Author response to anonymous referee #1

HESS-2021-105

Ruben Imhoff

Ruben.Imhoff@deltares.nl

May 21, 2021

Dear reviewer,

We would like to thank you for your interest in our work and the enthusiastic reaction to our manuscript. With four constructive and elaborate reviews, we think we are very well served by our reviewers. Your comments have been valuable and have helped us to improve the manuscript.

Below, we give a response to the given suggestions. We have placed the reviewer's comments in black font and below that, our response in blue font for clarity.

Sincerely,

Ruben Imhoff, Claudia Brauer, Klaas-Jan van Heeringen, Hidde Leijnse, Aart Overeem, Albrecht Weerts and Remko Uijlenhoet

General comments

1) Do you think, MFB could be improved by dividing the NL into spatial segments (like the classical moving window) or even depending on the distance to the radar, or is the density of automatic stations to low? (The argument that MFB is limited to one factor for a whole country does not hold in general.)

   Thanks for suggesting this. We have tried this before and it generally gave better results than with the country-wide MFB-adjustment factor. The question then remains how large these regions should be. With smaller regions, the local MFB factor is better able to correct for the local spatial (and distance-dependent) errors. That possibility of doing that depends, as the reviewer also indicated, on the local density of automatic gauges. Knowing that it frequently occurs that one or multiple gauges give no or unreliable values, the regions should not become too small in order to have sufficient gauges within the region. If the regions contain a low number of rain gauges, there will be the risk that none of the gauges catch rainfall (especially during convective events), and thus no adjustment factor will be derived. With a country-wide MFB adjustment factor, this probability is lower, but that comes at the price of not at all taking into account the spatial errors in the radar QPE.

   We think that this option is no matter what worth mentioning, so we propose to briefly mention it in the discussion section, where we will add it to the lines 243 – 245, as: “MFB adjustment of radar rainfall fields is still the most frequently applied adjustment method (Holleman, 2007; Harrison et al., 2009; Thorndahl et al., 2014; Goudenhoofdt and Delobbe, 2016). The results indicate that this choice may be reconsidered, at least for the Netherlands and in case a country-wide or large-region adjustment factor is applied. More regionalized MFB adjustments are possible, but depend on the density and availability of the automatic gauge stations.”

2) Does the RA adjustment eliminate spatial dependencies of the resulting QPE? To be more precise: Is the quality of RA depending on the distance to the radar site? And if so - would CARROTS allow for the derivation of an improved adjustment procedure?

   Given the dense manual rain gauge network, 1 gauge per 100 km^2, used in the spatial adjustment on a daily basis, we expect spatial dependencies of the resulting QPE in R_A are relatively small, especially on a daily basis. In earlier work (Overeem et al., 2009a), we noticed underestimation of extreme rainfall for short durations, which can be attributed largely to remaining errors in radar data. The combination of a daily spatial adjustment and an hourly mean-field bias adjustment is probably not sufficient to remove these errors completely on subdaily timescales. Rain-induced attenuation typically results in underestimation of heavy rainfall at longer range from the radar, but not systematically at the same locations. Changes in the vertical profile of reflectivity, which can result in systematically lower rainfall estimates at further range from the radar, are likely easier to adjust for using rain gauge data. It is typically associated with longer lasting stratiform events for which the daily spatial adjustment factor is more suitable compared to more localized short-duration convective events. Overall, the quality of R_A could definitely be improved for sub-daily rainfall, but large systematic differences in space are not expected.

3) What’s the effect of temporally present spokes (positive and negative) in the historical data set?

   That is very minor, mostly due to the 10-yr mean and the moving window of a month. We hope that the presented sensitivity analysis gives an indication of the stability of the factor.

Nevertheless, we have also tested the effect of including 2008 in the factor derivation. This year gave an odd result in the KNMI products, as also described in Sec. 2.1: “The year 2008 is actually the first year in the KNMI archive of both data sets, but it was left out of the analysis here. R_U for this year showed a significantly different behaviour than the other years, especially during the first half year in which the product rarely underestimated and frequently even overestimated the rainfall sums. The reason for this behaviour is not yet fully understood. KNMI (2009) reported that spring was exceptionally dry in the north of the country and that the months January and May were among the warmest on record. On some days with overestimations, clear bright band effects were visible in the radar mosaic, which may have contributed to the systematic differences.” This significantly different first half year does impact the factor derivation and resulted in somewhat lower correction factors for the first six months. Although the effect is not a major one, we could observe it in the results, while similar effects are not present for e.g. the (extreme) dry year 2018.

4) I propose to add a figure showing the pixel-based differences between MFB and RA as well as RC and RA on a 5-min-basis, e.g. box-whisker (NL mean or catchments) and map with median and percentiles. This would help understanding the effects on discharges in different regions.

