Short-duration rainfall extremes in Sweden: a regional analysis

Jonas Olsson, Johan Södling, Peter Berg, Lennart Wern and Anna Eronn

ABSTRACT

We investigate short-duration rainfall extremes in a 22-year long (1996–2017) time series of 15-min values from 128 stations in Sweden. Two different approaches are used to fit extreme value distributions (EVDs) to the data: generalized extreme value distribution with annual maximum values and generalized Pareto distribution with peak-over-threshold values. First, EVDs are fitted to the data from each station, and the extreme value parameters are used together with the station coordinates in a cluster analysis. Based on the clustering results, a division of Sweden into four regions is proposed. Following the station-year method, the time series from all stations in each cluster are pooled into regional series after which depth–duration–frequency (DDF) statistics are calculated. Averaged over durations and return periods, the highest DDF values are found in region south-west, being ~25% higher than the lowest ones found in the northern region. A six-parameter DDF function is proposed which, together with a simple empirical expression for the associated confidence interval, allows for easy DDF estimation at arbitrary durations and return periods, including uncertainty. No tendencies of change in magnitude or frequency of the short-duration extremes are found during the period.

Key words | cloudburst, depth–duration–frequency, precipitation, statistics

Jonas Olsson (corresponding author)
Peter Berg
Anna Eronn
Research and Development,
Swedish Meteorological and Hydrological Institute,
601 76 Norrköping,
Sweden
E-mail: *jonas.olsson@smhi.se*

Johan Södling
Professional Services,
Swedish Meteorological and Hydrological Institute,
601 76 Norrköping,
Sweden

Lennart Wern
Core Services,
Swedish Meteorological and Hydrological Institute,
601 76 Norrköping,
Sweden

INTRODUCTION

Intense short-duration rainfall ('cloudburst') is a form of extreme weather phenomenon that causes considerable problems and costs for society. In Sweden, the consequences are generally limited to material damage of buildings and infrastructure, but life-threatening situations may arise, e.g., when debris flow is triggered. Recently, cloudbursts have been in the spotlight in connection with some cases of urban flooding, notably in Copenhagen 2011 (just a few

miles from Malmö, the third largest city in Sweden) and in Malmö 2014, the latter with estimated economic damage of 31.5 million EUR which makes it the worst urban flooding on record in Sweden (Hernebring *et al.* 2015). Further, rainfall extremes are expected to intensify as a consequence of global warming (e.g., Stocker *et al.* 2013; Olsson & Foster 2014). For (re-)designing and 'climate proofing' society, updated and accurate short-duration rainfall extreme statistics are crucial.

In Sweden, national depth–duration–frequency (DDF) statistics were long based on the approach of Dahlström (1979). There, regional statistics were derived using a relationship to total rainfall during May, July and August.

In 2010, this approach was rejected in favour of one nationally applicable DDF equation, based on cloud-physical reasoning and calibrated to observations (Dahlström 2010), which is recommended in today's design guidelines (Svenskt Vatten 2011). The calibration was done mainly against data from 15 municipal tipping-bucket gauge records of varying lengths, most of them located in southern Sweden, with some additional analyses of data from the national meteorological station network. In this network, 15-min precipitation has been measured in ~130 stations since 1996. An analysis of the first 10 years of data from these stations was performed by Wern & German (2009), who estimated DDF statistics and compared them to the Dahlström (2010) equation. Overall, the statistics agreed rather well but differences were found both for short durations at long return periods and for long durations at short return periods. Further, Wern & German (2009) found rather distinct regional differences, with generally higher extremes in the south than in the north. Hernebring *et al.* (2012) related DDF statistics in Europe to mean summer temperature and total summer precipitation, and found an overall similar variation over Sweden.

Now, with more than 20 years of data from the 15-min stations, we have performed a comprehensive regional analysis with the objective of providing DDF statistics with particularly three advancements, as compared to today's guidelines and DDF statistics:

- Regional values: robust DDF statistics for climatologically similar sub-regions of Sweden, in contrast to today's national statistics.
- Long return periods: estimated depths for return periods >10 years, which is the maximum in today's DDF statistics.
- Quantified uncertainty: upper and lower confidence limits, which is missing in today's DDF statistics.

The analysis is based on established statistical methods and concepts, which to our knowledge are partly applied in new ways. The intention has been to keep the analysis empirical, data-driven and transparent. Besides the national 15-min data, municipal tipping-bucket gauge data and gauge-adjusted weather radar data are employed in supporting analyses.

STUDY AREA AND DATA

The main data set used in this study, hereafter referred to as 'SMHI data', originates from the national network of 128 automatic meteorological stations operated by the Swedish Meteorological and Hydrological Institute, SMHI (Figure 1). Precipitation is measured by Geonor weighing gauges with a time step of 15 min. The gauge uses a precision vibrating wire transducer to weight and determine the precipitation collected. A thin layer of oil is added to impede any evaporation. If properly applied, this oil layer can essentially eliminate evaporation even during long periods without maintenance. Further, the gauges are equipped with a
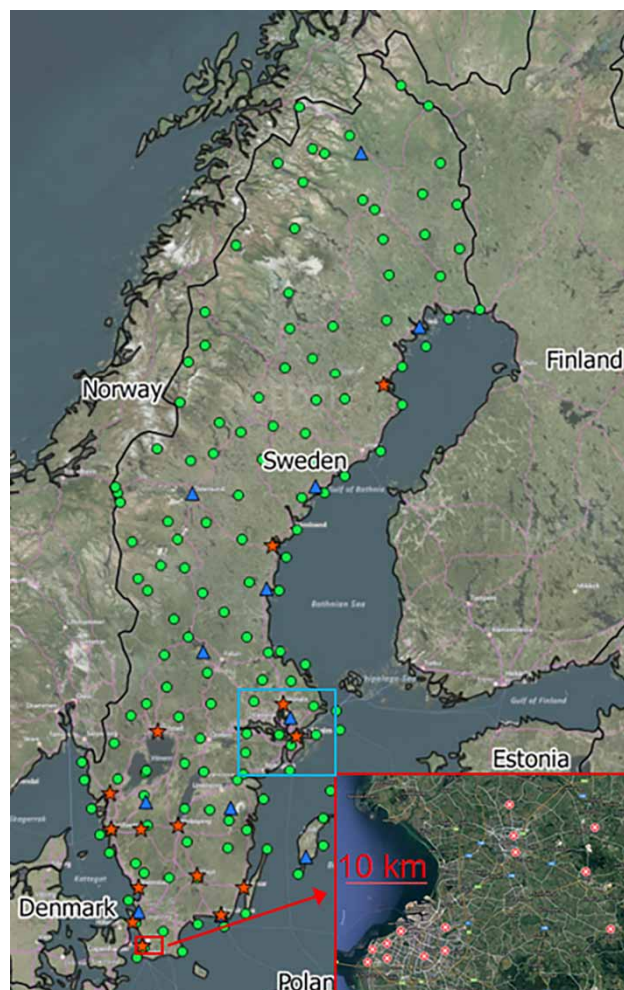
Figure 1 | Overview of the observational data sets used in this study. SMHI automatic stations (circles), municipal gauges (stars) and weather radars (triangles). The larger rectangle (bottom right) shows the local municipal gauge network and the smaller rectangle shows the area in which HIPRAD data were used.

wind screen in the form of metal or plastic plates (or leaves) that minimize precipitation losses. Still, some wind-induced undercatch is likely to occur, e.g., caused by wind gusts and updrafts associated with intense convective activity. The magnitude of this undercatch is however difficult to quantify. Aerodynamic simulation of the wind fields around precipitation gauges has indicated only very minor losses, even at moderate intensities (e.g., Nespor & Sevruk 1999). Here, we do not make any further undercatch correction.

The network of automatic meteorological stations started in 1995–1996 and 22 years of data were available for this study (1996–2017). There are some missing periods at some stations but overall the data coverage is ∼90% and the combined length of the time series is almost 2,500 years. Initially, the observations are automatically quality controlled in the database system of SMHI, and each observation receives a 'quality flag'. This control involves, e.g., comparison of rainfall depths sampled at different accumulation intervals. Suspicious values are manually checked and either corrected or rejected. As the observations used in this study all occurred in the period May to September they represent liquid precipitation and we therefore refer to them as rainfall.

