

Dear Ryan Teuling, dear Editor,

We thank you for continuing the review process of our manuscript. Please find below the list of the points raised by the two reviewers (*italicized*) together with the answers (in normal font). We sincerely hope that the revisions meet your expectations.

Kind regards,

Michel Journée  
Edouard Goudenhoofdt  
Stéphane Vannitsem  
Laurent Delobbe

## Responses to RC1

### General comments

#### *General comment 1: Rainfall product description*

The description of the method has been improved. It is important to note that the methods cannot be described in full details due to their complexity. However, a new appendix has been added in the revised version where radar experts will find more information on the significant improvements regarding the treatment of static non-meteorological echoes.

#### *General comment 2: Description of the methods*

We carefully considered the possibility to move all the methods in a dedicated section. We think, however, that our paper does not fit well on the typical "Data-Methods-Results" structure as we do not really propose any methodological novelty. Regarding the data analysis section (Section 3), most methodological descriptions are furthermore short (except for the NMF, but the background information about NMF is now moved to the Appendix) and most methods are standard/generic (e.g., inverse distance weighted interpolation, cross validation). We therefore prefer to keep the methodological descriptions close to the discussion of their results. We are also convinced that going to a typical "Data-Methods-Results" structure will impair the readability of our manuscript.

#### *General comment 3: Application of areal averages (section 3.4)*

The Section 3.4 has been complemented with an analysis to illustrate how the difference between the IDW of rain gauge data and the radar product evolves when they are aggregated in space and time. For this, we derived precipitation values for 67 catchments of various sizes and for accumulation durations from 1 to 72 hours based on the IDW and RADFLOOD21 datasets. Figure 14 shows in terms of MAE how the discrepancies between both datasets decrease when the rainfall data are aggregated in space over larger domains and accumulated over longer periods.

In addition, a figure with the cumulative precipitation sums derived from the timeseries in Figure 12 has been added (see Figure D4).

Regarding the suggestion to apply an hydrological model for simulating discharges and analyzing the results making use of the two rainfall datasets, we agree that it would be a very

interesting extension of the present study. However, we consider that it is outside the scope of this paper which is really focused on documenting rainfall. Furthermore, such additional study is clearly outside our field of expertise.

## Specific comments

*Lines 27 – 29: What does this exactly mean for hydrology and/or the flooding that took place?* These measurements bring additional information on the event and its dynamics and complexity. The following sentences have been added in the introduction of the revised version: "The increase of the seismic noise during the event is induced by the rising stream turbulence, sediment and debris transport. On 15 July at 00:15 UTC a sudden increase of the seismic noise coincides with a detailed testimony reporting a sudden roaring in the valley before the arrival of a flash flood, described as a "tsunami", 3 km downstream of the geophysical station. The gravimeter is installed 48m underneath the surface and signal variations are induced by water accumulating above the gravimeter. The evolution of the gravity measurements along the event shows increasing subsoil's saturation with less and less water accumulation and increasing runoff."

*Lines 70 – 74: What was the density (per km<sup>2</sup>) of the gauging network and how does this density differ per region. I.e., how is the density in a heavily hit catchment such as the Vesdre, for instance? It's sometimes hard to estimate that from the figure.*

A map of Belgium with the definition of the discussed catchments and the locations of the rain gauges has been added in the Introduction. In complement, the following sentences have been added in Section 2.1:

"The spatial distribution of these 308 sites is illustrated in Figure 1 by the grey dots. On average over Belgium, this represents a density of 10 measurement sites per 1000 km. This density is somewhat lower for the Meuse basin (9 per 1000 km) and slightly larger for the Vesdre basin (11.5 per 1000 km)."

*Lines 81 – 82: How did this gap filling exactly work? Were gaps replaced with the data of the closest station, was some kind of spatial interpolation method used or something else?*

Weather observations are routinely checked at RMI against errors. This data quality control (QC) task is performed manually with the help of data interfaces that include statistical plausibility tests, time series plots, maps, etc. In the case of missing rain gauge data, several estimations are suggested by the QC interface: data from the nearest rain gauge, inverse distance weighted interpolation (IDW) from neighboring rain gauges, radar-based QPE, etc. In the present case, the estimations were derived by inverse distance weighted interpolation (IDW) from neighboring rain gauges. This is now specified in the revised manuscript:

"Estimations derived by inverse distance weighted interpolation (IDW) from neighboring weighing rain gauge data have been considered to correct the 2 periods of erroneous data as well as to fill the gaps."

