Validation of the present day annual cycle in heavy precipitation over the British Islands simulated by 14 RCMs

A. Schindler,[1] D. Maraun,[2] and J. Luterbacher[1]

[1]   The representation of the annual cycle of heavy daily precipitation events across the United Kingdom within 14 regional climate models (RCMs) and the European observation data set (E-OBS) over the 1961–2000 period is investigated. We model extreme precipitation as an inhomogeneous Poisson process with a non-stationary threshold and use a sinusoidal model for the location and scale parameter of the corresponding generalized extreme value distribution and a constant shape parameter. First we fit the statistical model to the UK Met Office 5 km gridded precipitation data set (UKMO). Second the statistical model is fitted to 14 reanalysis driven 25 km resolution RCMs from the ENSEMBLES project and to E-OBS. The resulting characteristics from the RCMs and from E-OBS are compared with those from UKMO. We study the peak time of the annual cycle of the monthly return levels, the relative amplitude of their annual cycle and the relative bias of their absolute values. We show that the performance of the RCMs depends strongly on the region. The RCMs show deficits in modeling the characteristics of the annual cycle, especially in modeling its relative amplitude and mainly in Eastern England. However the peak time of the annual cycle is adequately simulated by most RCMs. E-OBS exhibits considerable biases in the absolute values of all monthly return levels, but the relative amplitude and the phase of the annual cycle of heavy precipitation are well represented. Our results imply that studies which rely on the explicit annual cycle of simulated heavy precipitation should be carefully considered.

1.   Introduction

[2]   Extreme precipitation events may cause severe impacts on infrastructure, ecosystems and agriculture, [e.g., *Intergovernmental Panel on Climate Change* (*IPCC*), 2012, and references therein]. Many natural cycles, such as the growing season of crop, do not follow the simple partition into four seasons. Crop yields are most likely to suffer if excess of precipitation occurs during critical development stages such as the early stages of plant reproduction [*Rosenzweig et al.*, 2001; *Parry et al.*, 2005].

[3]   Future scenario simulations with global climate models indicate an increase in intensities of extreme daily precipitation for the northern hemisphere [*Kharin et al.*, 2007; *Semenov and Bengtsson*, 2002; *Meehl et al.*, 2007; *IPCC*, 2012, and references therein]. For Europe, an increase of the intensity of extreme daily precipitation is expected for the north, and a decrease for the south [*Beniston et al.*, 2007; *Frei et al.*, 2006; *Christensen et al.*, 2007]. For Central Europe an increase in extreme precipitation is projected even for regions where mean precipitation is projected to decrease [*Christensen and Christensen*, 2003; *Fowler et al.*, 2007]. For the United Kingdom, especially winter and spring extreme precipitation is projected to increase [*Fowler et al.*, 2005; *Fowler and Ekström*, 2009; *Fowler et al.*, 2010].

[4]   A more detailed analysis of the timing and amplitude of the annual cycle of extreme precipitation is of major importance for different sectors, such as agriculture, health, forestry, tourism, and water management [e.g., *IPCC*, 2012, and references therein]. For rain gauge data from the United Kingdom a corresponding analysis of the annual cycle of daily extreme precipitation has been carried out by *Maraun et al.* [2009a] and *Rust et al.* [2009]. They identified the northwest of the UK and East Anglia as two regions with a strong annual cycle of the monthly maxima. The Midlands are described as a region with a low annual cycle of the monthly maxima. They also showed that the peak times of these annual cycles differ by regions: the annual cycle peaks in late summer in the eastern UK and in late fall or winter in the western UK.

[5]   Prior to investigation of future changes in the seasonality of extreme precipitation based on numerical model integrations, the representation of the corresponding aspects

[1]Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus Liebig University of Giessen, Giessen, Germany.

[2]GEOMAR, Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.

Corresponding author: A. Schindler, Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University of Giessen, Senckenbergstrasse 1, DE-35390 Giessen, Germany. (anne.schindler@geogr.uni-giessen.de)

Table 1. Regional Climate Models Used in This Study[a]

RCM	Institute	Grid (Number of Points Over UK)	Reference
RACMO2	KNMI	rotated pole (393)	*van Meijgaard et al.* [2008]
HadRM3Q0	UKMO	rotated pole (393)	*Collins et al.* [2006]
HadRM3Q3	UKMO	rotated pole (393)	*Collins et al.* [2006]
HadRM3Q16	UKMO	rotated pole (393)	*Collins et al.* [2006]
HIRHAM5	DMI	rotated pole (393)	*Haugen and Haakenstad* [2006]
HIRHAM	METNO	rotated pole (393)	*Christensen et al.* [1996]
RPN_GEMLAM	EC	rotated pole (393)	
CLM	ETHZ	rotated pole (393)	*Böhm et al.* [2006]
REMO	MPI-M	rotated pole (393)	*Jacob* [2001]
RCA	SMHI	rotated pole (393)	*Samuelsson et al.* [2011], *Kjellström et al.* [2005]
RCA3	C4I	rotated pole (393)	*Kjellström et al.* [2005]
PROMES	UCLM	regular lat-lon (344)	*Sánchez et al.* [2004]
ALADIN	CHMI	regular lat-lon(368)	*Pal et al.* [2007]
REGCM3	ICTP	regular lat-lon(353)	*Giorgi and Mearns* [1999]

[a]First column: the acronyms of the RCMs used in this study; second column: the acronym of the institute; third column: the grid they are operated on and in brackets the number of grid points over the UK; fourth column: relevant reference.

in these models for the present-day climate against observational data has to be rigorously validated.

[6] On an annual or seasonal basis, *Fowler et al.* [2007], *Buonomo et al.* [2007], *Rivington et al.* [2008], and *Fowler and Ekström* [2009] have validated the representation of extreme precipitation in the UK by a selection of Regional Climate Models (RCMs).

[7] In this study we evaluate how well the annual cycle of extreme precipitation in the UK is represented by 14 RCMs for the control run period 1961–2000. To this end we estimate extreme value distributions from the output of the RCMs (one for each grid point and for each RCM). We then compare characteristics of these distributions to the characteristics of the corresponding distributions estimated from observational data in order to validate the performance of the RCMs with respect to different aspects of moderate extreme precipitation (i.e., extreme events with a return period of less than 100 years).

[8] This paper is organized as follows: section 2 introduces the observation data set from the UK Met Office (UKMO) [*Perry et al.*, 2009], the European observation data set (E-OBS) [*Haylock et al.*, 2008], as well as the simulated data sets from 14 RCMs from the ENSEMBLE project [*van der Linden and Mitchell*, 2009]. Section 3 is devoted to the statistical method we use to extract distributions for extreme events from the various data sets and to the validation methods. In section 4 we present results for a sample grid point located in East Anglia and discuss the spatial patterns of the annual cycle of heavy precipitation throughout the UK.

2. Observed and Modeled Data

[9] We compare the distributions derived from simulated data with the distributions derived from observed data for the time period 1961–2000. The observation data set containing the largest number of observation time series serves as a reference for precipitation.

[10] As reference data set we use the gauge-based data provided by the UK Met Office (UKMO) with a resolution of 5 km [*Perry et al.*, 2009]. This data is rotated by bilinear interpolation and averaged to a horizontal grid spacing of ≈25 kilometers as described by *Maraun et al.* [2009a], to match the different grids used by the RCMs (see Table 1).

[11] We validate the simulated precipitation of the control run of 14 RCMs against the reference data. The RCMs were driven by ERA-40 reanalysis data [*Uppala et al.*, 2005] in order to isolate RCM errors from general circulation model errors (perfect boundary conditions [*Christensen et al.*, 1997; *Frei*, 2003]). In Table 1 we list the RCMs validated in this study with information on the underlying grid and further reference. Additional summarized information and further references are given by *Christensen et al.* [2010].

[12] We also compare the European observation data set (E-OBS) [*Haylock et al.*, 2008] for the UK region to UKMO. Since it comprises only 137 rain gauges over the UK (compared to 3,000–5,000 rain gauges in the UKMO data set), it is less useful as a reference data set [*Hofstra et al.*, 2009; *Maraun et al.*, 2012]. We investigate whether the E-OBS data set contains sufficient information to characterize variations of daily extreme precipitation throughout the year. If this is the case, confidence would be provided in the ability to analyze the annual cycle of extreme precipitation throughout Europe at least in regions with similar station density, topography and climate.

3. Methods

[13] Extreme value statistics has been widely used in precipitation studies, e.g., by *Katz et al.* [2002], *Maraun et al.* [2009a], *Toreti et al.* [2010], *Kallache et al.* [2011], *Buonomo et al.* [2007], and *Tomassini and Jacob* [2009]. In a first step we fit a statistical model based on extreme value theory to observed and simulated data (section 3.1). In a second step selected characteristics of all RCMs and of E-OBS are validated against the reference characteristics (section 3.2): all data sets are compared to the high resolution gridded UKMO [*Perry et al.*, 2009].

[14] The extreme value analysis of each gridded data set is carried out independently for each grid point. This implies fitting the marginal distributions to nearly 400 time series of daily precipitation with 40 years of continuous data for each of the data sets.