For complementary analyses, two other sources of precipitation observations are used: municipal tipping-bucket stations and gauge-adjusted radar observations. Municipal observations (time step 1 min) are available from two databases, which were both carefully quality controlled in this study.

1. Time series from 15 Swedish cities (Figure 1) (Hernebring 2006). The length of the series varies between 5 and 24 years (average: 14). In the following, this data set will be referred to as the 'national municipal data'.
2. Time series from a local network of 15 stations in southern Sweden (Figure 1). These observations span the period 2000–2015. In the following, this data set will be referred to as the 'local municipal data'.

Radar observations are available through the HIPRAD (hIgh-resolution precipitation from gauge-adjusted weather radar) product (Berg *et al.* 2016). The current version of HIPRAD differs from that described in Berg *et al.* (2016), mainly in that precipitation estimates originate from the BALTEX Radar Data Centre (BRDC) radar composite (Koistinen & Michelson 2002). In order to improve the

description of particularly long-term precipitation accumulations, which are sometimes heavily influenced by location-dependent bias in the BRDC composites, the BRDC data are adjusted with monthly scaling factors towards a data set of gridded gauge observations (see further, Berg *et al.* 2016). The national data set is based on daily observations that have been optimally interpolated by including information about wind speed and direction as well as altitude (Johansson & Chen 2003). The latest version of HIPRAD covers the period 2000–2014 and has a temporal resolution of 15 min and a spatial resolution of 2×2 km². In this study, we use HIPRAD data from a square region over the city of Stockholm in central Sweden (Figure 1).

METHODS

Rainfall extremes and adjustment

Extremes of different duration are extracted from the rainfall time series in the conventional way, by using a moving time window to identify the maximum 'block rain' within the recorded events (Figure 2(a)). A limitation of the rainfall time series from the automatic stations is that they have a fixed 15-min time step, representing each quarter (00–15, 15–30, 30–45, 45–60) every hour. This leads to an underestimation of the extremes, as illustrated in Figure 2(b) where the moving ('true') maximum 15-min depth is almost 2 mm higher than the maximum depth obtained with fixed 15-min time steps. To quantify this effect for extreme values, a preparatory analysis was performed using the national municipal data set (see 'Study area and data'). These 15 series were processed as shown in Figure 2(b), i.e., event maximum depths were calculated using both fixed 15-min time steps and a moving 15-min time window. For accumulation durations (AD) >15 min, the moving time window was thus moved in either 1-min or 15-min steps. Table 1 shows the mean (over all 15 gauges) ratio between the maximum depths obtained from the original 1-min and the aggregated 15-min data, respectively, expressed as adjustment coefficients. The effect is pronounced for durations 15 and 30 min, for which the 'fixed extremes' need to be increased by 18% and 8%, respectively. For longer durations the effect is only a few per cent. These values are
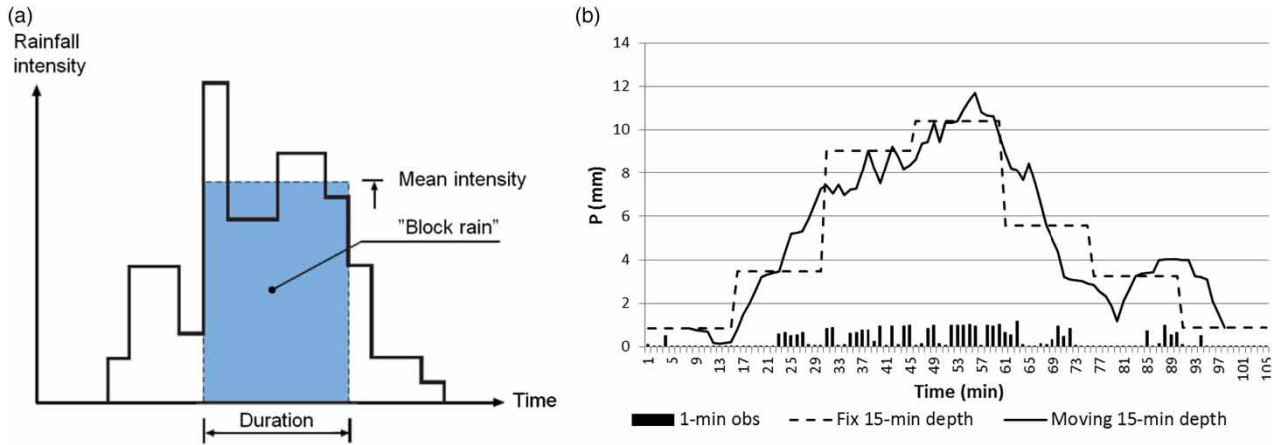
Figure 2 | Principle of obtaining the maximum 'block rain' in a recorded event (a). Example illustrating the impact of a fixed time step on the maximum value in an event (b).

Table 1 | Multiplicative coefficients to compensate for the fixed 15-min observation time step

AD	15 min	30 min	45 min	1 h	2 h
Coefficient	1.18	1.08	1.041	1.036	1.029

overall similar to previous estimates of this effect (e.g., WMO 2009). In all results of this study, this compensation for fixed time step is included.

Extreme value analysis, parameterization and uncertainty

Two standard approaches to statistically describe short-duration rainfall extremes according to extreme value theory are used in this study; (i) the generalized extreme value (GEV) distribution fitted to annual maximum (AM) values and (ii) the generalized Pareto (GP) distribution fitted to 'peak-over-threshold' (POT) values (also termed 'partial duration series' PDS) (e.g., Coles 2001). We will denote these approaches AM and POT in the following. These two approaches have been used complementarily in several earlier studies, e.g., Wang (1991), Madsen *et al.* (1997) and Chang *et al.* (2015). The GEV cumulative distribution function is expressed as:

$$
F(x; \mu, \sigma, \xi) = \begin{cases} \exp\left(-\left(1 + \xi * \left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right)\right) & \xi \neq 0 \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & \xi = 0 \end{cases} \quad (1)
$$

where μ is the location parameter, σ the scale parameter and ξ the shape parameter. GEV represents a family of distributions depending on the value of ξ: Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$). The GP cumulative distribution function is expressed as:

$$
F(\mu, \sigma, \xi)(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \xi = 0 \end{cases} \quad (2)
$$

where the location parameter μ corresponds to the threshold used to generate the POT values. According to extreme value theory, if AM values are approximately GEV-distributed, POT values will be approximately GP-distributed, notably with the same ξ-value (e.g., Serinaldi & Kilsby 2014).

In both approaches, the distributions are fitted to observed extreme depths D (mm) for (accumulation) durations AD 15, 30, 45, 60, 120, 180, 360 and 720 min. Identifying the most accurate method of parameter estimation has been the topic of previous studies and here we use maximum likelihood (e.g., Coles 2001), even though it has been questioned for small samples (e.g., Madsen *et al.* 1997). In the AM approach, the extremes are independent by construction and the number of values equals the number of years in the series. In the POT approach, a time separation of 2 h (for $AD \leq 2$ h) or AD (for $AD > 2$ h) was required between two POT values to obtain sufficient de-clustering. This is a common approach in Sweden (e.g., Hernebring 2006) and covers quite well the range of time

separations generally used (e.g., Dunkerley 2008). A key issue in the POT approach is the choice of threshold, which governs the number of values available for the GP fitting. With too few values the fitting risks become unstable and the parameters uncertain; with too many values the fitting risks become overly influenced by less-than-extreme value (see, e.g., Beguería 2005). Here, the number of POT values used was either as many as the number of years of data (POT_1; without subscript POT1 is assumed), to agree with the AM approach, or ten times POT_1 (POT_{10}), to investigate the impact of including also less extreme values (ten values per year have been suggested for Sweden by Hernebring 2006). Goodness of fit was assessed primarily by the *p*-value from one-sample Kolmogorov–Smirnov (KS) testing (e.g., Wilks 1995) complemented with visual inspection.

After distribution fitting to extremes of different duration, empirical DDF functions are used to describe the depths D estimated by the fitted distribution (e.g., Chow *et al.* 1988). These functions are commonly approximated by a mathematical expression with a number of free parameters; see, e.g., Di Baldassarre *et al.* (2006) who evaluated a number of general expressions with two or three parameters. In an earlier investigation of the national municipal data (see 'Study area and data') by Hernebring (2006), commonly applied two- and three-parameter expressions were found insufficient for accurately describing DDF functions over the wide range of durations considered here and, instead, a six-parameter expression was proposed. Here, we use a similar approach and express the DDF functions as:

$$D = k(T)*AD^{p(T)} \tag{3}$$

where T denotes return period (years) and $k(T)$ and $p(T)$ are functions expressed as:

$$k(T) = a_1*T^{a_2} + a_3 \tag{4}$$

$$p(T) = b_1*T^{b_2} + b_3 \tag{5}$$

where a_1–a_3 and b_1–b_3 are parameters that were optimized by least squares.