Thanks for this suggestion. We do agree that the QPE validation can be more elaborate. Instead of the reviewer’s suggestion, we propose to make two different changes: (1) a change to Fig. 5 including the absolute error with the reference for all catchments. See below (our response to line 178) for the renewed figure and corresponding text in the results. (2) A scatter, similar to Fig. 2, showing the performance of both R_{MFB} and R_C, and a table showing the Fractions Standard Error (FSE) based on the hourly rainfall sums of all QPE products and the reference for the land surface area of the Netherlands. We will show this per year and season to give an indication of the (standardized) seasonal and annual variability in the rainfall estimation error. We refer to our response to reviewer #4 for the specifics.

5) What is the performance in heavy rain situations? Despite mean numbers, please comment on the effects.

We agree that we can say a bit more about heavy rain situations. We propose to change “As mentioned in the previous paragraph, the climatological adjustment factor is not calculated for the current meteorological conditions and resulting QPE errors, which could lead to considerable errors during extreme events. Nonetheless, this is also the case for the MFB adjustment technique (Schleiss et al., 2020). The absolute errors for the 10 highest daily sums in this study for the Aa and Hupsel Brook catchments (one of the largest and the smallest catchment in the study) are similar for the MFB and climatological adjustment methods, with on average a 20% difference with the reference (this would have been 50–60 % without corrections). Note that for individual events in these twenty extremes, the errors can still reach 48% for the QPE adjusted with CARROTS and 64% for the MFB adjusted QPE.” into:

“As mentioned in the previous paragraph, the climatological adjustment factor is not calculated for the current meteorological conditions and resulting QPE errors, which could lead to considerable errors during extreme events. Nonetheless, this is also the case for the MFB adjustment technique (Schleiss et al., 2020). The absolute errors for the 10 highest daily sums based on the reference in this study for the Aa and Hupsel Brook catchments (one of the largest and the smallest catchment in the study) are similar for the MFB and climatological adjustment methods, with on average a 20% difference with the reference (this would have been 50–60 % without corrections). In most of these events, both R_C and R_{MFB} underestimated the true rainfall amount. However, for a small number of these top 10 events, the QPE products overestimated the true rainfall amount. This occurred more frequently with CARROTS (25% of the cases) than with the MFB adjustment (15% of the cases). Note that for individual events in these twenty extremes, the errors can still reach 48% for the QPE adjusted with CARROTS and 64% for the MFB adjusted QPE.”.

6) Could a dbz-dependent factor improve CARROTS?

That is an interesting idea! We can imagine that especially for extreme (summer) rainfall events, a dBZ-dependent factor can significantly improve CARROTS, because in these situations the QPE error is generally higher (Schleiss et al., 2020), while the CARROTS factor is based on the average error for that time of the year, which is partly (mostly even) based on less extreme events. The remaining question is then whether the 10-year dataset contains enough (relatively) extreme events for a stable and reliable derivation of a dBZ-dependent factor for high intensity rainfall. There will be sufficient lower-intensity stratiform events, but we expect that the current seasonality in the factor already gives a reasonably good correction for the errors for such (winter) events.

We propose to add a sentence at the end of the paragraph at lines 265 – 271: “A way to better correct for biases during extreme events could be to derive either different Z-R relationships, depending on the type of rainfall, or dBZ-dependent correction factors, which could be derived in a similar way to the CARROTS derivation method. Whether this works or not for extreme events depends on the number of extreme events in the available historical dataset.”

Specific comments and technical corrections

Lines 62-63: I would expect the need to adjust the analysis BEFORE spatially dislocate the reflectivities to avoid adjustment with climate correction factors that are not specific to the original measurement location. Please clarify, why a post-processing should be preferable. I don't see any advantage in that.

Thanks for this remark. This statement was based on the previous paragraph where we stated: “when the adjustment method changes the spatial structure of the original radar rainfall fields (kriging and Bayesian methods), this may impact the continuity of the rainfall fields over time and thereby also the radar rainfall nowcasts (Ochoa-Rodriguez et al., 2013; Na and Yoo, 2018).” However, for MFB adjustments and CARROTS you are absolutely right. Hence, we propose to change “(2) is available in real time so that it can be used operationally for postprocessing of radar-based rainfall forecasts, such as nowcasting” into “(2) is available in real time so that it can be used operationally for radar-based rainfall forecasts, such as nowcasting”.

Line 87 - “distance-weighted interpolation”: Please comment a bit more detailed.

