RMetS

Royal Meteorological Society

The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics

D. Maraun,[a]* H. W. Rust[b] and T. J. Osborn[a]

[a] *Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *Physics Institute, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany*

ABSTRACT: The annual cycle of extreme 1-day precipitation events across the UK is investigated by developing a statistical model and fitting it to data from 689 rain gauges. A generalized extreme-value distribution (GEV) is fit to the time series of monthly maxima, across all months of the year simultaneously, by approximating the annual cycles of the location and scale parameters by harmonic functions, while keeping the shape parameter constant throughout the year. We average the shape parameter of neighbouring rain gauges to decrease parameter uncertainties, and also interpolate values of all model parameters to give complete coverage of the UK. The model reveals distinct spatial patterns for the estimated parameters. The annual mean of the location and scale parameter is highly correlated with orography. The annual cycle of the location parameter is strong in the northwest UK (peaking in late autumn or winter) and in East Anglia (where it peaks in late summer), and low in the Midlands. The annual cycle of the scale parameter exhibits a similar pattern with strongest amplitudes in East Anglia. The spatial patterns of the annual cycle phase suggest that they are linked to the dominance of frontal precipitation for generating extreme precipitation in the west and convective precipitation in the southeast of the UK. The shape parameter shows a gradient from positive values in the east to negative values in some areas of the west. We also estimate 10-year and 100-year return levels at each rain gauge, and interpolated across the UK. Copyright © 2008 Royal Meteorological Society

KEY WORDS   observed climate; daily precipitation; extreme value statistics; UK climate; annual cycle

Received 17 March 2008; Revised 23 June 2008; Accepted 23 September 2008

1. Introduction

The strongest weather impact on agriculture, economy and society results from rare extreme events, as, for instance, heat waves, heavy storms and flooding due to intense rainfall. The occurrence of precipitation extremes already shows persistent trends in many regions of the world (Trenberth et al., 2007) and is projected to further increase under anthropogenic global warming (Meehl et al., 2007).

For the UK, Osborn et al. (2000); Osborn and Hulme (2002) and Maraun *et al*. (2008 hereafter M2008), studied the decadal variability of the daily precipitation intensity distribution, and found trends towards heavy precipitation during winter, and to a lesser extent also during spring and autumn. Fowler and Kilsby (2003), however, found no overall trend in the annual 1-day precipitation maximum during recent decades.

Regional climate model simulations according to the IS92a and IPCC SRES scenarios project an increase in extreme precipitation across the UK (Jones and Reid, 2001; Ekström et al., 2005).

One method for analysing the occurrence rate of extreme weather events is extreme-value statistics (EVS)

(Embrechts et al., 1997; Coles, 2001). This branch of statistics has been widely applied in hydrology and climatology (Brown and Katz, 1995; Coles and Tawn, 1996; Coles and Casson, 1998; Katz, 1999; Katz et al., 2002; Naveau et al., 2005). For the popular block maxima approach, one divides the observed time series into blocks and aims to model the distribution of the maxima of these blocks. Most analyses of heavy precipitation consider either annual or seasonal maxima. As many applications require knowledge about return levels corresponding to return periods of the order of decades or even centuries, disregarding month-to-month variations is a seemingly justified simplification. Using this approach, one furthermore avoids the difficulties of explicitly modelling the annual cycle.

Precipitation in the UK, however, does show a pronounced annual cycle, and for the assessment of agricultural and hydrological impacts it is important to know when during the year precipitation extremes are expected to occur. While an increase in mean precipitation alone during the growing season tends to increase the agricultural yield, heavy precipitation might damage crops, especially in their juvenile stage. Furthermore, grain crops are highly vulnerable to flooding, and the intensity and timing of rainfall influence the persistence and efficiency of pesticides (Rosenzweig et al., 2001). The annual cycle of extremes is also important for flooding and erosion:

---

* Correspondence to: D. Maraun, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ UK. E-mail: d.maraun@uea.ac.uk

droughts reduce the capability of soil to absorb water; heavy rains following a period of drought – a situation more likely during summer than in winter – thus result in increased runoff and a higher potential for flooding (Rosenzweig et al., 2001). Likewise, the erosion of soil by heavy rainfall may depend on the time of year that the event occurs: erosion may be more likely if the soil is initially dry (Yu et al., 2006), or if there is little vegetation, such as on agricultural fields during winter (Favis-Mortlock, 2006). The latter effect is likely to increase, because more winter precipitation is expected to fall as rain, rather than snow, due to global warming (Nearinga et al., 2005).

Additionally, changes in extreme precipitation have not been homogeneous throughout the year. As stated above, M2008 found sustained trends towards heavier precipitation in winter, but during summer no clear trends manifested themselves. This observation is consistent with model projections of future climate (Jones and Reid, 2001; Ekström et al., 2005). Consequently, knowledge about the present-day occurrence of extreme precipitation during the year is essential to assess the impact of future changes. Also, the above-mentioned discrepancy between the results of M2008 and Fowler and Kilsby (2003) might become clearer by studying the changing occurrence of heavy rainfall during the year: if the annual maxima considered in the latter study occur predominantly during summer time (where no trends occur), the overall trend might be vanishing although precipitation extremes during other seasons become more likely.

Finally, there are valid reasons to investigate the annual cycle of extreme precipitation even if the focus is actually on annual maxima. For example, the block maxima approach requires identically distributed random variables within a block. If the occurrence rate of high magnitudes is changing from season to season, this condition is in general not fulfilled. By explicitly modelling the annual cycle we may reduce the violation of this condition.

Therefore, we develop a model for the annual cycle of precipitation extremes, based on EVS. Considering monthly maxima instead of annual maxima, we make better use of the rather limited data. We include the annual cycle in the form of a harmonic function that modulates the distribution of monthly maxima in the course of a year. The parametric (harmonic) form of the model reduces the uncertainty of the parameter estimates. This approach is based on a more general method described in Coles (2001) and Katz et al. (2003), who include external influences as covariates in their extreme-value analysis. We analyse a set of 689 rain gauge records across the UK, revealing coherent spatial patterns of extreme precipitation characteristics and their annual cycle.

In Section 2, we present the data used in this study, and in Section 3 we briefly introduce the concepts of extreme-value statistics, maximum likelihood estimation and covariates. The actual statistical model of the annual cycle is developed in Section 4. Results are presented in Section 5 and discussed in Section 6.

2. Data

The selection of daily precipitation data we use in this study is based on the (Met Office Integrated Data Archive System) MIDAS land surface observation data, provided by the British Atmospheric Data Centre (BADC, www.badc.ac.uk).

We choose a subset identical to the one in M2008, which is itself an update of earlier work first presented in Osborn et al. (2000). Our selection comprises 689 stations covering the whole UK, selected according to their overall length of record and a low number of missing values. The spatial coverage is dense in England, with fewer stations elsewhere, especially in the north of Scotland.

We use all data in the range of 1 January 1900– 31 December 2006; most gauges, however, commenced recording in January 1961, and for some stations no recent values (e.g. for the last decade) are available. For a detailed discussion of the selected rain gauges, including a list of stations, please refer to M2008 and the corresponding supplementary material. Dry days of zero precipitation are not removed in our analysis.

3. Methods

To analyse the annual cycle of extreme precipitation, we employ EVS. This branch of statistics aims to describe the occurrence rate of extreme values in a sequence of random numbers, that is, the tail of a probability distribution. The central theorem of EVS (Fisher-Tippett, or Three-Types Theorem) states that for increasingly large values, the tails of most probability distributions can be approximated using the general extreme-value distribution (GEV) (Embrechts et al., 1997; Coles, 2001). A standard approach of EVS describes the probability distribution of the most extreme values within a block of consecutive data points, the so-called block maxima approach, which we pursue here. We estimate the parameters of the GEV from a series of maxima with the maximum likelihood approach (Coles, 2001).

3.1. The generalized extreme-value distribution

Assume a physical process (here, precipitation) is represented by a sequence of n random variables, X_t ($t = 1..n$), which are independent and identically distributed (iid), with unknown distribution. Denote the maximum of this sequence as

$$M_n = \max\{X_1, \ldots, X_n\} \tag{1}$$

The Fisher-Tippett theorem states the following: if the probability distribution of the properly rescaled maximum converges for increasing block length ($n \to \infty$) to a

limiting distribution $G(z)$, then $G(z)$ belongs to the family of GEV distributions:

$$G(z; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (2)$$

defined on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The parameter μ is called the location parameter and determines the position of the distribution, the scale parameter σ determines the width, and ξ, the shape parameter, determines the decay of the distribution for large values of z: for $\xi < 0$, the tail has a finite upper value (Weibull distribution); for $\xi > 0$, the tail is long with a power law decay (Fréchet distribution). In the limit $\xi \to 0$ one obtains a Gumbel distribution with an exponential decay, i.e. a short tail (Embrechts et al., 1997; Coles, 2001). Leadbetter et al. (1983) have shown that the iid condition can be relaxed such that the Fisher-Tippet theorem holds also for a wide class of stationary, but not necessarily independent, stochastic processes.

