



Regional frequency analysis of extreme rainfall in Belgium based on radar estimates

Edouard Goudenhoofdt¹, Laurent Delobbe¹, and Patrick Willems²

¹The Royal Meteorological Institute of Belgium, Brussels, Belgium

²Department of Civil Engineering – Hydraulics Division, University of Leuven, Leuven, Belgium

Correspondence to: Edouard Goudenhoofdt (edouard.goudenhoofdt@meteo.be)

Received: 15 March 2017 – Discussion started: 2 May 2017

Revised: 12 July 2017 – Accepted: 15 September 2017 – Published: 27 October 2017

Abstract. In Belgium, only rain gauge time series have been used so far to study extreme rainfall at a given location. In this paper, the potential of a 12-year quantitative precipitation estimation (QPE) from a single weather radar is evaluated. For the period 2005–2016, 1 and 24 h rainfall extremes from automatic rain gauges and collocated radar estimates are compared. The peak intensities are fitted to the exponential distribution using regression in Q-Q plots with a threshold rank which minimises the mean squared error. A basic radar product used as reference exhibits unrealistic high extremes and is not suitable for extreme value analysis. For 24 h rainfall extremes, which occur partly in winter, the radar-based QPE needs a bias correction. A few missing events are caused by the wind drift associated with convective cells and strong radar signal attenuation. Differences between radar and gauge rainfall values are caused by spatial and temporal sampling, gauge underestimations and radar errors. Nonetheless the fit to the QPE data is within the confidence interval of the gauge fit, which remains large due to the short study period. A regional frequency analysis for 1 h duration is performed at the locations of four gauges with 1965–2008 records using the spatially independent QPE data in a circle of 20 km. The confidence interval of the radar fit, which is small due to the sample size, contains the gauge fit for the two closest stations from the radar. In Brussels, the radar extremes are significantly higher than the gauge rainfall extremes, but similar to those observed by an automatic gauge during the same period. The extreme statistics exhibit slight variations related to topography. The radar-based extreme value analysis can be extended to other durations.

1 Introduction

Localised rainfall extremes can have a strong impact on human activities especially in urban areas (Ootegem et al., 2016). For flood management applications (e.g. sewer system and dam design) it is needed to know the probability that rainfall exceeds a given amount. This probability is often expressed as the rainfall level which, on average, will be exceeded once over a given period of T years, which is defined as the return period. For infrastructure design applications, one is interested in return periods from 50 to 100 years. Such long return periods often exceed the available observation period and a model is needed.

Extreme values are often extracted from a time series using block maxima, typically over one year (AM) for meteorological data. The performance of the statistical modelling applied to AM data is limited by the number of years available. The peak-over-threshold (POT) method, where values exceeding a given threshold are kept, allows to increase the number of samples. The extreme value theory showed that under some hypotheses – including independence – of the random variables, the AM and POT series can asymptotically converge only to the 3-parameter distributions known as GEV and GPD, respectively.

Different fitting methods to the extreme value distributions have been developed in the literature. The maximum likelihood estimator (MLE) is the most widely used fitting method but for small samples it can lead to unrealistic parameter estimates. This problem is partially addressed with the generalized MLE proposed by Martins and Stedinger (2000) or the L-moments method (Overeem et al., 2009). The above methods do not focus on the tail of the distribution, which is the

most relevant for risk analysis. For this goal, Willems et al. (2007) proposed a method based on regression in Q-Q plots.

To reduce the uncertainty associated with the limited number of data at a single site, regional frequency analysis (RFA) methods have been proposed (Svensson and Jones, 2010). The RFA is characterized by the selection of the regions and the parameter estimation approach applied to each region (Buishand, 1991). There are numerous studies of RFA for rainfall extremes based on rain gauge datasets. The index flood approach, which considers that only the location parameter varies in the region, is very popular (Gellens, 2000; Sveinsson et al., 2001; Rulfova et al., 2014). Uboldi et al. (2014) used a bootstrap technique to randomly select data from neighbouring locations with a probability depending on the distance and altitude difference with the target location. The combined use of POT and RFA methods is recommended by Roth et al. (2015).

One of the challenges in RFA is the intersite dependence (e.g. Hosking and Wallis, 1988). Even for a 1 h duration, rainfall maxima exhibit a spatial correlation (Vannitsem and Naveau, 2007). Using the sum of the length of all sites is common but causes an underestimation of the extremes (e.g. Bardet et al., 2011). Several approaches have been proposed to deal with this problem (e.g. Castellarin, 2007; Weiss et al., 2014).

To obtain the rainfall statistics at any given point, spatial models have been developed using geographical and climatological covariates (e.g. Cooley et al., 2007). In Belgium, Van de Vyver (2012) derived a spatial GEV model depending linearly on the altitude. Rulfova et al. (2014) found for 6 h rainfall in the Czech Republic that the assumption of a linear model might be too restrictive, especially for convective precipitation.

The rain gauge network can capture rainfall extremes for widespread situations. However, they can only catch a small part of rainfall extremes caused by convective storms, which exhibit strong spatial variations over short distances. The use of high-resolution gridded rainfall datasets to study rainfall extremes is still in its infancy. This can be explained by their unavailability, their processing requirements and their limited quality. Precipitation estimations from satellite offer global and relatively long records suitable for extreme value analysis (Marra et al., 2017) but still suffer from large uncertainties (Sapiano and Arkin, 2009). The best potential is currently provided by radar-based quantitative precipitation estimation (QPE) products. It should be noted that the radar estimates represent the averaged precipitation over a given area (typically a square of 1 km). While this area is much bigger than the gauge area, we will consider it as representative for small-scale precipitation. It has been shown that the sub-pixel variability of rainfall extremes is significant, especially for short durations (Peleg et al., 2016). The relatively short record of radar datasets is an issue if the extreme statistics depend only on time (i.e. are completely dependent spatially). While this is a reasonable assumption for larger du-

rations (e.g. 1 day), it is difficult to prove for short durations (e.g. 1 h). In case of significant climatic variations, a short record will be more representative of the extreme statistics.

In a pioneering work, Overeem et al. (2009) showed that a 11-year radar dataset is suitable to derive depth-duration-frequency (DDF) curves for the Netherlands. But some differences with rain gauge results were found for short durations. Based on a unique 23-year radar dataset in Israel, Marra and Morin (2015) found that the DDF curves were generally overestimated but 60% of them lay within the rain gauge DDF confidence intervals. In Ontario (Canada), Paixao et al. (2015) demonstrate the potential to integrate radar (Digital Precipitation Array product) to rain gauge analysis, especially to identify homogeneous regions of extreme rainfall. Saito and Matsuyama (2015) used a 26-year radar-gauge dataset (without RFA) to study the spatial variation of hourly rainfall extremes in Japan. They found significant spatial patterns but also large uncertainties in the radar datasets. Different index flood approaches were tested by Eldardiry et al. (2015) in Louisiana, who defined a region as a square window of 44 km size. They found for Louisiana (USA) that the relatively short period (13 years) explains the high uncertainty of the analysis, that the index flood method is recommended and that a systematic underestimation is associated with the radar products (its spatial resolution is 4×4 km). Haberlandt and Berndt (2016) found that the operational DWD product is only suitable for studies on longer durations after bias correction. Using a 10-year high-resolution radar rainfall dataset, Wright et al. (2014b) performed a RFA using stochastic storm transposition. They found that the radar-based intensity-duration-frequency (IDF) estimates generally reproduce conventional gauge-based IDF estimates but overestimate these for longer return periods and shorter durations.

The potential of the radar data can be fully exploited by studying the extremes of the mean (or maximum) rainfall over areas. With the goal of deriving alert thresholds for 159 regions in Switzerland, Panziera et al. (2016) studied the areal rainfall maxima (with sizes from the pixel to the region). Using RFA on squares, Overeem et al. (2010) derived areal rainfall DDF curves for the Netherlands. Wright et al. (2014b) applied a similar methodology but on different catchments in Louisiana.

In this study, we want to demonstrate the potential of high-resolution radar-based QPE to derive rainfall extreme statistics at a given location. To our knowledge none of the previous studies combine high-quality radar-based QPE with a high-quality reference rain gauge measurements. At the Royal Meteorological Institute of Belgium (RMIB), a QPE has been derived from the reprocessing of raw volumetric radar measurements. This dataset has been used for various applications such as case studies and model verification. The methodology to derive this dataset has been verified for the period 2005–2014 against an independent rain gauge network (Goudenhoofd and Delobbe, 2016). RMIB also has a

unique 40 year dataset of 10 min rain gauge measurements, which has been used in extreme value studies (Vannitsem and Naveau, 2007; Van de Vyver, 2012).

Unlike existing radar studies, we select our data using the POT approach and use a regression in Q-Q plot (QQR) fitting method. Radar-based extreme statistics for 1 and 24 h durations are compared with the ones derived from rain gauge data covering the same period. We propose a new RFA which makes use of independent radar data in a predefined neighbourhood. The results are compared with those obtained using the long-term rain gauge network. Finally, the regional approach is applied at each radar pixel on the whole of Belgium to study the spatial variations of the rainfall extremes.