3.1. Extreme Value Analysis

[15] Three different approaches exist in univariate extreme value statistics. In the block maxima approach, the most common one, maxima of a block of data are modeled, e.g.,

the one-year-maxima or one-month-maxima of daily observations, [*Fowler and Kilsby*, 2003; *Maraun et al.*, 2009a]. The second approach models the excesses above a high threshold, often called peaks over threshold approach [e.g., *Toreti et al.*, 2010]. The third approach is a combination of both: extreme events are modeled by using threshold excesses but the distribution is expressed in terms of block maxima [e.g., *Tomassini and Jacob*, 2009; *Kallache et al.*, 2011]. This third approach will be applied in the present study as it makes use of more data than the block maxima approach and at the same time is independent of the threshold. Thus the parameters of different grid points are comparable and a non-stationary threshold can be used in order to make better use of the data.

[16] To describe extreme events we need to specify the occurrence rate as well as the intensity of the events. Here we follow the approach by *Coles* [2001]. The occurrence of rare events is assumed to be Poisson distributed and its intensity is described by the generalized Pareto distribution (GPD). The exceedances of the threshold u are modeled as a Poisson point process where the parametrization is independent of u and the parameters are equivalent to those of the corresponding generalized extreme value distribution (GEV):

[17] To select rare events a high threshold $u \gg 1$ is fixed, such that the assumptions hold. Then the intensity measure of the Poisson process on $[0, 1] \times [z, \infty]$, for $z > u$, is given by

$$\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}, \tag{1}$$

where μ, σ, ξ are called *location*, *scale* and *shape* parameters.

[18] The threshold choice therefore only affects the inference, especially the bias-variance trade off.

[19] A characterization of precipitation extremes by a single number is the quantile z_p which is exceeded with probability p, called the *return level* to return period $\frac{1}{p}$:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1 - p)\}^{-\xi}\right] & , \text{ if } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1 - p)\} & , \text{ if } \xi = 0 \end{cases} \tag{2}$$

Loosely speaking, in a stationary climate z_p is expected to be exceeded once every $1/p$ blocks, e.g., once every $1/p$ years if the block length equals one year. These quantiles are easily interpretable and will in the following be used for model validation (section 4.2).

[20] Daily precipitation as well as extreme precipitation in the UK show an annual cycle [*Barrow and Hulme*, 1997; *Maraun et al.*, 2009b] which needs to be incorporated into the model [see *Katz et al.*, 2002; *Smith*, 1989]. In order to include the non-stationarity of the data in the stationary formulation of the Poisson point process (equation (1)), it is necessary to adapt the location and also the scale parameter while the shape parameter remains unchanged [*Rust et al.*, 2009]. The non-stationary parameters then read

$$\mu(t) = \mu_0 + \mu_1 \sin\left(\frac{2\pi t}{365.25}\right) + \mu_2 \cos\left(\frac{2\pi t}{365.25}\right), \tag{3}$$

$$\sigma(t) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi t}{365.25}\right) + \sigma_2 \cos\left(\frac{2\pi t}{365.25}\right), \tag{4}$$

for $t = 1, \ldots, 365.25 \cdot 4$ (to account for leap years). Substituting equations (3) and (4) into the equation for the probability (equation (1)) we arrive at an inverse problem for the return levels (equation (5)), which we can solve numerically.

$$1 - \frac{1}{p} = Pr\big(\max(X_1, .., X_n) \leq z_p\big) \approx \prod_{i \in \{1, \ldots, n_p\}} p_i \tag{5}$$

where

$$p_i = \begin{cases} 1 - n_p^{-1}\big(1 + \xi(z_p - \mu_i)/\sigma_i\big)^{-1/\xi}, & \text{if } 1 + \xi(z_p - \mu_i)/\sigma_i \geq 0 \\ 1 & , \text{ otherwise} \end{cases}$$

with the non-stationarity of the parameters denoted as $\mu_i := \mu(t_i)$ and $\sigma_i := \sigma(t_i)$ for $i \in \{1, \ldots, n_p\}$ and n_p being the number of observations per block. To calculate the annual return level z_p the block length is set to 365, i.e., $n_p = 365$. To calculate monthly return levels, $z_p(m)$, (non-stationary setting) the product over the index set is restricted to a specific month, m. This can be achieved by choosing $i \in \{1, \ldots, n_p\}$ in equation (5) such that $t_i \in m$ (January: $i \in \{1, \ldots, 31\}$; February: $i \in \{32, \ldots, 60\}$; etc.).

[21] These monthly return levels are actually annual return levels conditional on the month of their occurrence. For instance the January-return level $z_{1/25}(Jan)$ is on average exceeded in January once every 25 years.

[22] Due to its flexibility to easily incorporate covariates in the likelihood formulation the maximum likelihood approach is used for the parameter estimation. For the realizations z_i of the random variable Z, for i = 1, … N_{obs}, where N_{obs} is the number of observations, one can maximize the likelihood function with respect to the parameter $\theta := (\mu_0, \mu_1, \mu_2, \sigma_0, \sigma_1, \sigma_2, \xi)$. In general it is easier to minimize the negative log-likelihood function than to maximize the likelihood function. With $\delta_i = 1$ if $z_i > u_i$ and $\delta_i = 0$ else we hence obtain

$$\begin{aligned} l_n(\theta) &= \log\left(\exp\left(-\frac{1}{n_p} \cdot \sum_{i=1}^{N_{obs}}\left(1 + \xi\frac{u_i - \mu_i}{\sigma_i}\right)^{-1/\xi}\right)\right) \\ &\quad + \log\left(\prod_{i=1}^{N_{obs}}\left(\frac{1}{\sigma_i} \cdot \left(1 + \xi\frac{z_i - \mu_i}{\sigma_i}\right)^{\frac{-1}{\xi}-1}\right)^{\delta_i}\right) \\ &= -\frac{1}{n_p} \cdot \sum_{i=1}^{N_{obs}}\left(1 + \xi\frac{u_i - \mu_i}{\sigma_i}\right)^{-1/\xi} \\ &\quad + \sum_{i=1}^{N_{obs}} \delta_i \cdot \log\left(\frac{1}{\sigma_i} \cdot \left(1 + \xi\frac{z_i - \mu_i}{\sigma_i}\right)^{(-1/\xi - 1)}\right) \end{aligned}$$

with $u_i = u(t_i)$ being the threshold at time $t_i \in \{1, \ldots, 365.25 \cdot 4\}$ for t_i being the time of the observation z_i in the 4-year cycle. For numerical stability the number of observations per block is set to one, $n_p := 1$. Subsequently the parameters are transformed to correspond to the GEV parameters of the distribution of monthly maxima [see *Katz et al.*, 2005].

[23] The asymptotic normality of the estimates is used for estimating the standard errors of the maximum likelihood estimates, $\hat{\theta}$. This theory for uncertainties also allows to adapt the estimation errors to account for serial dependence in the time series [*Smith*, 1990; *Fawcett and Walshaw*, 2007]. This is necessary since the requirement of independence in

the time series is not necessarily met by precipitation data [e.g., *Coles et al.*, 1999; *Katz et al.*, 2002]. We simulate from the multivariate normal distribution of the parameter estimates and calculate the return levels, z_p. The 0.025- and 0.975-quantile envelope of these $n = 1000$ simulated return levels is set as the point-wise 95% confidence interval for z_p.

[24] Threshold choice is a variance-bias trade-off. In order to use as many data as possible in the analysis a low threshold is chosen, but high enough so that above this threshold the asymptotic marginal and dependence properties still appear to be stable [*Eastoe*, 2007]. A means to achieve this goal are diagnostic plots such as the mean residual life plot and plots of the estimated GPD parameters across a range of thresholds. Since this visual approach is not feasible for roughly $19 \cdot 400$ time series, we use a pragmatic, iterative approach.

[25] As a starting value we set the threshold u such that the highest 5% of the data exceed u. To account for sub-annual variation these 5% of the data are selected relative to the mean precipitation amount at each day of the year (days of zero precipitation are not excluded from the calculation). This threshold is smoothed with a 60-day moving-average. Since the main features, such as the annual return levels, are similar for 10-day, 30-day and 60-day smoothing (not shown) we choose the least complex threshold for the subsequent analysis. In the next step we estimate θ and assess the goodness-of-fit. If the fit is good, see below, the threshold is kept; if not, it is increased by 1% and the parameters are estimated once more and the goodness-of-fit is assessed again, [e.g., *Kallache et al.*, 2011; *Toreti et al.*, 2010]. The iteration halts when the fit is good or increasing the threshold does not improve it.

[26] In this study we assess the goodness-of-fit via quantile-quantile plots. Quantile-quantile plots compare the empirical quantiles with the theoretical quantiles. When covariates are present, the random variables Z_t are transformed to standard Gumbel scale [*Coles*, 2001]. We simulate confidence intervals for the theoretical quantiles and count the data points outside the 95% confidence interval. The goodness-of-fit is considered bad if 5% of the empirical 4th quartile of the data lies outside the simulated 95% confidence interval.

[27] A higher threshold than the starting one was suggested by the goodness-of-fit in less than 5% of the grid points. Since we assumed a 95% confidence interval, we expect that 5% of all points are outside the confidence interval. Thus the original threshold was kept.

