Hydrol. Earth Syst. Sci. Discuss., 12, C619–C632, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C619/2015/

Hydrology and Earth System Sciences
Discussions
Open Access

Interactive comment on "5 year radar-based rainfall statistics: disturbances analysis and development of a post-correction scheme for the German radar composite" *by* A. Wagner et al.

A. Wagner et al.

andreas.wagner@geo.uni-augsburg.de

Received and published: 21 March 2015

We wish to thank the reviewer for her/his constructive remarks and will try to implement them. However, it seems to us that the referee exclusively considers the correction scheme, which covers only half of our paper.
The paper comprises two parts, as clearly reflected by the title and the structure: The first part is the error analysis on a nationwide radar composite. We are not aware of a similar analysis, and if there was one, the analysis over Germany would still be new. Only the second part is the development of a post-correction scheme, as referee#1 stated.

We regret that the referee seems not to be fully convinced by our approach. We believe there are legitimate reasons for the presented approach. So, first, we clarify the reasons for our approach, the intended use of the correction algorithm and why we use these methods ("Significance"). Then we address and answer the referees concerns and further elaborate the justification of our approach.
The questions are inserted below (italic) followed by our responses:

REFEREE#1: Significance

As for the scientific significance of the manuscript, I have mixed feelings. It is, in fact, hard to address errors and artifacts after variables from a radar network (reflectivity or precipitation) have been composited on a joint Cartesian grid. Data from a single radar station in polar coordinates allows for the application of more efficient and more targeted correction procedures.
From a practical perspective, it would thus make sense to provide a workflow that can be used to reduce errors after the actual composition took place. This might be helpful for users of composite data who do not have access to the original "raw" data or the capabilities to process these. However, we need to admit that such a post-composition correction will never be more than a kludge. From a research perspective, it is kind of a dead end: you know that you could do better in terms of precipitation estimation and quality control if you had access to the "raw" radar data in polar coordinates. So, the problem is not a scientific one, but rather one of data accessibility and the capabilities of radar data processing.
Hence, I would like to maintain that the potential scientific significance of "post-composition correction schemes" is limited. Nonetheless, I acknowledge that there might be users, in research or applications, who could benefit from an efficient and straightforward post-composition correction procedure. Unfortunately, I have a couple of serious concerns that are specific to the procedure suggested by Wagner and

colleagues in their present manuscript. I would like to point out the major ones, though not exhaustively, in the following section.

ANSWER: Since DWD radar data are a very valuable hydrological data source, but not all potential users might be aware of pitfalls and shortcomings, we think it is of utmost importance to communicate the specific findings presented in this study. From our experience, hydrological radar data users may tend to believe that Doppler filters remove clutter 100%, that the polar data was flawless after all the correction of raw data that is operational in DWD, and that their composite data was homogeneous unless individual radars are missing. It is important to us to show, analyse and quantify particularly these effects that might not commonly be known. Additionally, the statistical results of our analysis have been used e.g. to help identify wind turbines in radar data. We see our contribution in pointing out and investigating into the processing effect of RX compositing. To the best of our knowledge, this has not been shown earlier. So we have no doubt about the scientific value of this hydrologically motivated radar data analysis.

Based on the results of the statistical analysis, we developed a post-correction scheme because of two reasons: the first one was that we were able to identify stable dependencies, e.g. the functional height dependence, and tested if they can be used within a simple correction scheme; the second one was, that our main concern is the analysis of long-term patterns in radar data and that there was a lack of suitable correction schemes for that purpose.

As a possible costumer of the usage of our correction scheme, insurance companies for example are interested in hazard maps, for instance to know regions of enhanced hail risk or severe precipitation. Therefore, the cell-tracking product CONRAD ("VX-composite") shall be analysed. The main requirements for this data basis are to ensure the comparability of measurements at any distance from the radar site and to remove systematic deviations. Short-term variations or limitations within radar data that do not influence the mean (annual) precipitation patterns are of minor interest. Accordingly, the crucial question was how to realise a correction algorithm that focuses

on mean precipitation patterns and that is also suitable for secondary radar-products such as the CONRAD composite. E.g., some deduced parameters of the CONRAD composite such as the intense precipitation warnings are based on accumulated radar precipitation amounts over 30 minutes. There was no correction scheme for this kind of data available.