2 Rainfall data

2.1 Rain gauge measurements

Over the years, Belgium (Fig. 1) has been covered by several rain gauge networks for different purposes.

Since the end of the 19th century, RMIB maintains a network (CLIM) of non-recording rain gauges from which rainfall measurements are taken at 08:00 LT. The data are carefully controlled and used for climate applications (Journée et al., 2015).

A Hellmann–Fuess pluviograph has been in operation in Uccle (RMIB) from 1898 to 2008 and has enabled the compilation of a continuous time series of 10 min rainfall (Demarée, 2003). The 10 min rainfall values had to be manually extracted from line graphs on papers. Starting from the fifties, additional rain gauges were installed to constitute a network (BUL) for hydrological research. Since the rain gauges underestimate the rainfall by 5–10 % due to its mechanism, its records have been calibrated using a collocated gauge from the CLIM network.

For weather forecast purposes, the RMIB maintains a network of automatic weather stations (AWS) in Belgium. These stations provide rainfall measurements at 10 min temporal resolution. The tipping-bucket gauges are progressively replaced by weighted gauges (the first one was installed in Uccle on 10 February 2009). The data are available since 2002–2004 and have been quality controlled.

The hydrological service of the Walloon region (SPW) maintains a dense network of hourly (every 5 min since 2012) rainfall measurements. The tipping-bucket gauges are progressively replaced by weighted gauges since 2015. The data have been quality controlled by RMIB since April 2004.

It is important to know the limitations of the respective rain gauges in case of extreme rainfall. It is known (Nystuen, 1999; Duchon and Biddle, 2010) that tipping buckets underestimate high rainfall rates. The use of weighted gauges for extreme rainfall is discussed in Colli et al. (2012). Every 10 mm, the pluviograph has to be emptied which results in an underestimation in case of extreme rainfall. The calibration

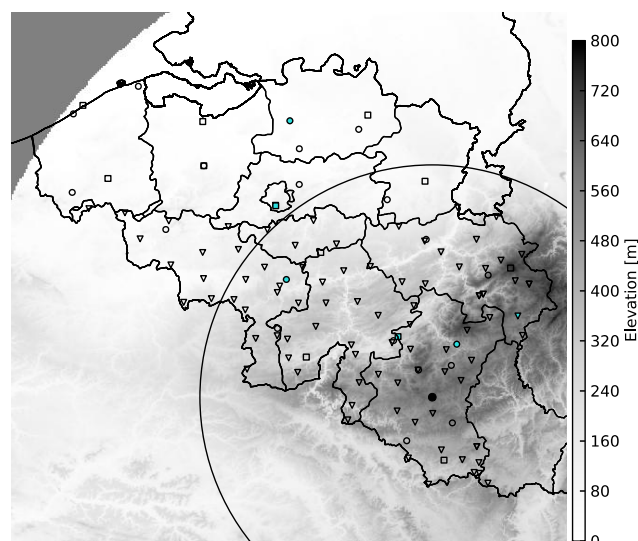


Figure 1. Elevation map centred on Belgium with the Wideumont radar (black dot) covering a 240 km range (the black circle denotes the 120 km range) with AWS (square), SPW (triangle) and BUL (circle) rain gauge networks. The gauge locations selected in this paper are in cyan. Country borders with France, Luxembourg, Germany and the Netherlands are also displayed.

of the pluviograph is probably not sufficient for sub-daily extremes. Finally, the quality controls, albeit conscientious, can never be considered as perfect.

2.2 Radar estimation

The QPE available on a 1 km grid every 5 min is made using an elaborated processing chain from the radar volumetric reflectivity measurements. The quality of the radar volume is controlled using several algorithms:

- a static clutter map: pixels with unrealistically high probability of rainfall are identified as clutter
- a beam blockage map: the percentage of the beam blocked by topography is computed using a simple propagation model
- a first clutter identification based on reflectivity differences between radar beam elevations
- a second clutter identification based on strong deviations of a pixel from its neighbourhood and unrealistic lines
- a third clutter identification for radar echoes in cloud-free areas determined by satellite observations

A maximum threshold for reflectivity is set to 55 dBZ to mitigate higher reflectivity values due to hail. The rainfall rate estimates are obtained using stratiform-convective classification, a 40 min averaged vertical profile of reflectivity

(VPR), a bright band identification and a specific transformation to rain rates for the two precipitation regimes. The detailed procedure is described in Goudenhoofdt and Delobbe (2016). As a reference for the QPE product, the CAP product is defined as the interpolation at 800 m above the radar level. It makes use of a standard $Z-R$ relationship, which comes from the hypothesis that the drop size distribution follows the distribution of Marshall–Palmer, as discussed in Uijlenhoet and Pomeroy (2001).

Consecutive rain rate estimates are integrated to obtain 10 min accumulations (5 min gaps are tolerated) to match the lowest resolution of the rain gauge data. Hourly accumulations are combined with the SPW gauges using a mean field bias correction. This method applied to the QPE product is referred to as the MFB product from now on. A more complex merging method (i.e. external drift Kriging) was tested but found to be unstable for some time moments.

It is important to mention the limitations of the radar products in case of extreme precipitation. The most important impact of the QPE processing on extreme values is the 55 dBZ reflectivity threshold used to mitigate hail. Using the convective $Z-R$ relationship, this corresponds to a maximum rainfall rate of 80 mm h^{-1} . Higher values of about 100 mm h^{-1} are possible when the standard $Z-R$ relationship is used for stratiform areas. This can only happen close to the radar where convective precipitation can not be identified. Slightly higher thresholds have been used by Overeem et al. (2009) (100 mm h^{-1}) and Wright et al. (2014b) (105 mm h^{-1}). A higher threshold is used by Marra and Morin (2015) (150 mm h^{-1}) but for a Mediterranean climate. Only half of the AWS gauges recorded up to 3 times more than 100 mm h^{-1} in 10 min. Given the sub-pixel spatial variability, one can assume that this threshold will never be exceeded for the pixel average. This threshold can only partly correct for the overestimation due to hail. The second most important error is related to signal attenuation especially in the case of well organised convective systems. This is why extremes might be underestimated the further the distance from the radar. In addition, the increasing radar sample volume will produce an underestimation of small-scale extremes. The uncertainty in the $Z-R$ relation is another important source of error.

2.3 Comparison framework

In this study, we will only consider validated rain gauge data. Given that the SPW network is used for merging, the radar dataset for 2005–2016 is used. To perform a direct comparison, the gauge data of AWS and SPW for the same period are used. For comparison against the reference BUL network, the gauge data for the period 1965–2010 are used. The time series of the BUL and CLIM networks have been tested for homogeneity by Van de Vyver (2012) and a selection of useful stations has been made. Gellens (2000) and Vannitsem and Naveau (2007) found that the vast majority of the CLIM

and BUL time series are stationary for summer rainfall. However, the existence of a multi-decadal oscillation in rainfall extremes has been found in the Uccle time series (Ntegeka and Willems, 2008; Willems, 2013).

The 10 min rainfall accumulation from the gauge networks (AWS, BUL) and radar products (CAP, QPE) are summed to obtain sliding 1 h rainfall accumulations. Such a duration is associated with convective storms, which can only be properly seen on radar images. The hourly bias obtained by the MFB method could be applied to the 10 min accumulations. However, it will not be used due to the possible risk of representativity errors related to convective storms and the small benefits expected.

The hourly rainfall from the SPW network and the radar products (CAP, QPE, MFB) are summed to obtain sliding 24 h rainfall accumulations. The SPW network is preferred to the AWS network because it is denser and more homogeneous. Such a duration is mainly associated with widespread precipitation for which the benefit of merging methods is clear. The risk of instability with MFB (e.g. in case of strong spatial variation of the bias) is tolerated given the significant expected benefit for widespread precipitation events.

It should be noted that using the lowest available duration for each network would result in an underestimation of the extremes due to the discrete time sampling (Marra and Morin, 2015). Additionally, random errors and time sampling difference can be compensated by performing the sum. For both the radar and the gauge, no missing data is tolerated in the sum to avoid underestimation. Furthermore, only timestamps with both radar and gauge data are kept.

Due to the amount of stations, it is not possible to analyse in detail the results at each station. Therefore, a few stations are picked at different distances from the radar (see Table 1 and Fig. 1). The Uccle station is chosen because it is included in the three networks, which makes inter-comparison possible. The availability of the 1 h accumulation data is about 95 % for the radar products and close to 100 % for the AWS gauges. The radar availability of the 24 h accumulation is lower than the 1 h accumulation because a significant part of the intervals without data are short. The availability of the SPW gauges is around 90 % but this is mainly due to the removal of snow events, when no extreme precipitation is expected. The availability of the BUL stations for the period 1965–2010 is highest at Uccle with 96.3 %, then about 86 % at Deurne and Gosselies. The station of Nadrin has only 60 % of availability (for the period 1965–2010) because it was started in 1978.

3 On-site frequency analysis

3.1 Methodology

It has been shown by Pickands III (1975) that the extreme values converge asymptotically to a generalized Pareto dis-

Table 1. Rain gauge stations used for comparison and availability of the extreme rainfall datasets. The last column is the percentage of time when both radar and gauge data are available. DNG is the *Deuxième nivellement général*.