In existing DDF statistics for Sweden (Dahlström 2010; Svenskt Vatten 2011), no uncertainty estimation is included

which is considered a limitation. Here, we use a pragmatic approach to approximate the uncertainty span around each depth D by a simple mathematical expression. First, based on previous experience, it is assumed that (i) the confidence interval is symmetric around an arbitrary depth D and (ii) its upper and lower bounds may be expressed as fractions $CI_\%$ of D, i.e., $D \pm CI_\%D$. It is further assumed that whereas $CI_\%$ may be considered independent of AD it is likely to depend nonlinearly on T, suggesting some form of power law relationship between $CI_\%$ and T.

Spatial correlation of short-duration rainfall extremes

In a preparatory analysis supporting the application of the station-year method (see 'Regional clustering and the station-year method') to the SMHI data (see 'Study area and data'), the spatial correlation of short-duration rainfall extremes in Sweden is investigated. The main objective is to estimate the decorrelation distance for rainfall extremes of different duration, in order to assess whether extremes from different stations can be considered statistically independent.

In the analysis, a POT approach is applied to convert rainfall time series into binary series with 1 representing a rainfall depth D above a threshold t and 0 representing $D < t$, and with t set high enough for the 1-values only to represent extremes. The correlation between series from two locations is then analysed using the φ coefficient, which is the binary equivalent of Pearson's correlation coefficient (e.g., Wilks 1995). The φ coefficient may be expressed as:

$$\phi = \frac{n_{11}*n_{00} - n_{10}*n_{01}}{\sqrt{n_{1*}*n_{0*}*n_{*0}*n_{*1}}} \tag{6}$$

where the *n*-values represent the entries of a contingency table (Table 2). The value $\varphi = 1$ (which can only be attained under certain conditions, see, e.g., Warrens 2013) indicates perfect (positive) correlation and $\varphi = 0$ no correlation.

Regional clustering and the station-year method

In order to identify homogeneous regions in Sweden with respect to short-duration rainfall extremes, K-means clustering was applied (MacKay 2003). In this analysis, each SMHI

Table 2 | Contingency table

	y = 1	y = 0	TOT
x = 1	n_{11}	n_{10}	n_{1*}
x = 0	n_{01}	n_{00}	n_{0*}
TOT	n_{*1}	n_{*0}	n_{**}

station (see 'Study area and data') was represented by a five-dimensional vector with the first three dimensions representing the station's estimated GEV or GP parameters and the last two dimensions representing the station's x and y coordinates in the national coordinate system (RT90; Lantmäteriet 2018). By this strategy, clusters of stations that have both similar parameter values and are nearby geographically may be identified. To make the dimensions comparable, they were normalized by subtracting the mean and dividing by the standard deviation. Further, weights were introduced in order to explore the impact of letting distribution parameters and geographical location, respectively, have unequal importance in the clustering. Stations with less than 15 years of data were excluded from the analysis. The number of clusters was varied from three to six.

The final clusters were evaluated using silhouettes (Rousseeuw 1987). The silhouette value indicates how similar an object (station, in our case) is to the cluster to which it has been assigned, compared to other clusters. The silhouette value ranges from −1 to +1, with a high value indicating a good correspondence to the cluster properties and a negative value indicating that it would be better assigned to another cluster. Many low or negative values imply that the number of clusters used is sub-optimal.

Following the identification of regional clusters, the station-year method was applied to estimate regional statistics (e.g., Buishand 1991; Overeem *et al.* 2008; Burn 2014). In this method, the observations from all stations within the region are pooled into one data set under the assumption that the different time series can be considered as independent realizations of the same rainfall climate. Even if the station-year method can be (and has been) applied to observations from nearby stations, statistical independence between extremes from the stations is generally considered an advantage (e.g., Buishand 1991; Steward *et al.* 1999). This prompted the above spatial correlation analysis (see 'Spatial correlation of short-duration rainfall extremes').

RESULTS AND DISCUSSION

Local statistics

Some key results from fitting the AM and POT methods to the SMHI data (see 'Study area and data') are given in Table 3. The values have been averaged over all stations (98) with at least 20 years of data. In terms of goodness of fit as assessed by the KS test, the hypothesis that the AM values follow a GEV distribution was not rejected for any station at 5% significance level. The hypothesis that the POT values follow a GP distribution was rejected for only one single station at one single duration. The *p*-values indicate generally better fits for AM, with an average value of 0.82, than POT, with 0.70 for POT_1 and 0.62 for POT_{10} (Table 3). The difference between POT_{10} and the two

Table 3 | Results from the AM and POT methods applied to the SMHI data

	KS *p*-value			Shape parameter (ξ)					
				Mean value			Fraction >0 (%)		
	AM	POT_1	POT_{10}	AM	POT_1	POT_{10}	AM	POT_1	POT_{10}
15 min	0.80	0.67	0.41	0.191	0.021	0.133	82	57	89
30 min	0.82	0.67	0.55	0.198	0.061	0.168	79	59	97
60 min	0.82	0.72	0.62	0.194	0.070	0.173	83	58	93
120 min	0.82	0.71	0.68	0.170	0.066	0.148	76	56	88
180 min	0.82	0.69	0.71	0.115	0.004	0.130	67	59	88
360 min	0.82	0.71	0.76	0.093	−0.013	0.084	64	49	79
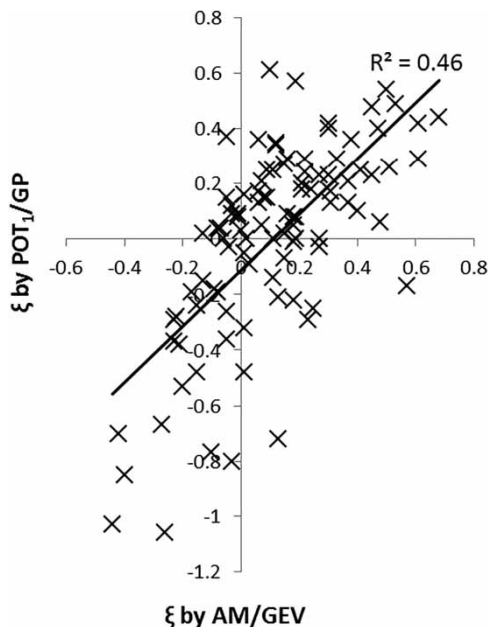
Figure 3 │ Comparison between ξ-values for all stations (crosses) obtained by AM and POT$_1$, respectively, at duration 3 h. The solid line was fitted by linear regression.

up to 2 h ~80% of the stations have a positive value, being on average almost 0.2, but for longer durations the values are ~65% and ~0.1, respectively. This indicates less heavy-tailed distributions with increasing duration. The POT$_{10}$ approach overall agrees with the AM results, although with a higher fraction of stations with $\xi > 0$ but a slightly lower average ξ.

A limited analysis of ξ-values obtained from AM and POT$_1$, respectively, was performed. As shown in Table 3, the ξ-values from AM are systematically higher and overall the correspondence is rather weak. Figure 3 shows the duration (3 h) with the highest R²-value of a linear regression line (0.46); for the other durations R² ranges between 0.29 and 0.38.