We agree that we can elaborate a bit more on how this method was applied. The original description can be found in Overeem et al. (2009b). We have tried to concisely describe this procedure and we propose to change lines 85 – 90 to:

“The same 31 automatic rain gauges are used for the MFB adjustment method, which will be introduced in Sec. 2.2.1. In contrast to the spatially uniform hourly MFB adjustment, the observations from the manual rain gauges are used for daily spatial adjustments, based on distance-weighted interpolation of these observations (Barnes, 1964; Overeem et al., 2009b). A spatial adjustment factor is derived per grid cell as follows (for a more elaborate description, see Sec. 3 in Overeem et al., 2009b):

$$F_S(i,j) = \frac{\sum_{n=1}^{N} w_n(i,j)*G(i_n j_n)}{\sum_{n=1}^{N} w_n(i,j)*R_U(i_n j_n)},$$

with *N* the number of radar-gauge pairs, $G(i_n j_n)$ the daily rainfall sum for manual rain gauge n at location (*i_n*, *j_n*) and $R_U(i_n j_n)$ the unadjusted daily rainfall sum for the corresponding radar grid cell. $w_n(i,j)$ is a weight for gauge *n*, based on the following function:

$$w_n(i,j) = e^{-\frac{d_n^2(i,j)}{\sigma^2}} .$$

Here, $d_n^2(i,j)$ is the squared distance between gauge *n* and the grid cell for which the factor is derived. σ determines the smoothness of the adjustment factor field. It was set to 12 km by Overeem et al. (2009b), based on the average gauge spacing in the Netherlands.

Finally, to spatially adjust the hourly MFB-adjusted rainfall fields, two more steps are followed. First, the hourly MFB-adjusted rainfall fields (see Sec. 2.2.1 for the MFB adjustment method) are accumulated to day sums. For each grid cell, a new adjustment field is then determined:

$$F_{MFBS}(i,j) = \frac{R_S(i,j)}{R_{MFB}(i,j)},$$

with $R_S(i,j)$ the spatially-adjusted day sum for grid cell (*i,j*) and $MFB(i,j)$ the MFB-adjusted day sum for grid cell (*i,j*). Second, the 1-h or higher frequency (5-min in this study) MFB-adjusted rainfall fields are multiplied with adjustment factor $F_{MFBS}(i,j)$.

This product is considered as a reference rainfall product in the Netherlands and it is therefore also regarded as reference here (referred to as R_A in this study). The R_A data is not available in real time (available with a delay of one to two months, because it only uses quality-controlled and validated rain gauge observations), but it is archived and can therefore be used for ‘offline’ methods. Both RA and RU have a 1-km2 spatial and 5-min temporal resolution.”

Line 93 - “even”: Why “even”?

Normally the unadjusted QPE (R_U) underestimates the true rainfall amount. For the first half year of 2008 this was not the case. In fact, the QPE (even) overestimates the true rainfall amount frequently.

Line 99 - “with”: by

Thanks, we will change this.

Line 138 - “Delft-FEWS system”: ?

Reviewer #2 also asked to give a little more information about this system. We propose to change the sentence “Most of the involved water authorities use these (lowland) rainfall-runoff models either operationally or for research purposes, often embedded in a Delft-FEWS system (Werner et al., 2013).” into “Most of the involved water authorities use these (lowland) rainfall-runoff models either operationally or for research purposes, often embedded in a Delft-FEWS system, which is a data-integration platform, used world-wide by many hydrological forecasting agencies and water management organizations, that brings data handling and model integration together for operational forecasting (Werner et al., 2013).”

Line 139 - “For this reason, most models were already calibrated”: calibrated to what input data? Does this have an impact on the results?

This was mentioned by multiple reviewers, thanks for pointing this out. We agree that we should better clarify this procedure. Most models, except for the catchments Roggelsebeek and Dwarsdiep, were already calibrated and are part of the operational systems of the involved water authorities. Calibration took place, in most cases, with local rain gauge data for a short period of one to a couple of years. The actual calibrations of the systems generally took place not that long ago – it is different per catchment - and uses a subset of the time period used in this study (2009 – 2018). The catchments Roggelsebeek and Dwarsdiep were calibrated with the reference data (R_A) for the periods 2013 – 2014 (Roggelsebeek) and 2016 – 2017 (Dwarsdiep). The choice for these periods was based on discharge observation availability and quality.

In the validation procedure, we are using the model runs with the reference data (R_A) as ‘observation’. Hence, in any case, this validation setup will favor the model runs that are fed by QPE products that are closer to the reference rainfall product.

