

# Comments to "Multi-radar performance in the Midwestern United States at large ranges"

Edouard Goudenhoofdt

August 4, 2017

## 1 General comments

This paper analyses the performance of several rainfall estimation algorithms from 3 polarimetric radars in the midwestern united states. Such verification studies are not particularly original but constitute a good contribution to the field. Unfortunately the limited amount of data used in the present study reduces its value.

The scientific quality of the paper is poor due to a limited literature review, questionable methods and results which are not sufficiently discussed and lack of significant conclusions.

The presentation is neither clear nor concise and some inconsistencies appear in the results. The text is particularly difficult to read due to inappropriate use of the language.

Despite those flaws I think interesting results on the performance of dual-pol radars can be shared with the community. Therefore major comments are given below to help the authors to reach an acceptable level of quality to consider publication.

## 2 Major comments

### 2.1 Does the paper address relevant scientific questions within the scope of HESS?

The study of the performance of radar-based precipitation estimation is clearly in the scope of HESS. However the motivation for this study should be better explained in the introduction. What are the possible applications ?

The specific objectives of the paper are detailed at the end of the introduction. While studying the effect of range is interesting, do you think it justifies the whole paper? Is the added value of dual-pol variables not the main focus of the paper? What do you exactly mean by testing the best radar algorithm? On average? Did you consider comparing the performance of the 3 radars since calibration of the dual-pol measurements is crucial. The rationale for separating the data in warm and cool seasons is not sufficiently explained. What about grouping the data by month or weather regimes?

## **2.2 Does the paper present novel concepts, ideas, tools, or data?**

There is nothing particularly new in the paper but long-term verifications of precipitation estimates from dual-pol radars are welcome.

The authors present the paper as an extension of a previous work with more data but it is limited to the year 2014. Since they are easily available in the US, why not using the radar data from 2013-2016? Using 4 years of data would decrease the statistical biases and be more representative of the weather in Missouri. I understand that only the hours when all radar and gauge data are available are taken but what explains the very low availability (13%) and how many of the 1100 hours are dry episodes? This results in a limited dataset subject to statistical biases.

## **2.3 Are substantial conclusions reached?**

In general the limitations in the data and presented results do not allow the derivation of substantial conclusions.

The paper concludes that errors come mainly from mismatch with rain-gauges either in winter or summer. For "all" data the  $R(Z, ZDR)$  is the best and  $R(ZDR, KDP)$  the worst while for "hits" data the best is  $R(Z)$  up to 100 km,  $R(KDP)$  from 135 km and the algorithm based on the type of precipitation in between.

Important results from the paper are not reported in the conclusions like the bias and error in function of the range. The strong impact of false and missed detection is not sufficiently discussed. It is often unclear if the results are valid for all algorithms or a selection of them. The fact that the best algorithm is different if you consider "all" or "hits only" data is not sufficiently discussed. There is not much on the added value of dual-pol compared to single-pol. Most results (not only in the conclusions) are presented in a qualitative way (best, worst, increase) while information on the magnitude of the (relative) performance is missing.

## 2.4 Are the scientific methods and assumptions valid and clearly outlined?

Some parts of the methodology are questionable.

The rain gauge network is small and not homogeneous and the gauges tend to be far from the radars. This makes the statistical analysis difficult especially the study of range effects. These limitations are not properly discussed through the text.

The information on the quality of the radar and rain gauge measurements used in this study is not sufficient and too general. Did the authors perform any complementary control on the data? The quality of the datasets is important to obtain robust results and avoid artefacts.

The presentation of the rainfall estimation algorithms is unclear. Do you accumulate the dual-pol variables over one hour? What is the rationale of using 11 different algorithms (looking at the table some algorithms have similar parameters)? Is it necessary to make all the combination of smoothed and original fields? The "RREC" algorithm is not well described. Is it fair to use this advanced algorithm?

You focus only on the rainfall estimation from radar variables. However processing weather radar data includes important steps like clutter mitigation and correction for the height of the measurements. Is this not available in WDSS? Did you consider using operational rainfall estimation products as reference? Did you consider compositing your 3 radars to obtain a better overall performance? Did you consider using multi-sensor products?

The verification methodology is questionable :

- Why not using only the normalised statistics ?
- Did you consider conditional statistics (e.g. only values above 1 mm) ?
- How do you deal with the non-gaussian behaviour of precipitation. Did you consider statistics based on ratios?
- Why do you use range statistics for all data and scatter plots with R2 for "hits only" data and not the opposite.
- Why some statistics are not shown in the figures (e.g. PoFD)?
- You only show the best and worst algorithms which exhibit huge performance differences making the figures difficult to read. There is hardly no information on the performance of the other 53 algorithms. You could show the best for each general type of algorithm.

- There is no comparison between the radars at gauges located at similar distances. This could highlight calibration issues.
- Why not showing the summer results for all radars?

## 2.5 Are the results sufficient to support the interpretations and conclusions?

In general the results are not sufficient to support the interpretations and there is a lack of discussion. It is particularly difficult to interpret the results when only the best and worst algorithm (out of 55) are shown.

You associate peaks in bias/error at specific ranges between 100 km and 150 km with bright band effects. I don't see why the overestimation due to bright band effect would not be spread over the range. The peaks appear strongly correlated with the falsely detected precipitation from the radar. In case of bright band the gauges should still measure precipitation. I think other radar errors or even gauge errors should be considered. Are the peaks caused by a few values? I would detect the largest errors and check visually for their cause. How are the results with "hits only" data?.

