Authors response to Referee 2 comments

In this paper a peak-over-threshold method is used to perform an extreme rainfall analysis and to derive return levels from weather radar and rain gauges in Belgium. The importance of this work is high, as radar archives are nowadays long enough to permit the development of extreme rainfall analyses which are of fundamental importance for many applications, but the common annual maxima 5 approach needs even longer time series. However, some important explanations and discussions, in addition to those already highlighted by the first review by F. Marra, are missing in the manuscript, and need to be provided before the article can be accepted for publication.

We would like to thank Luca Panziera for underlying the importance of our work. The detailed comments allowed us to improve the presentation and the discussion of our results.

Thank you a lot for the work you have done to improve the clarity of this manuscript. The text, thanks also to the comments of the other reviewers, is now much clearer. I still have a few comments regarding this work.

Replies to original major comments

What is new?
It is somehow difficult to understand which new contribution this paper brings with respect to previous studies, and I think that this should be better highlighted in the text. To my understanding, the main novelty of this paper is the use of a POT method for an extreme rainfall analysis for weather radar data.

As pointed out by the first reviewer, the originality of our work did not appear clearly. The use of the POT method together
with the other novelties of our approach are now highlighted in the text. Please see page 3, line 30.

Thank you.

Temporal Declustering
As rainfall data need to be declustered in order to remove the temporal correlation in the time series before GPD parameters estimation, the authors choose an interval of 12 hours (for 1-hour rainfall) and 3 days (for daily rainfall) in order for two threshold exceedances to be considered as independent. The choice of these intervals, which should be referred to as run length or run parameters according to the literature, seems reasonable, but it could potentially have a big impact on the derived return levels, as it shapes the exceedances time series whose maxima are used for the parameters estimation. If the data are temporally clustered, such temporal lags could not be long enough to remove dependency, but if the temporal clustering occurs rarely, they could actually lead to a significant bias of the return levels estimates. What do the authors mean as temporal independence? How did the authors choose such temporal lags? Did the authors investigate the effect of changing these values on the parameters estimation and final return levels? The subjective choice of these values should be motivated and discussed in the text.

How to deal with declustering is indeed a crucial point to address in extreme rainfall analysis. It is now properly discussed from page 7, line 25.

Thank you for the explanations. My only comment is the following: at pag. 7 line 29 you state that using 3 days temporal lag hardly affects the selection of the 1 hour extremes. Could you please provide some numbers (not necessarily you have to add them in the text)? Did you also try to select a temporal lag shorter than 12 hours (e.g. 6 hours) or longer than that (e.g. 24 hours)? You should also investigate the effects of the different temporal lags on the GPD parameters estimation. You might want not to insert this in the paper, but I would appreciate to see how the GPD parameters change.

Exponential distribution
As the choice of a null shape parameter is fundamental for this work, I think that it should be motivated more in the text. Therefore, I suggest to briefly report and discuss the main results ofWillems (2000), in order to better understand the motivation of this choice The text states also that this choice was taken because of the short period. However, with a POT approach the shortness of the period should not be a limiting 5 factor, as many events are considered. It should also be discussed if this is the

best choice for both 1-hour and 24-hours accumulations. Did the authors try to estimate also the shape parameters, to see if from the data a value different from 0 could be derived?

The short period remains a limiting factor to model the tail of the GPD. The choice of the Exponential distribution is further justified from page 8, line 3. We therefore did not try to estimate the shape parameter.

Thank you.

Radar and gauge comparison

The authors present an interesting comparison between the radar and gauges extremes, for 1-hour and 24-hours accumulations.
Despite this being very interesting and instructive, the implications for this study are not very clear. I suggest the authors discuss at least qualitatively the influence of this investigations on the overall results of the study.

The implications of the radar and gauge comparisons have been added from page 9, line 16 and from page 10, line 18.

Pag. 9, from line 17, you state that due to the randomness of the cell boundary effect, this should not affect the results. And also that the missing data are expected to have only a minor impact on the statistics. I am not totally sure about this, but probably there is not much to do to avoid these problems. You might want to mention this in the text.

Regional frequency analysis

The regional frequency analysis needs also to be better explained and the choices which were taken need to be motivated and discussed. How did the authors choose the 20-km radius for the analysis? How the resulting return levels at a given pixel should be interpreted, as they stem from exceedances in rainfall values which were observed all around it? Does it still make sense to speak about point measurement? How are the maps of GPD parameters affected by the choice of the 20-km radius circles?

Our methodology should be better explained indeed. An extended literature review is given from page 2, line 24. Our methodology is discussed from page 11, line 5. We think this explains why we can still speak about point measurements. For the derivation of spatial maps, please see from page 12, line 29.

Thank you for the better explanation.

Return levels maps

I guess the final goal of the study is to derive maps of return levels with relative uncertainty for Belgium. Despite the return levels are shown for given rain gauge locations, it would be desirable to show also maps of return levels for selected return periods. Would it be possible to insert a map or two of the return levels? How would those maps be affected by the 20-km radius selected for the regional frequency analysis? How these maps should be interpreted? Since you are using a constant shape parameter (equal to 0), and the longest return levels are shaped by it, long return periods map will tend to produce maps less variable in space with respect to short return periods. This should be discussed in the text.

Two return level maps have been added and discussed. Please see page 13, line 4.

Thank you for adding the maps of the return levels. if I am correctly interpreting the maps, I think that they are strongly influenced by the selected circle for the RFA, and I think that this does not appear clearly from the discussion of page 13, lines 4-21. I see this as a major drawback of the proposed method, and, this should be clearly mentioned in the text.

Do you also have maps for the 24 hours RFA? Can you show them as well?

Other comments

1. Pag. 2 line 16: delete the parenthesis from the citation Willems et al. (2007)

2. Pag. 2 line 32: I would delete “perfectly”: even for widespread rainfall situation, local phenomena can lead to gradients in rainfall rarely measured by the rain gauges.
3. Pag. 3 line 5 at the end: “It has been shown”, please correct.
4. Pag. 3, line 31: to be correct, also the work of Panziera et al. (2016) used a radar QPE combined with rain gauges. Are you sure that all the other mentioned papers did not also consider a combination of radar and rain gauges?
5. Pag. 8, line 8. Why the EXP model is less affected by observational errors?
6. Table 2: the event 7 should not be the 6th, as 25.17 (radar, event 7) is larger than 24.35 (radar, event 6)?
7. Pag. 11, line 14. I don’t think that mentioning here my 2016 paper is required.
8. Pag. 13, line 13.I think that the hail contamination is more sensitive to the radar sampling volume than the height of the beam, since a larger sample volume means also a smaller probability that it is dominated by hail, but also other hydrometeors types are present. I think that hail contamination is more relevant close to the radar.
9. Pag. 14, line 23-32 and 34-35 look like a plan for future work, so you might consider to move them at the end of the Conclusions. I also suggest to reorganize the Conclusions sections, as methodological information and the results are mixed together making this sections not easy to read.