3.2. Validation Methods

[28] For validation purposes we compare the fitted marginal distributions against the reference marginal distribution grid point by grid point, [e.g., *Brown et al.*, 2008], as well as summaries of the spatial patterns of the marginal distributions, [e.g., *Maraun et al.*, 2012]. As characteristics of the marginal distributions, $c = char_k$, we choose the GEV parameters (offset, phase ϕ and amplitude A for the location and scale sinusoids and the shape parameter), the annual return levels ($z_{1/2}, z_{1/5}, z_{1/10}, z_{1/20}, z_{1/50}, z_{1/100}$) and the monthly return levels, i.e., the annual return levels conditional on the month of their occurrence, $z_p(m)$ for each month, m, and $p \in \{1/2, 1/5, 1/10, 1/25, 1/50\}$ as well as their annual cycle. The annual cycle of $z_p(m)$ is characterized by the *offset* (the mean of $z_p(m)$ over all m), the *phase* (the month for which $z_p(m)$ is

largest) and the *relative amplitude* (half the range of $z_p(m)$ divided by the offset).

[29] In this study we focus on the analysis of the monthly return levels and their annual cycle. Summarized information on the GEV parameters and annual return levels is presented in Appendix A.

[30] We take an arbitrary grid point (in East Anglia, $52°59'11.6''$N, $0°56'98.046''$E) to exemplify the characteristics of the marginal distributions, especially the annual cycle of the monthly return levels: in Figure 1 we show a radial plot in polar coordinates. The radial coordinate corresponds to the value of precipitation in mm and the angular coordinate corresponds to the time of the year. If considered as a clock face the first hour can be identified with January, the second with February and so on. In this figure we plot a selection of the characteristics of the reference marginal distribution at one example grid point: the annual return levels $z_{1/2}, z_{1/5}, z_{1/10}, z_{1/20}, z_{1/50}$ are represented by the gray circles centered on the origin; the monthly return levels $z_{1/25}(m)$ for $m \in \{1, \ldots, 12\}$ are represented by the black polygon; the phase and the amplitude of $z_{1/25}(m)$, half the range of $z_p(m)$, are represented by the black point. Additionally 40 years of daily observations (dark gray rays) are plotted, where high precipitation events (events exceeding $z_{1/2}$) are marked with a small point. The monthly return levels $z_{1/25}(m)$ of the other data sets (colored polygons) together with their phase and amplitude (colored points) are depicted as well.

[31] In the following the statistics used to summarize and compare the information at all grid points are presented. These do not account for spatial dependencies, since the order of the grid points is not taken into account. The prefix "pattern" emphasizes the spatial domain of the statistics as opposed to time series statistics. For all characteristics under consideration we calculate the pattern mean, $\bar{c} := pmean(c)$, the mean value over all grid points and the *pattern standard deviation* over all grid points, denoted by $psd(c)$, summarizing the variation from grid point to grid point.

[32] To compare the respective characteristics derived from one RCM against the reference characteristics not only point-wise, we also calculate the regionalized absolute or relative bias respectively and the anomaly pattern accordance [*Wilks*, 2006]. The relative bias of $c_{i,\mathcal{M}}$, a characteristic of a model $\mathcal{M}$ at grid point i, is defined as the quotient $bias_{i,rel} = c_{i,\mathcal{M}}/c_{i,ref}$, while the absolute bias of $c_{i,\mathcal{M}}$ is defined as the difference $bias_{i,abs} = c_{i,\mathcal{M}} - c_{i,ref}$. The regionalized relative/absolute bias is defined as $median_{i \in region}(bias_{i,rel/abs})$. The UK is divided into nine regions according to *Gregory et al.* [2007] and *Wigley et al.* [1984] depending on mean precipitation: North Scotland (NS), East Scotland (ES), South Scotland (SS), Northwest England (and North Wales) (NWE), Northeast England (NEE), Northern Ireland (NI), Southwest England (and South Wales) (SWE), Central and East England (CEE), and Southeast England (SEE). *Fowler and Kilsby* [2003] have demonstrated that these regions can also be used to classify extreme precipitation.

[33] To investigate the anomaly pattern accordance of a model $\mathcal{M}$ with the reference, we calculate the *pattern correlation*,

$$pcor(c_{\mathcal{M}}, c_{ref}) = \sum_{i=1}^{N} \left(\frac{(c_{i,\mathcal{M}} - \bar{c}_{\mathcal{M}})(c_{i,ref} - \bar{c}_{ref})}{psd(c_{\mathcal{M}}) \cdot psd(c_{ref}) \cdot N} \right) \quad (6)$$
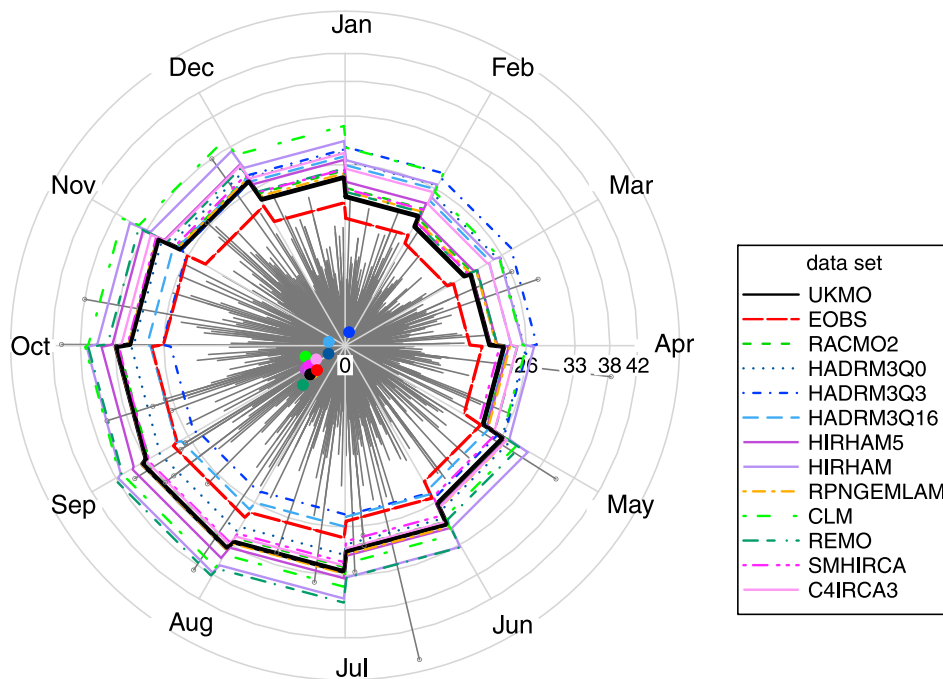
Figure 1. Forty years of gridded daily observations of precipitation (gray rays) and monthly return level estimates to probability 1/25 (colored polygons) at a grid point in East Anglia (52N 59′11.6″, 0E 56′ 98.046″). Time of the year as angular coordinate, daily precipitation amount in mm as radial coordinate and estimated annual 2-,5-,10- and 20-year return levels in mm as radial labels. The information is two-folded: (1) the fit of the statistical model (gray rays compared to black polygon) and (2) the validation of the representation of the monthly 1/25-return levels by the RCMs (colored polygons compared to black polygon). The colored points near the pole correspond to the centroids of the polygons, summarizing the information of the amplitude and phase of the annual cycle of the monthly return levels. Small points at the end of the gray rays indicate an observed 2-year event.

and the *pattern centered root mean square distance*,

$$pcrmsd(c_{\mathcal{M}}, c_{ref}) = \sqrt{\frac{\sum_{i=1}^{N}\left((c_{i,\mathcal{M}} - \bar{c}_{\mathcal{M}}) - (c_{i,ref} - \bar{c}_{ref})\right)^2}{N}}.$$

(7)

The anomaly pattern accordance is measured by the ratio of the pattern standard deviations, the pattern correlation and the pattern centered root mean square distance. It can be summarized by a single point in a two-dimensional diagram, the Taylor diagram [*Taylor*, 2001]. The distance of the point from the origin of the diagram represents the quotient $psd(c_{\mathcal{M}})/psd(c_{ref})$, the angular coordinate of the point indicates the arcus cosine of $pcor(c_{\mathcal{M}}, c_{ref})$ and $pcrmsd(c_{\mathcal{M}}, c_{ref})$ is proportional to the distance from (1, 1). Taylor diagrams in general do not display the uncertainties associated with the estimated values.

4. Results

[34] We first compare the fit of the statistical model with the observed data from the UK Met Office [*Perry et al.*, 2009] at one arbitrarily chosen example grid point located in East Anglia (Section 4.1). We then validate the estimated return levels from E-OBS and from the simulated data against the estimated return levels from UKMO, the reference data: at the specific grid point and also with respect to their spatial patterns and to their absolute and relative bias respectively (Section 4.2).

[35] The example grid point is located in the eastern rain shadow areas with a less marked annual cycle in mean precipitation. There the summer precipitation can make the highest contribution to the annual total and is likely to be of convective origin [*Barrow and Hulme*, 1997]. This grid point exemplifies typical problems of the RCMs in simulating the annual cycle of heavy precipitation.

4.1. Statistically Modeled Versus Observed Intensities

[36] To illustrate the ability of the statistical model to represent the annual cycle of extreme precipitation, we arbitrarily choose an example grid point in East Anglia. At this location we demonstrate the suitability of the statistical model visually. We then perform a likelihood ratio test for all grid points.