For a set of empirical data, an infinite sequence is of course not possible. However, in many cases a reasonable approximation by the GEV can be reached already for finite values of n, depending on the auto-correlation and distribution of X_t. Furthermore, for the estimation of the GEV parameters, a sufficiently large number of observed maxima is needed. Therefore, one divides the time series into blocks and considers the maxima in these blocks. Here, one has to trade-off block length and number of maxima, that is, bias and variance (uncertainty). For many purposes in climatology, annual maxima are a preferred choice, not only because of the block length but also to avoid an explicit modelling of the seasonal cycle. In some cases, such as precipitation, monthly maxima can already be approximated sufficiently well with the GEV (see Section 5.1).

For risk assessment, one is interested in the probability of the observed variable (here, daily precipitation) exceeding a certain level. These levels are expressed as return levels r_T for a certain return period T; r_T is defined as the level which is exceeded with probability

$$P(z > r_T) = 1 - G(r_T; \mu, \sigma, \xi) = \frac{1}{T} \quad (3)$$

i.e. on average, once every T blocks. For further details on extreme-value statistics we refer the reader to the excellent introduction by Coles (2001) and the comprehensive book by Embrechts et al. (1997).

3.2. Maximum likelihood estimation

From an observed series, x_t, $t = 1..n \cdot m$, divided into m blocks of length n, one can extract a block maxima series z_i, $i = 1..m$, according to Equation 1. From this, one can estimate the parameters of a GEV distribution describing this series. One possible approach is maximum likelihood estimation (Edwards, 1992; Embrechts et al., 1997; Coles, 2001).

The likelihood for the parameters given a maxima series z_i is

$$L(\mu, \sigma, \xi \mid z_i) = \prod_{i=1}^{m} g(z_i; \mu, \sigma, \xi) \quad (4)$$

where $g(z; \mu, \sigma, \xi)$ is the density function of the distribution $G(z; \mu, \sigma, \xi)$. The likelihood is a function of the parameters μ, σ and ξ, for a given set of maxima z_i; as such, it is not a probability density function. It is, however, proportional to the probability that data z_i would occur given the parameters μ, σ and ξ. For a discussion of the difference between a probability and a likelihood refer to, for instance, Edwards (1992).

The idea of maximum likelihood estimation is now to adjust the parameters such that $L(\mu, \sigma, \xi \mid z_i)$ attains a maximum. The resulting vector

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \arg\max_{\mu,\sigma,\xi}\{L(\mu, \sigma, \xi)\} \quad (5)$$

is the maximum likelihood estimator (MLE) for the parameters. We solve Equation 5 by numerical optimization. The covariance matrix measuring the uncertainty of the estimates is calculated from the estimated Fisher information matrix. The latter is the second derivative of the log-likelihood function $\ell(\mu, \sigma, \xi \mid z_i) = \log L(\mu, \sigma, \xi \mid z_i)$ and is a measure for the curvature of the likelihood at the maximum (Coles, 2001).

Maximum likelihood estimation assumes that the data are a typical realization from the distribution with parameters $\hat{\mu}, \hat{\sigma}, \hat{\xi}$. Therefore, it works very well for large datasets; for very limited data, one might prefer other strategies, e.g. probability-weighted moments (Hosking et al., 1985; Embrechts et al., 1997).

However, the MLE has an advantage which we will use in the following: GEV parameters that depend on time or external variables can straightforwardly be included in the model.

In climatology, the characteristics of extremes are often non-stationary and depend on changes in large-scale processes, seasonality or long-term trends. Consequently, the GEV parameters are no longer constants but functions of a driving process or time. For instance, the constant μ (or σ or ξ) can be replaced by the function

$$\mu = \mu(t) = \mu_0 + a_\mu \cdot c(t) \quad (6)$$

The $c(t)$ can either be a function in time, for instance, a parametric trend or a harmonic function, or it can be the observation of a process that influences the extremes, for instance, a large-scale weather index. In the latter case, $c(t)$ is called a covariate (Coles, 2001; Katz et al., 2003).

These new parameters, μ_0 and a_μ (or the corresponding parameters for σ and ξ), can be estimated in the same way as μ, σ and ξ using Equation 5.

4. Developing the statistical model of the annual cycle

4.1. Exploratory fit

To investigate how the distribution of extreme precipitation events in the UK changes during the year, we first carried out an exploratory study. We separated each station's time series into 12 sub-series, 1 for each month. Within these sub-series, we chose the length of the month as the block length, n. The difference in lenghts of the months of 1, or maximally 3 days, is disregarded. In Section 5.1, we will demonstrate that this block length is already sufficient to yield an adequate approximation of the maxima distribution with the GEV. We also repeated the analysis with a block length of 2 months (combining the months of 2 consecutive years) and found no significant change in the results. A potential reason is that the distribution of daily precipitation (commonly modelled as gamma, e.g. Osborn and Hulme (2002)) is already close to a GEV distribution. Furthermore, the low autocorrelation of the precipitation process does not slow down the convergence of the maxima distribution towards the GEV.

Examples from three typical rain gauges are shown in Figure 1, 3 and 5. The black crosses with error bars depict the monthly estimates for μ, σ, ξ, and the coefficient of variation, σ/μ, as well as the 0.1 quantile and the 0.01 quantile, with the corresponding 95% confidence intervals. The solid line is a sine fitted to the estimates for the single months.

Throughout the text, we will refer to these exploratory estimators as the exploratory (EX) model. In the following, the EX model will be confronted with a sinusoidal model.

4.2. The sinusoidal model

The exploratory study showed that, especially for μ and σ, a sine function with amplitude and phase as parameters
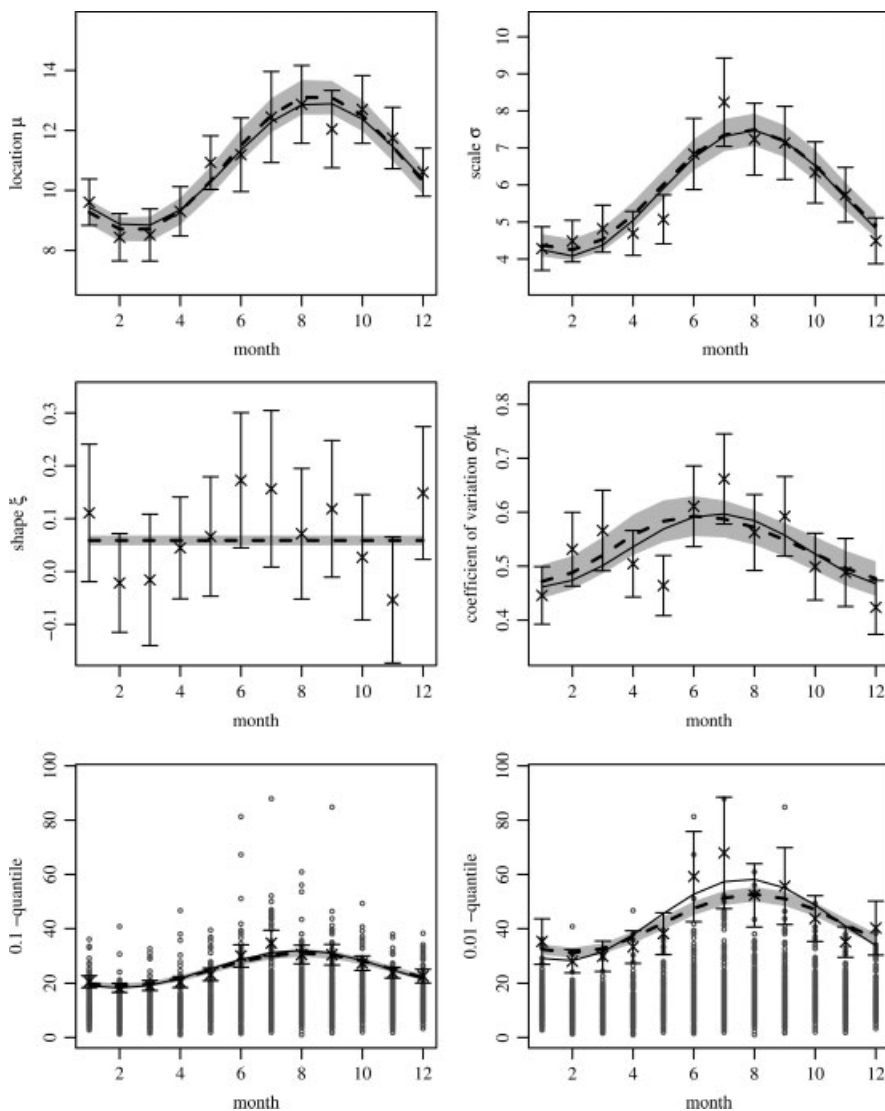