Spatial correlation of short-duration extremes

Figure 4 illustrates the results from the correlation analysis through the φ coefficient (see 'Spatial correlation of short-duration rainfall extremes') applied to the local municipal data as well as the HIPRAD data (see 'Study area and data'). In the analysis different values of the threshold t, used in the conversion to binary series, were tested, representing percentiles between 90 and 99.9 in each data set. With a too low value of t there is a risk of including also

others may partly be attributed to the increasing power of the test for increasing samples. The POT$_{10}$ fits gradually improve with increasing duration.
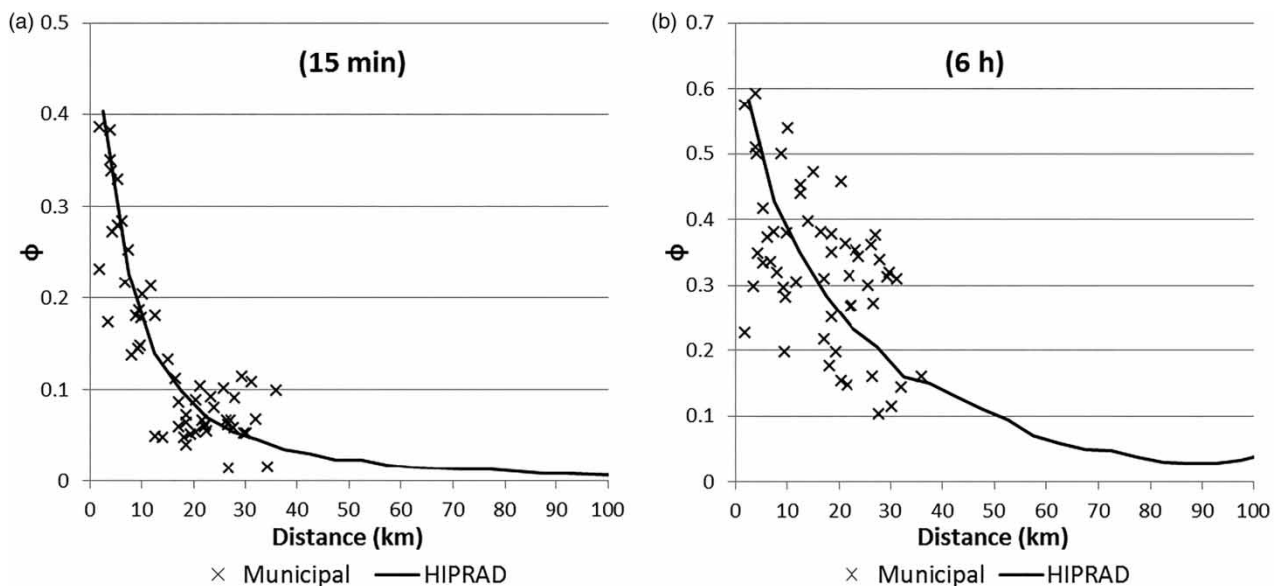
For the AM approach, the shape parameter ξ is generally positive but less so with increasing duration. For durations

Figure 4 │ The φ coefficient based on the municipal data and HIPRAD for durations 15 min (a) and 6 h (b).

non-extremes, and with a too high value there will be locations with only zeros. Further, as the frequency distribution of precipitation in the two data sets are different, e.g., because of the different sensors used (gauge and radar, respectively), the same percentile will represent different absolute *t*-values. Overall, the impact of the *t*-value is however small and Figure 4 is well representative of all results obtained.

For duration 15 min (Figure 4(a)), the φ coefficient estimated from the local municipal data decreases from almost 0.4 at the shortest distance (2 km) to ~0.05 at the longest distance (35 km). This pattern is very well matched by the HIPRAD data, for which the φ coefficient has been averaged over bins in order to reduce statistical scatter. The corresponding *t*-values are in the range 2–4 mm/15 min. We note that Madsen *et al.* (2002) in a similar analysis found an overall similar result for 10-min data in Denmark, which is not unexpected as the rainfall climate is similar to southern Sweden. For duration 6 h (Figure 4(b)), the φ coefficient from the municipal data is markedly higher, ranging from almost 0.7 at the shortest to ~0.2 at the longest distances. Again, the results for HIPRAD agree well and in this case *t* is in the range 15–20 mm/6 h. The tendency towards higher correlation with longer duration reflects that short-duration extremes are generally associated with spatially smaller-scale atmospheric phenomena (e.g., convective cells) and longer-duration extremes with larger structures (e.g., cyclones), as discussed in Madsen *et al.* (2002).

Very generally, a statistical rule of thumb is that correlations below ~0.2 can be considered very weak or negligible (e.g., Evans 1996). Here, this implies that 15-min (6-h) extremes are uncorrelated at distances > ~10 km (~30 km). As the average distance between two stations in the SMHI data set (see 'Study area and data') is ~60 km, we conclude that the short-duration rainfall extremes can be considered statistically independent and that the station-year method is applicable to station clusters.

Regional clustering

The regional clustering (see 'Regional clustering and the station-year method') was performed on parameter sets estimated by applying both the AM and the POT approach to the SMHI data. Overall the two approaches generated similar clusters with only single or few stations differing in some cases. As the AM approach generated (i) the best fits to the station time series (Table 3) and (ii) somewhat more stable clusters over the range of durations considered, in the following, results from this approach are presented.

If using four clusters in the K-means procedure (see 'Regional clustering and the station-year method'), at 15-min duration one cluster is formed in the south-east of Sweden (SE), one in the south-west (SW), one in central Sweden (C) and one in northern Sweden (N) (Figure 5(a)). Generally, the conditions for convection and corresponding intense short-duration rainfall are more favourable in the south, e.g., by the southward movement of moist and warm air masses from continental Europe. Also, frontal systems may be associated with strong convection, e.g., in cold frontal troughs, and these frontal systems often move in over Sweden from the south-west. This brings intense rainfall to the part of south-western Sweden that borders the Kattegat Sea and less rainfall further east. Further north, by the influence of the Scandinavian mountains, the most intense rainfall produced by eastward-moving fronts or air masses generally hits Norway instead of Sweden. Another situation potentially associated with intense rainfall events is when dry and warm air masses from Finland and Russia move in from the west or north-west, taking up moisture over the Baltic Sea. These events are likely to occur mainly in the eastern parts of southern and central Sweden but also along the northern part of the east coast bordering the Bothnian Bay. Overall, the four clusters in Figure 5(a) reflect the different regimes of short-duration rainfall extremes based on general meteorological reasoning. See further 'Regional statistics' and 'Magnitude vs. frequency: Spatio-temporal variability' below for quantitative assessment of the different clusters.

The clustering based on duration 1 h (Figure 5(b)) is nearly identical to the one for 15 min, except that one station in the C cluster is reclassified into cluster SE. Also the 6-h clusters (Figure 5(c)) are similar to the 15-min and 1-h clusters. Only a few stations are classified differently, mainly close to the east coast where the SE cluster extends slightly further north and where some of the stations in cluster N are reclassified into cluster C. This indicates that the impact of air masses from the east is more prominent for longer durations.
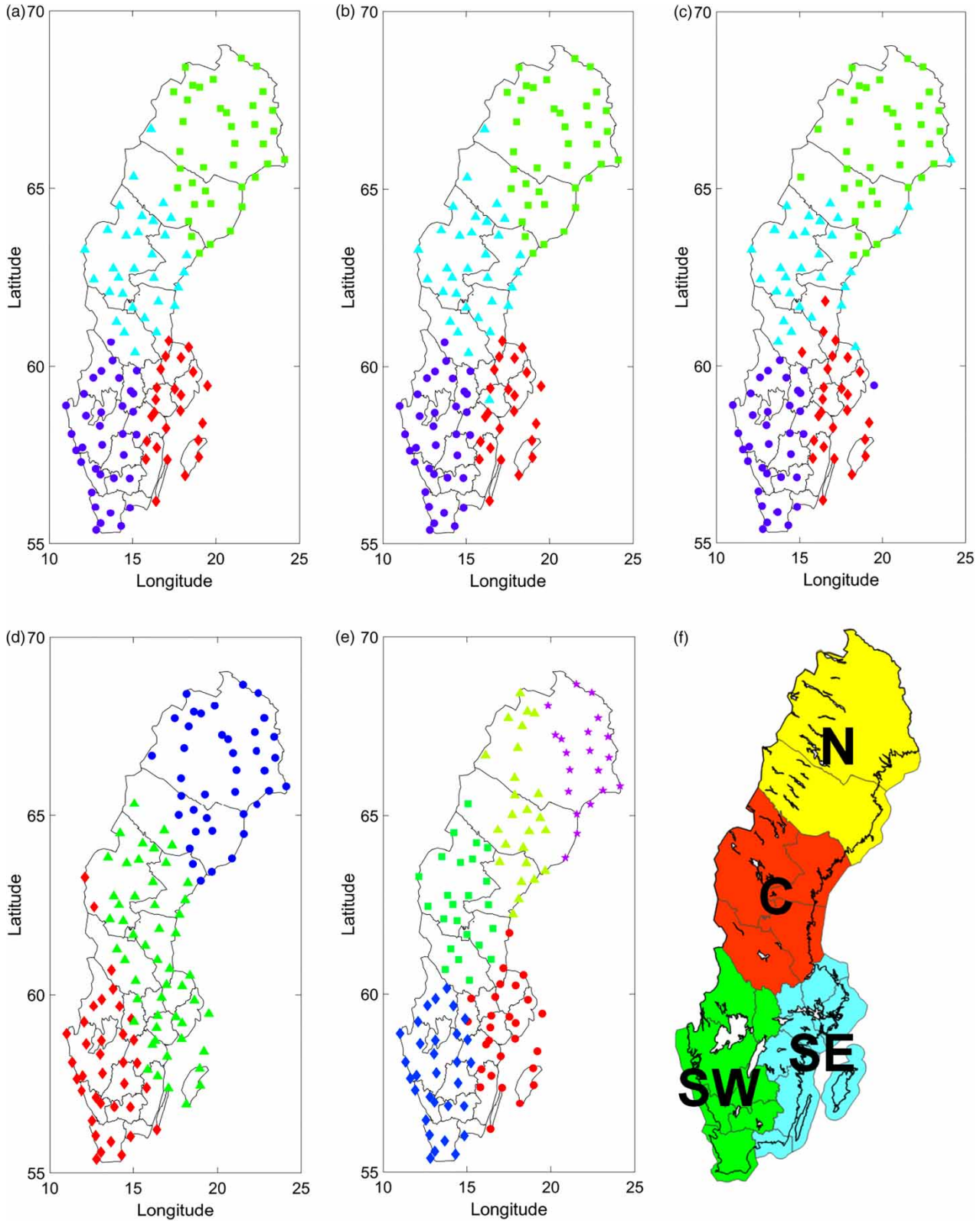
Figure 5 │ K-means results with four clusters for durations AD = 15 min (a), 1 h (b) and 6 h (c). K-means results with three clusters for AD = 15 min (d) and five clusters for AD = 1 h (e). Final regions (f).