The RX-product is widely used and it is also the basis for the secondary radar product CONRAD composite. Obviously, any systematic errors in RX will produce systematic errors in follow-up products such as CONRAD statistics. This was the reason for us to investigate the RX-product for limitations and to develop a correction algorithm. The algorithm should minimize mean systematic deviations within this data on a longer temporal scale with the aim also to transfer it to other products such as the results of the CONRAD analysis. As reprocessing was not possible, we have to work with the existing data.

The analysis of CONRAD data is future work. Here, in our presented study, we emphasize the error analysis and the hydrological point of view.

We are aware that the proposed scheme is not a substitute for common correction algorithms on single radar images (p. 1770, ll. 5-8); but it has clear advantages regarding the correction of systematic limitations.

Referee#1 major methodological concerns:

REFEREE#1: The "hybrid" nature of approach:

I already pointed out that it would be good to have an efficient (straightforward) procedure to enhance the quality of precipitation estimates on a composite grid. This is because local radar data are sometimes unavailable (with local I refer to observations from one radar station in native polar coordinates). However, the authors mainly use

local data in order to investigate the error structure of observed reflectivity. Then they transfer this error structure to the composite grid. But if local data was available, other correction procedures could and should be applied.
Then, the quality of the local radar observations can be used e.g. as a weighting criterion for composition (see e.g. Peura, 2010, but there are many more). I am aware that the PX product is only a qualitative product. But why not use the DX product then? I really have to say that the entire procedure appears unnecessarily twisted
ANSWER: We acknowledge the work of Peura (2010); here, we are interested in the mean precipitation patterns and their derivates (intense precipitation, hail, etc.). Our main problems within composite data are the differences between single radar areas and overlapping areas and the “altitude” effects. Usually, they are only visible on a longer temporal scale. So for our approach we also have to analyse these effects on a longer temporal scale.
A main aspect of the correction scheme is the altitude correction. The full coverage of one radar site is necessary, firstly, to deduce a reliable behavior of rain amounts/reflectivities with height, and secondly to detect clutter effects. Therefore, the additional use of single radar data is meaningful.
The DX-product is based on the same scan as the PX-product but with 256 reflectivity classes and polar coordinates. We are aware of the handicap imposed by the use of classified products; higher resolution data would certainly be desirable for detailed investigation, but our data base would be very scarce if more than six classes were analyzed. We chose to fit fewer classes based on safe statistics, and feel justified by the result: With this data base, we were able to estimate the average bias due to beam elevation, clutter, and beam shading. Additionally, the PX-product and RX-product are both projected in Cartesian coordinates. So the transfer of detected spokes and pixels affected by clutter from one product to the next is easier.
No question we can do better with better data. Undoubtedly, polar or even raw data from a single radar allow more efficient correction than any kludge of a post-composition correction, as the reviewer correctly states. It may well be argued why

there is no VRP correction, and why composition is not done in a way to avoid sharp transition in the overlapping area, but this is not the scope of our study.
As a matter of fact, the composite RX data is available for users and correspondingly being used in various applications; users in hydrology must be aware of limitations and pitfalls. As shown, the "physical" correction schemes that are routinely applied in DWD were not able to remove all of the errors. We do NOT state or recommend that the suggested procedure is an alternative to correction procedures that "could and should be applied" early in the processing chain. This indeed would seem unnecessarily twisted. But we see the necessity to correct existing data for errors that have not, or not completely, been handled earlier. To derive this type of correction, it is deemed appropriate to exploit all information that is still available (to us, but not to the end user), including single radar products such as the PX-product.