We propose to change “For this reason, most models were already calibrated (e.g. Brauer et al., 2014b; Sun et al., 2020).” into “For this reason, most models were already calibrated using interpolated rain gauge data (e.g. Brauer et al., 2014b; Sun et al., 2020). The calibration period was based on the availability and quality of discharge observations for that basin, but it was generally one to two years within the period considered in this study (2009 – 2018). The WALRUS models for catchments Roggelsebeek and Dwarsdiep were not calibrated prior to this study and were therefore calibrated with the reference data (R_A) for the periods 2013 – 2014 (Roggelsebeek) and 2016 – 2017 (Dwarsdiep). The choice for these periods was based on discharge observation availability and quality.”

Line 144 - “Kling-Gupta Efficiency (KGE)”: Please explain briefly the idea of the score. What is a good score - the higher the better?

This was also mentioned by multiple reviewers. Again, thanks for pointing this out. In our attempt to keep the text as brief as possible, we have overlooked the need to introduce the KGE metric more elaborately. We plan to elaborate the sentence with: “The resulting discharge simulations were validated for the same period and 5-min timestep using the Kling-Gupta Efficiency (KGE) metric (Gupta et al., 2009):”

$$“KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2},$$

$$\alpha = \frac{\sigma_S}{\sigma_O},$$

$$\beta = \frac{\mu_S}{\mu_O},$$

with r the linear correlation between observed and simulated discharge, α the flow variability error between observed and simulated discharge and β the bias factor between mean simulated (μ_s) and mean observed (μ_o) discharge. σ_s and σ_o are the standard deviation in the simulated and observed discharge. The KGE metric ranges from -∞ to 1.0, with 1.0 being a perfect agreement between observations and simulations. In this study, the discharge simulated with R_A as input was regarded as the observation."

Line 152 - "has one of the highest biases": In the discharge or the QPE itself?

Both actually, but the bias in the discharge simulations is a result of the bias in the QPE.

Line 139 – title: Better: Seasonal and spatial variability

Good suggestion, we will change the title to Seasonal and spatial variability.

Line 162 – "south and east": South and East (please correct all appearances in the text)

Thanks for mentioning this. We looked this up and it should only be capitalized when it is part of the name (e.g. south Africa), but for wind directions it should not be capitalized.

Line 171 – degree symbol: Replace by K (Kelvin)

We will do so, Kelvin is indeed better.

Line 175 – 120 km: What is the range of the radars? The figure suggests 100 km?

The range is longer than the indicated 100 km. In the radar domain, a maximum range of 200 km around the radar in De Bilt was used (hence, that is what the domain is based on), so more than the entire land surface of the Netherlands is covered by the radar domain. Note that the maximum range of each radar is 320 km. However, in literature the 100 km range is often used as an indication of the maximum distance up to where the QPE is expected to be reliable. We decided to follow this 'standard' with the hope that it would not confuse the reader.

To avoid confusion for the reader, we propose to extend the sentence "The two grey circles indicate a range of 100 km around the radars in Den Helder (DH) and Herwijnen (H)." in the figure caption with: "The two grey circles indicate a range of 100 km around the radars in Den Helder (DH) and Herwijnen (H). Note that the used range in the composite was more than 100 km, but 100 km is often regarded as the distance up to where the radar QPE is expected to be reliable."

Line 177 – dependence: Change to 'dependency'.

We will do so, thanks.

Line 178 – Section 3.2: The text seems to refer to mean values that are not presented in Fig. 5. Please add the mean values (including the spread) to the Figure and/or clearly indicate, what you're referring to.

Thanks for pointing this out, this was indeed missing. Also in line with the suggestions of reviewer #2, we decided to adjust the figure and add another subfigure with the annual mean absolute error between the QPE product and the reference (R_A) per catchment. This changes the figure caption as well as the text in Sec. 3.2. The proposed changes are (in addition, note that we have added more to this section after the comments of reviewer #4. We refer to our responses to this reviewer for the full adjustment and added text to this section):