Figure 1. Oxford (England); GEV fit for monthly maxima of daily precipitation. Crosses plus bars: 12 separate fits, one for each month (EX model) and the corresponding 1.96σ confidence intervals; Dashed line and shading: The sinusoidal model for the annual cycle (SIN model), including 1.96σ confidence bands. Dots in the lower panels are the observational maxima for each month over the whole record.

may be an appropriate model for the annual cycle. To obtain a model which is linear in the parameters, we modelled μ and σ as a combination of sine and a cosine with a 1-year period, evaluated at the centre-day of each month, c_i:

$$
\begin{aligned}
\mu_i = \mu_0 &+ a_\mu \sin\left(\frac{2\pi c_i}{365.25}\right) \\
&+ b_\mu \cos\left(\frac{2\pi c_i}{365.25}\right), \quad i = 1\ldots 12
\end{aligned} \quad (7)
$$

σ_i being modelled correspondingly. In the following, we will refer to the sinusoidal model as the SIN model.

In comparison to the EX model, which considers each month separately, the SIN model of the location and scale parameter has got the advantage that it combines information from all months, and uses all available data to estimate seven parameters ($\mu_0, a_\mu, b_\mu, \sigma_0, a_\sigma, b_\sigma, \xi$). In the former case, we have to estimate 36 parameters (μ, σ, ξ for each month) with the same amount of data. This implies a significant reduction in degrees of freedom of the model, and thus, reduces the uncertainty in the derived quantities, e.g. return levels. This advantage, however, will only be realized if the simplified model is able to capture the annual cycle behaviour in an appropriate way (see Section 4.3 for a discussion of our choice of a constant shape parameter at each location).

For every rain gauge, we estimate the parameters of the SIN model, combined into the vector $\theta = (\mu, a_\mu, b_\mu, \sigma_0, a_\sigma, b_\sigma)$, by the maximum likelihood approach Equation 5. Again, three typical examples are shown in Figures 1, 3 and 5. The result for the SIN model is shown as a dashed line, with shading depicting the 95% confidence bands. The agreement of the SIN model (dashed line) with the sine (solid line) fitted to the results from the EX model demonstrates the suitability of the model assumption. The climatological discussion will follow in Section 6.

4.3. Modelling the shape parameter

Figures 1, 3 and 5 show that a SIN model is not easily justified for the shape parameter. The estimated values of ξ for the EX model do not clearly follow a simple sine, and the confidence bands are so wide that the individual monthly values are compatible with a constant value throughout the year, or in some cases even with zero. Thus, we investigated whether a constant model for ξ is suitable, or whether an annual cycle (ξ_0, a_ξ, b_ξ) has to be taken into account. For this purpose, we employed two criteria: the consistency of the spatial structure of ξ_0, a_ξ and b_ξ; and, additionally, a likelihood ratio test (Cox and Hinkley, 1994).

For ξ, a coherent spatial pattern only emerges for the constant offset (see Figure 7(g)), confirming our assumption not to enforce $\xi = 0$. However, no coherent pattern appears for the amplitude and phase of the annual cycle in ξ (not shown). The likelihood ratio test, performed for each station, assesses whether the annual cycle in ξ results in a significant improvement of the estimation. A significant number of stations (that is, more stations than expected by chance on the 1, 5 and 10% levels) showed a significant improvement of the likelihood with an annual cycle in ξ; but also here, no spatially coherent structure occurred. This result suggests that the annual cycle in ξ may compensate for minor model mis-specifications at some stations, but has no physical relevance.

Based on this outcome, we decided to enforce a constant shape parameter throughout the year, estimated individually for each station. Investigations of several randomly selected examples (not shown) revealed that the omission of the annual cycle in ξ introduces only a marginal change in the estimated quantiles, but considerably reduces the parameter uncertainty.

4.4. Smoothing and refitting

The spatial coherence of many of the estimated parameters suggests to combine the information of single stations. As parameters vary slowly in space, smoothing the values of neighbouring stations would gain a reduction in variance to the cost of only a little bias: some small-scale structures, perhaps due to orography, are not accounted for ('smoothed out'). In addition to the reduction in variance, smoothing allows one to interpolate from a network of stations to an arbitrary point in the UK. The accuracy of this interpolation obviously depends on the density of the network and, again, on the smoothness of the parameter changes. In a mountainous region, where small-scale orography affects the local precipitation, a narrower smoothing window (and thus, a denser station coverage) is necessary, compared to a flat region. Unfortunately, stations cluster in England, as opposed to a sparse network in the Scottish Highlands. Thus, our interpolations, especially for Northern Scotland, should be regarded as a guidance to highlight large-scale patterns only, local characteristics are not resolved.

We smoothed ξ_0, the spatially coherent parameter estimated with the highest uncertainty. To keep calculations simple and comprehensible, we decided against a more sophisticated kriging algorithm (Cressie, 1991) but choosing instead a kernel smoother,

$$
\overline{\xi}_{0,i} = \sum_{j=1}^{N_s} w_{i,j} \hat{\xi}_{0,j} \quad (8)
$$

Here, the weights $w_{i,j}$ between a point i in the UK and a station j are chosen as a Gaussian kernel with a cross-section width of 30 km, with $\sum_{j=1}^{N} w_{i,j} = 1$. The number of stations N is 689. The smoothed $\overline{\xi}_{0,i}$ now has a greatly reduced standard error $\sigma_{\overline{\xi}}$.

Ideally, the averaging would be done simultaneously with the maximum likelihood estimation. This, however, would involve large computational cost and probably also cause numerical problems. Thus, we decided to smooth the maximum likelihood estimate $\hat{\xi}_0$ of the SIN model, and then re-estimate all other parameters

with a fixed $\xi_0 = \overline{\xi}_0$ for every station (i.e. each station's maximum likelihood estimate of ξ_0 is replaced by the value from the smoothed field for that location). The reduced parameter vector of this model is then $\theta_{-\xi_0} = (\mu_0, a_\mu, b_\mu, \sigma_0, a_\sigma, b_\sigma)$, leading to a new maximum likelihood estimate. A Monte Carlo study (not shown) indicated that the standard error of $\overline{\xi}$, $\sigma_{\overline{\xi}}$, was small enough not to affect the (reduced) covariance matrix of the re-estimated parameters.

The increase in bias of these estimates compared to the fit including a (variable) parameter ξ_0 is low, but the variance is considerably reduced. Smoothing the quite accurately estimated scale and location parameter would not gain much reduction of variance; hence, we forbear from any further spatial smoothing.

5. Results

In the following, all results from the SIN model are obtained with a spatially smoothed and fixed value of the shape parameter $\overline{\xi}_0$ (without an annual cycle), and re-estimated location and scale parameters, each with a sinusoidal annual cycle.

5.1. Example stations

A typical example of an observation where daily precipitation extremes are well modelled by a sinusoidal annual cycle (i.e. where the SIN model fits well) is the record from Oxford (Met Office Source Identifier 606). We used data from 1 January 1900 to 31 December 2006. One can clearly see the strong annual cycle in the location and scale parameter μ and σ (upper row in Figure 1).

The suitability of the SIN model becomes evident from the good agreement between the results for the individual months (the EX model, crosses), and the dashed curve depicting the SIN model. One also sees the reduction of the error bars by about 50%, when using the SIN model instead of the EX model. The middle row shows the fit of the shape parameter ξ and the coefficient of variation σ/μ. The shape parameter apparently exhibits a bi-annual cycle (crosses), but all values for the single months are compatible with the constant value we included in the SIN model. In some studies, the coefficient of variation is modelled as a constant to further simplify the statistical model. Our results indicate that here this assumption is not justified: the annual cycle for μ and σ are slightly out of phase, which results in an asymmetric sine-like shape of the coefficient of variation (though its amplitude is reduced relative to the annual cycle of μ and σ).