[37] For the validation of the statistical model at the selected grid point, we focus on the dark gray rays and the black polygon in Figure 1. The radial labels mark the annual 2-, 5-, 10- and 20-year return level estimates of the statistical model. The outermost circle represents the estimated 50-year return level (≈48 mm). These return levels are exceeded by observations in 40 years 26, 10, 5, 1 and 0 times respectively, which is consistent with the expected exceedance numbers (20, 8, 4, 2 and 0).

[38] The black polygon corresponds to the estimated monthly return levels, $z_{1/25}(m)$, which is in one month on average exceeded once every 25 years. The mean of these monthly values, 26.8 mm, corresponds to the annual return

level with probability 0.48, that is an annual return period of 2.2 years.

[39] Focusing on the observed 2-year events, i.e., the gray rays exceeding $z_{1/2} = 26$ mm (marked with a small point), and their occurrence in the course of the year, the annual cycle in high precipitation events is visible: of the 26 exceedances during the time period 1961–2000 there has been zero observed 2-year events during the winter months (DJF) while during summer at least two events occurred each month. This sub-annual variation has to be represented in the statistical model and the monthly return levels, $z_{1/25}(m)$, should display a similar behavior as the observed 2-year events.

[40] As inferred from Figure 1 this annual cycle is captured by the statistical model: the monthly return level estimates $z_{1/25}(m)$ are smaller for the winter months and higher for the summer months, accounting for the sub-annual variation of the intensity of the precipitation events. Therefore the return period, i.e., the probability of exceeding $z_{1/25}(m)$, does not depend on the time of the year. This visual examination shows that the statistical model fits and captures the annual cycle in observed heavy precipitation at this location.

[41] At this grid point the associated uncertainties, the simulated, point-wise 95% confidence intervals, with the monthly return levels vary between 16% and 23% of $|z_p(m)|$ for all p and for all m. Throughout all grid points the uncertainties vary between 10% and 25% of the monthly return levels. The uncertainties, since given as a percentage of $z_p(m)$, neither depend on m nor on the location of the grid point. As expected the uncertainties tend to be larger for higher return periods.

[42] For all grid points we perform a likelihood ratio test, to check if the inclusion of the annual cycle in the parameters is necessary for describing extremes on a sub-annual scale. The likelihood ratio test (for block length one) suggests that the incorporation of the annual cycle in the location parameter improves the model significantly ($\alpha = 5\%$) at all grid points but one; the further incorporation of the annual cycle in the scale parameter only significantly improves the model at 85% of the grid points. Nonetheless we include the non-stationarity in both parameters in order to have one statistical model for the whole of UK. We believe that the model can be further improved by using second-order harmonics or splines as covariates for the location or scale parameter.

4.2. Simulated Versus Observed Intensities

[43] In the following paragraphs we investigate the dependence of the intra-annual variation (relative amplitude), the time of the maximum (phase) and the absolute values of the annual cycle of the monthly return levels on the location. The resulting spatial patterns of E-OBS and of the RCMs are compared and validated against the spatial pattern of the UKMO data set. We start the investigation with an example grid point to illustrate the characteristics under investigation and the possible problems of the RCMs.

[44] *Grid point.* For the validation of the monthly return levels estimated from simulated data, at this single grid point, we focus on the spider diagram in Figure 1. It consists of the colored polygons ($z_{1/25}(m)$) and their centroids, the colored points near the pole. The centroid summarizes the information of the phase and the amplitude of the annual cycle of $z_{1/25}(m)$: the angular coordinate indicates the phase, i.e., the time of its maximum value and the radial coordinate indicates

the amplitude, i.e., (max − min)/2. The offset, that is the annual mean, is not explicitly shown but can be retrieved from the monthly values.

[45] The following discussion of the estimates for the example grid point all refer to Figure 1.

[46] The annual return levels of the UKMO data set range from ≈26 mm (2-year)(first radius) to ≈53 mm (100-year) (not shown). The values of $z_{1/25}(m)$ range from 19.8 mm in February to 33.7 mm in August (black polygon), with an amplitude of ≈7 mm (radial coordinate of black centroid). This amplitude of the annual cycle amounts to 25.88% of the annual mean (7/26.8 ≈ 0.2588), the relative amplitude. The phase is given by the month of the maximum value (August, angular coordinate of black centroid).

[47] The return levels from the E-OBS data are under-represented: the annual return levels range from ≈22 mm (2-year)(not shown) to ≈47 mm (100-year)(not shown) as well as $z_{1/25}(m)$, ranging from ≈17.0 mm (February) to ≈28.5 mm (August, red polygon). The amplitude is under-represented as well with ≈5.8 mm (radial coordinate of red centroid). For E-OBS the under-representation of the offset and the under-representation of the absolute amplitude results in a good representation of the relative amplitude, 25.19%: the intra-annual variation is represented well by the E-OBS data set. The phase is also well represented: the maximum occurs in August (angular coordinate of red centroid).

[48] The annual return levels derived from the RCMs are either slightly under-represented, i.e., range from 24 mm to 50 mm (HADRM3Q16), or overrepresented, i.e., range from 29 mm to 91 mm (HIRHAM). Considering the annual return levels, the range is best captured by RACMO2 (from ≈26 mm (2-year) to ≈56 mm (100-year))(not shown).

[49] Focusing on the absolute values of the monthly return levels to $p = 1/25$, we see different behaviors throughout the year and throughout the RCMs: for the winter months all RCMs over-represent the return levels, but for the summer months three RCMs over-represent the return levels and six RCMs under-represent it, respectively.

[50] The amplitude of $z_{1/25}(m)$ is under-represented by most RCMs. Almost all RCMs agree on late summer as the time of the maximum return level, i.e., the phase of the annual cycle is, apart from HADRM3Q3 and Q16, well simulated at this grid point in East Anglia.

[51] At this specific grid point in the UK, the intensity of a heavy precipitation event depends on the time of the year, both for observations and simulations.

[52] *Relative amplitude.* The relative amplitude, as defined in section 3.2, measures the strength of the annual cycle of the monthly return levels relative to the annual mean. In Figure 2 the maps of the relative amplitude of $z_{1/25}(m)$ derived from each data set are presented. The map of the relative amplitude of the reference data set UKMO shows a strong annual cycle of the monthly return levels in Northwest Scotland and East Anglia with an amplitude of the annual cycle that exceeds 25% of the annual mean. In between those two regions there is a transition zone where the amplitude of the annual cycle is less than 10% of the annual mean. For the majority of the grid points the relative amplitudes range between 12% and 25% of the annual mean throughout the UK.

[53] For the UKMO data set the spatial patterns of the relative amplitude to different return periods are similar: the
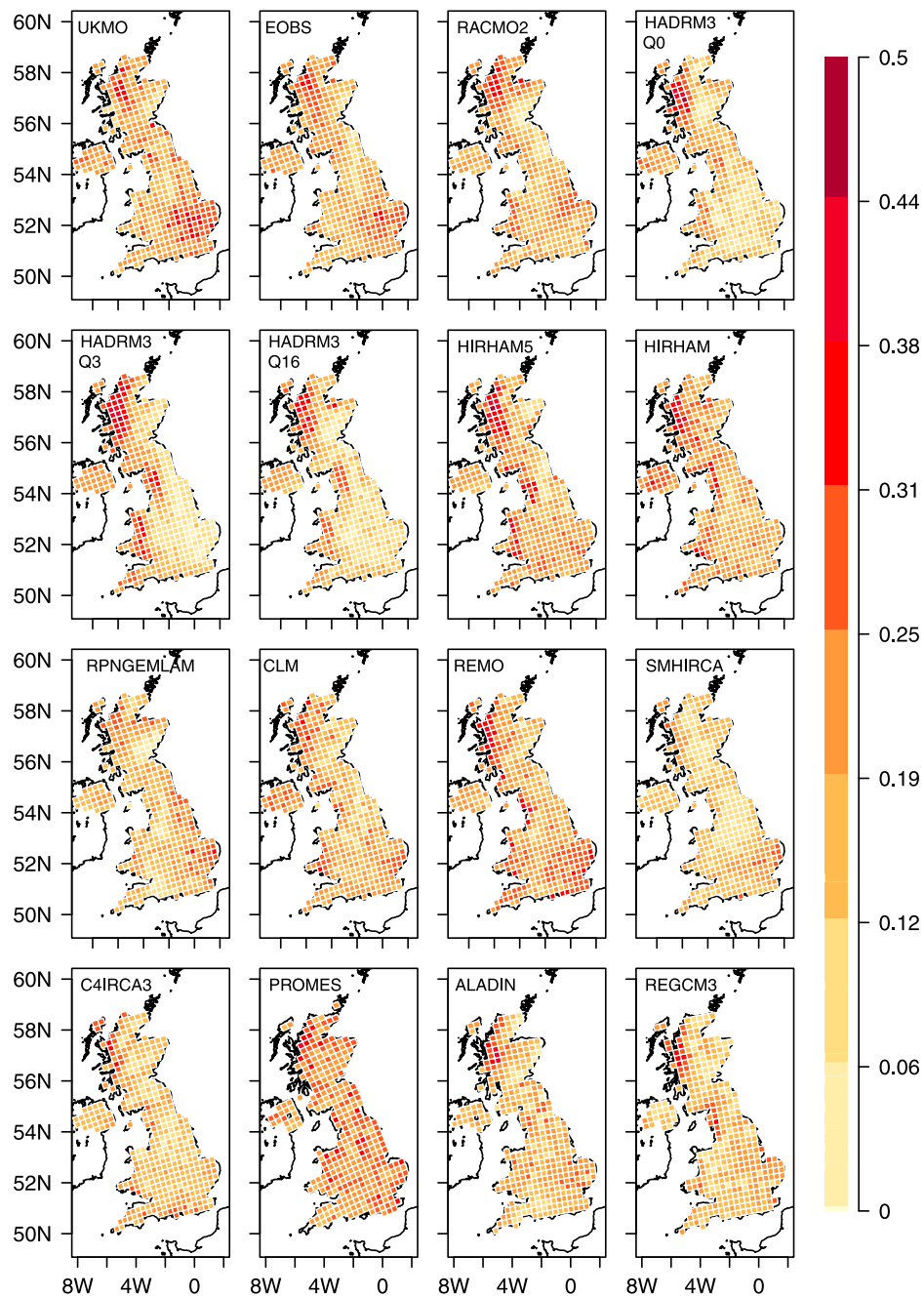
Figure 2. Maps of the relative amplitude (unitless) of the annual cycle of the monthly 1/25-return levels, estimated from gridded (first plot) UKMO data, (second plot) E-OBS and (third through sixteenth plots) 14 regional climate models.

size of the regions with a strong annual cycle are smaller for smaller return periods and similar for larger return periods (not shown).