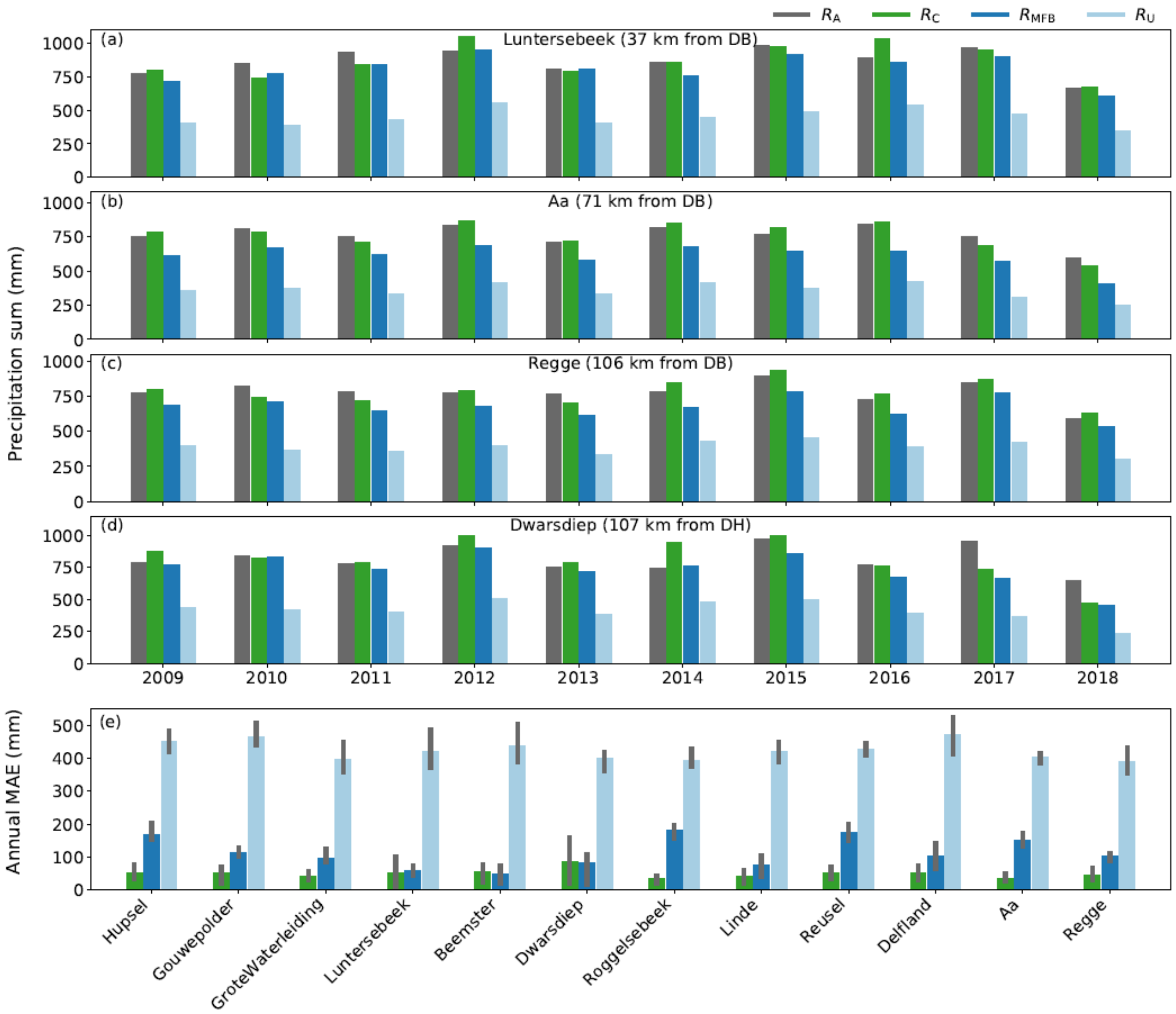

Figure 6. Effect of the adjustment factors on the catchment-averaged annual rainfall sums. (a – d) The results for a sample of four catchments that are spread over the country (and thus the radar domain): (a) Luntersebeek, (b) Aa, (c) Regge and (d) Dwarsdiep. Shown are R_A (grey), the estimated rainfall sum after correction with the CARROTS factors (R_C; green), the estimated rainfall sum after correction with the MFB adjustment factors (R_{MFB}; dark blue) and the rainfall sum with the unadjusted radar rainfall estimates (R_U; light blue). The distance between the catchment center and the closest radar in the domain is given in the title of subfigures a -d (DH is Den Helder and DB is De Bilt). The radar in Herwijnen, which replaced the radar in De Bilt in 2017, is not included here, because this radar was operational for the shortest time in this analysis. (e) the mean absolute error of the annual precipitation sum between the QPE products and the reference rainfall sum (R_A). The vertical grey lines, per bar, indicate the IQR of the MAE based on the ten years.

3.2 Annual rainfall sums

An advantage of the MFB adjustment is that it corrects for the circumstances during that specific day and thus also for instances with overestimations (Fig. 4a). On a country-wide level, this is clearly advantageous, also compared to CARROTS (Fig. 5). The negative effect of the spatial uniformity of the factor, however, becomes apparent in Fig. 6, which compares the annual precipitation sums of the two adjusted radar rainfall products with the reference and R_U for the twelve basins. For all basins, both adjusted products manage to significantly increase the QPE towards the reference. However, for nine out of twelve basins, R_C outperforms R_{MFB} (Fig. 6e). Exceptions are Beemster, Luntersebeek and Dwarsdiep, where the performance of both products is not that different.