REFEREE#1: Very specific to the situation in Germany

Juggling with the RX/PX/DX product terms already implies that the entire procedure is highly specific to the data situation in Germany. This makes transferability to other "environments" virtually impossible. I think that this is a problem for an international journal such as HESS.
ANSWER: Yes, these products are specific for Germany; as our study deals with 16 radars in total and the product covers an entire country, we think the study is of high relevance. Also because other countries have single radar data or products, compositing products or secondary radar products with comparable problems. Basic radar problems are discussed and corrected, such as beam widening, increasing beam-height with distance from the radar site, Clutter effects or negative spokes. The terrain-following precip scan is indeed very specific for Germany. But the altitude correction also works for one scan with the same elevation angle such as at the radar site of Emden in the north-western part of Germany. Even for a CAPPI a comparable

correction can be deduced. But then the decrease of rainrate with distance from the radar-site will probably more depend on beam widening or attenuation effects than on altitude. The correction algorithm includes some static correction factors or maps which obviously have to be identified if used somewhere else. Our specific analysis and correction are meant to describe errors in the data under investigation and to prepare the way for further analysis (see below “usefulness”), not to correct future or foreign data sets.

REFEREE#1: Is the data up-to-date?

Why restrict the analysis to data from 2005-2009? I can understand that such studies and their publication necessarily imply substantial time lags, but six years appears very much. This raises another concern: As I heard, the German Weather Services now rountinely performs a "re-analysis" of radar data with the most up-to-date RADOLAN processing chain. I would expect that this also involves an update of RX data. Have the RX data in this study been produced with the latest, or at least a recent, RADOLAN re-analysis? The evaluation of "outdated" RX data would appear unfair. Speaking of RX: Why did the authors not use the RY product which, as far as I understand, involves a better quality control and is not produced by using the "Push" mechanism. Altogether, I have the strong impression the the data used in this analysis is not up-to-date.

ANSWER: The main reason for the analysis of the time-span 2005 to 2009 was to use a homogeneous data basis with minimal changes in scan strategy, availability of radar systems or maximum detected range. In 2010, the maximum range of single radar sites within the composite was extended from 128 km to 150 km (p. 1790, l. 22). This significantly influences the size and the amount of overlapping areas. Additionally, the radar systems were replaced and partly relocated in the following years with a subsequent modification in scan strategy and quality control, creating a break in data homogeneity. Compositing using a quality flag has been introduced by DWD well AFTER our data

time span, even the example cited (Peura, 2010) comes well after the time span of the data we investigated.

We do not consider our data to be outdated nor the analysis of "historic" data as unfair. The RADOLAN processing chain indicates that there is an interest in historic data, but as the reanalysis does not (fully) cover the above time span, and none of the products investigated here (RX and PX) will be reanalyzed at all, we see the need for additional corrections.

As already mentioned, the RX data is widely used and it is the basis for the CONRAD composite data. So it is not meaningful for us to analyse the RY-product.

REFEREE#1: Inconsistent analysis periods

Why is the analysis with the PX data conducted with data from 2000-2006? Why not use the same period as for the RX data analysis? Is this to guarentee independency? Why should it be indenpendent?

ANSWER: We extended the results of Wagner et al. (2012) in this study, covering the years 2000 through 2006. The RX was not yet available at that time. Consistent analysis periods might look nicer, but we see no real problem here. Our approach depends on safe statistics which is directly linked to the amount of measurements used as the longer timespan provides more measurements.

The PX-data and the calibration period for the RX-data show a time overlap of two years (2005 and 2006). A time overlap of the PX-data with the validation period is not recommended. So this period between 2000 and 2006 is a compromise between long-time series (safe statistics), temporal independency from the validation period (2007 and 2008) and temporal dependency on the calibration period (2005, 2006 and 2009).

Additionally, if the correction factors we derived for the time-span 2000 to 2006 are also applicable for the years 2005 to 2009 this supports the transfer to other periods, as long as the data is homogeneous.

REFEREE#1: Inhomogeneity

The authors do not present a solution to deal with inhomogeneities in the data time series (relocations of radar, changes in processing, changes in scanning strategy, changes in calibration, ...).