If forcing the K-means procedure towards three clusters, effectively clusters SE and C are merged into one (Figure 5(d)). With five clusters, cluster N is divided into two clusters (and somewhat extended further to the south-east) (Figure 5(e)). With six clusters (not shown), a new cluster is formed in between clusters SW and SE. We take these findings, i.e., that clusters are either merged or divided when changing the number, as a confirmation that the clustering procedure can be considered stable and physically consistent. In the final regionalization we use four clusters, which is a compromise between on the one hand representing the main regional patterns found and on the other hand having enough (and a similar number of) stations in each cluster for stable and consistent statistics (e.g., Burn 2014). For practical reasons, the final clusters follow the borders between Swedish counties (Figure 5(f)) but these 'administrative' borders are overall very well in agreement with the 'physical' borders indicated in the results of the clustering (Figure 5(a)–5(c)). The number of station-years is 624 in cluster SW, 450 in SE, 617 in C and 749 in N. These numbers are smaller than but reasonably close to the target group size 780 recommended by Burn (2014).

The silhouettes of the four clusters overall indicate that they are distinct, thus supporting the results from the K-means analysis. Figure 6 shows the silhouettes corresponding to the clustering based on 15-min values (Figure 5(a)). We note first of all the absence of negative values, i.e., no station would fit better in another cluster. Further, cluster SW is the most distinct one, with almost 75% of the stations having a silhouette value >0.5. In cluster N, the corresponding number is ~60% (but a few stations have very low values) and in cluster C, ~50%. Cluster SE
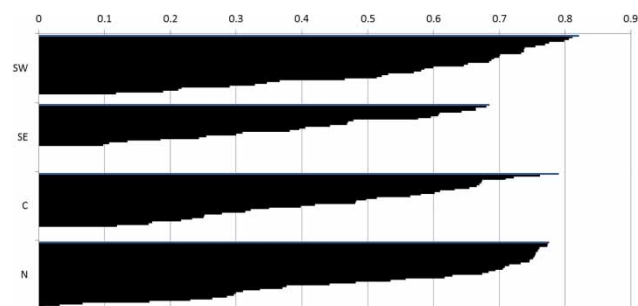

**Figure 6** │ Silhouettes for the four 15-min clusters in Figure 5(a).

is the least distinct with less than 40% of the stations having a silhouette value >0.5. Silhouettes for other durations are overall similar to Figure 6, although sometimes with single negative values.

Regional statistics

In Figure 7, the AM and POT methods are applied to the regional series of 1-h values, where the time series from all stations in the region have been pooled. On the *x*-axis (return period *T*) the data have been averaged over logarithmically spaced bins. Concerning the empirical values (points in Figure 7), these are identical for $T > 10–20$ years but for shorter return periods the POT values are higher than the AM values, to an increasing degree with decreasing *T*. This illustrates the limitation of the AM method, where maxima from years without real extremes are included instead of real (but non-maximum) extremes from other years.

Below $T \sim 3$ years, the fitted GP distribution is located above the GEV distribution, in line with the difference between the AM and POT values. From 3 up to ~40 years' (regions N and C) or ~100 years' (SE and SW) return period, the two distributions are virtually identical. For longer return periods, the GEV distribution is generally located above the GP distribution, especially in regions N and C (Figure 7(a) and 7(b)). In region N, the GEV distribution overestimates the depths for T > 100 years, whereas in region C the fits are equally representative from visual inspection. In regions SE and SW, the fits are similar for T > 20, with slightly better visual agreement for the GEV distribution in region SE.

For the regional series, the KS test indicates a markedly better fit of GEV to the AM data (average KS *p*-value over all durations and regions: 0.58) than of GP to the POT data (0.35). This difference was found to be caused by the discrete nature of the data, i.e., that they are given as multiples of the rainfall depth resolution in the Geonor gauges (0.1 mm). This makes the lowest POT values consist of many identical values, which has a negative impact on the KS test in this probability range (approximately the bottom third); over the rest of the range the accuracy of the fits are similar. We do not consider this to have any practical importance, but based on the visual inspection (of all durations), as
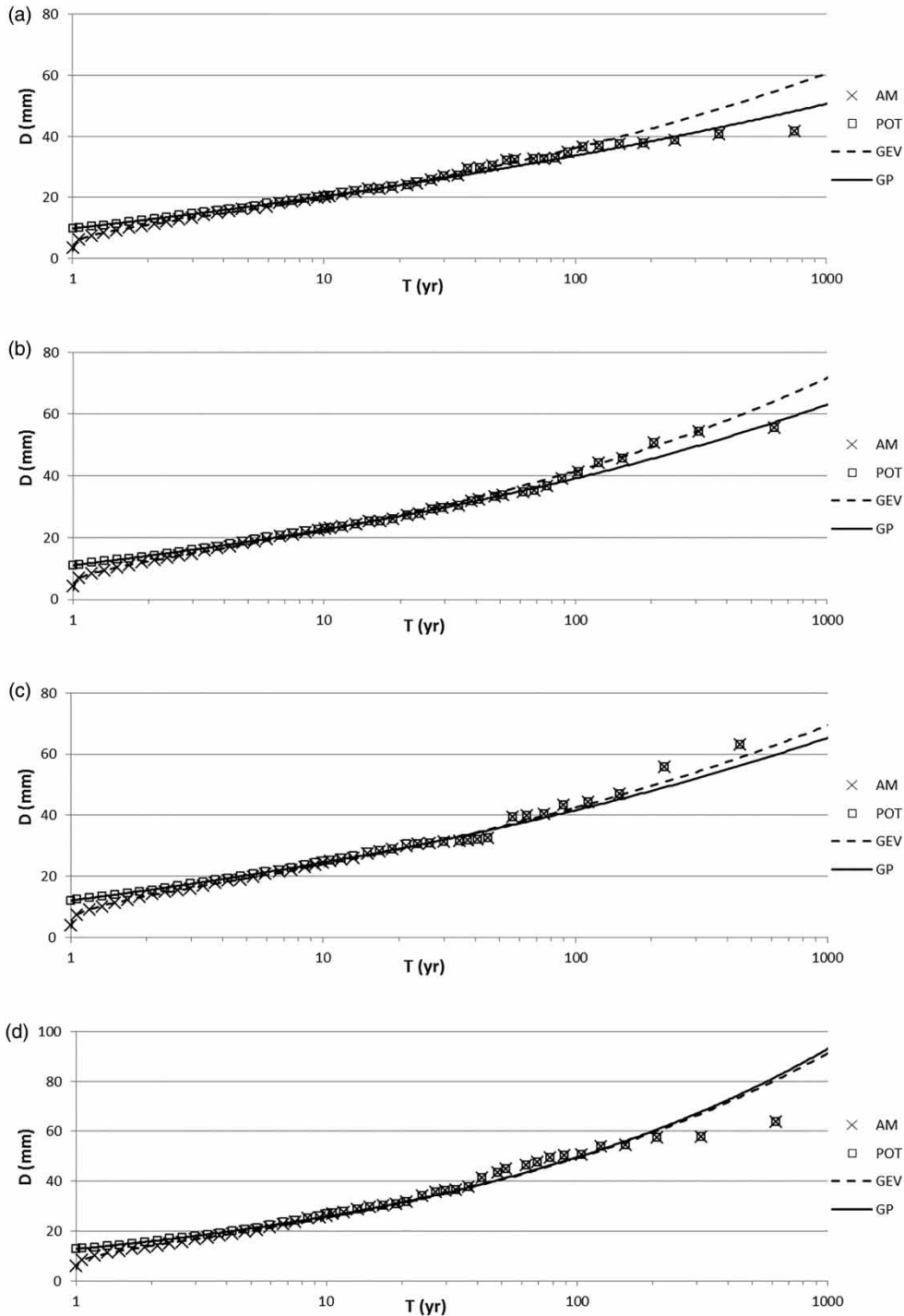
Figure 7 │ Empirical 1-h depths from the AM and POT methods and fitted GEV and GP distributions for regions N (a), C (b), SE (c) and SW (d). The plotting position is Weibull (Makkonen 2006).