The MFB adjusted QPE performs better for the Beemster polder, Dwarsdiep polder (Fig. 6d) and Luntersebeek catchment (Fig. 6a) due to their location in the radar mosaic. The Luntersebeek catchment (central Netherlands, Fig. 1) is located closer to both radars. There, R_{MFB} generally performs better and sometimes even overestimates the true rainfall, which is consistent with Holleman (2007). The performance of R_{MFB} for the Dwarsdiep catchment is similar as its performance for the Linde catchment (both in the North of the country), but R_C shows more variability in the error from year to year for the Dwarsdiep catchment (Fig. 6d), leading to a better relative performance of R_{MFB}. The CARROTS QPE tends to overestimate the rainfall amount of the three aforementioned basins (Beemster, Dwarsdiep and Luntersebeek) for some years (e.g. with 16% for the Luntersebeek in 2016). Overall, the performance of R_C and R_{MFB} are not that different for these three basins, with on average just a lower MAE for R_{MFB} than for R_C for the polders Beemster and Dwarsdiep (Fig. 6e).

Summarizing, the CARROTS factors have a clear annual cycle, with generally higher adjustment factors further away from the radars (Sec. 3.1). On average for the Netherlands, the MFB adjusted QPE outperforms the CARROTS correct QPE. However, the spatial variability in the CARROTS factors, in contrast to the uniform MFB adjustment, results in estimated annual rainfall sums for the twelve hydrological basins that are generally closer to the reference (for nine out of twelve basins) than with the MFB adjusted QPE, especially for the east and south of the country. This effect is expected to become more pronounced when the adjusted QPE products are used for discharge simulations.

Line 187 – "better for the Dwarsdiep polder (10% underestimation) and Luntersebeek catchment (6% underestimation)": This does not hold for all the years. Does it refer to the mean?

This indeed referred to the mean. After the changes to Fig. 5 (see above), the mean is shown including the spread. We propose to change this line to: "better for the Dwarsdiep polder (on average 10% underestimation) and Luntersebeek catchment (on average 6% underestimation)".

Line 196 – "for regions close to the edges of the radar domain": not true for Northern NL.

That is indeed not true for northern NL, we will change the sentence to: "This results in estimated annual rainfall sums that are closer to the reference than with the MFB adjusted QPE for the east and south of the country."

Line 213 – "The CARROTS QPE outperforms R_{MFB}, when this product is used as input for the twelve rainfall-runoff models": How can it be, when QPE is worse? Better in day-to-day corrections? Please give more detailed explanation on this! In addition, the sentence is not correct as stated in the next sentences that it does not hold for Beemster. Please be more precise.

With the renewed figure 5, it becomes clear that the CARROTS QPE outperforms the MFB-adjusted QPE for most catchments. We hope that this gives part of the explanation. With regard to the sentence about the Beemster, that is correct. Reviewer 2 also suggested to change this. We propose to change the sentence as follows: "The exception to this is the Beemster polder. The Beemster is mostly upward seepage driven leading to a predictable baseflow for all models runs. In addition, the model is located close to an automatic weather station and is located in between both operational radars, which makes the MFB adjustment more beneficial for this region. The difference in performance between the hydrological model simulations is small, with a KGE of 0.92 (using R_C) versus 0.96 for R_{MFB}, as compared to the reference run."

Line 234 – "generally outperformed": Be more specific - this sounds too positive to me regarding the annual variability for the annual sum at two of the shown catchments.

We propose to change the sentence into: “The method and resulting QPE product outperformed the mean field bias (MFB) adjustment, that is used operationally in the Netherlands, for catchments in the east and south of the country. When the QPE products were used as input for hydrological model runs, the method outperformed the MFB adjustment method for all but one basin.”

Line 236 – “The main difference with the MFB adjustments”: to?

We meant the main difference with the CARROTS method. We propose to change it to: “The main difference that distinguishes the CARROTS method from the MFB adjustment is the presence of a high-density network of (manual) rain gauges in the reference dataset, a dataset that is not available in real-time.”

Lines 236 – 242: To my opinion, also the timeliness of the method makes it very valuable for a first guess in real-time that should be available much earlier than products depending on gauge data.

This is a good remark and we think something that is worth mentioning at the end of this paragraph. We propose to add the following to the end of this paragraph: “An additional advantage of the method is the real-time availability of the correction factors, which is independent of the timeliness of the rain gauge data.”.

Lines 243 – 247: To my knowledge, the MFB is also applied for limited areas depending on the number of gauges. Please comment on this.

That is true, see also our response to the first general comment. We are planning to add the following sentence to the end of the paragraph: “Note, however, that the MFB adjustment is uniformly applied over the entire country in the Netherlands. It is possible to apply a factor per region when the region contains enough gauges. This could partly address the spatial errors in the radar QPE.”.