Station	Altitude (DNG)	Distance to radar (km)	Duration	Avail. gauge (%)	Avail. radar (%)	Avail. both (%)
Humain (AWS)	296	36	1 h	98.5	94.8	93.5
Uccle (AWS)	100	128	1 h	99.9	94.8	94.7
Uccle (SPW)	100	128	24 h	90.6	86.0	78.2
St-Vith (SPW)	456	61	24 h	89.2	86.0	76.7
Deurne (BUL)	12	161	1 h	86.0	–	–
Uccle (BUL)	100	128	1 h	96.3	–	–
Gosselies (BUL)	187	97	1 h	85.7	–	–
Nadrin (BUL)	403	30	1 h	59.3	–	–

tribution (GPD):

$$F_{(\xi, \mu, \sigma)}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } \xi = 0, \end{cases} \quad (1)$$

with ξ , μ and σ commonly defined as the shape, location and scale parameters, respectively. The special case when the shape parameter is equal to zero is defined as the exponential distribution (EXP).

The choice of the threshold has an important impact on the estimation of the distribution parameters. When the number of selected values increases, the variance naturally decreases but the bias increases (due to the deviation from the theoretical distribution). It is more practical to use a threshold rank instead of a threshold value to control the sample size.

To apply the theory, the extreme values have to be independent but successive peaks within the same time window can be observed due to the nature of precipitation. For the 1 h duration, two peaks are considered dependent if the time interval is less than 12 h as proposed by Ntegeka and Willems (2008). This choice is consistent with the characteristics of convective storms analysed in Goudenhoofdt and Delobbe (2013). Jakob et al. (2011) used a separation time of 24 h but found little sensitivity when taking lower or higher values. We also found that using 3 days hardly affects the selection of the 1 h extremes. For the 24 h duration, we use a time interval of 3 days, which is the typical scale of synoptic regimes. These choices are consistent with Roth et al. (2014) who empirically found a temporal dependence of 1 and 2 days for winter and summer precipitation, respectively. In practice, a peak is kept if it is the maximum compared to its dependent peaks (if any).

The type of the distribution can be derived by looking for the Q-Q plot where the extremes behave in an asymptotic linear way. Willems (2000) found for the Uccle series that the tail of the distribution has an exponential behaviour for all durations. In the gauge datasets used in this study, we also found a tendency for the EXP distribution. The EXP distribution is preferred for short periods since estimating the shape parameter is very uncertain. Blanchet et al. (2015) found that GPD fails to robustly estimate the tail of the distribution be-

cause of a lack of data and unrealistic return levels for very long return periods (when the shape parameter is positive). An additional argument for the EXP model is that it is less affected by observational errors, which plays an important role here.

In this study we use the QQR method based on regression in Q-Q plots proposed by Willems et al. (2007). The exponential Q-Q plot is the extremes x versus minus the logarithmically transformed exceedance probability $1 - G(x)$. The EXP distribution appears as a line in this plot, with slope equal to the scale parameter σ :

$$x = x_t - \sigma \ln(1 - G(x)), \quad (2)$$

where x_t is the threshold level. The same properties hold for the plot of the return level x_T against the return period T when the latter is plotted on a logarithmic scale:

$$x_T = x_t + \sigma \ln(T \cdot M/n), \quad (3)$$

where M is the number of extremes and n the length of the time series.

The estimators for the slope are based on linear regression in the Q-Q plot above the specific threshold level x_t . Amongst the available estimators for σ we used an unconstrained and unweighted linear regression.

The optimal threshold rank t is found by minimization of the mean squared error (MSE) of the calibration. With our datasets, this rank is chosen between 18 and 30 considering the uncertainties and the relatively short period, respectively. Confidence intervals for the scale parameter are computed using a parametric bootstrap technique. The fitted distribution is used to generate 1000 extreme value series with a size corresponding to the optimal rank. The fitting procedure is applied to each of the 1000 series to obtain 1000 simulated scale parameters. The 10 and 90 percentiles of the simulated parameters are taken as the 10 and 90 % confidence interval bounds for the true scale parameter.

3.2 Comparison of 1 h extremes

The extreme events as seen by both the radar and the gauge are compared in Table 2. Since the focus is on the tail of

Table 2. Comparison of the 10 highest 1 h rainfall extremes from the gauge (AWS) and radar (QPE) at Humain and Uccle stations. The events with a high probability of hail have their number in bold. The events are ordered by the maximum of the gauge and radar values.

Humain				
Event	Date yyyy-mm-dd	Time	Gauge (mm h ⁻¹)	Radar (mm h ⁻¹)
1	2016-06-07	18:50	57.65	45.25
2	2005-07-30	00:40	28.60	11.62
3	2014-04-24	15:40	27.00	20.35
4	2014-06-10	21:40	15.60	26.40
5	2007-06-14	01:20	25.80	16.32
6	2009-05-25	13:10	24.10	25.17
7	2008-05-14	17:40	13.10	24.35
8	2015-07-19	01:00	22.87	15.47
9	2009-06-27	14:30	20.40	19.83
10	2009-07-22	21:20	19.80	12.08
11	2010-07-14	15:40	19.80	–
12	2012-06-12	22:20	18.30	15.61
13	2013-03-23	07:40	–	17.30
14	2005-06-28	22:20	–	16.74
Uccle				
Event	Date yyyy-mm-dd	Time	Gauge (mm h ⁻¹)	Radar (mm h ⁻¹)
1	2016-06-07	15:20	18.08	38.21
2	2011-08-23	08:40	35.50	23.22
3	2009-10-07	18:40	30.79	33.32
4	2012-05-20	16:30	12.37	29.79
5	2005-09-10	19:40	29.10	17.54
6	2011-08-18	15:50	28.98	14.77
7	2007-06-14	14:50	21.90	25.88
8	2011-09-03	22:40	25.34	18.46
9	2016-06-11	18:50	–	24.88
10	2005-07-29	19:10	24.29	–
11	2010-07-14	15:20	24.15	–
12	2014-07-29	16:10	20.10	18.17
13	2013-07-27	22:20	20.07	–
14	2008-07-26	10:40	16.60	18.30

the distribution, only the 10 highest values from either the gauge or the radar data are selected. The events for which the probability of hail is high (i.e. when the threshold was applied) are highlighted. An event is considered as problematic if the corresponding radar or gauge extreme rank is below 30. For these events, the underlying precipitation patterns are analysed using the radar images. This comparison allows for identifying the weaknesses of the gauge and radar datasets.

The maximum at Humain has been observed by both the radar and the gauge on 7 June 2016. This relatively high value can be due to randomness and the short period of records. But it is also possible that the other quantiles are underestimated (the maximum was recorded by the new weighted gauge). There is generally a good match between

the radar and the gauge quantiles except for the following events:

- event 2: the radar underestimates globally
- event 7: the gauge is located at the boundary of the convective cell
- event 11: the radar signal is strongly attenuated by a mesoscale convective system
- event 13: there was probably snow in the gauge
- event 14: the gauge is located at the boundary of a convective cell.

The second highest quantile at Uccle has been observed by both the radar and the gauge on the 7 October 2009. There is generally a good match between the two datasets. A few events are problematic:

- event 1, 4: the gauge is at the boundary of a cell
- event 9: there is a stationary storm underestimated by the gauge
- event 10: the gauge is at the boundary of a cell and the radar is attenuated (same as event 2 in Humain)
- event 11: the radar signal is strongly attenuated (same as event 11 in Humain)
- event 13: the radar is attenuated.

The problems with cell boundaries are easily explained: the radar estimation is taken at a given height above ground and the rain is subject to wind drift before reaching the ground. This effect increases with the distance to the radar. Due to its randomness, it should not affect the statistics. The other problematic events can be considered as missing data. Since the level of missingness is limited, the impact on the statistics is expected to be small.

Figure 2 shows the results of the extreme value analysis for the 1 h rainfall accumulation. The return levels are obtained using formulas from Willems et al. (2007), which are based on the Weibull plotting position. Numerical values of the temporal independence, the optimal rank, the location parameter and the scale parameter can be found in Table 3. The percentage of independent peaks (among peaks exceeding the threshold) is around 20 % for both the radar and the gauges at the two locations. This low value is mainly due to the fact that five consecutive values at 10 min resolution are correlated.

The empirical quantiles of the QPE product are systematically slightly lower than those for the AWS gauges. This may be expected as we compare point rainfall observations with rainfall averaged on a 1 km square. However, the underestimation of very high rainfall rates by tipping-bucket gauges can compensate for this effect. One also notes small groups

Table 3. Results of the extreme value distribution fitting at two locations of the AWS network. The tables successively show the temporal independence, optimal rank, the location parameter and the scale parameter. A value is indicated as missing when its extreme rank is below 30.

temporal independence (%)				
Station	Gauge	CAP	QPE	MFB
Humain	25.6	20.7	22.6	–
Uccle	20.8	19.4	21.0	–
optimal rank				
Station	Gauge	CAP	QPE	MFB
Humain	30	30	28	–
Uccle	29	23	30	–
location parameter (mm h^{-1})				
Station	Gauge	CAP	QPE	MFB
Humain	12.2	11.0	10.7	–
Uccle	12.3	13.9	12.3	–
Station	Gauge	CAP	QPE	MFB
Humain	7.5	8.9	6.6	–
Uccle	6.8	10.8	6.4	–

of similar values for both the radar and the gauge, which are mainly associated with hail events. This can be explained by the effect of hail threshold and the rainfall rate limit, respectively. The extremes tend to be heavy tailed but this can be at least partially explained by the observation biases described above.