*Lines 82 - 84: How were the daily values adjusted? This, as well as my previous comment, is quite relevant information for the dataset you are providing.*

If a manual observation is made later than 08:00 local time and if it is raining between 08:00 and the actual observation time, then the daily accumulation of the preceding day will be

overestimated while the accumulation of the following day will be underestimated. When such a case is detected, the daily accumulations are adjusted to follow the evolution of the daily totals of neighboring rain gauges. The following sentence has been added in Section 2.1 "This adjustment is based on the sequence of daily totals of neighboring rain gauges."

Globally, the data quality control implied very few interventions on the data, and none for the rain gauges located in the most severely affected area (e.g., Vesdre Basin). This is now stated in the revised manuscript:

"Globally, the QC analysis led to few interventions on all available rain gauge data and none concerning the rain gauges located in the most severely affected areas (i.e., Vesdre basin)."

*Line 104 – "different kind of technology": do you mean that the hardware of the radars is different?*

The radars are all C-band dual-polarization weather radars except the Wideumont radar which was still a single-polarization radar in 2021. Next to this difference, the radars also differ by their configuration and data processing. This has been clarified in section 2.2.1 of the revised manuscript.

*Lines 104 – 105: This is indeed a major challenge. It would be interesting for the reader if the authors can briefly describe how RMI handles this to make their composite.*

This is complex and technical. The main approach to reduce inhomogeneities between radars is the use of a static clutter map and a long-term gauge bias correction for each radar. These two steps are described briefly in sections 2.2.2 and 2.2.4 and in a new appendix.

*Lines 109 – 112: Could the authors describe in a little more detail what this basic radar estimation method is (just the Marshall-Palmer reflectivity to rain rate transformation, or more?) and how the comparison with the rain gauges took place. Regarding the latter, we often compare gauge location by corresponding grid cell, but it is also possible to take the rainfall advection into account (as the radar and rain gauges measure at different elevations), etc.*

The description has been extended in section 2.2.2. The advection error is not taken into account. This error is relatively random and can be mitigated by taking median values.

*Lines 113 – 120: I think this is a major improvement of the radar rainfall estimation in hilly regions. Actually, this improvement makes the provided and described dataset unique and different from what was already available at the time. Although I am familiar with the method used, I do not think that the average reader is. Hence, can I ask the authors to put more emphasis on this procedure and describe it in more detail.*

In the present case, this improvement is indeed very important for producing high-quality rainfall estimates at ground. The importance of the method is already stressed in section 2.2.2 of the manuscript. The method is now described with more details in the appendix.

*Lines 123 – 124: What was considered abnormal or unrealistic, i.e. was there a threshold or is this qualitative?*

The clutter removal techniques are extensively described in Goudenhoofdt and Delobbe (2016). Concerning the identification based on the vertical profile of reflectivity, the technique is based on thresholds and it is described in the reference paper as follows : "A measurement at a given elevation is considered as clutter if the gradient between its value and the corre-

sponding (horizontally) interpolated value on a higher (lower) elevation exceeds in magnitude  $-20 \text{ dBZ km}^{-1}$  ( $+10 \text{ dBZ km}^{-1}$ ). Because of variations from signal fluctuations, a minimum absolute difference of 5 dBZ between two corresponding values at different elevations is required for clutter identification”.

*Lines 129 – 130: How is this identification of the precipitation type exactly done and what Z-R relationship is used? I know most of this is described in Goudenhoofdt and Delobbe (2016), but either refer explicitly to this paper or (/and) describe it here.*

Some modifications have been applied since the publication in 2016. The relations for rain, relevant for this event, are now explicitly described in section 2.2.3.

*Lines 136 – 137 - “The rainfall accumulation over 5 minutes is obtained by computing the movement of precipitation using optical flow techniques”: What kind of optical flow techniques were used?*

The combined local-global method is used. Note that the flood case was not characterized by strong winds. Temporal sampling effects (e.g. jumping cell or ripple effects on the accumulation maps) were therefore very limited. This information has been added in section 2.2.4.

*Lines 139 – 143: If any methods from other papers were used in the KED method and setup, then some referencing is necessary here.*

Although no specific external references have been used, Hudson and Wackernagel (1994) is now cited for background information about KED. More details are also provided in appendix B.