Lines 248 – 251: What is the experience with De Bilt?

Good point, that is something we can comment on. As stated in Sec 2.1, “Between September 2016 and January 2017, both radars were replaced by dual-polarization radars and the radar in De Bilt (‘DB’ in Fig. 1) was replaced by a new one in Herwijnen (‘H’ in Fig. 1). The radar renewals and relocation have had a limited impact on the QPE product, mainly because the operational products are not yet (fully) using the additional information from the dual-polarization (Beekhuis and Holleman, 2008; Beekhuis and Mathijssen, 2018).” Hence, a change like this has had a minor impact on the CARROTS derivation, given the historical dataset of 10 years it is based on. We do expect that when the dual-polarization potential is fully used or when an extra radar is added to the composite in e.g. the south or east (where the biases are generally highest), this no longer holds and the factors need to be recalculated using an archive based on the new situation.

We propose to add a few sentences to this paragraph in the discussion section: “However, the proposed CARROTS method has to be recalculated for every change in the radar setup, calibration, additional post-processing steps (e.g. VPR corrections, Hazenberg et al., 2013) or final composite generation algorithm. For instance, including a new radar in the composite would require a recalculation of the adjustment factors, thereby assuming the presence of an archive of the new composite product. This could potentially limit the usefulness of the proposed method. As mentioned in Sec. 2.1, the replacement of both Dutch radars by dual-polarization radars in combination with the replacement of the radar at location De Bilt to location Herwijnen (Fig. 1) between September 2016 and January 2017 has only had a limited impact on the operational products, and thereby on the CARROTS derivation. The operational products are not yet (fully) making use of the dual-polarization potential. We expect that the factors will have to be recalculated when the additional information from the dual-polarization radars is used to improve the products or when e.g. the German and Belgian radars close the Dutch border are added to the composite.

That CARROTS is relatively insensitive to such minor changes in the composite or the year-to-year variability of rainfall, is likely a result of the ten-year archive that has been used. The sensitivity analysis in Sec. 3.4 has shown that leaving individual years out of the archive hardly influences the CARROTS factors. Nevertheless, based on the current analysis we cannot conclude what the minimum number of years in the archive has to be to obtain stable CARROTS factors that are similar to the factors derived in this study. This is a recommendation for future research as such a sensitivity analysis can point out what the minimum archive length has to be to calculate the CARROTS factors. In the case of a new radar QPE product, it is also recommended to recalculate the archive (if possible), to make sure new CARROTS factors can be derived."

Line 257 – "computationally expensive": I don't think that this is the major limitation in real-time adjustment. Computational efficient methods with spatially non-uniform factors are also common. Please change your statements from 'black-and-white' and allow for 'grey'.

We agree. Instead of this statement, we think it is worth referring back to your earlier statement about the timeliness of the CARROTS method, which makes it independent of the arrival time of the rain gauge observations in real time.

We propose to change the sentences: "A disadvantage of geostatistical and Bayesian merging methods is that they are computationally expensive and require the real-time availability of a dense network of rain gauges. Instead, we consider the proposed climatological radar rainfall adjustment method as a benchmark for the development and testing of operational radar QPE adjustment techniques." to "A possible disadvantage of these real-time methods (MFB, geostatistical and Bayesian merging) is the dependency on the timely availability of rain gauge data, which is not the case for CARROTS. Altogether, we consider the proposed climatological radar rainfall adjustment method as a benchmark for the development and testing of operational radar QPE adjustment techniques."

Line 262 – "in time": What do you mean?

This is indeed an unnecessary addition to the sentence. We propose to remove these words from the sentence, as the meaning of the sentence remains the same without it.

Line 295 – "factors": annual sums?

Indeed, we did mean the annual sums obtained after correcting the QPE with the CARROTS factors. Suggested change: "This bias is almost absent for the CARROTS factors" to "This bias is almost absent for the annual rainfall sums after correction with the CARROTS factors".

Line 300 – "for hydrological applications": Does this hold for extreme events? It is not stated very precisely in the next sentences. What about overforecasting in non-extreme situations?

Thanks for mentioning this, because it indeed can depend on the type of event taking place. We propose to change the focus to reconsidering the use of MFB adjustments (in this way) operationally in the Netherlands. So, the sentence would change from "For the Netherlands, these results indicate that the operationally used MFB adjustment performs worse than the proposed climatological adjustment factor for hydrological applications." to "For hydrological applications in the Netherlands, these results indicate that the current operational use of a country-wide MFB adjustment may be reconsidered as it often performs worse than the proposed climatological adjustment factor, which can be seen as the minimum benchmark to outperform."