For all data the best and worst algorithm are  $R(Z,ZDR)$  and  $R(ZDR,KDP)$ , respectively. This is interesting but not sufficiently discussed. What explains the bad performance of  $R(ZDR,KDP)$ ? I am surprised that  $R(Z,ZDR)$  is the best since from the literature it is difficult to calibrate ZDR.

As expected, the seasonal results reveal a significantly higher contribution from summer precipitation to the statistics. There is no need to expand too much on this. However, a comment on the range performance is welcome. I don't see the winter data contributing more to PoFD than summer (line 312).

The results of Figures 10-12 for the "hits" data are interesting and allow some comparison between the algorithms. It clearly shows the added value of  $R(KDP)$  at long ranges. Again showing only the best and worst algorithm limits the interest. What is the performance of  $R(Z,ZDR)$  here? You don't discuss the bad correlation of some gauges which correspond to the peaks in the "all data" results.

## 2.6 Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

The results of your previous study are not presented or discussed.

What is exactly done by WDSS? Also the rainrate estimation and the statistics?

## 2.7 Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

The verification methodology is not clearly described :

- Some statistics are not defined and there are no references.
- The justification for NMB sounds inappropriate to me. You simply divide the mean bias by the sum of gauges values. Is it not to allow comparison between the gauges?
- The definition of the probabilities of detection between the radar and the gauge and related amounts are unclear. You are mixing probability and rainfall amounts in the explanation. Why would you scale the probability by the amount of precipitation? A reference from the forecast verification literature can help.
- You should already describe here how you select the data (all data or hits only).
- How do you define zeros in the data ? Do you apply any kind of thresholds? What are the implicit thresholds in the measurements?

The presentation of the results is unclear :

- How do you determine the overall worst and best algorithm?
- The text and figure caption are confusing. Which algorithm are shown in the figures 2-8? Best/Worst overall, per stats or per gauges? Also for PoD I guess?
- There are no units on the figure itself and the scale variations between the figures make comparison difficult.
- The text would be clearer if you compare the radar performance in one paragraph.
- Which algorithm is used on Figure 9?

## 2.8 Does the title clearly reflect the contents of the paper?

The title is not good:

- "Multi radar" suggests that a composite is used
- there is no reference to dual-polarization
- "at large ranges" is too limited (I would not mention range at all)

## **2.9 Does the abstract provide a concise and complete summary?**

There are several issues :

- the introduction is too long
- the methodology is not well described
- the main results from the conclusions are not mentioned

## **2.10 Is the overall presentation well structured and clear?**

The overall presentation is relatively well structured but the section and subsections titles are confusing (e.g Study area and methods, Individual radar performance).

## **2.11 Is the language fluent and precise?**

The text is difficult to read :

- there is a lack of structure and logic at the level of the paragraphs (e.g. lines 53-58)
- the use of the terminology is sometimes inappropriate or lacks precision (e.g. line 68 and line 403)
- the style is relatively poor and inconsistent
- there are several grammatical and spelling errors

## **2.12 Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?**

The definition of the statistics and especially the probabilities are imprecise. The NMB formula seems incorrect (do you need to divide by N)? The acronym FAR (False Alarm Ratio) is used once on line 281 but not defined.

## **2.13 Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?**

In general the text should be rewritten to improve concision and clarity.

The first parts of the introduction (i.e. description of dual-pol technology) can be made shorter.

The "results and discussion" section should be reorganised. Parts of the methodology (e.g. selection of the data) is described in this section and should

be moved in the corresponding section. There are too much figures of the same kind. You analyse the 3 radars successively with a decreasing level of details. There is apparently no logic for that. I suggest to put the results of the 3 radars on the same figure and comment it in the same paragraph. I suggest to put the results of winter and summer data on the same figure.

The conclusions section should be reorganised. You don't have to repeat the literature review at the beginning. There is not a lot of general discussion with similar studies. The outlook at the end of the conclusions is too limited and general.

## **2.14 Are the number and quality of references appropriate?**

The current literature review is limited and not well structured.

The review of existing verification studies of dual-pol radars is limited. While the operational polarimetric radar starts in 2012 in the US, there are not many references from the last years. The review is focused on the US while the dual-pol technology is available in other parts of the world, especially in Europe (e.g., Figueras i Ventura et al., 2012).

You mentioned only briefly the source of errors affecting all type of radars. However you refer to these errors often when you interpret the results (e.g. bright band). There is an extensive literature on the topic but you should at least cite a few key papers (e.g., Uijlenhoet and Berne, 2008).

You should also reference long-term verification studies of rainfall estimates from single-pol radars.

## **References**

- Figueras i Ventura, J., Boumahmoud, A.-A., Fradon, B., Dupuy, P., and Tabary, P.: Long-term monitoring of French polarimetric radar data quality and evaluation of several polarimetric quantitative precipitation estimators in ideal conditions for operational implementation at C-band, *Quarterly Journal of the Royal Meteorological Society*, 138, 2212–2228, doi:10.1002/qj.1934, 2012.
- Uijlenhoet, R. and Berne, A.: Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction, *Hydrology and Earth System Sciences*, 12, 587, doi:10.5194/hess-12-587-2008, 2008.