The fit of the EXP distribution is relatively good for the two locations with a relatively low MSE (not shown). The scale parameter tends to be higher for the gauge data than the radar data. In general, the uncertainty for the scale parameter remains high and this results in wide confidence intervals for higher return periods.

When using the CAP product, the higher quantiles are overestimated especially for Uccle. This can be mainly attributed to the effect of hail. This results in an overestimation of the scale parameter.

3.3 Comparison of 24 h extremes

The comparison of the 10 highest extremes from either the radar (MFB) or the gauge (SPW) can be seen in Table 4. For Uccle, most extreme values occurred during summer and are therefore associated with convective storms. There is a good match between the gauge and the radar except for a few events:

- event 8, 11: the gauge is at the boundary of a convective cell

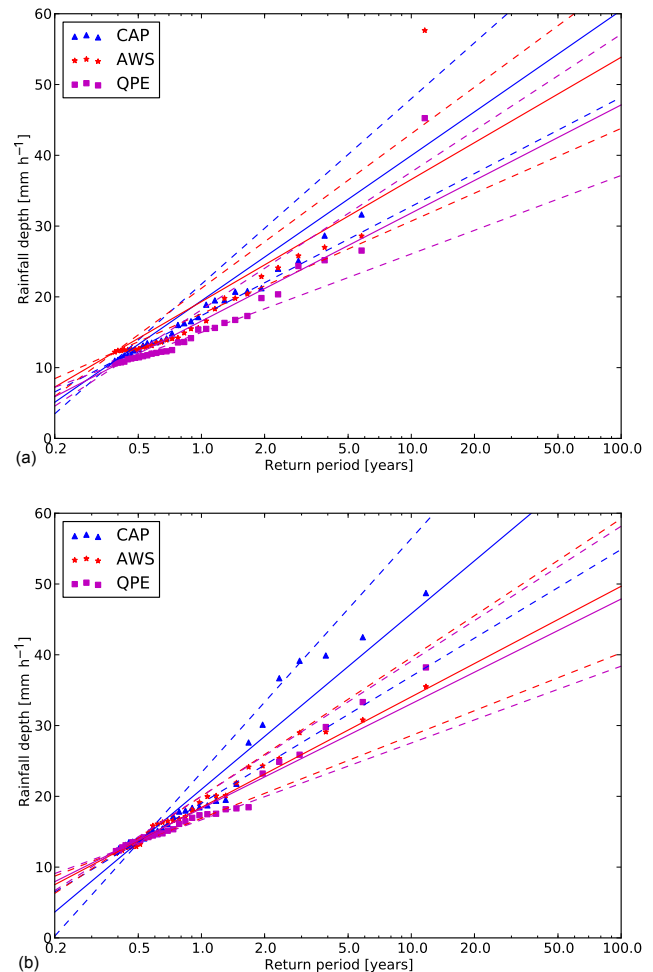


Figure 2. Return levels for 1 h duration at location Humain (a) and Uccle (b) of the AWS gauge (red stars) compared to CAP (blue triangles) and QPE (magenta squares) radar products. The extreme value distribution (solid line) fitted to the extremes and its confidence intervals (dashed line) are also displayed.

- event 13: strong radar attenuation by a mesoscale convective system
- event 14: snow episode probably underestimated by the radar.

For Saint-Vith, the extreme values occurred either in summer or in winter with therefore a mix of convective and widespread precipitation episodes. The match is very good except for the following events:

- event 2: at the boundary of a cell (probably with hail)
- event 3: slight overestimation due to snow melting (QPE); overestimation due to non-uniform bias (MFB)
- event 13: at the boundary of a cell.

Table 4. Comparison of the 10 highest 24 h rainfall extremes from the gauge (SPW) and radar (MFB) at Uccle and Saint-Vith stations. A value is indicated as missing when its extreme rank is below 30. The events are ordered by the maximum of the gauge and radar values.

Uccle				
Event	Date	End time	Gauge mm day ⁻¹	Radar mm day ⁻¹
1	2010-08-16	23:00	63.30	48.99
2	2009-10-07	23:00	52.50	61.83
3	2011-08-23	15:00	59.31	61.00
4	2006-08-03	23:00	43.00	58.44
5	2016-05-30	23:00	35.30	53.34
6	2014-08-26	15:00	45.30	48.51
7	2012-10-04	08:00	34.60	45.63
8	2012-06-12	11:00	–	44.87
9	2016-06-12	17:00	31.30	39.45
10	2011-09-04	21:00	38.70	26.10
11	2015-08-16	03:00	–	37.75
12	2007-06-15	11:00	36.99	33.91
13	2014-07-10	04:00	36.90	–
14	2016-01-16	02:00	36.30	–

Saint-Vith				
Event	Date	End time	Gauge mm day ⁻¹	Radar mm day ⁻¹
1	2007-01-18	16:00	74.60	56.88
2	2009-07-03	16:00	37.90	61.68
3	2011-12-16	23:00	–	56.62
4	2012-07-28	21:00	53.60	46.72
5	2012-10-04	12:00	49.70	39.86
6	2007-08-22	19:00	47.50	48.73
7	2010-08-16	03:00	45.80	55.50
8	2006-08-05	06:00	43.70	41.10
9	2007-12-03	08:00	43.40	46.09
10	2007-09-28	08:00	42.40	38.87
11	2014-09-21	14:00	34.00	40.71
12	2016-05-31	02:00	40.01	33.44
12	2016-07-23	21:00	40.00	–

The problematic events not related to boundary effects can be considered as missing data. Since they are limited it is expected that they only slightly affect the statistics.

Figure 3 shows the results of the extreme value analysis for the 24 h rainfall accumulation. Numerical values can be found in Table 5. The percentage of independent peaks (amongst peaks exceeding the threshold) is between 6 and 9 % for the two locations and for all datasets. This is what we expect from a 24 h accumulation available every hour.

For Uccle there are not many differences between QPE and MFB because most events are associated with convective storms. Compared to the gauge quantiles, the radar quantiles are lower below 1-year and higher between 1- and 5-year return periods. This can be attributed mainly to hail overestimation by the radar and gauge losses. It results in a higher

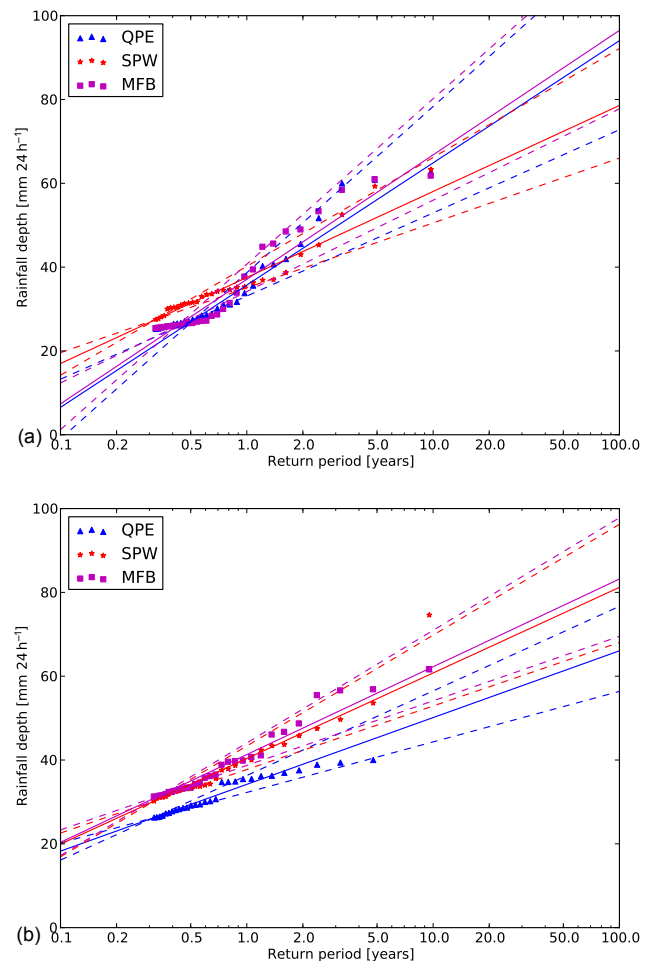


Figure 3. Return levels for 24 h duration at location Uccle (a) and Saint-Vith (b) of the SPW gauge (red stars) compared to QPE (blue triangles) and MFB (magenta squares) radar products. The extreme value distribution (solid line) fitted to the extremes and its confidence intervals (dashed line) are also displayed.

scale for the radar, which is close to the upper bound of the gauge confidence interval.