[54] The map to the E-OBS data exhibits most regional features of the relative amplitude, with correct relative strength of the annual cycle, except for the Welsh mountains (Figure 2). There the grid points at higher elevations show a weaker annual cycle than the surrounding grid points, which is opposite to the behavior of the reference data. These could be effects of the interpolation since the gauge stations

contributing to the data set are mainly located along the coast or along the border to England and not in the Welsh mountains.

[55] Figure 2 illustrates that the representation of the relative strength of the annual cycle of heavy precipitation by the RCMs varies considerably. Almost all RCMs show good ability to identify Northwest Scotland as a region with above average relative amplitude. For the second region with a strong annual cycle, East Anglia, this is true only for a few RCMs. There are RCMs which simulate no intra-annual
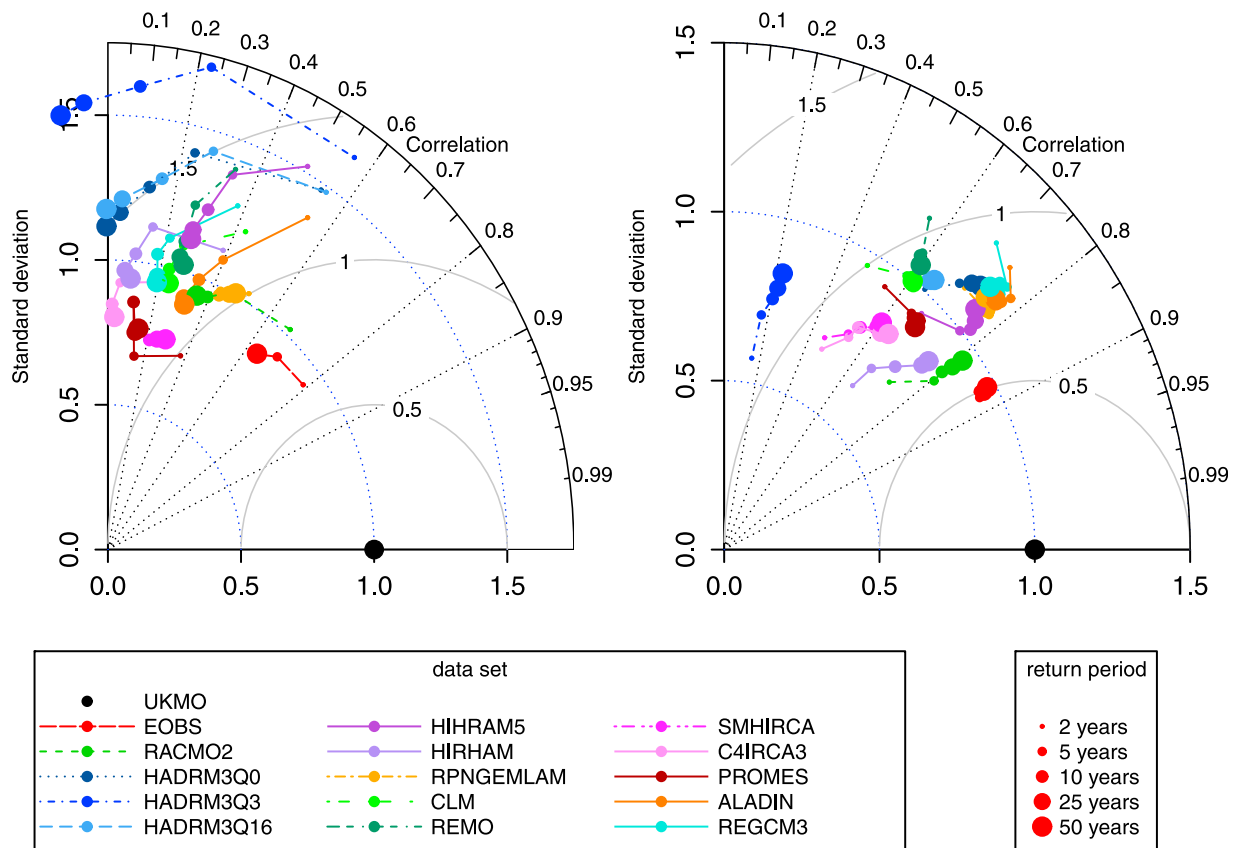
Figure 3. Taylor diagrams for the maps (left) of the relative amplitude of $z_{1/p}(m)$ and (right) of the month of the maximum return level derived from UKMO, E-OBS and 14 regional climate models. Each point represents the pattern accordance of the anomaly pattern of one data set relative to the anomaly pattern of the reference (UKMO). The angular coordinate represents the pattern correlation. The radial coordinate represents the ratio of the standard deviations. The distance to the black reference point (1, 1) is proportional to the pattern centered root mean square distance. The diameters of the points correspond to the return periods.

variation in this region. Both the local features and correct magnitude of the relative amplitude are not represented correctly by any RCM.

[56] The relative amplitude is mostly under-represented by the RCMs. For most regions and for most RCMs the bias amounts to less than 20% (not shown). However the bias depends on the region: it is smaller for western regions and larger for eastern regions (not shown). For instance for NEE, CE and SEE the spread between RCMs is large, the bias ranges between 15% and 70% (not shown).

[57] In Figure 3 (left) we compare the anomaly patterns of the relative amplitude of the RCMs to the reference across all probabilities. The pattern deviation against the reference pattern (black point) is depicted by the colored points.

[58] The anomaly pattern to the E-OBS data set (red point) correlates considerably better to the reference pattern across all probabilities than the RCMs' anomaly patterns (Figure 3, left). Since E-OBS is observation-based, this result is not surprising. But regarding other characteristics of daily extreme precipitation E-OBS does not necessarily perform better than the RCMs [*Hofstra et al.*, 2009; *Maraun et al.*, 2012].

[59] The increasing distance of the points from (1, 1) with increasing return period (increasing diameter of points)

shows that the representation of the relative amplitude worsens with decreasing probabilities across most RCMs. Especially the pattern correlation of the anomaly patterns is decreasing to small positive or even negative values for most RCMs suggesting no accordance to the reference data. In part this might be related to increasing uncertainties of the estimates with smaller probabilities not represented in the Taylor diagrams.

[60] The behavior of the three Hadley Centre RCMs is noteworthy, because they only differ in some parameterizations, resulting in different climate sensitivities. All three versions represent the spatial patterns of the relative amplitude of the annual cycle badly but their representation varies considerably. The spatial pattern to the standard (Q0) climate sensitivity version is the closest to the reference pattern. The spatial patterns of the low (Q3) and high (Q16) climate sensitivity versions are very close to the spatial patterns of the elevation (Figures 2 and 3).