well as recommendations from previous work in Denmark (e.g., Madsen *et al.* 2002), we choose to use the POT approach for the new national statistics. Concerning the GP shape parameter ξ, this is clearly larger in region SW (average value over all durations: 0.23) than in the other regions (~0.15).
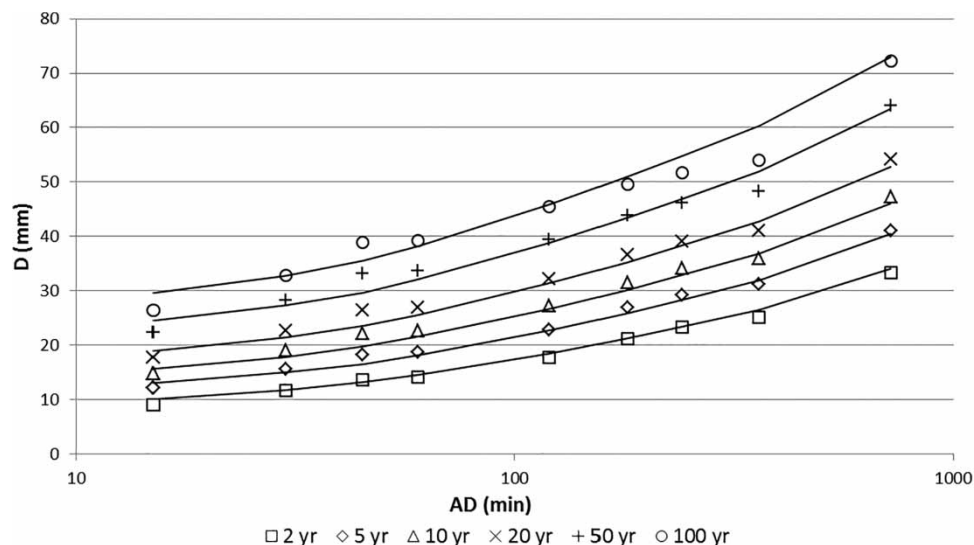
Figure 8 | The generalized DDF function (Equation (3); solid lines) fitted to empirical values (symbols) in region C.

In Figure 8, the generalized DDF function (Equation (3)) has been fitted to depths estimated by applying the POT method to different durations between 15 min and 12 h in region C. For T up to 10 years the fit is very close to the values, whereas there are larger deviations for longer return periods which we attribute mainly to statistical scatter. The result is similar for the other regions. It is difficult to formulate any objective criterion for acceptable fits but from visual inspection we find them satisfactory. We may note that the mean absolute percentage error (E%) of the fits for sub-daily durations between return periods 2 and 50 years is ∼7% for the four regions, which is about half the error of the best-performing corresponding DDF function in Di Baldassarre *et al.* (2006; Figure 10).

Figure 9(a) indicates that the assumption of a symmetric confidence interval (see 'Extreme value analysis, parameterization and uncertainty') holds well up to $T \sim 100$ years, above which it becomes more and more asymmetric. The level of asymmetry was further quantified by the relative difference between the distance from the fitted distribution to the upper and lower confidence intervals, respectively (CI_u and CI_l in Figure 9(a)). Averaged over all durations and return periods up to 100 years, this relative difference is 6.2% which we consider small enough for approximating them by their mean value CI, which is then expressed as a fraction of the associated depth D, i.e., $CI_\%$ (see 'Methods' above). Even though $CI_\%$ exhibits a slight dependence on

duration, it is relatively well approximated by a constant value $\overline{CI_\%}$ (Figure 9(b)). The nonlinear dependence of $\overline{CI_\%}$ on return period is illustrated in Figure 9(c), where $\overline{CI_\%}$ ranges from a few per cent at short return periods up to 15–20% for $T = 100$ years. There is some regional variation, most clearly for long return periods, with the highest $\overline{CI_\%}$-values in region SW and the lowest in region N, but overall we consider the regional mean value to be an acceptable approximation. The mean value is well described by the expression:

$$\overline{CI_\%} = 2T^{0.45} \tag{7}$$

as shown in Figure 9(c). This equation can thus be used together with Equation (3) to generate DDF statistics including uncertainty for arbitrary combinations of duration and return period.

Some examples of final DDF statistics obtained by Equations (3) and (7) are given in Table 4. There are distinct regional differences with the highest depths in region SW followed by SE, C and N. Averaged over all durations and return periods, the depths in region N are 25% lower than in region SW. It may be noted that the values for region SW agree well with corresponding values estimated for eastern Denmark, located just west of region SW (Figure 1). Further, the values for region C well match the corresponding values for central Finland, east of the Baltic Sea
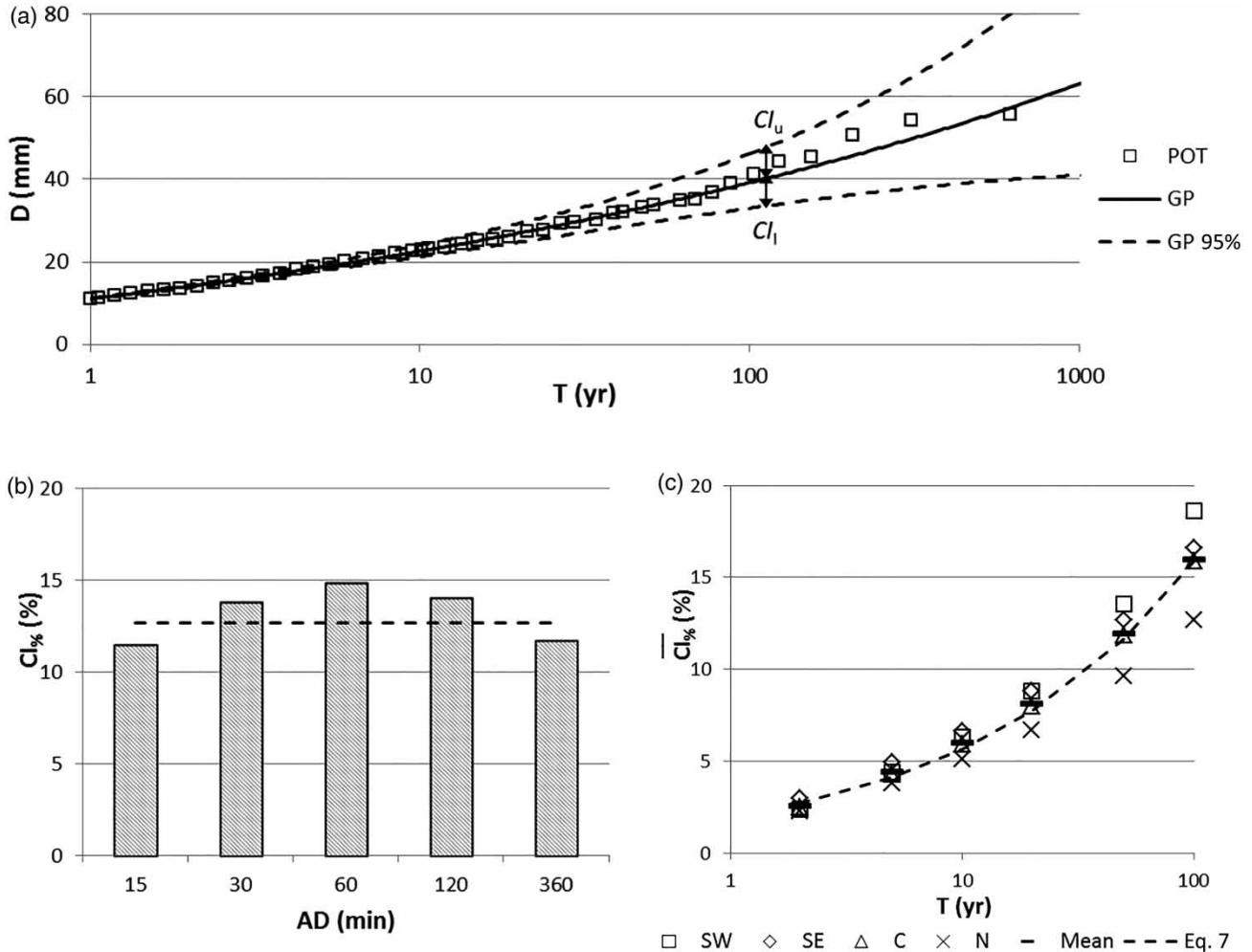
Figure 9 | Empirical 1-h depths from the POT method and fitted GP distribution with 95% (Wilson score) confidence interval for region C (a), $CI_\%$ as a function of duration for return period 50 years in region SE, with dashed line showing the mean value $\overline{CI_\%}$ (b), $\overline{CI_\%}$ as a function of return period for the regions, their mean value and by Equation (7) (c).