The bottom row of Figure 1 shows the 0.1 and 0.01 quantiles, that is, the magnitude exceeded with a probability of 10 and 1% in a certain month, respectively. Good agreement between the EX and the SIN model is evident again. However, for the 0.01 quantile, a deviation becomes apparent: for the EX model, some extremely high values in June, July and September lead to high estimates of the shape parameter for these months, and thus, also to high quantiles (crosses). In the SIN model, the time-independent shape parameter ξ is estimated from values of the whole year. In this representation, the record maximum precipitation values in June, July and September are seen as quite improbable 'outliers'; hence the estimates of the corresponding quantiles in the SIN model are considerably lower (dashed line) than those from the EX model. Nevertheless, in both models the highest extreme precipitation events for Oxford are to be expected in August.

The diagnostic plots in Figure 2 confirm the suitability of the SIN model. The left panel shows a probability–probability (P-P) plot, that is, the empirical frequency distribution of the data against the probability distribution of the fitted model. The right panel shows a quantile–quantile (Q-Q) plot, that is, the empirical quantiles against the quantiles of the fitted model. Due to the time-dependent location and scale parameter, the distribution varies throughout the year. Thus, both the data and the probability distribution of the model have been re-normalized to a time-independent Gumbel distribution. (For each month, the distribution is transformed to have zero location, unit scale, and vanishing shape parameter. This enables one to combine all results into one diagnostic plot instead of twelve separate ones (Coles, 2001)). In both cases, the bisector indicates a
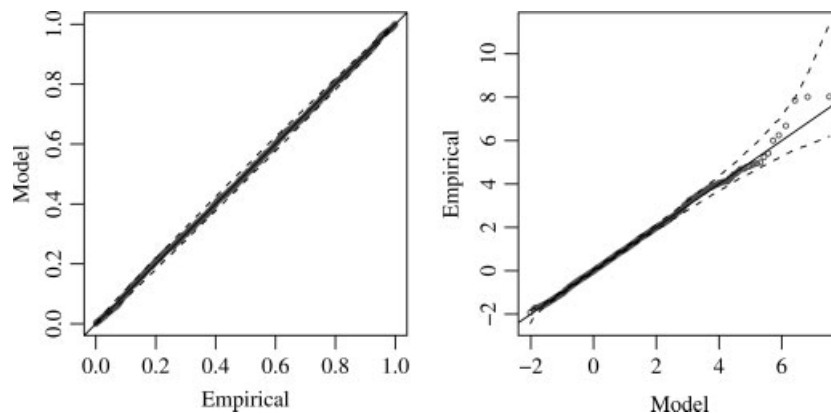

Figure 2. Oxford (England); diagnostic P-P plot (a) and Q-Q plot (b) of the SIN model. The fitted time-varying distribution (and the data, respectively) is re-normalized to a time-independent Gumbel distribution. 95% confidence intervals (dashed) are obtained from 999 realizations of this Gumbel distribution.

perfect agreement between data and model. The P-P plot focuses on the mass of the distribution, the Q-Q plot reveals more information about the tails. The latter one complements the quantile plots shown in the bottom row of Figure 1. The three extremely high values occurring in June, July and September lead to an upswing of the data quantiles compared to the model quantiles in the extreme tail of the Q-Q plot. However, except this slight discrepancy involving just three monthly maxima from the 107-year record, the overall agreement is high.

Kinlochewe in the northwestern Scottish Highlands (Figures 3 and 4, Source Identifier 66, data for 1 January 1961–1 December 2006) is another example for a good fit of the SIN model. Here, the annual cycle in ξ is even less pronounced. Unlike in Oxford, the most extreme precipitation is expected in December and January. The diagnostic plots in Figure 4 shows that the SIN model in general performs very well; only some very extreme values in February and March, as well as in September and October, are not captured by the SIN model. This causes the slight deviation of the Q-Q plot for very large values.

The time series of Aberporth (Wales, Source Identifier 1198, data for 2 January 1941–31 December 2006) shows some extremely high values between June and October, which are not well captured by the SIN model (see Figure 5). In the EX model, these values result in high estimates of the shape parameter during this season. The constant shape parameter in the SIN model hence leads to a much lower estimation of the return levels compared to the EX model (although the resulting curve (dashed) still lies within the confidence intervals of the EX model). This discrepancy could indicate that the SIN model is misspecified for some stations where the shape parameter changes throughout the year. However, the discussion of return level estimates in Section 5.3 suggests that the SIN model, in combining information of the whole year, simply reduces the uncertainty in the shape parameter estimation.

The discrepancies apparent from the quantile plots in Figure 5 are reflected in a deviation of the empirical quantiles from the SIN model quantiles for extreme
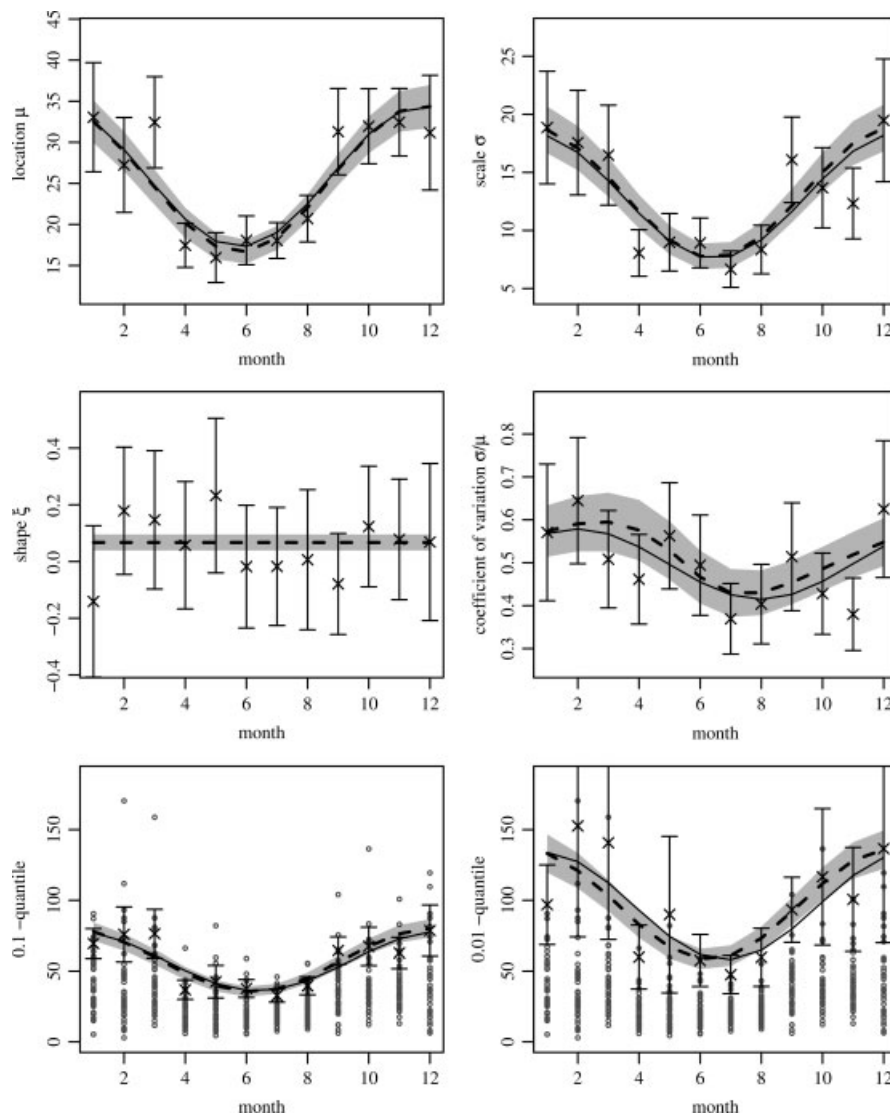