*Lines 160 – 161: Although I fully agree, what can be said about the quality of the radar here? I.e., is the radar missing any precipitation in space, or overestimating, in this region? As radar estimates in hilly and mountainous areas are challenging, this is to be expected. It would be very interesting to have an idea of this too. This could also point out why and where the combination of both data sources is going to be crucial for a good analysis of this event. The answer to this comes already partially back in Figure 4, so good to give some more details here.*

It is difficult to assess the quality of the radar product regarding the spatial structure of the most extreme precipitation because the density of rain gauges is rather low with respect to the size of the area affected by the most extreme 3-day accumulations (e.g., above 200mm). We nevertheless complemented the discussion about uncertainty/accuracy with the following sentences:

”By taking a closer look on the rightmost point of the scatter plots of Figure 5 (i.e., manual rain gauge with the largest 3-day accumulation), we find an underestimation of almost 25% by the radar product, while the IDW interpolation of neighboring rain gauges underestimates that specific observation by 35%. A similar analysis but focused on daily rainfall totals also highlights that the estimation of the largest values is challenging but better addressed by the radar product. Daily precipitation values above 80mm, which have been recorded 17 times by manual rain gauges during the 3-day period, are on average underestimated by 16% by the radar product and by 22% by the IDW interpolation of neighboring rain gauges.”

A paragraph about uncertainties in rain gauge data has also been added at the end of Section 2.1.

*Lines 164 – 168: This actually belongs in the methods section, which is right now missing for the analysis part of this paper.*

For the sake of readability of the paper, we prefer not to move these lines (see our answer to the 2nd general comment).

*Lines 175 – 176: Could the authors briefly describe the used method (and reason to use that method) from Van de Vyver (2012, 2013) in the methods section and then refer to the paper?*

The following sentences have been added to briefly describe the considered extreme precipitation statistics:

”In order to assess the exceptionality of the precipitation amounts during the event, extreme precipitation statistics derived for Belgium by Van de Vyver (2012) and Van de Vyver (2013) have been considered. These statistics result from a spatial generalized extreme value (GEV) distribution fitted on annual rainfall maxima derived for various accumulation durations and several locations in Belgium from historical precipitation data. Thanks to this spatial approach, extreme precipitation statistics are estimated for any location in Belgium and are currently considered as the reference for, e.g., the design of hydraulic structures. Return periods can be associated from these statistics to the 3-day precipitation accumulations, as illustrated in Figure 6 for the rainfall distribution estimated by the radar product (i.e., Figure 4b).”

*Lines 206 – 208 – “For longer accumulation durations (from 6 hours to 3 days), a clear trend towards extreme values can be noticed with, in parallel, an increase in the size of the severely affected areas.”: I think the authors can put some more emphasis on this finding. It would be interesting to relate this to literature findings (either here on in a discussion section) for similar extreme flooding events or similar rainfall patterns. I.e., is this what you expect or is this unique to this event?*

This finding illustrates the highly unusual nature of this event. In Belgium, extreme summer rainfalls are generally caused by local convective storms and, consequently, extreme return periods are obtained for short durations and concern relatively small areas. In surrounding regions with similar climatic conditions, this type of event is also very rare. A relatively similar event with mixed convective and stratiform rainfall occurred in the Netherlands and the bordering part of Germany in August 2021 (Bauer et al., 2011). Over an area of 740 km<sup>2</sup> more than 120 mm of rainfall were recorded in 24 h. Maximum amounts for the whole episode reached 160 mm, which is still lower than the extreme values obtained in the present case.

Long duration rainfall events, affecting large areas and producing flash floods in several catchments are relatively common in the Mediterranean and Alpine-Mediterranean regions as shown in several inventories of European flood events (Gaume et al., 2009; Marchi et al., 2010 and Amponsah et al., 2018). These events occur mostly in Autumn under specific climatic forcing and triggering mechanisms that are not present in the climatic region of interest here.

The above comments have been included in Section 3.2 near Fig. 8. in the revised version.

*Lines 221 – 243: This belongs in the methods section.*

The background information about NMF was moved to the Appendix C.

*Section 3.3: I am not familiar with the NMF method, but can I ask the authors to explain a*

*bit more elaborately what you are trying to reach with this method and why you focus in the results on a rank-3 NMF (and not more or less)? Why do we need the NMF method? Is it to be able to split the mesoscale rainfall field into more regional scale stratiform and convective rainfall, or is there another reason? See also my comment below related to figure 10a.*

Our objective with the NMF analysis is to illustrate that the temporal distribution of rainfall was not everywhere the same. NMF decompositions of various ranks have been derived, but the rank-3 NMF analysis provides the clearest decomposition of the data into distinct spatio-temporal patterns as shown in Figure 10. Figure 10 provides a very synthetic view of the event with 3 maps and 3 timeseries. It allows to derive the following qualitative descriptions of the event:

In the south-east part of the country, rainfall were rather continuous on 13th and 14th July. 15th July as rather dry in that area.

