thank the referee for the time and effort spent in reading the initial manuscript and for making many clear and constructive suggestions for improvement. This helped a lot to improve the manuscript.

Referee's comment

Specific comments

1. I very much appreciate the honesty of the authors about removing low-quality data from the dataset that resulted from radar issues. Because of this, and because of the focus on only the warm season, I doubt whether mentioning "climatology" in the title would be suitable. Please reconsider this, or at least add to the title that this is about warm-season precipitation (which is still very valuable).

Authors' response

We agree with the comment. We decided to change the title to be more appropriate considering the length of the dataset used. The new proposed title is "Applicability of dual-polarization weather radar quantitative rainfall estimation for climatological purposes".

Referee's comment

2. On line 72, it is stated that the Italian rain gauges have a resolution of 0.2 mm and 1 minute. This means that, if the gauges report rainfall intensities, the minimum rainfall intensity that these gauges would be able to record is 12 mm h^{-1} . Or do the gauges record total rainfall accumulation with a 1-minute time resolution (in which case there is no issue with the total accumulations)?

Authors' response

This is the recording resolution, measurement resolution is higher. Following the referee comment to make it more clear the sentence was rearranged as follows:

„The temporal resolution of the gauges network is 1-minute. The Arpa Piemonte weather stations are equipped with CAE PMB2 tipping-bucket rain gauges. Their resolution (0.2 mm) is the amount of precipitation for one tip of the bucket. The working range of measures is from zero mm to 300 mm/h with underestimation for high precipitation intensities. Such errors are corrected according to results of WMO Field Intercomparison of Rainfall Intensity Gauges (Vuerich et al., 2009)."

Referee's comment

3. On lines 105-107, the computation of Φ_{DP} and K_{DP} from raw Φ_{DP} is mentioned, along with the fact that "carefully tuned parameter values according to data specifics" are used for this. It would be interesting and highly relevant to include a more thorough description of this in the paper, especially since K_{DP} is a key variable in this paper. I think a one- or two-sentence summary of the method would be nice, along with a short description of the parameters and how they were determined.

Authors' response

Added a few sentences to describe the derivation of K_{DP}: "With default parameter values the rays where differential propagation phase folding occurred did not unfold correctly and thus the function did not produce correct specific differential phase values. In order to fix the folding issue function parameters *self_const* (self-consistency factor) and *low_z* (low limit for reflectivity – reflectivity below this value is set to this limit) had to be tuned. The default values were 60000.0 and 10.0 respectively and after testing with various combinations of various values the values 12000.0 and 0.0 were found to produce optimal results and therefore were chosen for final calculations."

Referee's comment

4. On lines 107-110, the self-consistency method for re-calibrating Z_H is discussed, where Z_{DR} is also used. Later, on lines 124-125, it is mentioned that the use of Z_{DR} for quantitative precipitation estimation is not recommended for C-band radars. I think it should be discussed here why Z_{DR} can be used for re-calibration of Z_H.

Authors' response

Agreed. Following the referee comment the following sentence was added to Section 2.2:

„Z_{DR} is not suitable for QPE on C-band radars, but it can be used in this calibration methodology after applying strict restrictions on the data used for this purpose."

Referee's comment

5. On line 127, the threshold for switching between $R(Z_H)$ and $R(K_{DP})$ is defined to be 25 dBZ for Z_H. Using Eqs (1) and (2), this threshold translates to $R \approx 1$ mm h−1 and $K_{DP} \approx 0.015°$ km−1. These values are much lower than what is cited from the literature ($R = 50$ mm h−1 and $K_{DP} = 0.5 - 1°$ km−1). What is the reason for using this much lower threshold? I think this should be explained in the paper.

Authors' response

We agree that an explanation would be suitable in the paper. Various thresholds were tested on our data and this performed the best. Following the referee comment the following sentence was added to the manuscript:

„The Z_H threshold value was selected after testing with various reflectivity levels. The threshold level is considerably lower than some of the thresholds used in the literature but on our datasets it performed the best."

Referee's comment

6. In Section 2.2, there is no mention of attenuation that could affect the $R(Z_H)$ estimates. This attenuation could be corrected for using Φ_{DP}. Is there a specific reason why attenuation correction is not carried out?

Authors' response

We agree that it should be mentioned in Section 2.2 and explained why it is not used. Following the referee comment the following was added to the manuscript:

"The QPE of $R(Z_H)$ can be affected by attenuation on C-band radars especially in heavy precipitation and at long distances. While this can be corrected using Φ_{DP} in our study it was not applied to the reflectivity data in order to not introduce another possible source of error between the results of Estonia and Italy that could not be easily quantified. Effectiveness of attenuation correction using Φ_{DP} is hampered by its temperature, shape and size distribution dependence which affect the accompanying error (Vulpiani et al., 2008)."