ANSWER: Indeed, there are several parameters that may cause additional inhomogeneities. The section p. 1790, ll. 14-24 addresses this issue. Static factors or static corrections have to be re-adjusted when important factors change. As indicated above ("Is the data up-to-date?"), we chose homogeneous time spans. Archived data of at least one year should be available to re-adjust these static correction factors or maps.

REFEREE#1: Usefulness

The verification (or "evaluation", as the authors put it) is not sufficient to demonstrate the usefulness of the suggested approach. This is mostly a matter of termporal scale: The authors compare the corrected and uncorrected radar-based precipitation estimates to rain gauge observations in order to compute the RMSE. This is done for annual precipitation depths. But is annual precipitation depth really a variable that I would want to quantify be using radar observations? I strongly doubt it. The authors argue with "climatological investigations" (see e.g. conclusions). In my opinion, though, looking at (mean) annual rainfall is not a priority in "radar climatology" at all.

Radar climatology is a highly topical research field that aims at identifying statistical properties of precipitation mostly at short duration (typically for convective storms), taking into account the spatial dimension (in additon to intensity and duration). This is also what the authors mention as their motivation in the introductory section. I think

that demonstrating an error reduction at an annual scale is not sufficient to justify the effort. This points to an overall weakness of the manuscript: it is not clear for which applications the correction procedures would actually be useful. Given that HESS aim at the field of hydrology, there should at least be an implication for which kind of hydrological analyses the product is recommended (as compared to e.g. an interpolated field of annual precipitation depths from rain gauges)

ANSWER: We agree with referee 1 that radar climatology often aims at identifying statistical properties of precipitation at short duration. But here we are interested in the total of precipitation, intense precipitation or hail patterns on a longer temporal scale. We believe that a statistical analysis of mean precipitation patterns is an equal part of a climatology besides the investigation of precipitation at short duration. A verification on a shorter temporal scale is not aimed at nor possible under this approach. The comparison of annual rain amounts with rain gauges is our focus.

The application of this correction scheme will be for patterns of different kinds of meteorological parameters derived from radar data on a longer temporal scale such as CONRAD analyses, as we already mentioned, or even for a map of mean annual rain amounts over Germany based on radar data in fig. 11.

Referee#1 is concerned that our study has only implicit relevance for hydrological applications. We are convinced that this study may help to make hydrologists aware of crucial limitations within radar composite data when used for a longer period.

REFEREE#1: Flawed verification

Apart from the fact the the evaluation could not really demonstrate the added value, I have the impression that the verification approach is methodologically flawed. First, the difference between the so-called "application period" and the "validation period" does not become clear. Second, and more important, it appears that rain gauges that had been used for the rain gauge adjustment were also used for the computation of *the RMSE. The authors confirm this assumption in ll. 3 of p. 1787. This approach makes the verification basically pointless, even if the authors did only use the median rainfall over the radar domain in order to derive adjustment factors. Verification of adjustment procedures requires rigorous cross-validation techniques. And the visual comparison of precipitation patterns in Fig. 11 does not proof the validity of the approach. The authors did not even validate the procedure for the interpolation of rain gauge observations (which would also require a cross-validation). Finally, the contribution of the single correction steps should be demonstrated: not only the "end product" should be verified, but the intermediate products along the procesing chain in order to show the adequacy of each step - in particular the correction for altitude and spokes (for clutter, it will be more difficult)*

ANSWER: The application period is based on data from 2005, 2006 and 2009. The static values and maps for the correction scheme are tuned for this period, such as the correction factors for the altitude correction (p. 1782, ll. 22-24) and the static map of mean allocation (p. 1785, l. 15; p. 1780, ll. 20-26). These static values and maps are adopted for the validation period of the years 2007 and 2008. We will clarify that further.

It is correct that we used a subset of rain gauges for both the adjustment of radar data and for the verification. The rain gauges for the adjustment of radar data only lead to mean adjustment factors for complete radar areas as the referee already mentioned. We want to demonstrate the agreement of patterns of radar data with observed point measurements for the greatest possible basis of pairs of values. But we agree with the referee that this is statistically not correct. We will clarify this and separate the rain gauges used for adjustment and those used for validation.