(Figure 1). Comparison with bordering areas in Norway is less meaningful because of its highly heterogeneous rainfall climate. More comprehensive evaluation and statistics are given in Olsson *et al.* (2017).

The national DDF (10-year) statistics in today's Swedish design guidelines (Svenskt Vatten 2011) agree well with the statistics obtained for region SW in this study. This is expected as the national statistics were derived using observations mainly from south-western Sweden. Consequently, the national statistics overestimate the statistics in regions SE, C and N, to a gradually increasing degree.

Table 4 | Selected final DDF statistics

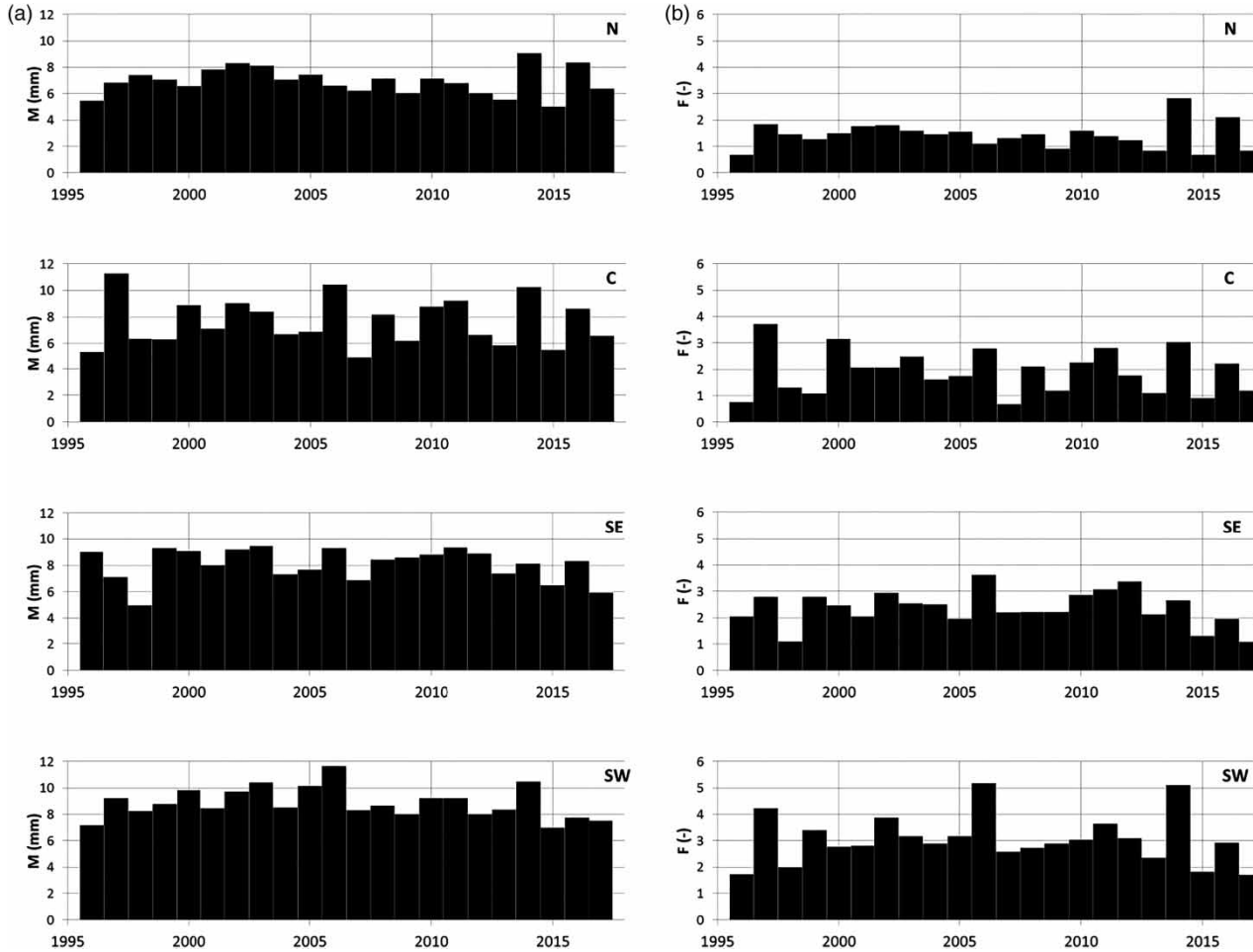|	Region SW		Region SE		Region C		Region N	
	T = 10	*T* = 100	*T* = 10	*T* = 100	*T* = 10	*T* = 100	*T* = 10	*T* = 100
15 min	18.0 ± 1.1	35.1 ± 6.1	16.3 ± 1.0	28.1 ± 4.9	15.7 ± 0.9	29.7 ± 5.2	14.2 ± 0.8	26.3 ± 4.6
60 min	24.5 ± 1.4	45.2 ± 7.9	22.6 ± 1.3	38.3 ± 6.7	21.6 ± 1.3	38.2 ± 6.7	19.0 ± 1.1	32.6 ± 5.7
180 min	34.1 ± 2.0	60.2 ± 10.5	32.0 ± 1.9	53.5 ± 9.3	30.2 ± 1.8	50.9 ± 8.9	26.1 ± 1.5	42.1 ± 7.3

Figure 10 │ Temporal variation of magnitude M (a) and frequency F (b) in the regions.

Magnitude vs. frequency: spatio-temporal variability

Further insight into the differences between the different regions may be attained from the time series shown in Figure 10, representing the magnitude and frequency of rainfall extremes, respectively. Magnitude is represented by the regional average annual maximum 15-min depth M, i.e., the average value of the annual maxima in all the regions' stations, calculated as:

$$M_r^y = \frac{\sum_{i=1}^{n_r} amax_i^y}{n_r} \tag{8}$$

where y denotes year, r region, n number of stations and *amax* the annual maximum 15-min depth in station i. Looking first at region N, M varies between 9.1 mm

(in 2014) and 5.0 mm (in 2015) with an average value $\bar{M}$ of 6.9 mm (Figure 10(a)). $\bar{M}$ gradually increases towards the south, being 7.6 mm in region C, 8.1 in SE and 8.9 in SW. The value in region SW is thus ~30% higher than in region N. The range of variation is however relatively similar in all regions.