For Saint-Vith, there is a clear effect of the bias correction (MFB) to remove the underestimation of the QPE product. As for Uccle, the radar quantiles are higher for return periods higher than 2 years but the effect is limited because less convective storms are involved. The final result is a good match of the two distributions for this station.

For the two stations, no significant instability in the MFB values have been found.

For Uccle, the CAP product overestimates the scale parameter and underestimates the location parameter due to hail and VPR errors, respectively. For Saint-Vith, the quantiles (not shown) are similar to QPE except for a very high unrealistic maximum.

Table 5. Results of the extreme value distribution fitting at two locations of the SPW network. The tables shows successively the temporal independence, optimal rank, the location parameter and the scale parameter.

temporal independence (%)				
Station	Gauge	CAP	QPE	MFB
Uccle	7.1	6.0	6.6	6.7
St-Vith	7.4	8.4	9.0	8.4
Station	Gauge	CAP	QPE	MFB
Uccle	30	26	19	23
St-Vith	30	30	30	28
Station	Gauge	CAP	QPE	MFB
Uccle	27.2	25.0	27.2	27.5
St-Vith	30.2	25.8	26.3	31.5
scale parameter (mm day ⁻¹)				
Station	Gauge	CAP	QPE	MFB
Uccle	9.0	13.5	12.7	12.9
St-Vith	8.9	8.2	6.9	9.1

4 RFA

4.1 Methodology

As in Overeem et al. (2009) and Wright et al. (2014b) we consider that the extreme statistics are the same within the region. The region should be sufficiently large to have a large sample size (many extremes) and small enough to neglect extreme statistic variability. No strong variability is expected in Belgium because it is a relatively flat country. Therefore we define our region as the radius of 20 km around the target location. A similar size has been used in other radar studies (e.g. Overeem et al., 2009; Wright et al., 2014b; Eldardiry et al., 2015).

We also consider that the extremes observed within the 20 km radius during a time window of 12 h are dependent. As in Wright et al. (2014b), we keep only the maximum amongst dependent values. We therefore implicitly assume that the regional maximum follows the same distribution as the local extremes. The possible benefit of taking one extreme value at random is an open question. It is important to remember that we are interested in the extreme statistics of any given pixel in the region. This is different from studying the extreme statistics of the maximum rainfall over the region as in Panziera et al. (2016). We also tested the hypothesis that 1 h extremes are independent after a certain distance which is set to 10 km. This distance corresponds to the maximum expected size of a convective cell (Goudenhoofdt and Delobbe, 2013). If this is true it allows to reduce the uncertainty of the

analysis. In the text, we will refer to these datasets by the names REG and R10, respectively.

Due to the spatial dependence, the effective length n_{eff} of the pooled time series is smaller than the total length of the records. The total length is obtained by multiplying the number of years n by the number of pixels N :

$$n_{\text{max}} = n \times N. \quad (4)$$

In this study n_{eff} is computed by multiplying n_{max} by the fraction of spatially independent peaks, amongst peaks exceeding the threshold. The latter is obtained by dividing the number of independent peaks by the total number of peaks. It can be shown that this is the same as the method based on the averaged exceedance rate found in Wright et al. (2014b) and explained in detail by Weiss et al. (2014). The large number of peaks available from the radar data allows us to choose a higher threshold rank. This increase in sample size leads to a more reliable extreme value analysis, which is the final goal of this radar-based RFA. Accordingly the QQR method is applied for threshold ranks between 30 and 100 and the optimal rank is found.

4.2 Comparison with rain gauges

Figures 4 and 5 show the results of the RFA for 1 h rainfall accumulation at the four locations selected from the BUL network. The results of the on-site frequency analysis for the gauge and collocated radar pixels are shown as reference. Numerical values can be found in Table 6. The percentage of temporally independent extremes for the gauge is close to 30 % for Deurne and Uccle, while it is slightly above 20 % for the two other stations. This suggests that there are larger clusters which might be related to altitude. Above the threshold, the percentage of spatially independent extremes (REG) ranges from 1.1 % (Uccle) to 2.6 % (Nadrin). The effective period length of the pooled dataset is then between 200 and 500 years. Using a decorrelation distance of 10 km results in twice as much data, which is more than one expects from randomness. It suggests that convection can be organized on large spatial scales.

The radar images associated with each maximum of the radar-based RFA is analysed:

- Deurne and Uccle (28 July 2006): several supercells in the whole of Belgium
- Gosselies (22 August 2011): a squall line moving parallel to the flow
- Nadrin (26 July 2008): a stationary convective cell.

The highest extremes exhibit abrupt variations in the form of steps for both the gauge and radar. This could be explained by the siphonage of the gauge and hail threshold, respectively. Since Nadrin is close to the radar, the standard $Z-R$ relationship is used instead of the convective $Z-R$ relationship. This permits higher rain rates (i.e. 100 mm h⁻¹).

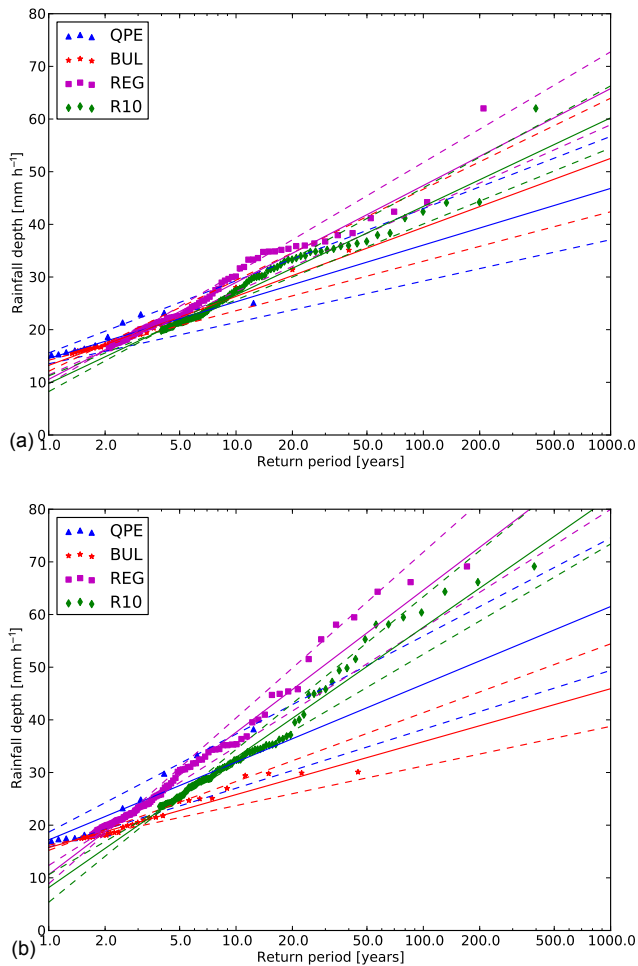


Figure 4. Return levels for 1 h duration at location Deurne (a) and Uccle (b) from the BUL gauge data (red stars) compared to the on-site QPE (blue triangle), REG (purple square) and R10 (green diamond) radar data. The extreme value distribution (solid line) fitted to the extremes and its confidence intervals (dashed line) are also displayed.

The gauge extremes are relatively low at Deurne and Uccle compared to Gosselies and Nadrin. The radar extremes are lower at Deurne compared to the other stations. This can be at least partially attributed to the large sample volume at this range. The match between the gauge and the radar (REG and R10) is good except at Uccle with much higher radar extremes. The REG exhibits higher extremes than R10 suggesting some dependence beyond 10 km. Indeed the results should be similar if the hypothesis of independence after 10 km was valid.

This can be partially attributed to hail but the similar four highest extremes suggest a gauge limitation. It is also striking that half of the 20 highest gauge extremes occurred during the period 1999–2008 (not shown). This positive trend for Uccle is possibly related to the urban heat island effect (Hamdi and Van de Vyver, 2011). The uncertainty of the

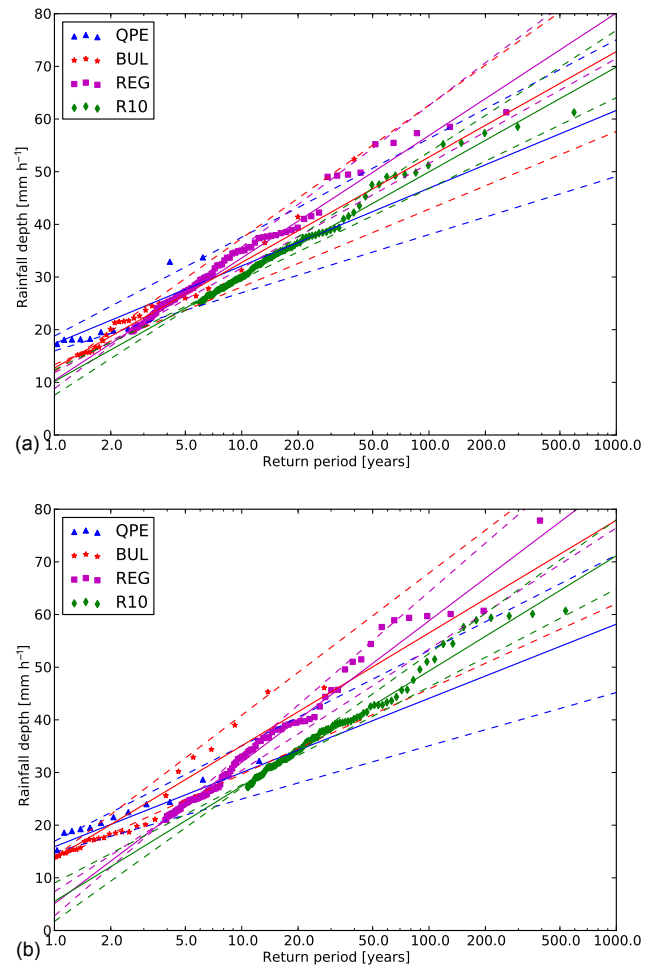


Figure 5. Return levels for 1 h duration at location Gosselies (a) and Nadrin (b) from the BUL gauge data (red stars) compared to the QPE (blue triangle), REG (purple square) and R10 (green diamond) radar data. The extreme value distribution (solid line) fitted to the extremes and its confidence intervals (dashed line) are also displayed.

radar fit is low because of the larger sample size, due to which a higher rank can be chosen. Furthermore, the fit is less impacted by the potentially large errors of the highest extremes. The location parameter (corresponding to the threshold) increases with the sample size of the products.