[61] As we have seen in Figures 2 (left) and 3 (left), the magnitude of the annual cycle of the monthly return levels as well as its spatial pattern are not well represented by any RCM. Few RCMs are able to identify both regions with a strong annual cycle, but even then the relative strength is not well simulated.
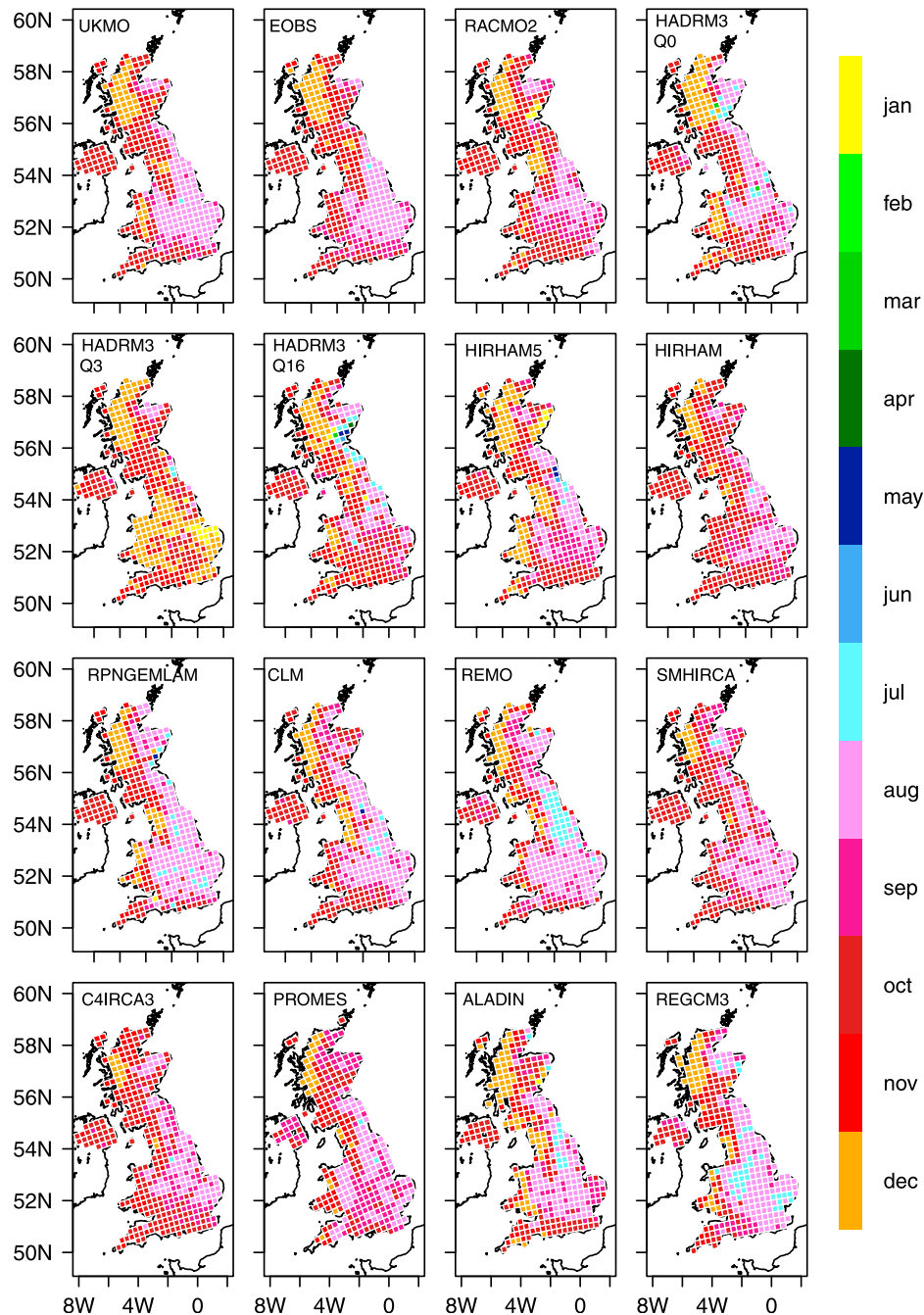
Figure 4. As in Figure 2, but for the month of the maximum monthly 1/25-return level.

[62] *Peak time of the annual cycle.* The maximum monthly return levels of the UKMO data set reveal spatiotemporal dependencies (Figure 4). The map of the occurrence time of the maximum monthly return levels reveals an east-west gradient following the orography. In the lower regions of CEE the maximum return level occurs in August propagating to the mountainous regions in the west with the peak of the annual cycle during December.

[63] The E-OBS data set shows again a good agreement with the reference data set except for Wales where the maximum occurs in September and October (Figure 4). This is possibly related to the same interpolation problems in the E-OBS data as hypothesized for the relative amplitude.

[64] Most RCMs are able to reproduce the east-west gradient despite some difficulties in representing local characteristics and in modeling the spatial extent of homogeneous regions (Figure 4). In regions, such as the south-east coast of Scotland, the simulation of the phase of the annual cycle spreads widely and seems to vary randomly spatially, e.g., HADRM3Q16, and between models. The influence of the highlands of Scotland reaches farther to the north–east, e.g. HADRM3Q3 or not far enough to the east, e.g., C4IRCA3, than in the reference data.

[65] Apart from the spatial inhomogeneity (local estimation problems and misrepresentations of the range of a uniform area) there are two additional types of misrepresentations in

England: REMO shows an additional homogeneous subregion in NWE where the maximum return level occurs one month early (July) but the surrounding regions are modeled correctly (Figure 4). In HADRM3Q3 the maximum return levels in CEE occur in January and December (not in August) (Figure 4) as already seen at the example grid point (Figure 1, navy blue centroid).

[66] These biases do not depend on the return period (not shown). For all return periods the bias in the peak time is largest for grid points located in CEE or SEE: there for most RCMs the spatially averaged bias ranges from one month to four months late (not shown). In these regions the observed peak occurs during summer and is most likely of convective origin. Convective precipitation is not resolved by the RCMs used in this study and needs to be parameterized. This leads to uncertainties and errors in the simulation of especially summer precipitation [*Maraun et al.*, 2010, and references therein].

[67] The spatial anomaly patterns of the peak time of the annual cycle is well represented in the E-OBS data set and well simulated by all RCMs but one (Figure 3, right). For all RCMs the pattern variability depends on the return period. It is well represented for high return periods and less well for small return periods. This is in contrast to the behavior of the pattern accordance with respect to the relative amplitude.

[68] For few RCMs and few grid points the deviation from the peak time of the annual cycle might arise from simulating the annual cycle too weak (Figure 2). But this is not the explanation for the difficulties of most RCMs and at most grid points (Figures 4 and 2).

[69] *Monthly return level.* The difficulties of the RCMs in capturing the annual cycle can be examined by studying the monthly return levels for each month separately. Exemplarily for all months we do this for February- and August-return levels, since their difference determines the amplitude of the annual cycle in CEE, which is the worst simulated region (in terms of the relative amplitude, Figure 2).

[70] The maps of the intensities of $z_{1/25}(m)$ based on the reference data reveal an east-west gradient throughout the year (e.g., Figures 5 and 6), with lower values in the east and higher in the west. The gradient is strong in the winter months.

[71] In the respective maps based on the E-OBS data, we see the east-west gradient in the winter months but the gradient in the summer disappears (Figures 5 and 6, second map).

[72] The performance of the RCMs in modeling heavy precipitation patterns is better for the winter months than for summer (Figures 5, 6 and 7). The east-west gradient appears in most return level maps for the winter months but is not prevalent in the summer. In most RCMs the modeled gradient is too weak, except for the PROMES RCM in which it is too strong (see Figure 5).

[73] In Figure 7 we examine the anomaly pattern accordance of the absolute intensities of the monthly return level maps: the four Taylor diagrams show the dependency of the accordance of the anomaly patterns of the February- and August-return levels on the return periods (Figures 7a and 7b) and the dependency of the anomaly pattern accordance of the monthly 1/5- and 1/25-return levels on the time of the year (Figures 7c and 7d).

[74] The accordance of the February-return level maps exhibits only a slight dependency on the return period (Figure 7a). The accordance of the August-return level maps however depends strongly on the return period for all RCMs: smaller return periods result in stronger pattern correlations and better *pcrmsd* values (Figure 7b). The accordance of the monthly 1/5-return level maps depends on the time of the year (Figure 7c). This dependence is even stronger for the accordance of the monthly 1/25-return level maps (Figure 7d). The pattern accordance is better for winter months (DJFM) than for extended summer months. This might be due to the fact that for summer months part of the precipitation is of convective origin, which is not resolved by the RCMs. It also varies considerably for different return periods. The performance for small return periods varies less throughout the year and is better correlated than the performance for high return periods. This is to be expected since return levels to small return periods are closer to mean precipitation, which is reasonably well modeled (including its annual cycle). Further, the estimation of higher return periods is more uncertain leading to large confidence intervals which are not taken into consideration when calculating the pattern correlation.

[75] The relative bias of the monthly return levels $z_{1/25}(m)$ focuses on the performance of reproducing the amount of precipitation. The median of the relative bias over all grid boxes in one region shows the under- and over-representation of the intensity summarized by regions and resolved for each month (Figure 8). The shaded area in Figure 8 originates from the simulated 95% confidence intervals of the UKMO return levels: the upper bound of the confidence interval as well as its lower bound are divided by the reference estimate for each grid point and summarized by the median over the regional grid points. Note that this is only a visual guidance on the average uncertainty of the UKMO return level estimates throughout the region and not a confidence interval for the regional median.

[76] The monthly return levels based on the E-OBS data (red line) are under-represented by 4% (NI) to 25% (NS) and the relative bias shows no intra-annual variation (Figure 8).

[77] The RCMs tend to under-represent $z_{1/25}(m)$ by different amounts depending on the region and on the month under consideration (Figure 8). In the northern regions of the British Island most RCMs largely under-represent the monthly return levels with no intra-annual variation. In CEE and SEE the relative bias exhibits an annual cycle with over-representation of the monthly return levels for the winter months by most RCMs. Only REMO shows no pronounced annual cycle in the relative bias, it even tends to have a smaller bias in summer months. On the other hand HADRM3Q3 shows a very pronounced annual cycle: the monthly return levels from April until November are under-represented and the winter months over-represented by 25%.

[78] *Discussion by regions.* The different performances of RCMs, as revealed by biases and maps, are summarized by regions.