Figure 3. Kinlochewe (Scotland); GEV fit for monthly maxima of daily precipitation, as in Figure 1.
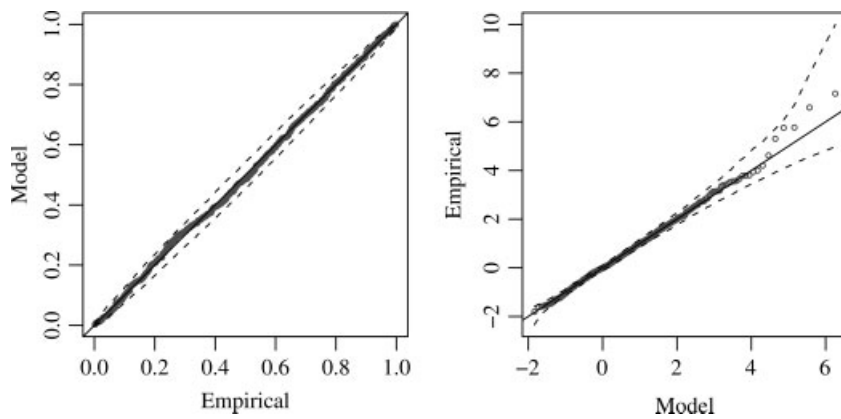
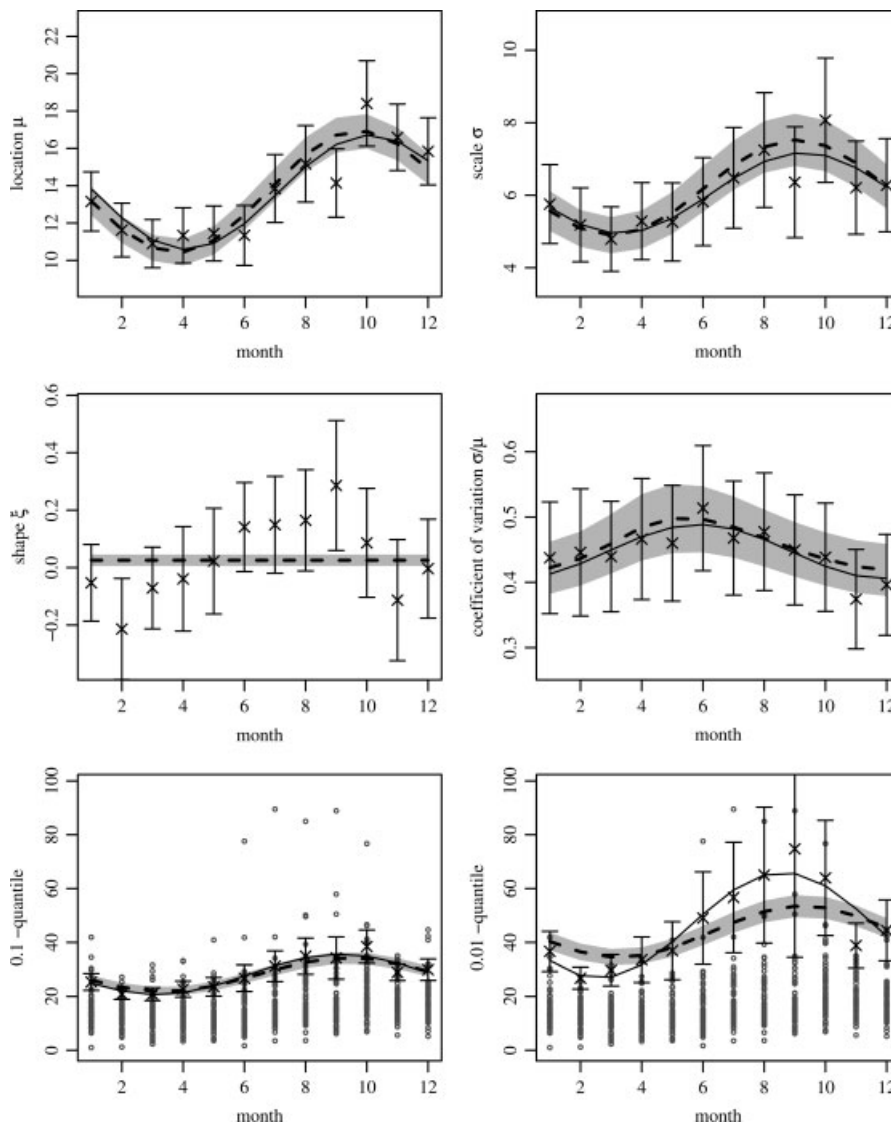Figure 4. Kinlochewe (Scotland); diagnostic P-P plot (a) and Q-Q plot (b); as in Figure 2.

Figure 5. Aberporth (Wales); GEV fit for monthly maxima of daily precipitation, as in Figure 1.

values, see the right panel in Figure 6. The five extremely high values occurring from June to October are clearly visible.

The presented examples demonstrate the benefits and limits of the SIN model. On the one hand, the model helps to significantly reduce the uncertainty of the estimated

parameters. One the other hand, the model might be misspecified for individual situations.

For Oxford and Kinlochewe, small but significant deviations of the scale parameter from a perfect sine are visible (Figures 1 and 3). Here, including higher-order harmonic functions might improve the performance of
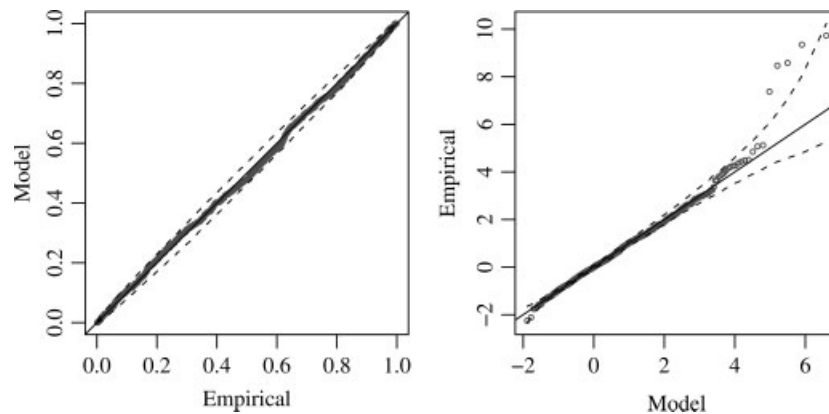
Figure 6. Aberporth (Wales); diagnostic P-P plot (a) and Q-Q plot (b); as in Figure 2.

the model. A similar argument holds for the shape parameter in Aberporth (Figure 5). However, as we aim to extract the most relevant physical characteristics of the annual cycle of heavy precipitation, and its spatial characteristics, we keep the model as simple as reasonably possible.

Aberporth has been selected as an example of a small number of stations, where some extremely high values are not well captured by the SIN model, as visible in the Q-Q plot (Figure 6). Apart from the points discussed above, the discrepancy could be explained by the presence of multiple distributions (see, e.g. van den Brink *et al.*, 2004; Vannitsem, 2007). However, as these 'outlier' stations do not cluster in certain regions, there is no further evidence for such a case. More likely, the problem could also be caused simply by finite sampling: analysing 689 finite realizations, some of them will almost certainly be untypical such that the proposed model does not fit perfectly.

5.2. Spatial structure

Figures 7 presents the spatial distribution of the estimated GEV parameters obtained for the SIN model. The circles depict the results for individual stations, after re-estimating the parameters of the reduced SIN model (see Section 4.4); the coloured background is a kernel interpolation as described in Section 4.4 to highlight large-scale spatial patterns. The rain gauges were chosen according to their completeness and length, to provide a reasonable amount of data for each month. As a result, the spatial coverage in some regions, especially in Scotland, is quite sparse. We would like to point out that, in consequence, small-scale structures due to small-scale orographic effects are not represented in this study.

Figure 7(a−c) shows the spatial pattern of the annual cycle in the location parameter $\mu_i = \mu_0 + a_\mu \sin(2\pi i/12) + b_\mu cos(2\pi i/12)$, namely, the annual mean μ_0, the relative amplitude $\sqrt{a_\mu^2 + b_\mu^2}/\mu_0$, and the phase (the date when the sine wave reaches its maximum and is plotted).

A distinct feature is the strong east–west gradient. Lowest values of the location parameter offset μ_0 occur in East Anglia, the highest values along the west coast,

especially in Cornwall, South Wales, the Lake District as well as the western Scottish Uplands and the southwestern Scottish Highlands. Interestingly, the location parameter is relatively low in the northern Highlands, the Outer Hebrides, the Orkney and Shetland Islands and Northern Ireland though in some of these areas the rain gauge network is sparse and is dominated by stations at lower elevations.

Relative to the annual mean, the annual cycle in the location parameter is, as expected, strongest in West Scotland including the Hebrides, Orkneys and Shetlands, the Lake District, Cornwall and North Wales around Snowdonia. However, high values also occur in some areas around the south and east coasts, especially in East Anglia and along the English Channel.