Except for the Uccle station, the scale parameter is the lowest for the QPE dataset due to the bias as a result of the small sample size. The scale parameter of the pooled radar datasets is slightly higher at Deurne and significantly higher in Uccle. For Gosselies and Nadrin, the R10 and BUL data have similar scales while it is slightly higher for the REG data. The fit to the REG and R10 data is within the uncertainty bound of the fit to the BUL data. For those two stations, the fit to the BUL data is even in the small uncertainty bound of the fit to the REG data.

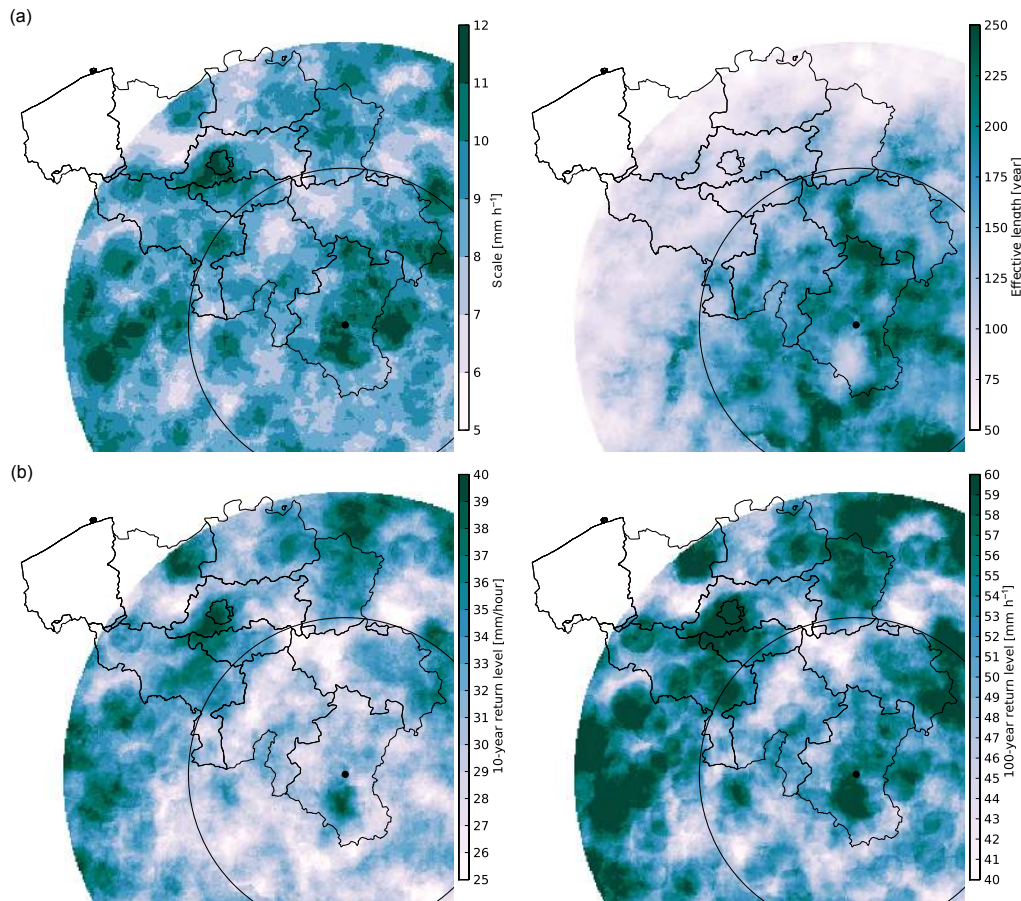


Figure 6. Results of the regional frequency analysis (RFA) for 1 h duration applied over Belgium up to 180 km from the radar. The scale parameter and the effective length are shown in (a). The levels corresponding to a 10- and 100-year return periods are shown in (b). A circle with a radius of 100 km centred at the radar is also drawn.

4.3 Spatial maps

We apply the RFA described above for 1 h duration to all pixel locations in Belgium with some modifications. We use a smaller radius of 10 km to reduce the computation cost and consider that all pixels are spatially dependent. This smaller radius improves the resolution of the maps at the expense of a higher uncertainty. Several pixels in the radar dataset are affected by permanent non-meteorological echoes. They can be identified by an unrealistic high frequency of extremes. In practice one looks at the distribution of the number of values exceeding 12 mm h^{-1} . The pixels with more than 50 exceedances have been found as outliers and removed. To make the comparison easier, we choose a fixed threshold rank of 60. No larger ranks have been considered due to computational limitations.

Figure 6 shows the results of the RFA applied to Belgium. The provinces of Belgium are also displayed to help comparison between the maps. No values are shown beyond the 180 km range because the quality of the radar QPE is significantly reduced. The return periods are computed using

Eq. (3) and therefore depend on the scale parameter and the effective length. The higher the scale the higher the difference between the 10- and 100-year return levels.

Some artifacts due to the radar and the regional approach can be seen on the maps. The effective length decreases significantly beyond 100 km meaning that the spatial dependence increases. This is due to the fact that the actual radar sample is larger than the 1 km pixel at those ranges. Circular patterns appear on the maps due to the influence of the pixels located at their centres. The high values are caused by pixels contaminated by non-meteorological echoes (e.g. at the German border) and hail. A stronger filter for non-meteorological echoes is not used because it could remove actual precipitation information. The circular effect might be reduced by using a larger radius or a higher threshold rank but this is computationally expensive. Areas with a 10-year return level exceeding 30 mm are mainly located beyond 100 km. This is probably due to an increased contamination by hail with the distance to the radar (and the height of the measurements). The small-scale variability in the study area can be explained by uncertainties due to the sample size.

Table 6. Results of the extreme value distribution fitting for the RFA. The tables shows successively the independence (temporal or spatial), the optimal rank, the location parameter and the scale parameter.

independence (%)				
Station	QPE	BUL	REG	R10
Deurne	–	27.5	1.4	2.6
Uccle	–	28.0	1.1	2.6
Gosselies	–	22.2	1.7	3.9
Nadrin	–	19.9	2.6	7.0
optimal rank (%)				
Station	QPE	BUL	REG	R10
Deurne	28	22	100	99
Uccle	30	30	70	88
Gosselies	29	30	96	90
Nadrin	23	30	100	91
location parameter (mm h^{-1})				
Station	QPE	BUL	REG	R10
Deurne	10.8	16.7	16.5	20.0
Uccle	11.5	17.5	21.1	24.2
Gosselies	11.9	15.2	20.4	26.5
Nadrin	12.2	12.9	21.0	29.0
scale parameter (mm h^{-1})				
Station	QPE	BUL	REG	R10
Deurne	4.7	5.7	8.0	7.3
Uccle	6.4	4.4	11.7	10.7
Gosselies	6.4	8.7	10.1	8.6
Nadrin	6.1	9.3	11.7	9.5

There is some correlation between the 10-year return level and the scale parameter. Therefore the spatial patterns between the two return periods are similar. Within the 100 km radius, the maps are only slightly influenced by the topography and the mean annual rainfall (Journée et al., 2015). This suggests that applying our regional approach is valid, at least for 1 h duration. Van de Vyver (2012) obtained slightly lower values for the 10-year return level but a slightly higher 100-year return level due to the positive shape parameter. One notes that the scale is very high around the Brussels region where the Uccle station is located.

5 Conclusions

5.1 Results

The potential of a radar-based precipitation dataset to study extreme rainfall at a given location is evaluated. The QPE is obtained by a careful processing of the volumetric reflectivity measurements from a single weather radar in Belgium. The

radar dataset covers the period 2005–2016, has a resolution of 1 km and is available every 5 min.