[79] 1. Northern Ireland is the only region with no serious misrepresentations by almost all RCMs. This region is represented quite well in terms of relative amplitude and phase. The relative bias of the monthly return levels is negligible for most of the RCMs. The good performance of the RCMs could be related to the relatively homogeneous
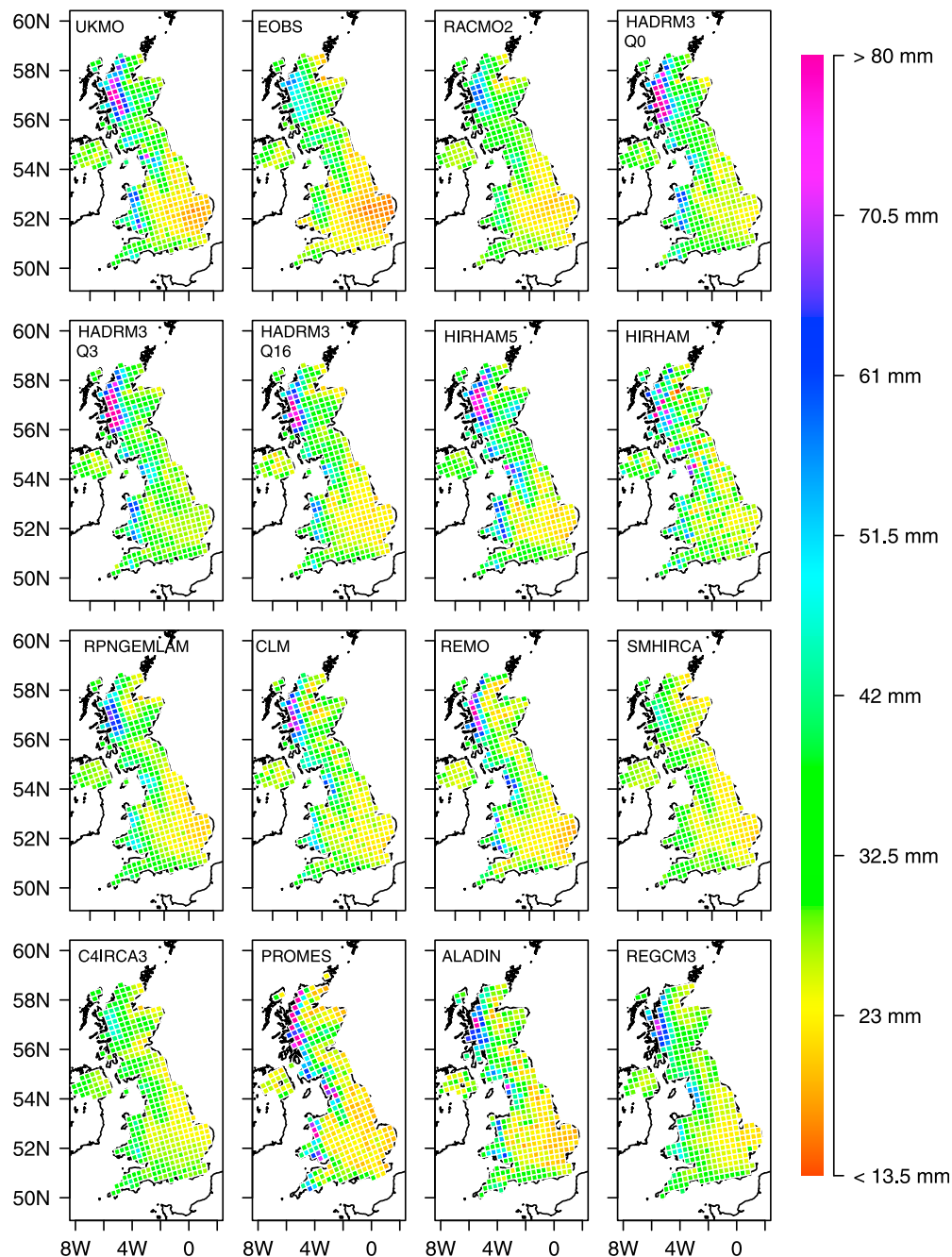
Figure 5. As in Figure 2, but for February-return levels to the probability 1/25, $z_{1/25}(Feb)$.

topography and mostly frontal origin of precipitation in Northern Ireland.

[80] 2. In Scotland the relative amplitude is under-represented by some and over-represented by other RCMs. The peak times are well simulated, only in ES one or two months too early by few RCMs. Northwest Scotland's distinctive features, that is high absolute values, strong annual cycle, maximum values late in the year, are also captured by most RCMs. The spatial expansion of these features and their exact values on the other hand are not well represented. The relative bias of the monthly return levels is large, only few RCMs have a bias of less than 10%, the other RCMs under-represent the monthly return levels by up to 35%. Even

though precipitation in West Scotland is also mostly of frontal origin, the representation of the position of the storm tracks might explain the large bias in the absolute values (contrary to Northern Ireland).

[81] 3. West England and Wales are not consistently represented across different RCMs and across different characteristics. The relative amplitude is too strongly represented by some RCMs and not strong enough by others. The peak time is well represented by some RCMs and one month too late by others. The monthly return levels are mostly under-represented, some RCMs show a slight annual cycle in the relative bias.
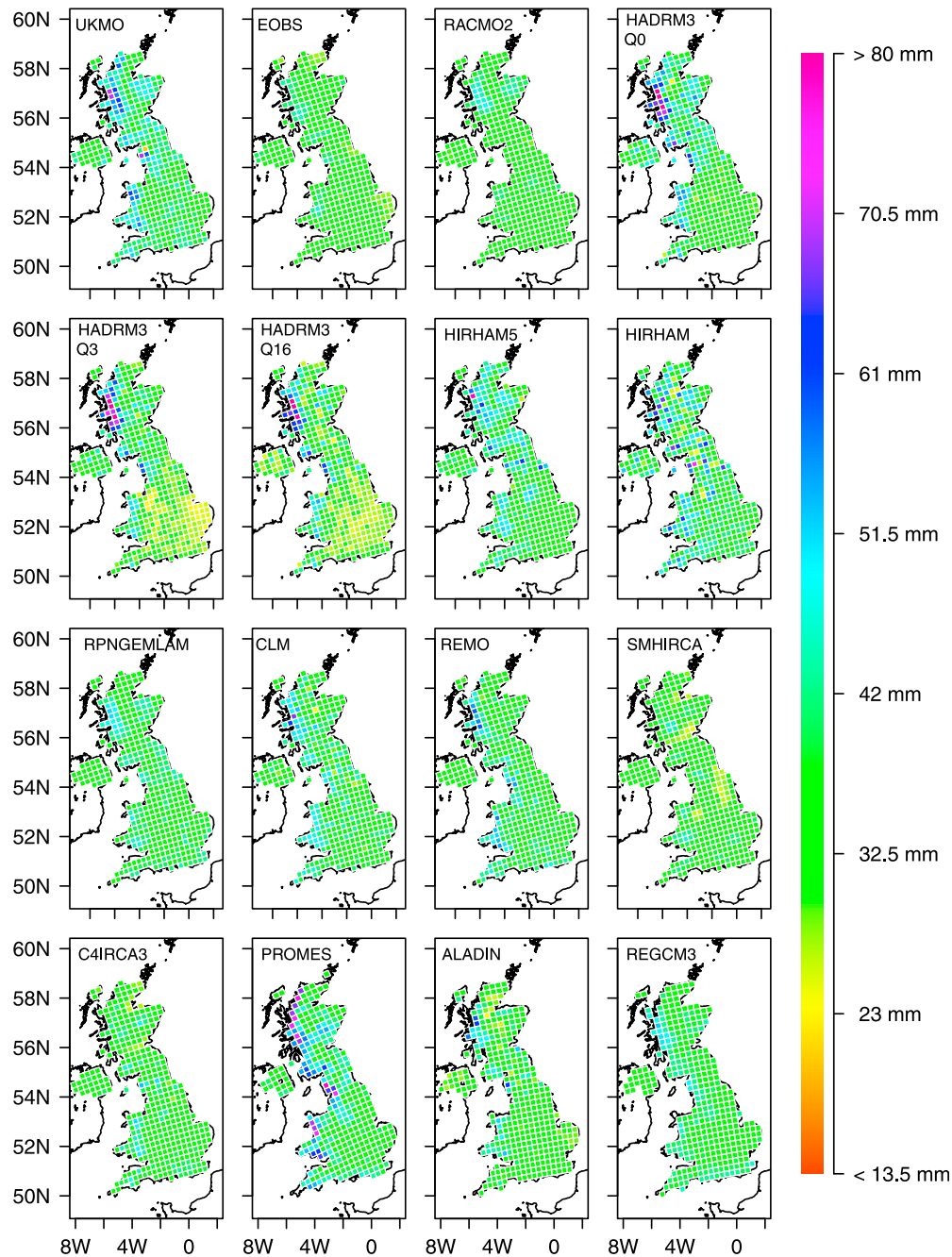
Figure 6. As in Figure 2, but for August-return levels to the probability 1/25, $z_{1/25}(Aug)$.

[82] 4. East England is the region with the most serious misrepresentations by almost all RCMs. The relative amplitude is under-represented by almost all RCMs in CEE and SEE and by most in NEE. In East Anglia, a region with a strong relative amplitude, the simulated relative amplitude by some RCMs amounts to less than 10%. Only some RCMs represent the timing of the maximum value well, some RCMs misrepresent the peak time by up to four months late. The relative bias of the monthly return levels for CEE and SEE shows an annual cycle, over-representing the return levels of the winter months and under-representing the return levels of the summer months.