The phase of the annual cycle finally shows a clear east–west gradient, with highest values in August in the heartland of England, and highest values in November along the south coast, the west coast, most of Wales and Northern Ireland and Scotland, with the exception of the northern and eastern Scottish coasts, where the highest values occur 1 or 2 months earlier.

A similar pattern emerges for the scale parameter, see Figure 7(d−f). The basic structure for the offset resembles that of the location parameter, although the values in the northeast of Scotland, the Outer Hebrides, the Orkneys and Shetlands are notably lower compared to the rest of the UK. But as this observation is based on a low number of rain gauges, it should not be overemphasized.

The overall pattern of the relative amplitude of the scale parameter annual cycle is similar to that of the location parameter, with two pronounced differences. First, along the west coast, the relative amplitude is high only in Scotland; in Wales and Cornwall it is weak. Second, the area of a high relative amplitude in the southeast is even more clearly marked.

The phase-pattern of the scale parameter is almost indistinguishable from that of the location parameter, but its annual cycle peaks earlier in the heart of England (already in July) and even later in the southwestern Scottish Highlands (in December).
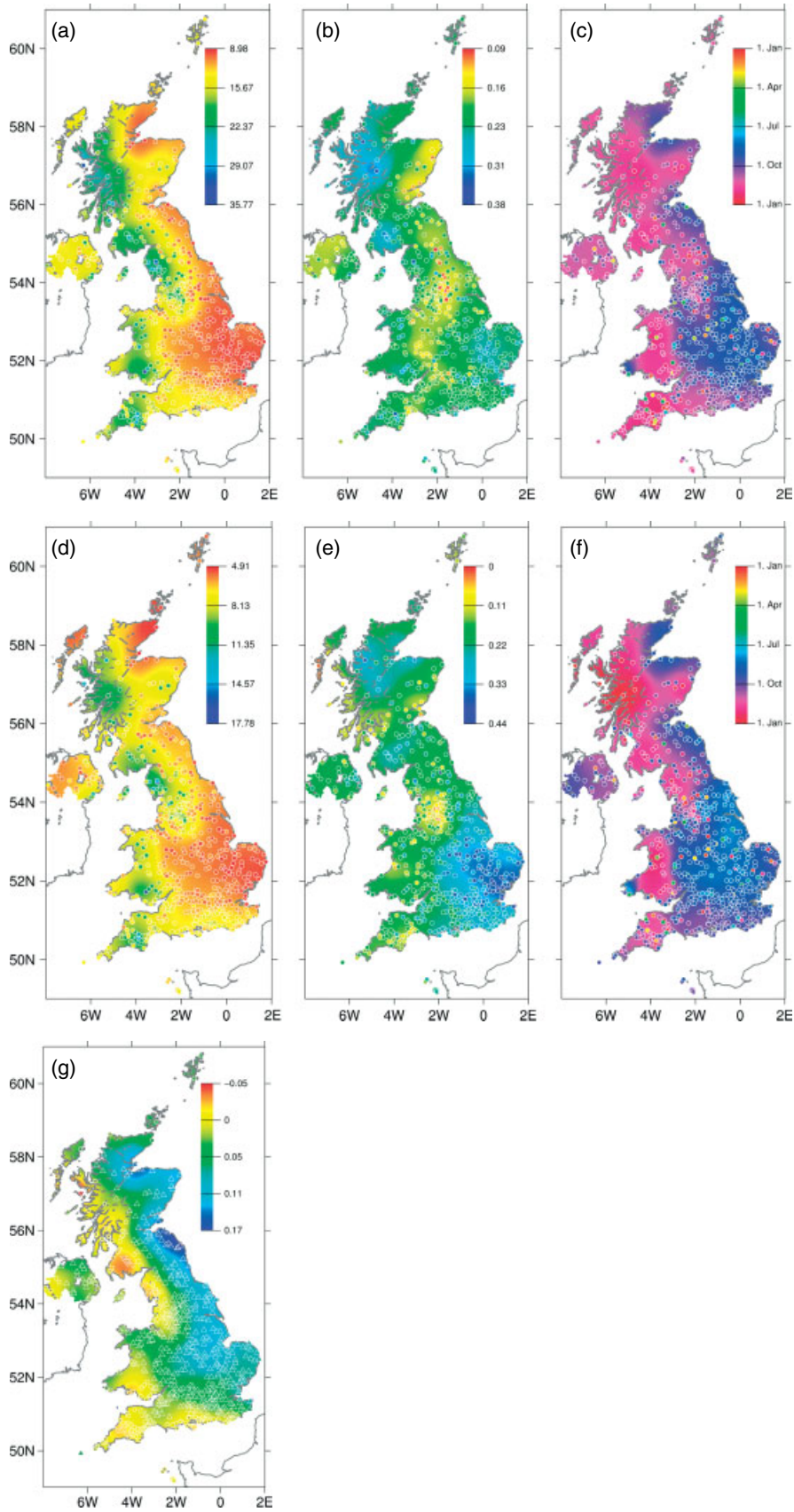
Figure 7. (a−c) Location parameter, (d−f) scale parameter. <u>Circles:</u> values for single stations; background: kernel average. (a) and (d) offset μ_0 and σ_0, (b) and (e) relative amplitude $\sqrt{a_\mu^2 + b_\mu^2}/\mu_0$ and $\sqrt{a_\sigma^2 + b_\sigma^2}/\sigma_0$ of the annual cycle, (c) and (f) phase of the annual cycle (position of the sine's maximum). (g) Shape parameter ξ. Triangles and Circles: values for single stations; background: kernel average. Upper (lower) triangles depict positive (negative) estimates not compatible with zero, circles depict estimates compatible with zero on a 95% level.
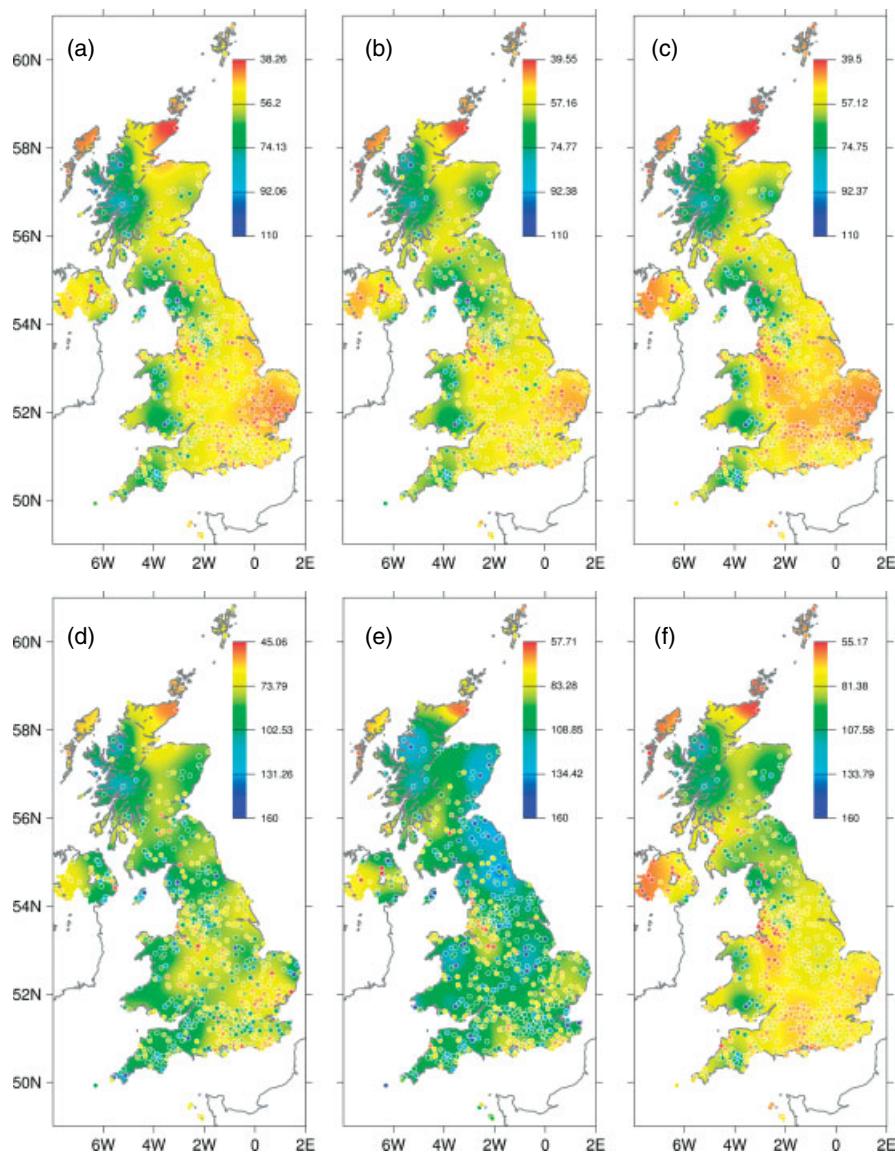
Figure 8. (a–c) 10-year return level, derived from (a) annual maxima, (b) the EX model and (c) the SIN model. Values lower than 35 mm (larger than 110 mm) have been plotted in red (blue). (d–f) 100-year return level, derived from (d) annual maxima, (e) the EX model and (f) the SIN model. Values lower than 40 mm (larger than 160 mm) have been plotted in red (blue).