Referee's comment

7. In Section 2.2, it would be good to mention that the effect of VPR will be limited in the analyses because only data from the warm season will be used, and that only data close to the radars (70 km and 30 km for Estonia and Italy, respectively) will be used.

Authors' response

Agreed. Following the referee comment the following sentence was added to Section 2.2:

"The QPE of $R(Z_H)$ can also be affected by the effect of non-uniform vertical profile of reflectivity (VPR). In the current study the effect of VPR will be limited because only data from warm season was used and distance limits to the radar data were set (70 km for Estonia and 30 km for Italy, respectively)."

Referee's comment

8. In Section 2.2, it is not explicitly mentioned how precipitation accumulations are computed. I assume (also based on the rest of the paper) that they are computed by simply adding subsequent instantaneous radar QPE values, without any space-time interpolation. It would be good to mention that here explicitly.

Authors' response

Agreed. The clarification seemed to suit better to Section 2.3 where it was added to the earlier description of accumulation:

„Radar-based QPEs have been accumulated to 1-hour duration and longer durations have been calculated based on these accumulations. Accumulations were calculated by adding subsequent instantaneous radar QPE values without any space-time interpolation."

Referee's comment

9. Is my interpretation of Fig. 1 correct if I say that in Estonia only a circular area around the radar is used (up to 70 km range), while in Italy a rectangular area (60 × 60 km2) around the radar is used? If this is correct, is there an explanation of why two different areas have been used? This should be included in the paper.

Authors' response

We agree with the referee and thus the following explanation was added to the manuscript:

"As can be seen from Fig. 1 circular area around the radar is used in Estonia but in Italy rectangular area is used. The reason for this is that Orography in Piemonte is very complex ranging from flat plains in the Po valley (about 100 m a.s.l.) to the Alps up to more than 4,000 m a.s.l. The Bric della Croce weather radar is located on Torino hill that is about 30 km from the Alps. Therefore, the elegant and simple limitation in range by some kilometers from the radar site does not work. To avoid mountainous areas, where partial and total beam-blocking, heavy ground contamination increases, a rectangle area, that extends towards flat grounds, has been preferred."

Referee's comment

10. Equation (3) for Pearson's correlation coefficient is incorrect. It should be: CC = Xn i=1 (ri − rm) (gi − gm) vuutXn i=1 (ri − rm) 2 vuutXn i=1 (gi − gm) 2 .

Authors' response

We would like to thank the referee for pointing that out. The Equation (3) was corrected accordingly in the manuscript.

Referee's comment

11. For the definition of the normalized mean error in Eq. (4), the multiplication with 100% needs to be omitted in order to make it consistent with the results presented in Tables 1-6. I would also like to suggest renaming this statistic to the "normalized mean absolute error", which in my view is closer to what it actually is.

Authors' response

We would like to thank the referee for pointing the error out and we agree with the suggestion of renaming the statistic. Manuscript was edited according to the suggestions.

Referee's comment

12. The authors could consider to also normalize the RMSE in Eq. (6) with the mean gauge rainfall. In this way, all statistics will be dimensionless. This is of course just a choice, and I would also be perfectly fine with leaving the definition as it is.

Authors' response

We thank the referee for the suggestion but decided to leave the definition as it is.

Referee's comment

13. On lines 192-193, the cause for the more severe underestimation of R from Z_H in Italy than in Estonia is said to be the fact that there is more intense precipitation. However, doesn't this mean that the employed Z − R relation is not suitable? Differences in raindrop size distribution (DSD)

climatologies between Estonia and Italy may also cause differences. So it would be good to comment here on the suitability of the retrieval relations (Eqs (1) and (2)) for both regions.

Authors' response

We agree that the retrieval relations are definitely a cause for differences among the two regions. The rationale behind using the same relations for both regions was the fact that rainfall retrieval relations always entail errors with them anyway and we wanted to keep the comparison as straightforward and homogeneous as possible. Following the referee comment the following sentences were added to Section 3.1:

„Another cause of differences between the two countries might be differences in the drop size distribution climatologies. Rainfall retrieval relations also entail errors and to keep the comparison as uniform as possible we decided to use the same relations for both Italy and Estonia."

Referee's comment

14. On lines 198-199, it is stated that using different time intervals can help in understanding the effect of temporal sampling differences between radar and gauges. While this is certainly true, it should also be noted that using longer accumulation intervals will also lead to less severe errors (compensating underestimates and overestimates; the $R(K_{DP})$ curve in Fig. 3 is a good example of this). I think a remark about this should also be added to the text. The same holds for line 219.

Authors' response

Agreed. Following the referee comment the following sentence was added to Section 3.3:

„Using longer accumulation intervals leads to less severe errors as the longer period compensates for both underestimates and overestimates."