Fig. 11 shows a qualitative factor not a quantitative one as in our approach we are interested in mean patterns. So, we think this visual impression gives valuable information about the performance of the correction algorithm.

The contribution of single steps is difficult to realise. In Wagner et al. (2012), dealing with a single radar only (Munich), we presented figures for each step of the

correction scheme. But in this paper it is different. The most interesting part of the correction algorithm now is the correction and adjustment of the overlapping areas. The correction of spokes and the altitude correction within overlapping areas lead to several layers (p. 1784, l. 20 – p. 1785, ll. 14). These layers are combined within the next correction steps “internal and external adjustment”. This was the reason why we cannot show again the results of the individual correction modules step-by-step. It is possible to show each correction step for the single radar areas and also for the corrected composite data before the last step (correction of clutter) but in our opinion this is of minor interest for the reader.

REFEREE#1: Lack of comprehensibility

A serious problem that extends throughout the paper is that the applied methods are not adequately described. I reckon that this problem might be resolved. However, the presence of inadequate and fuzzy documentation is striking, and it is not acceptable. Just a few examples: p. 1773, ll. 12-13; p. 1776, ll. 18 ff.; p. 1782, ll. 21 ff. This shortcoming basically applies to the documentation of all the correction steps in the entire section 4.1. And in a similar way, it also holds for many of the figures (e.g. usage of inadequate color scales in figures 2, 4, 6, and 11; dotty plots in e.g. figures 3, 7, 8 where at least transparency would have been appropriate

ANSWER: Our correction algorithm is an extension and improvement of a correction algorithm presented in Wagner et al. (2012) for a single radar. Some methods/techniques were adopted from this study and are referred accordingly (e.g. p. 1773, l. 21; p. 1775, l. 6). Our approach is straightforward but not automated. For instance the separation of corrupted and uncorrupted pixels, discriminating thresholds or histograms, showing the frequencies of occurrence of radar reflectivities of a certain radar area are all done manually for best results

p. 1773, ll. 12-13: The separation of corrupted and uncorrupted pixels has already been described in Wagner et al. (2012). We will add one of the histograms mentioned with a short explanation here.

p. 1776, ll. 18 ff.: This again is a topic which has already been used and discussed in Wagner et al. (2012). We did not need an automatic algorithm to decide whether spokes still show meaningful precipitation patterns or not because of the reason already mentioned.

p. 1782, ll. 21 ff.: Admittedly, this passage about transferring and modifying the results from the single radar results to the composite is pretty short. The transfer is easy: first we calculate the altitude of each pixel within one single radar area based on the elevation angle of individual single radar images for the years 2005 to 2009 (the whole time-series were not available). If the elevation angles changed since 2006 or in between 2005 and 2009 the altitude of each pixel needs to be updated and therefore modified. The altitude correction factor derived from the analysis of the PX-data 2000 to 2006 is then applied to perform the altitude correction.

We will clarify the documentation of the applied methods and figures for the assumed flaws named by the reviewer. We are unsure as to the “dotty” appearance of the plots. The original figures have high quality (eps). We did not recognize this lack of quality after the production step of the paper.

We cannot follow the reviewer's remarks about the “color scales” and the “transparency”. The scale is linear only for the mean classes but covers the whole range of precipitation amounts. We kindly ask the reviewer for some more information.

Please find attached as a supplement a suggested flowchart showing the entire correction scheme which should help to understand the correction steps more easily. We would present it in Chapt. 4.1.1.

Literature

Peura, M.: The living composite, Proceedings of the Sixth European Conference on Radar in Meteorology and Hydrology (ERAD2010), Vol. 1 (Advances in Radar Applications), pp. 350–354, 2010.

Wagner, A., Seltmann, J., and Kunstmann, H.: Joint statistical correction of clutters, spokes and beam height for a radar derived precipitation climatology in southern Germany. Hydrol. Earth Syst. Sci., 16, 4101–4117, 2012.