The first evaluation is based on a comparison of the extreme statistics between the radar dataset and two automatic rain gauge networks with 10 min and 1 h resolution, respectively. For each network, two locations are chosen to study sliding 1 and 24 h extremes using the collocated radar estimation. A regression method in Q-Q plots is used to fit an exponential distribution to independent peaks. This method has the property to focus on the tail of the extreme value distribution, which is of interest when studying extremes. An optimal threshold rank is selected by minimizing the MSE of the regression.

The 10 highest 1 h extremes occurred in summer and are well captured by both the radar and the gauge. A few problematic events are caused by wind drift or severe radar signal attenuation and should be considered as missing data. Differences up to 30 % between the gauge and radar values are observed and can be explained by spatial sampling and estimation errors. The radar extremes tend to be lower than the gauge extremes especially for short return periods. This is consistent with the results of Peleg et al. (2016) on the small-scale spatial variability of extreme rainfall. In particular, tipping-bucket gauges underestimate the heavy rainfall rate and can be blocked by accumulated snow. The radar underestimates due to signal attenuation and overestimates in the case of hail. Additional radar uncertainties come from time sampling and the $Z-R$ relationship. Despite the uncertainties in the datasets, the fitting of the exponential distribution to the QPE product is within the large uncertainty bound of the AWS one. This result is in accordance with the fact that the temporal variability (related to the sample size) is higher than the spatial variability (Peleg et al., 2017).

For the 24 h accumulation there is a mix of summer and winter events, with more of the latter for stations with higher altitude. There is a clear benefit of bias correction for the highest station, making the distribution fits similar for both stations. For both 1 and 24 h accumulations, the basic radar product exhibits unrealistic high extremes, which results in an overestimated scale parameter. Such a product is therefore not suitable for an extreme value analysis.

In the second evaluation, a RFA is applied to 1 h radar data at the location of four pluviographs with recordings from 1965 to 2010. Spatially independent extremes within a circle of 20 km are selected using a novel approach. They are fitted with a maximum threshold rank extended from 30 to 100 thanks to the increased sample size. There is a good agreement between the radar and the gauge for the two closest stations. The most important result is that the uncertainty is significantly lower using the available radar data. The extremes are lower when a decorrelation distance of 10 km is assumed suggesting that this hypothesis is not valid. In Uccle, the radar extremes and therefore the scale parameter are significantly higher. This can be attributed partially to radar overestimation due to hail and gauge underestimations, but

the increasing urban heat island effect should not be ruled out. The decreasing tail of the radar extremes is at least partially caused by the hail threshold but a physical limit for the Belgium climate could play a role.

For each of the rain gauge networks, only a few stations have been selected and presented in this paper. The results from these stations are representative of the variability of the results obtained from the other stations.

The regional approach has been applied all over the study area using a 10 km radius and a fixed threshold rank of 60. The extreme statistics for the 1 h duration are slightly influenced by the topography. The reliability of the radar results beyond the 100 km range is questionable.

5.2 Prospects

There is still some room to improve the quality of the radar and gauge datasets. The recently installed weighted gauges are able to cope with intense rainfall and snowfall. One will have to wait a few decades before it can produce reliable statistics. Radar calibration errors can be mitigated by computing a monthly bias using rain gauges. The attenuation can be solved easily by using other radars when available. To avoid overestimation of the extremes, an advection correction can be used for the time sampling error. Dual-polarization radars can potentially provide a better estimation for high rainfall rates (Figueroa i Ventura and Tabary, 2013). However uncertainties related to the relation between the radar measurements and the rainfall rate remain, especially in the case of hail. In this study we considered all data as the amount of liquid water at the ground. For some applications it could be necessary to take the melting of snow and hail into account. Identification of hail at ground level is a challenging problem using radar and ground station networks (Lukach et al., 2017).

Since the paper focuses on the comparison between radar and rain gauges, the extreme value analysis has been kept simple. While the EXP distribution was found to fit generally well with the empirical data, the generalized Pareto distribution should be considered as well for the RFA. The analysis of longer durations can be refined by taking into consideration the effect of the type of rainfall (e.g. Rulfova et al., 2014; Panziera et al., 2016). A bias correction should also be considered for a proper handling of the asymptotic behaviour of the distribution (Willems et al., 2007).

The extreme value theory was applied to the radar datasets by removing the spatially dependent extremes in the region of analysis. This is performed using a simple technique based on a decorrelation distance. Evin et al. (2016) decided not to use such a method because it reduces the sample size. Better performance is expected using recently proposed statistical models (Buishand et al., 2008; Davison et al., 2012).

The radar-based RFA can be extended to other durations to derive IDF curves. Note that the hypothesis of constant parameter over the region might not be valid for longer du-

urations. In many applications in hydrology, it is the averaged rainfall over a given area which is relevant. A popular technique is to apply areal reduction factors to point-based statistics. The radar dataset can be used directly to derive areal rainfall statistics (e.g. Durrans et al., 2002; Overeem et al., 2010; Wright et al., 2014a).

Code and data availability. The code used in this study is part of the RMIB radar library. The rain gauge rainfall measurements and radar-based precipitation estimates are archived at the RMIB.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors would like to thank Francesco Marra, Luca Panziera and referee no. 3 for their very constructive comments which allowed us to significantly improve the quality of the paper. We thank the hydrological service of the Walloon region (SPW) for providing their rain gauge data. The comments of Michel Journee and Hans Van de Vyver are highly appreciated.

Edited by: Uwe Ehret

Reviewed by: Francesco Marra, Luca Panziera, and one anonymous referee

References

- Bardet, L., Duluc, C.-M., Rebour, V., and L'Her, J.: Regional frequency analysis of extreme storm surges along the French coast, *Nat. Hazards Earth Syst. Sci.*, 11, 1627–1639, <https://doi.org/10.5194/nhess-11-1627-2011>, 2011.
- Blanchet, J., Touati, J., Lawrence, D., Garavaglia, F., and Paquet, E.: Evaluation of a compound distribution based on weather pattern subsampling for extreme rainfall in Norway, *Nat. Hazards Earth Syst. Sci.*, 15, 2653–2667, <https://doi.org/10.5194/nhess-15-2653-2015>, 2015.
- Buishand, T.: Extreme rainfall estimation by combining data from several sites, *Hydrolog. Sci. J.*, 36, 345–365, 1991.
- Buishand, T. A., de Haan, L., and Zhou, C.: On spatial extremes: With application to a rainfall problem, *Ann. Appl. Stat.*, 2, 624–642, <https://doi.org/10.1214/08-AOAS159>, 2008.
- Castellari, A.: Probabilistic envelope curves for design flood estimation at ungauged sites, *Water Resour. Res.*, 43, W04406, <https://doi.org/10.1029/2005WR004384>, 2007.
- Colli, M., Lanza, L. G., and La Barbera, P.: An evaluation of the uncertainty of extreme events statistics at the WMO/CIMO Lead Centre on precipitation intensity, *AGU Fall Meeting Abstracts*, 2012.
- Cooly, D., Nychka, D., and Naveau, P.: Bayesian Spatial Modeling of Extreme Precipitation Return Levels, *J. Am. Stat. Assoc.*, 102, 824–840, <https://doi.org/10.1198/016214506000000780>, 2007.