Frequency is represented by the average number of exceedances of the threshold 5 mm/15 min, F, calculated as:

$$F_r^y = \frac{\sum_{i=1}^{n_r} thex_i^y}{n_r} \tag{9}$$

where *thex* denotes the number of threshold exceedances. The threshold 5 mm/15 min was selected to have enough exceedances for robust statistics, and it represents a return period T < 1 year. In region N, F varies between 0.7 and

2.8 with an average value $\bar{F}$ of 1.4. Also $\bar{F}$ increases towards the south, being 1.9 in region C, 2.4 in SE and 3.1 in SW (Figure 10(b)). The frequency in region SW is thus more than twice as high (∼115%) as in region N.

The period is too short for meaningful assessment of climatological trends, but we note that there is no tendency towards higher values of neither M nor F. In fact, if fitting linear trends to the data, these trends become slightly negative for all series in Figure 10. This is hardly surprising, considering that there is no clear increase in summer temperature during the period, and it does not contradict the general assumption of more intense rainfall in a warming climate. Investigations of longer data series in Sweden have obtained mixed results, including some signals of intensified extremes (e.g., Hernebring 2006).

CONCLUDING REMARKS

We have calculated new DDF statistics for Sweden, based on 22 years of 15-min rainfall observations from 128 stations. The statistics are given for four regions, for return periods up to 100 years and have a quantified uncertainty. We believe that the new statistics are beneficial for both engineering practitioners, e.g., for improved design of infrastructural components, and the R&D community, e.g., for evaluating the accuracy of new sensors and atmospheric models.

Three novel aspects in the results are worth highlighting:

1. Regional clustering of short-duration rainfall extremes based on the parameters of fitted extreme value distributions generated overall robust, consistent and physically sensible results.
2. The uncertainty of the DDF statistics could be quantified as a simple and practically applicable function of return period.
3. Spatial correlation of rainfall extremes as estimated from gauge-adjusted radar observations agreed well with the results from local gauge networks.

In the future, we intend to further evaluate our approach of using the station-year method in combination with 'parameter value clustering'. The regionalization may be compared with the results from other approaches, e.g.,

L-moments. Further, by more in-depth analysis it may be attempted to develop location-specific rather than regional statistics, using suitable geographical and/or climatological predictors.

ACKNOWLEDGEMENTS

REFERENCES

Beguería, S. 2005 Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology* **303**, 215–230.

Berg, P., Norin, L. & Olsson, J. 2016 Creation of a high resolution precipitation data set by merging gridded gauge data and radar observations for Sweden. *Journal of Hydrology* **541**, 6–13.

Buishand, T. A. 1991 Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal* **36**, 345–365.

Burn, D. H. 2014 A framework for regional estimation of intensity–duration–frequency (IDF) curves. *Hydrological Processes* **28**, 4209–4218.

Chang, K. B., Lai, S. & Othman, F. 2015 Comparison of annual maximum and partial duration series for derivation of rainfall intensity-duration-frequency relationships in Peninsular Malaysia. *Journal of Hydrologic Engineering* **21**, 05015013.

Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill, New York, USA.

Coles, S. 2001 *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag, London, UK.

Dahlström, B. 1979 *Regional Fördelning av Nederbördsintensitet – en Klimatologisk Analys (Regional Distribution of Precipitation Intensity – A Climatological Analysis)*. Report R18:1979, BRF, Stockholm, Sweden.

Dahlström, B. 2010 *Regnintensitet – en Molnfysikalisk Betraktelse (Rainfall Intensity – A Cloud Physical View)*. Report 2010-05, Svenskt Vatten Utveckling, Svenskt Vatten AB, Stockholm, Sweden.

Di Baldassarre, G., Brath, A. & Montanari, A. 2006 Reliability of different depth-duration-frequency equations for estimating

short-duration design storms. *Water Resources Research* **42** (12), W12501.

Dunkerley, D. 2008 Identifying individual rain events from pluviograph records: a review with analysis of data from an Australian dryland site. *Hydrological Processes* **22**, 5024–5036.

Evans, J. D. 1996 *Straightforward Statistics for the Behavioral Sciences*. Brooks-Cole Publishing, Pacific Grove, CA, USA.

Hernebring, C. 2006 *10 års-Regnets återkomst, Förr och nu (Return Period of the 10-Year Rainfall in the Past and in the Present)*. Report SVH 2006-04, Svenskt Vatten Utveckling, Svenskt Vatten AB, Stockholm, Sweden.

Hernebring, C., Dahlström, B. & Kjellström, E. 2012 *Regnintensitet I Europa med Fokus på Sverige – ett Klimatförändringsperspektiv (Rainfall Intensity in Europe with Focus on Sweden – A Climate Change Perspective)*. Report 2012-16, Svenskt Vatten Utveckling, Svenskt Vatten AB, Stockholm, Sweden.

Hernebring, C., Milotti, S., Steen Kronborg, S., Wolf, T. & Mårtensson, E. 2015 Skyfallet i sydvästra Skåne 2014-08-31 (The cloudburst in southwestern Scania 2014-08-31). *Journal of Water Management and Research* **71**, 85–99. (in Swedish with English abstract).

Johansson, B. & Chen, D. 2003 The influence of wind and topography on precipitation distribution in Sweden: statistical analysis and modelling. *International Journal of Climatology* **23** (12), 1523–1535.

Koistinen, J. & Michelson, D. B. 2002 BALTEX weather radar-based precipitation products and their accuracies. *Boreal Environment Research* **7**, 253–263.

Lantmäteriet 2018 https://www.lantmateriet.se/en/maps-and-geographic-information/GPS-och-geodetisk-matning/Om-geodesi/Transformationer/RT-90—SWEREF-99/.

MacKay, D. 2003 *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK.

Madsen, H., Rasmussen, P. F. & Rosbjerg, D. 1997 Comparison of annual maximum series and partial duration series for modeling extreme hydrologic events, 1, At-site modelling. *Water Resources Research* **33** (4), 747–757.

Madsen, H., Mikkelsen, P. S., Rosbjerg, D. & Harremoës, P. 2002 Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resources Research* **38** (11), 1239.

Makkonen, L. 2006 Plotting positions in extreme value analysis. *Journal of Applied Meteorology and Climatology* **45**, 334–340.

Nespor, V. & Sevruk, B. 1999 Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation.

Journal of Atmospheric and Oceanic Technology **16** (4), 450–464.

Olsson, J. & Foster, K. 2014 Short-term precipitation extremes in regional climate simulations for Sweden. *Hydrology Research* **45** (3), 479–489.

Olsson, J., Berg, P., Eronn, A., Simonsson, L., Södling, J., Wern, L. & Yang, W. 2017 *Extremregn I Nuvarande och Framtida Klimat (Extreme Rainfall in Present and Future Climate)*. Report Climatology No. 47, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden (in Swedish with English summary).

Overeem, A., Buishand, T. A. & Holleman, I. 2008 Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology* **348**, 124–134.

Rousseeuw, P. J. 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53–65.

Serinaldi, F. & Kilsby, C. G. 2014 Rainfall extremes: toward reconciliation after the battle of distributions. *Water Resources Research* **50**, 336–352.

Steward, E. J., Reed, D. W., Faulkner, D. S. & Reynard, N. S. 1999 The FORGEX method of rainfall growth estimation I: review of requirement. *Hydrology and Earth System Sciences* **3** (2), 187–195.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. M. 2013 *Climate Change 2013: The Physical Science Basis – Working Group I Contribution to the IPCC 5th Assessment Report*. Cambridge University Press, Cambridge, UK.

Svenskt Vatten 2011 *Nederbördsdata vid Dimensionering och Analys av Avloppsvatten (Precipitation Data for Design and Analysis of Sewer Systems)*. Publication P104, Svenskt Vatten, Stockholm, Sweden.

Wang, O. J. 1991 The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology* **129**, 263–280.

Warrens, M. J. 2013 On association coefficients, correction for chance, and correction for maximum value. *Journal of Modern Mathematics Frontier* **2** (4), 111–119.

Wern, L. & German, J. 2009 *Korttidsnederbörd I Sverige 1995-2008 (Short-Term Precipitation in Sweden 1995-2008)*. Report Meteorology No. 139, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

Wilks, D. S. 1995 *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, CA, USA.

WMO 2009 *Manual for Estimation of Probable Maximum Precipitation*. Publication No. 1045, World Meteorological Organization, Geneva, Switzerland.