Referee's comment

15. On lines 215-217, an important statement is made about the improvement that $R(Z_H, K_{DP})$ gives over the other methods. At first reading, I thought that this statement is too bold given the results presented in Table 1, but on second thought it is correct. What would have helped me is if something along the lines of "(i.e., each statistic is approximately as good at the best of the other two)" after "...other product's weak points" would have been included. You could consider including this here.

Authors' response

We agree that the explanation should be improved. Reworded the sentence as follows:

"$R(Z_H,K_{DP})$ shows considerable improvement by combining strong aspects of the two methods"

Referee's comment

16. In Fig. 5, it is interesting to see that of the 4 highest 1-hour accumulations measured by a gauge, 3 of them have significantly higher radar estimates for $R(Z_H, K_{DP})$ than either $R(Z_H)$ or $R(K_{DP})$. This means that for $R(Z_H, K_{DP})$, probably the best estimator of R is selected for most of the intervals (i.e., for at least one of the underlying 5-minute intervals $R(Z_H)$ is higher than $R(K_{DP})$, and it is correctly selected for $R(Z_H, K_{DP})$). I think this merits some more discussion in the paper, especially since this is the case for 3 of the 4 highest 1-hour accumulations.

Authors' response

We agree that pointing this out together with additional explanation would be useful. Following text was added to the manuscript:

"Although from Fig. 5 it can be noticed that of the four highest 1-hour accumulations measured by the gauge, three of them have significantly higher radar estimates for $R(Z_H,K_{DP})$ than either $R(Z_H)$ or $R(K_{DP})$. This could be explained by precipitation that was very variable in intensity and also in spatial coverage in these three cases which in turn caused unsteady behaviour of the precipitation estimates. Z_H underestimates high intensities, but with low intensities K_{DP} becomes noisy and the rainfall intensity estimation is not feasible. Finally, to reduce K_{DP} uncertainties range averaging is mandatory, leading to underestimation in case of very localized showers. By blending both $R(Z_H)$ and $R(K_{DP})$, a better rainfall estimation is expected."

Referee's comment

17. On lines 242-243, it is stated that the normalized bias is much smaller for the 24- hour accumulations than for the 1-hour accumulations. However, looking at the definition of the NMB in Eq. (5), there should be absolutely no difference between the two, if the same underlying samples have been used (i.e., it makes no difference whether you first sum over 24 hours, and then subtract gauge from radar sums, or if you compute the difference first and then sum over 24 hours because subtraction is a linear operation). So what is the cause of these differences? Is it because you use different underlying samples, possibly by taking only accumulations above 0.1 mm (see captions of Figs 4 to 7)? If this is the case, this stresses the importance of low-intensity rain for total rainfall accumulations. This should be explained clearly. The same holds for differences between 24-hour and 1-month accumulations.

Authors' response

The cause is most probably different underlying samples. The 0.1 mm threshold is applied after the accumulation as a last step before calculating the verification metrics. This means that the total accumulated precipitation sum is larger in 24h accumulation dataset than in 1h dataset (although the difference is not big, 10900 mm vs 10200 mm in case of Estonia and gauge measurements). Following the referee comment the following was added to the manuscript:

"By looking at the definition of NMB in Eq. (5) it can be seen that in case the same underlying samples are used NMB should be equal on all accumulation lengths. In our study the underlying samples were different as the 0.1 mm threshold was applied after the accumulation as a last step before calculating the verification metrics. This emphasizes the importance of low-intensity precipitation for total accumulations."

Referee's comment

18. If I compare the NMB presented for $R(Z_H)$ in Table 3 and the corresponding panel in Fig. 6, I'm surprized at the fact that the underestimation by the radar is so small. Is this because there is an extremely high density of points just above the black line close to 0 mm in Fig. 6?

Authors' response

Rechecking the dataset and recalculating the NMB gave the same results so the reason behind it must be high density of points above and near the black line on low accumulation values.

Referee's comment

19. Comparison of Figs 5 and 7 gives me the feeling that there may be an error in one of them. For example, if I roughly add all of the accumulations from $R(Z_H)$ in Fig. 5, the resulting amount of rain is

much smaller than when I roughly add all of the accumulations from $R(Z_H)$ in Fig. 7. Furthermore, the number of accumulations exceeding 0.1 mm given in the caption is higher for Fig. 7 than for Fig. 5. This is impossible unless a different dataset has been used. So I suggest to take another careful look at the figures and the results presented in the tables.