In the central part of the country, few rainfall were observed on 14th July, although it was the most severe period for the Vesdre basin. For the central part of the country, rainfall responsible for floods occurred on 15th July.

In the East part of the country (e.g., Vesdre basin), we can note a sequence of very intense precipitation peaks from the evening of 13th July to the morning of 15th July.

We find it interesting to highlight these very different spatio-temporal rainfall patterns, even if we do not have yet determined what the dynamical source of these patterns is.

The following sentences have been added in the revised manuscript (see Section 3.3):

”This NMF analysis will allow us to highlight that rainfall was distributed in time differently over the country. The NMF analysis will provide a synthetic view of the event from which qualitative descriptions can be derived.”

”Although NMF approximations of various ranks have been derived, the discussion will be focused on the rank-3 NMF that provides a clear decomposition of the data into 3 distinct spatio-temporal patterns as illustrated in Figure 10.”

*Figure 3: Here I would show the catchments that you are focusing on instead of the province of Liège.*

The Province of Liège is shown on Figure 4 just to highlight the area where the most extreme precipitation occurred. It allows mainly to link to zoomed maps (bottom line) to the maps of Belgium (top line).

*Figure 10a: The sub panel is showing the difference between the rank-3 NMF and RAD-FLOOD21 data, right? This is not yet clear from the caption. In addition, from the figure and the corresponding text it is not directly clear to me if a difference is good or bad and how big of a difference is considered acceptable. As mentioned earlier, I am not familiar with the NMF method, so that definitely plays a role too. Is a small difference and indication that with (just) three ranks, and corresponding regions, you capture most of the rainfall in space and time, or should I interpret it differently?*

The caption of Figure 10a has been clarified. Since NMF provides an approximation of the initial data, there is a difference on the 3-day total between the rank-3 NMF and RADFLOOD21 data, as shown on Figure 11a. The map on Figure 11a indicates where the discrepancies between both is the largest. In particular, there is a zone in central Belgium where the rank-3

NMF approximation underestimates the 3-day total by up to 50mm. This is a pattern not captured by the rank-3 NMF, but where total rainfall amounts remain rather moderate (i.e., lower than 100mm over the 3 days).

*Figure 12: Useful figure, but I think this figure fits better in the study area description. Hence, put it together with one of the first figures. I think it would fit well in figure 3.*

The map of Belgium with the definition of the discussed catchments has been moved to the Introduction section.

*Figure 13: It looks like the individual points in the graphs have relatively coarse steps (1, 2, 3, 6, 12, etc. hours), even though the data is based on hourly accumulations. Hence, I think the authors could put more data points in this graph, which more clearly identifies when differences between the two datasets occur. In fact, why not just go to 5-min steps and derive the cumulative sums on that time step?*

The Figure 13 in the revised manuscript illustrates now maximum precipitation accumulations derived for durations from 1 to 72 hours by steps on 1 hour. Regarding the use of the hourly radar product instead of the 5-min one, we consider the analysis based on hourly data as meaningful since the event starts to be extreme for accumulation durations larger than 6 or even 12 hours (see Figures 8 and 9).

*Figure A1: Just to double check, we are looking at the maximum x-hour window accumulation that occurred, right? So, for e.g. the grids cells in the Vesdre, we are looking at the statistics of the maximum x-hr accumulation that occurred during the 3-day period and not the average? Besides, do you look at a rolling window in which for every 5-min steps a new x-hour window is assessed or have the authors assessed it differently? Good to briefly mention this in the methods and the caption.*

In Figure A1 (now Figure D1), we indeed consider for each grid cell the maximum x-hr accumulation that occurred during the 3-day period. This maximum x-hr accumulation is derived from the hourly radar product for durations till 12 hours, and otherwise from daily-aggregated radar product. This seems to us already clearly mentioned in the text and caption:

Text, Section 3.2: "The analysis of Figure 8 can be extended to any location in Belgium based on the **hourly** radar product. For each 1-km pixel of the radar product, the **maximum rainfall accumulation is derived for durations from 1 hour to 3 days during the period** July 13th 06:00 UTC to July 16th 06:00 UTC."

Caption of Fig. A1 (now D1): "Return period estimated for the **maximum** precipitation accumulation between 13th July 06:00 UTC to 16th July 2021 06:00 UTC for durations from 1 hour to 3 days. The **maximum accumulations are derived from hourly (resp. daily) radar product data for durations till 12 hours (resp. of 1, 2 or 3 days).**"

## Technical corrections

*Line 107: Although true, I think this needs some referencing.*

A reference has been added.