Figure 5: Would be interesting to see the sum or mean over all the years (including the spread).

Good suggestion. See our response to Line 178 where we address this.

References

Barnes, S. L.: A technique for maximizing details in numerical weather map analysis, Journal of Applied Meteorology, 3, 396–409, 1964.

Beekhuis, H. and Holleman, I.: From pulse to product, highlights of the digital-IF upgrade of the Dutch national radar network, in: Proceedings of the Fifth European Conference on Radar in Meteorology and Hydrology (ERAD 2008), Helsinki, Finland, https://cdn.knmi.nl/system/data_center_publications/files/000/068/061/original/erad2008drup_0120.pdf?1495621011, 2008.

Beekhuis, H. and Mathijssen, T.: From pulse to product, Highlights of the upgrade project of the Dutch national weather radar network, in: 10th European Conference on Radar in Meteorology and Hydrology (ERAD 2018) : 1-6 July 2018, Ede-Wageningen, The Netherlands, edited by de Vos, L., Leijnse, H., and Uijlenhoet, R., pp. 960–965, Wageningen University & Research, Wageningen, the Netherlands, https://doi.org/10.18174/454537, 2018.

Brauer, C. C., Torfs, P. J. J. F., Teuling, A. J., and Uijlenhoet, R.: TheWageningen Lowland Runoff Simulator (WALRUS): application to the Hupsel Brook catchment and the Cabauw polder, Hydrology and Earth System Sciences, 18, 4007–4028, https://doi.org/10.5194/hess-18-4007-2014, 2014b.

Goudenhoofdt, E. and Delobbe, L.: Generation and verification of rainfall estimates from 10-Yr volumetric weather radar measurements, Journal of Hydrometeorology, 17, 1223–1242, https://doi.org/10.1175/JHM-D-15-0166.1, 2016.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Harrison, D. L., Scovell, R.W., and Kitchen, M.: High-resolution precipitation estimates for hydrological uses, Proceedings of the Institution of Civil Engineers - Water Management, 162, 125–135, https://doi.org/10.1680/wama.2009.162.2.125, 2009.

Holleman, I.: Bias adjustment and long-term verification of radar-based precipitation estimates, Meteorological Applications, 14, 195–203, https://doi.org/10.1002/met.22, 2007.

KNMI: KNMI - Jaar 2008: Twaalfde warme jaar op rij, https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2008/jaar, 2009.

Na, W. and Yoo, C.: A bias correction method for rainfall forecasts using backward storm tracking, Water, 10, 1728, https://doi.org/10.3390/w10121728, 2018.

Ochoa-Rodriguez, S., Rico-Ramirez, M., Jewell, S. A., Schellart, A. N. A., Wang, L., Onof, C., and Maksimovic, v.: Improving rainfall nowcasting and urban runoff forecasting through dynamic radar-raingauge rainfall adjustment, in: 7th International Conference on Sewer Processes & Networks, http://spiral.imperial.ac.uk/handle/10044/1/14662, 2013.

Overeem, A., Buishand, T. A., and Holleman, I.: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar, Water Resources Research, 45, W10424, https://doi.org/10.1029/2009WR007869, 2009a.

Overeem, A., Holleman, I., and Buishand, A.: Derivation of a 10-year radar-based climatology of rainfall, Journal of Applied Meteorology and Climatology, 48, 1448–1463, https://doi.org/10.1175/2009JAMC1954.1, 2009b.

Schleiss, M., Olsson, J., Berg, P., Niemi, T., Kokkonen, T., Thorndahl, S., Nielsen, R., Ellerbæk Nielsen, J., Bozhinova, D., and Pulkkinen, S.: The accuracy of weather radar in heavy rain: a comparative study for Denmark, the Netherlands, Finland and Sweden, Hydrology and Earth System Sciences, 24, 3157–3188, https://doi.org/10.5194/hess-24-3157-2020, 2020.

Sun, Y., Bao, W., Valk, K., Brauer, C. C., Sumihar, J., and Weerts, A. H.: Improving forecast skill of lowland hydrological models using ensemble kalman filter and unscented kalman filter, Water Resources Research, 56, e2020WR027 468, https://doi.org/10.1029/2020WR027468, 2020.

Thorndahl, S., Nielsen, J. E., and Rasmussen, M. R.: Bias adjustment and advection interpolation of long-term high resolution radar rainfall series, Journal of Hydrology, 508, 214–226, https://doi.org/10.1016/j.jhydrol.2013.10.056, 2014.

Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., and Heynert, K.: The Delft-FEWS flow forecasting system, Environmental Modelling & Software, 40, 65–77, https://doi.org/10.1016/j.envsoft.2012.07.010, 2013.