Figure 7(g) depicts the spatial pattern of the shape parameter. The triangles and circles represent the results for individual fits of the full SIN model. The circles mark all values compatible with zero, i.e. zero lies within their 1.96 standard error intervals (95% significance level). Upper (lower) triangles stand for positive (negative) values not compatible with zero. The background shows the kernel averaged results. The background values at the station locations have been used for the re-estimation of the reduced SIN model, see Section 4.4.

The shape parameter exhibits an interesting southwest-northeast gradient, which is quite different from the pattern of the other parameters. Values are significantly positive (upper triangles) over almost all the UK away from the south and west coasts, in Northern Ireland and North Wales, with higher values in the Midlands, the east coast and the northern Highlands. Maxima occur along the east coast around Inverness, and between Edinburgh

and Newcastle. Zero and negative values occur along the south and west coast, with clusters of negative values in Cornwall, the Lake District, and the southern Uplands of Scotland. This pattern clearly justifies the use of the GEV distribution instead of the Gumbel distribution (with $\xi = 0$).

5.3. Return levels

Return level estimates are often derived with the annual maxima approach. For many cases in risk assessment, one is interested in extreme levels of long return periods, and not in the detailed shape of the annual cycle. It may, therefore, appear to be a reasonable choice to consider only one value per year. The block maxima approach, however, assumes a stationary process. In general, the maximum of a non-stationary process is not well modelled by a GEV distribution. This means that the block maxima approach is only valid for a

process without, for instance, an annual cycle. A case where the GEV might be a reasonable approximation is a situation, where the extreme precipitation occurs only in a particular pronounced season. As demonstrated in the previous sections, UK precipitation extremes show a distinct, though spatially varying, annual cycle and, *a priori*, it is not clear whether the annual maxima approach provides reasonable results.

In the following, we compare return levels derived from the annual maxima approach with those from the EX model and the SIN model. The estimation of return levels from annual maxima is straightforward according to Equation 3. Return levels from the EX model and the SIN model have to be calculated numerically. When $G_i(z)$ $(i = 1, \ldots, 12)$ denotes the probability of the occurrence of a value smaller than z in the month i, then the T-year return level r_T can be calculated by solving the equation

$$\prod_{i=1}^{12} G_i(r_T) = 1 - \frac{1}{T} \qquad (9)$$

Corresponding confidence intervals for the SIN and the EX model can be derived by propagating the parameter uncertainties through Equation 9. These intervals will depend on the actual length of the observation and the desired return level. For a guideline of the expected range, we refer the reader to the 0.9 and 0.99 quantile plots for Oxford and Kinlochewe (Figures 1 and 3), representing the monthly 10-year and 100-year return levels for a long and a rather short observation, respectively.

We estimate return levels for all stations. By spatial interpolation, we highlight large-scale structures of return levels, disregarding small-scale orographic effects.

The return levels derived from annual maxima should suffer from the non-stationarity given by the annual cycle. The EX model explicitly, and without any constraints, models the annual cycle based on monthly maxima. It relies on the suitability of approximating monthly maxima with the GEV, and on the stationarity assumption within every month. As this model has the most degrees of freedom, it is afflicted with the widest confidence intervals. The shape parameter, estimated for every month separately, might be particularly sensitive to extreme 'outliers'. The SIN model provides return level estimates with considerably narrower error intervals, yet it relies on the suitability of the sine shape to model the annual cycle of the location and scale parameters, and the assumption that the shape parameter is invariant during the year. Figure 8(a−c) shows the 10-year return levels estimated from (a) annual maxima, (b) the EX model and (c) the SIN model. At first glance, the three panels are virtually indistinguishable.

The most notable exceptions are that estimates from the EX model are higher than those from the other approaches in two regions: southern East Anglia and southeast Scotland/northeast England. One can see the expected east–west gradient, with higher return levels along the west coast and lower values especially in the

southeast. Particularly high return levels prevail in the western Highlands and the Lake District.

Figure 8(d−f) shows the 100-year return levels estimated from (d) annual maxima, (e) the EX model and (f) the SIN model. A marked discrepancy stands out: values estimated from the EX model are larger than those derived from the SIN model or the annual maxima approach in most parts of the UK. This discrepancy is especially strong in the coastal stretch from Edinburgh to Newcastle where the shape parameter is positive and relatively large; a rather small discrepancy is found in the Lake District and the western Highlands where the shape parameter is compatible with zero or even negative (Figure 7(g)). Large positive shape parameter estimates can be a consequence of a few extremely high rainfall events. In the EX model, shape parameters are estimated separately for every month of the year which results in large positive shape parameter estimates for months with a few large events (see Figure 5). These months contribute strongly to the calculation of extreme return levels, while months with small shape parameters do not. This flexibility in the shape parameter to vary during the year is neither present in the SIN model nor in the annual maxima approach; shape parameter estimates are confined to a smaller value, closer to an annual average than to the extreme estimates in the EX model. As a consequence, the SIN model and the annual maxima approach yield smaller 100-year return levels. Without any further *a priori* knowledge about the model structure (e.g. physical reasons for a time-constant shape parameter), it is not clear which of the return level estimates are more realistic.

The differences between the 100-year return level estimates from the SIN model and the annual maxima approach are less pronounced. In general, the latter results show a higher spatial variability, a fact not surprising, since the former approach combines the shape parameters of neighbouring rain gauges. The relatively good agreement between the two approaches suggests that annual maxima mostly provide reasonable return level estimates, despite the previously discussed theoretical limitations. Equally, the agreement supports the specification of our SIN model.

The overall spatial pattern of the 100-year return level resembles that of the 10-year return levels: highest values are observed in the western Highlands and the Lake District, lowest values in the southeast.

6. Discussion and conclusions

We studied the annual cycle of heavy daily precipitation across the UK by means of EVS, and approximated the distribution of monthly maxima by the GEV distribution. To combine the information from individual months, we developed a parametric model which described the annual variations in the location and scale parameter of the GEV distribution as a phase-shifted sine function, and assumed a time-independent shape parameter.

This model proved to be suitable: the approximation of the monthly maxima distribution by the GEV and the annual cycle by a sine wave both appeared to be reasonable. The parametrization in form of a sine wave helped to considerably reduce the parameter uncertainties compared with a model that was fit to each month independently. The combination of the estimated shape parameter values from neighbouring rain gauges further reduced parameter uncertainties.

Our statistical results demonstrate that the scale parameter, in general, is not in phase with the location parameter (Figures 7(c) and 7(f)), although the spatial patterns of their annual means are highly correlated (Figures 7(a) and 7(d)). Thus, modelling the shape parameter as a constant multiple of the location parameter is not valid when looking at the seasonal variations of extreme precipitation. For a study of annual maxima, the phase shift is irrelevant and a combination of location and scale parameter might be useful, although their ratio might be spatially varying.

We compared return levels derived from annual maxima, an exploratory model of separately considered monthly maxima and the parametric sinusoidal model (see Figure 8). We found that, although the annual maxima approach relies on a stationarity assumption of daily precipitation, it provides, in general, reasonable results. The exploratory model is afflicted with the highest parameter uncertainties and, perhaps, systematically overestimates the shape parameter, and hence, the return levels of rare events (e.g. 100-year events). The sinusoidal model proved to be a good compromise between a bias due to a stationarity assumption and the uncertainty owing to too many parameters.

Our study provides detailed insight into the seasonal variations of extreme precipitation in the UK. Along the middle north-south axis, the amplitude of the annual cycle in the location parameter is less than 15% of the corresponding annual mean (Figure 7 (b)), in the scale parameter even just around 10% (Figure 7(e)). However, along the west coast and in the southeast, seasonal variations are strong, with up to 30% in the location parameter and 40% in the scale parameter.