*Lines 130 – 131: Good to add a reference here.*

The Z-R relationship is mentioned here [https://www.ou.edu/radar/z\\_r\\_relationships](https://www.ou.edu/radar/z_r_relationships).

pdf (see page 2) with a reference to [http://www.roc.noaa.gov/ops/z2r\\_osf5.asp](http://www.roc.noaa.gov/ops/z2r_osf5.asp) which is not anymore active. No public reference from NOAA seems currently available.

*Lines 275 – 278: From a hydrologist’s perspective, it is more useful to provide these numbers in mm instead of in km<sup>3</sup>. The same comment holds for the conclusion section.*

Total precipitation values in km<sup>3</sup> were requested by the Editor. Values in mm have been added between brackets.

*Figure 2: The most important components of this figure, the radar locations and their 100-km range, are relative hard to distinguish from the rest of the map with the using colors. My suggestion would be to use different colors and somewhat larger font and icon size.*

The figure has been modified accordingly.

*Figure 8: The figure has a title and a footnote, which are partially repeated in the caption. I would suggest to just keep the text in the caption and remove the title and footnote from the figure. In addition, this holds for some other figures, too.*

Most figures had internal titles and sometimes subtitles and footnotes in order to be comprehensible in themselves (e.g., in case they are reused outside the manuscript). We removed internal titles from several figures (Figures 6, 7, 8 and 9 in the revised manuscript). For other more complex figures, we think that summarized internal titles are still useful, in addition to a well-detailed caption.

*Figure A2: What is the highest value in the sub panels? It seems that the maximum value of the color scale could be somewhat lower, e.g. 50 mm, that would give somewhat more contrast. The largest 3-hourly values reach almost 55mm (on 2021-07-14 from 18:00 to 21:00 UTC). The colorscale has been adapted to give slightly more contrast.*

All language-related suggestions have been implemented in the revised manuscript.



## Responses to RC2

*I do not have much to add to the excellent review by Ruben Imhoff, except to say that personally I am not sure about the value of the principal component analysis. Rather than interpreting the complex spatiotemporal patterns in Figure 9, I would much prefer to see a catchment based analysis of the rainfall dynamics to get a better grasp of the hydrological processes (as also pointed out by Ruben Imhoff).*

Our objective with the NMF analysis is to illustrate that the temporal distribution of rainfall was not everywhere the same. NMF decompositions of various ranks have been derived, but the rank-3 NMF analysis provides the clearest decomposition of the data into distinct spatio-temporal patterns as shown in Figure 10. Figure 10 provides a very synthetic view of the event with 3 maps and 3 timeseries. It allows to derive the following qualitative descriptions of the event:

In the south-east part of the country, rainfall were rather continuous on 13th and 14th July. 15th July as rather dry in that area.

In the central part of the country, few rainfall were observed on 14th July, although it was the most severe period for the Vesdre basin. For the central part of the country, rainfall that causes floods occurred in 15th July.

In the East part of the country (e.g., Vesdre basin), we can note a sequence of very intense precipitation peaks from the evening of 13th July to the morning of 15th July.

We find it interesting to highlight these very different spatio-temporal rainfall patterns, even if we do not have yet determined what the dynamical source of these patterns is.

The following sentences have been added in the revised manuscript (see Section 3.3):

”This NMF analysis will allow us to highlight that rainfall was distributed in time differently over the country. The NMF analysis will provide a synthetic view of the event from which qualitative descriptions can be derived.”

”Although NMF approximations of various ranks have been derived, the discussion will be focused on the rank-3 NMF that provides a clear decomposition of the data into 3 distinct spatio-temporal patterns illustrated in Figure 10.”

*Maybe the authors could also add a comparison of rainfall and runoff time series for selected catchments, just to see what the observed rainfall amounts meant on the ground.*

We agree that it would be interesting to present such results. However, we consider that it is outside the scope of this paper which is really focused on documenting rainfall. Presenting the hydrological impact is also outside our field of expertise.

## References

- Hudson, G. and Wackernagel, H.: Mapping temperature using kriging with external drift: Theory and an example from scotland, *International Journal of Climatology*, 14, 77–91, <https://doi.org/https://doi.org/10.1002/joc.3370140107>, 1994.
- Van de Vyver, H.: Spatial regression models for extreme precipitation in Belgium, *Water Resources Research*, 48, <https://doi.org/10.1029/2011wr011707>, 2012.

Van de Vyver, H.: Practical return level mapping for extreme precipitation in Belgium, RMI scientific and technical publication, 62, 2013.