- Davison, A. C., Padoan, S. A., and Ribatet, M.: Statistical Modeling of Spatial Extremes, *Stat. Sci.*, 27, 161–186, <https://doi.org/10.1214/11-sts376>, 2012.
- Demarée, G.: Le pluviographe centenaire du plateau d'Uccle: son histoire, ses données et ses applications, *La Houille Blanche*, 4, 95–102, <https://doi.org/10.1051/lhb/2003082>, 2003.
- Duchon, C. E. and Biddle, C. J.: Undercatch of tipping-bucket gauges in high rain rate events, *Adv. Geosci.*, 25, 11–15, <https://doi.org/10.5194/adgeo-25-11-2010>, 2010.
- Durrans, S. R., Julian, L. T., and Yekta, M.: Estimation of depth-area relationships using radar-rainfall data, *J. Hydrol. Eng.*, 7, 356–367, 2002.
- Eldardiry, H., Habib, E., and Zhang, Y.: On the use of radar-based quantitative precipitation estimates for precipitation frequency analysis, *J. Hydrol.*, 531, 441–453, <https://doi.org/10.1016/j.jhydrol.2015.05.016>, 2015.
- Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., and Penot, D.: A regional model for extreme rainfall based on weather patterns subsampling, *J. Hydrol.*, 541, 1185–1198, <https://doi.org/10.1016/j.jhydrol.2016.08.024>, 2016.
- Figueras i Ventura, J. and Tabary, P.: The new French operational polarimetric radar rainfall rate product, *J. Appl. Meteorol. Clim.*, 52, 1817–1835, 2013.
- Gellens, D.: Trend and Correlation Analysis of k-Day Extreme Precipitation over Belgium, *Theor. Appl. Climatol.*, 66, 117–129, <https://doi.org/10.1007/s007040070037>, 2000.
- Goudenhoofdt, E. and Delobbe, L.: Statistical Characteristics of Convective Storms in Belgium Derived from Volumetric Weather Radar Observations, *J. Appl. Meteorol. Clim.*, 52, 918–934, <https://doi.org/10.1175/JAMC-D-12-079.1>, 2013.
- Goudenhoofdt, E. and Delobbe, L.: Generation and Verification of Rainfall Estimates from 10-Yr Volumetric Weather Radar Measurements, *J. Hydrometeorol.*, 17, 1223–1242, <https://doi.org/10.1175/JHM-D-15-0166.1>, 2016.
- Haberlandt, U. and Berndt, C.: The value of weather radar data for the estimation of design storms – an analysis for the Hannover region, *Proceedings of the International Association of Hydrological Sciences*, 373, 81–85, 2016.
- Hamdi, R. and Van de Vyver, H.: Estimating urban heat island effects on near-surface air temperature records of Uccle (Brussels, Belgium): an observational and modeling study, *Adv. Sci. Res.*, 6, 27–34, <https://doi.org/10.5194/asr-6-27-2011>, 2011.
- Hosking, J. R. M. and Wallis, J. R.: The effect of intersite dependence on regional flood frequency analysis, *Water Resour. Res.*, 24, 588–600, <https://doi.org/10.1029/WR024i004p00588>, 1988.
- Jakob, D., Karoly, D. J., and Seed, A.: Non-stationarity in daily and sub-daily intense rainfall – Part 1: Sydney, Australia, *Nat. Hazards Earth Syst. Sci.*, 11, 2263–2271, <https://doi.org/10.5194/nhess-11-2263-2011>, 2011.
- Journée, M., Delvaux, C., and Bertrand, C.: Precipitation climate maps of Belgium, *Adv. Sci. Res.*, 12, 73–78, <https://doi.org/10.5194/asr-12-73-2015>, 2015.
- Lukach, M., Foresti, L., Giot, O., and Delobbe, L.: Estimating the occurrence and severity of hail based on 10 years of observations from weather radar in Belgium, *Meteorol. Appl.*, 24, 250–259, 2017.
- Marra, F. and Morin, E.: Use of radar {QPE} for the derivation of Intensity-Duration-Frequency curves in a range of climatic regimes, *J. Hydrol.*, 531, 427–440, <https://doi.org/10.1016/j.jhydrol.2015.08.064>, 2015.
- Marra, F., Morin, E., Peleg, N., Mei, Y., and Anagnostou, E. N.: Intensity-duration-frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean, *Hydrol. Earth Syst. Sci.*, 21, 2389–2404, <https://doi.org/10.5194/hess-21-2389-2017>, 2017.
- Martins, E. S. and Stedinger, J. R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, 36, 737–744, <https://doi.org/10.1029/1999WR900330>, 2000.
- Ntegeka, V. and Willems, P.: Trends and multidecadal oscillations in rainfall extremes, based on a more than 100-year time series of 10 m in rainfall intensities at Uccle, Belgium, *Water Resour. Res.*, 44, W07402, <https://doi.org/10.1029/2007WR006471>, 2008.
- Nystuen, J. A.: Relative Performance of Automatic Rain Gauges under Different Rainfall Conditions, *J. Atmos. Ocean. Tech.*, 16, 1025–1043, [https://doi.org/10.1175/1520-0426\(1999\)016<1025:RPOARG>2.0.CO;2](https://doi.org/10.1175/1520-0426(1999)016<1025:RPOARG>2.0.CO;2), 1999.
- Ootegem, L. V., Herck, K. V., Creten, T., Verhofstadt, E., Foresti, L., Goudenhoofdt, E., Reyniers, M., Delobbe, L., Tuyls, D. M., and Willems, P.: Exploring the potential of multivariate depth-damage and rainfall-damage models, *J. Flood Risk Manage.*, <https://doi.org/10.1111/jfr3.12284>, online first, 2016.
- Overeem, A., Buishand, T. A., and Holleman, I.: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar, *Water Resour. Res.*, 45, W10424, <https://doi.org/10.1029/2009WR007869>, 2009.
- Overeem, A., Buishand, T. A., Holleman, I., and Uijlenhoet, R.: Extreme value modeling of areal rainfall from weather radar, *Water Resour. Res.*, 46, W09514, <https://doi.org/10.1029/2009WR008517>, 2010.
- Paixao, E., Mirza, M. M. Q., Shephard, M. W., Auld, H., Klaassen, J., and Smith, G.: An integrated approach for identifying homogeneous regions of extreme rainfall events and estimating {IDF} curves in Southern Ontario, Canada: Incorporating radar observations, *J. Hydrol.*, 528, 734–750, <https://doi.org/10.1016/j.jhydrol.2015.06.015>, 2015.
- Panziera, L., Gabella, M., Zanini, S., Hering, A., Germann, U., and Berne, A.: A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland, *Hydrol. Earth Syst. Sci.*, 20, 2317–2332, <https://doi.org/10.5194/hess-20-2317-2016>, 2016.
- Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., and Burlando, P.: Spatial variability of extreme rainfall at radar subpixel scale, *J. Hydrol.*, <https://doi.org/10.1016/j.jhydrol.2016.05.033>, online first, 2016.
- Peleg, N., Blumensaat, F., Molnar, P., Fatichi, S., and Burlando, P.: Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling, *Hydrol. Earth Syst. Sci.*, 21, 1559–1572, <https://doi.org/10.5194/hess-21-1559-2017>, 2017.
- Pickands III, J.: Statistical inference using extreme order statistics, *Ann. Stat.*, 3, 119–131, 1975.
- Roth, M., Buishand, T., Jongbloed, G., Tank, A. K., and van Zanten, J.: Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany, *Weather and Climate*

- Extremes, 4, 1–10, <https://doi.org/10.1016/j.wace.2014.01.001>, 2014.
- Roth, M., Jongbloed, G., and Buishand, T.: Threshold selection for regional peaks-over-threshold data, *J. Appl. Stat.*, 43, 1291–1309, <https://doi.org/10.1080/02664763.2015.1100589>, 2015.
- Rulfova, Z., Buishand, A., Kysely, J., and Roth, M.: Two-Component Extreme Value Distributions for Convective and Stratiform Precipitation, AGU Fall Meeting Abstracts, 2014.
- Saito, H. and Matsuyama, H.: Probable Hourly Precipitation and Soil Water Index for 50-yr Recurrence Interval over the Japanese Archipelago, *SOLA*, 11, 118–123, <https://doi.org/10.2151/sola.2015-028>, 2015.
- Sapiano, M. R. P. and Arkin, P. A.: An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data, *J. Hydrometeorol.*, 10, 149–166, <https://doi.org/10.1175/2008JHM1052.1>, 2009.
- Sveinsson, O. G. B., Boes, D. C., and Salas, J. D.: Population index flood method for regional frequency analysis, *Water Resour. Res.*, 37, 2733–2748, <https://doi.org/10.1029/2001wr000321>, 2001.
- Svensson, C. and Jones, D. A.: Review of rainfall frequency estimation methods, *J. Flood Risk Manage.*, 3, 296–313, <https://doi.org/10.1111/j.1753-318X.2010.01079.x>, 2010.
- Uboldi, F., Sulis, A. N., Lussana, C., Cislighi, M., and Russo, M.: A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution, *Hydrol. Earth Syst. Sci.*, 18, 981–995, <https://doi.org/10.5194/hess-18-981-2014>, 2014.
- Uijlenhoet, R.: Raindrop size distributions and radar reflectivity-rain rate relationships for radar hydrology, *Hydrol. Earth Syst. Sci.*, 5, 615–628, <https://doi.org/10.5194/hess-5-615-2001>, 2001.
- Van de Vyver, H.: Spatial regression models for extreme precipitation in Belgium, *Water Resour. Res.*, 48, W09549, <https://doi.org/10.1029/2011WR011707>, 2012.
- Vannitsem, S. and Naveau, P.: Spatial dependences among precipitation maxima over Belgium, *Nonlin. Processes Geophys.*, 14, 621–630, <https://doi.org/10.5194/npg-14-621-2007>, 2007.
- Weiss, J., Bernardara, P., and Benoit, M.: Modeling inter-site dependence for regional frequency analysis of extreme marine events, *Water Resour. Res.*, 50, 5926–5940, <https://doi.org/10.1002/2014WR015391>, 2014.
- Willems, P.: Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types, *J. Hydrol.*, 233, 189–205, [https://doi.org/10.1016/S0022-1694\(00\)00233-X](https://doi.org/10.1016/S0022-1694(00)00233-X), 2000.
- Willems, P.: Multidecadal oscillatory behaviour of rainfall extremes in Europe, *Climatic Change*, 120, 931–944, <https://doi.org/10.1007/s10584-013-0837-x>, 2013.
- Willems, P., Guillou, A., and Beirlant, J.: Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function, *J. Hydrol.*, 338, 221–236, <https://doi.org/10.1016/j.jhydrol.2007.02.035>, 2007.
- Wright, D. B., Smith, J. A., and Baeck, M. L.: Critical Examination of Area Reduction Factors, *J. Hydrol. Eng.*, 19, 769, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000855](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000855), 2014a.
- Wright, D. B., Smith, J. A., and Baeck, M. L.: Flood frequency analysis using radar rainfall fields and stochastic storm transposition, *Water Resour. Res.*, 50, 1592, <https://doi.org/10.1002/2013WR014224>, 2014b.