The season when precipitation extremes are most likely to occur depends strongly on the region as well: along the west coast, the heaviest precipitation is expected during late autumn and winter, whereas along the east coast and the Midlands, the maximum location and scale parameters occur during late summer (Figures 7 (c) and (f)).

The estimated return levels agree well with the results published in the Flood Estimation Handbook (Faulkner, 1999). Those results additionally account for small-scale variations by additional empirical knowledge. In that study, however, the accuracy of very high return levels might be limited because of the restriction to a Gumbel distribution, that is, a vanishing shape parameter. In our analysis, a coherent spatial pattern of the shape parameter emerged, indicating that different geographical situations might influence the shape parameter (see Figure 7(g)).

Our results are related to different processes of precipitation and driving mechanisms. The offset of the location parameter is correlated (Pearson) at 0.65 with the maximum elevation in the 10-km vicinity of a rain gauge (topography data from the USGS ETOPO30 dataset), clearly showing the effect of mountain ranges on orographic precipitation.

Regions of predominantly convective rainfall extremes can be identified by extremes occurring in summer. In the UK, these are, especially, the Midlands and the east of England. Here the location parameter is low without a strong annual cycle. The scale parameter in this region also has a low offset, but shows a pronounced annual cycle. In other words: in Central and East England, precipitation is generally low, but from time to time very heavy summer thunderstorms occur. These findings agree notably well with the thunderstorm climatology for the UK, developed by Perry and Hollis (2005, thunderstorm climatology, available at http://www.metoffice.gov.uk/climate/uk/averages/19712000/mapped.html). Convective precipitation and thunderstorms also accompany frontal precipitation during winter, especially along the remotest parts of the British west coast. These areas, dominated by frontal precipitation, are characterized by low values of the shape parameter, and high values of the location and scale parameter, peaking in winter.

Between the frontal dominated climate along the west coast with extremes occurring during winter, and the somewhat more continental climate with extremes arising predominantly from summer convection, a transition zone exists. Here, the overall annual cycle is weak, with convective rain contributing during summer, and frontal precipitation contributing during winter.

Our results help to assess the future impact of climate change. Climate models predict an increase in UK heavy precipitation throughout the year, though more consistently during winter than summer (Christensen et al., 2007). On the one hand, our study identifies regions and seasons, where extremes are already strong and might get even stronger (e.g. early winter in western UK). On the other hand, we showed regions and seasons, where extreme precipitation is weaker at present but might become significant in the future (e.g. winter in East Anglia). These results might prove important for agriculture and hydro engineering.

Future work could further investigate the different precipitation processes discussed above and try to incorporate, for instance, the elevation into the statistical model. Furthermore, it could be interesting to add a higher harmonic to the model to decrease possible misspecifications, and to study whether the shape of the annual cycle might be region dependent. A recently developed approach by Heffernan and Tawn (2004) can be used to assess the interdependencies between precipitation events at different stations, and provide estimates for the total volume of rainfall in a single event. Finally, our results might be used to evaluate the performance of

regional climate models at simulating the annual cycle of extreme precipitation.

Acknowledgements

References

Brown BG, Katz RW. 1995. Regional analysis of temperature extremes: spatial analog for climate change? *Journal of Climate* **8**(1): 108–119.

Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli RK, Kwon WT, Laprise R, Rueda VM, Mearns L, Menéndez CG, Räisänen J, Rinke A, Sarr A, Whetton P. 2007. Climate change 2007: the physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter Regional Climate Projections*. Cambridge University Press: Cambridge.

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics*. Springer: London.

Coles S, Casson E. 1998. Extreme value modelling of hurricane wind speeds. *Structural Safety* **20**(3): 283–296.

Coles S, Stephenson A. 2006. *Ismev: An Introduction to Statistical Modeling of Extreme Values*, R package version 1. 2.

Coles SG, Tawn JA. 1996. A bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society. Series C, (Applied Statistics)* **45**(4): 463–478.

Cox DR, Hinkley DV. 1994. *Theoretical Statistics*. Chapman & Hall: London.

Cressie NAC. 1991. Statistics for spatial data. *Wiley Series in Probability and Mathematical Statistics*. Wiley: London.

Edwards AWF. 1992. *Likelihood*. Johns Hopkins University Press: Baltimore.

Ekström M, Fowler HJ, Kilsby CG, Jones PD. 2005. New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 2. Future estimates and use in impact studies. *Journal of Hydrology* **300**: 234–251.

Embrechts P, Klüppelberg C, Mikosch T. 1997. Modelling extremal events for insurance and finance. *Applications in Mathematics*. Springer: Berlin.

Faulkner D. 1999. Flood estimation handbook. *Rainfall Frequency Estimation*, Vol. 2. Institute of Hydrology: Wallingford.

Favis-Mortlock D. 2006. *Encyclopedia of Soil Science*, chapter Erosion by Water. CRC Press: Boca Raton; 568–573.

Fowler HJ, Kilsby CG. 2003. A regional frequency analysis of United Kingdom extreme rainfall from 1961–2000. *International Journal of Climatology* **23**: 1313–1334.

Heffernan JE, Tawn JA. 2004. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B* **66**(3): 497–530.

Hosking JRM, Wallis JR, Wood EF. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**(3): 251–261.

Jones PD, Reid PA. 2001. Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *International Journal of Climatology* **21**: 1337–1356.

Katz RW. 1999. Extreme value theory for precipitation: Sensitivity analysis for climate change. *Advances in Water Resources* **23**(2): 133–139.

Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. *Advances in Water Resources* **25**(8–12): 1287–1304.

Katz RW, Parlange MB, Tebaldi C. 2003. Stochastic modeling of the effects of large-scale circulation on daily weather in the southeastern u. s. *Climatic Change* **60**(1–2): 189–216.

Leadbetter MR, Lindgren G, Rootzen H. 1983. *Extremes and Related Properties of Random Sequences and Processes, Springer Series in Statistics*. Springer: New York.

Maraun D, Osborn TJ, Gillett NP. 2008. United Kingdom daily precipitation intensity: Improved early data, error estimates and an update from 2000 to 2006. *International Journal of Climatology* **28**(6): 833–842, DOI 10. 1002/joc. 1672.

Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ, Zhao ZC. 2007. Climate change 2007: The physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter Global Climate Projections*. Cambridge University Press: Cambridge, New York.

Naveau P, Nogaj M, Ammann C, Yiou P, Cooley D, Jomelli V. 2005. Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience* **337**(10–11): 1013–1022.

Nearinga MA, Jettenb V, Baffautc C, Cerdand O, Couturierd A, Hernandeza M, Bissonnaise YL, Nicholsa MH, Nunesf JP, Renschlerg CS, Souchèreh V, van Oos K. 2005. Modeling response of soil erosion and runoff to changes in precipitation and cover. *Catena* **61**(2–3): 131–154.

Osborn TJ, Hulme M. 2002. Evidence for trends in heavy rainfall events over the UK. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* **360**: 1313–1325.

Osborn TJ, Hulme M, Jones PD, Basnett TA. 2000. Observed trends in the daily intensity of United Kingdom precipitation. *International Journal of Climatology* **20**: 347–364.

Perry M, Hollis D. 2005. The development of a new set of long term climate averages for the UK. *International Journal of Climatology* **25**(8): 1023–1039.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, ISBN 3-900051-07-0.

Rosenzweig C, Iglesias A, Yang XB, Epstein PR, Chivian E. 2001. Climate change and extreme weather events: Implications for food production, plant diseases, and pests. *Global Change and Human Health* **2**(2): 90–104.

Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Tank AK, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, Soden B, Zhai P. 2007. Climate change 2007: the physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter Observations: Surface and Atmospheric Climate Change*. Cambridge University Press: Cambridge, New York.

van den Brink HW, Konnen GP, Opsteegh JD. 2004. Statistics of extreme synoptic-scale wind speeds in ensemble simulations of current and future climate. *Journal of Climate* **17**(23): 4564–4574.

Vannitsem S. 2007. Statistical properties of the temperature maxima in an intermediate order quasi-geostrophic model. *Tellus A* **59**(1): 80–95.

Yu J, Shainberg I, Mamedov AI, Levy GJ. 2006. Effects of wetting by spray on concentrated flow erosion and intake rate. *Soil Science* **171**(12): 929–936.