Authors' response

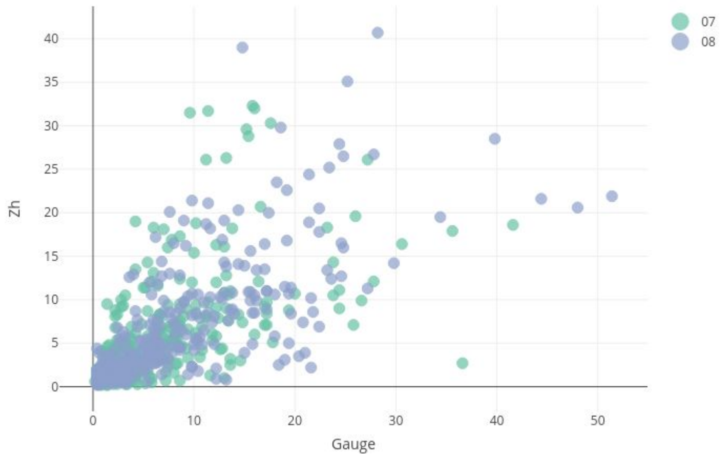
The comparison between radar and raingauge is carried out only if both the gauge measurement and the radar estimation are simultaneously greater than 0.1 mm; otherwise, this comparison loses meaning. Here, short duration and scattered precipitations are considered (i.e. few rain gauges record rainfall during an event). When the rainfall accumulation interval decreases, the number of valid couples (i.e. both greater than zero) tends to decrease. That's the reason because the number of samples is greater in Figure 7 than in Figure 5. As all invalid couples are discharged, it has no meaning to compare the total accumulation between these plots.

Referee's comment

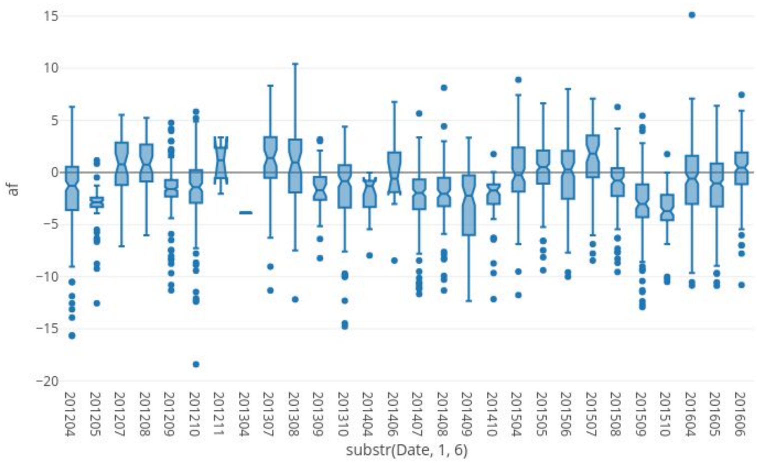
20. Figure 7 seems to show two regimes for $R(Z_H)$, where one overestimates and the other underestimates for higher rainfall accumulations. It would be interesting to discuss this in the paper. I'm interested to learn if these regimes are separable by some other variable such as time, temperature, or something else.

Authors' response

The reviewer is right. The Bric della Croce weather radar is located on a top of hill at 770 m asl and during the winter season a vertical profile reflectivity correction (VPR) is applied (Koistinen, 1991). This correction is manually switched on at the beginning of the cold season and it is switched off at the end. In case of convective precipitation, this correction may lead to rainfall overestimation. On the other hand, stratiform cold precipitation is heavily underestimated when VPR correction is switched off. So, the VPR correction leads to these regimes in daily comparison (Figure 7). The separation between the two regimes could be obtained by reducing the study area even more and limiting the study to June, July and August. Unfortunately, the corrected reflectivity (including VPR) is available for studied years only; later both corrected and uncorrected become available. The following Figure shows the same scatterplot as in the paper but limited to summer months July and August 2012-2016.



The double regime induced by VPR correction disappears. However, if we consider the logarithmic ratio between rainfall estimated by weather radar and measured by gauge ($af = 10\,log10(R/G)$), it is clearly visible a seasonality with underestimation (on average) for April, May and October and November and slight overestimation during warm months.



Cited references:

Koistinen, J.: Operational correction of radar rainfall errors due to the vertical reflectivity profile, in: Proceedings of the 25th Radar Meteorology Conference, American Meteorological Society, Paris, France, 91–96, 1991.

Vuerich, E., Monesi, C., Lanza, L., Stagi, L., Lanzinger, E.: WMO Field Intercomparison of Rainfall Intensity Gauges, Vigna di Valle, Italy, October 2007 - April 2009, WMO/TD- No. 1504; IOM Report-No. 99, available at: https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-84_Lab_RI/IOM-84_DataSheets/TippingBucket_Italy_CAE.pdf, 2009.

Vulpiani, G., Tabary, P., Parent du Chatelet, J. and Marzano, F.S.: Comparison of advanced radar polarimetric techniques for operational attenuation correction at C band, J. Atmos. Ocean. Technol., *25*, 1118-1135, https://doi.org/10.1175/2007JTECHA936.1, 2008.