[83] The annual cycle in the relative bias is prevalent in those regions and those RCMs where, respectively in which, the relative amplitude is not well represented.

[84] The ability of E-OBS to represent the annual cycle of monthly return levels is independent from the region (apart from a collection of grid points in Wales).

5. Conclusion

[85] In this study we evaluate the performance of 14 Regional Climate Models (RCMs) of the ENSEMBLES project in representing the annual cycle of daily heavy precipitation across the UK over the 1961–2000 time period.
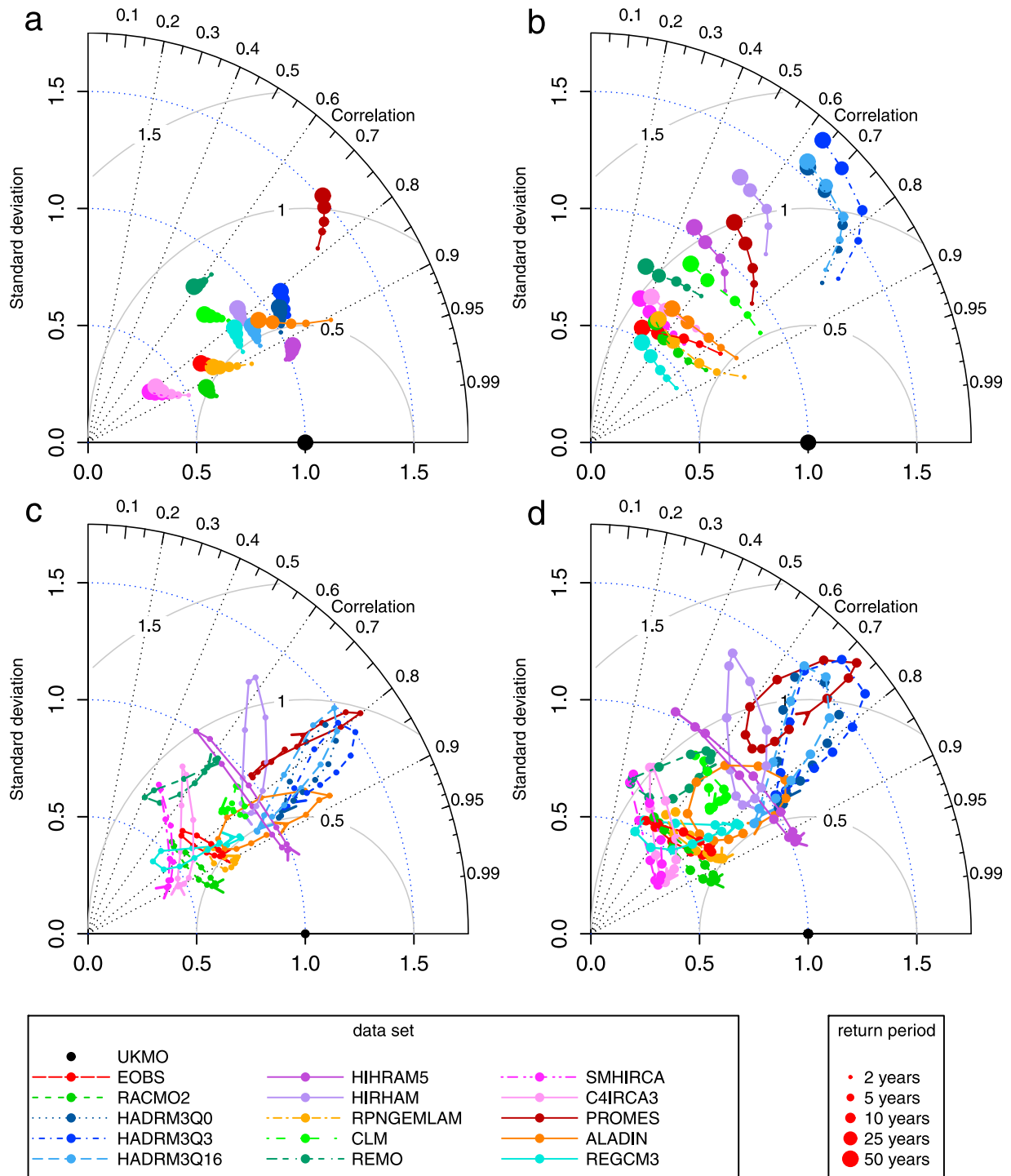
Figure 7. As in Figure 3, but for the maps of the monthly return levels: for (a) February and (b) August as well as return periods (c) 5 and (d) 25 years. The diameters of the points correspond to the return periods of the monthly return levels. The twelve points of each data set correspond to the twelve monthly return levels; the arrow indicates $z_{\{1/5,1/25\}}(Jan)$ and its direction the course of the year.
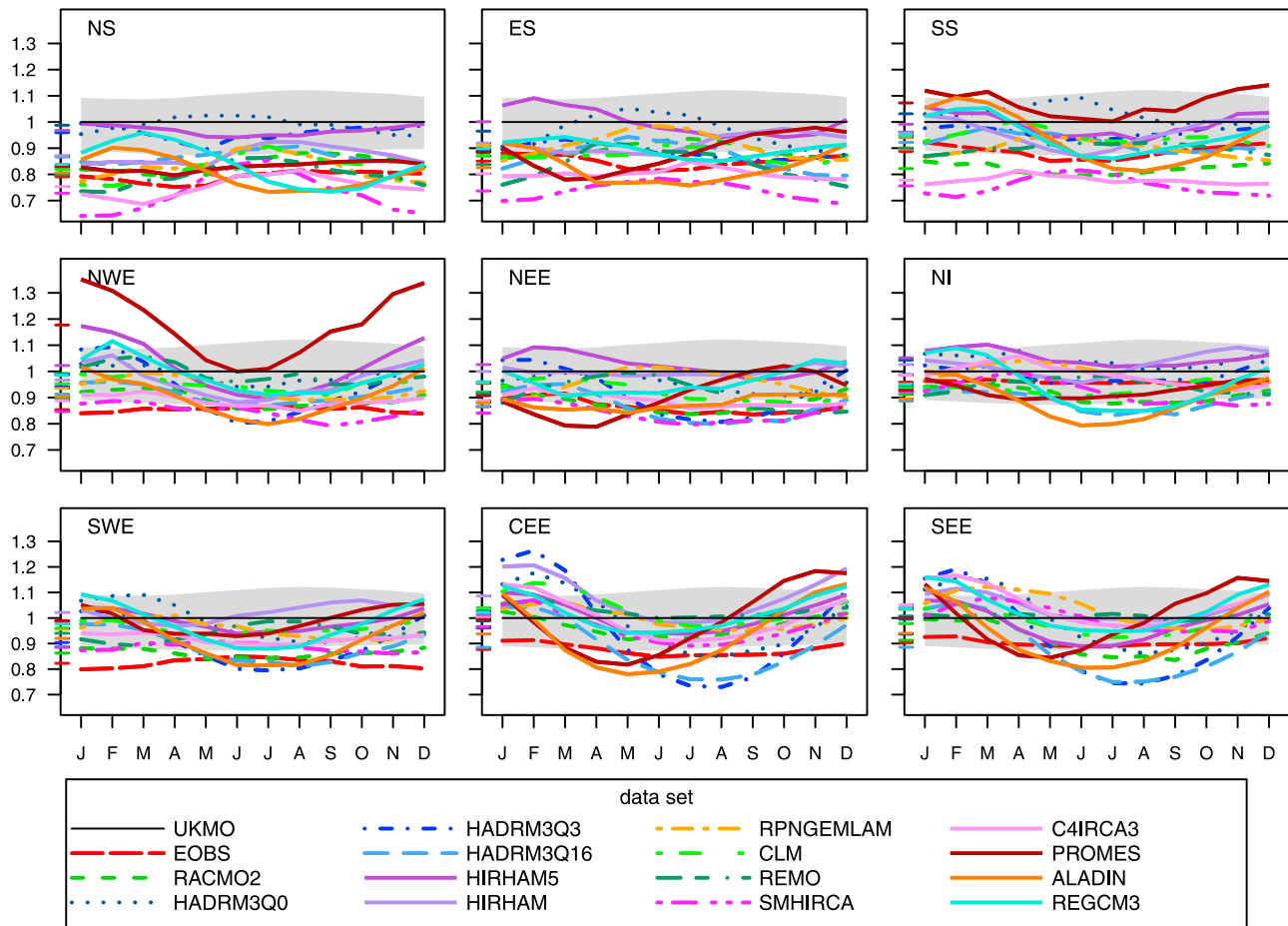
Figure 8. The relative bias of the monthly 1/25-return levels averaged over the grid points of one region, resolved by each month (colored lines) and annual mean (colored tick marks). UKMO estimates (black line at 1) and their regionally averaged uncertainties (shaded in gray).

[86] We estimate extreme value distributions from the output of the RCMs following the Poisson process approach. To describe the annual cycle of observed heavy precipitation events, it is necessary to use a sinusoidal model for the distributions' location and scale parameters.

[87] We show that the RCMs represent the annual cycle of heavy precipitation in the UK with different quality. The representation of the monthly return levels, i.e., the annual return levels conditional on the month of their occurrence, depends strongly on the region, the time of the year and slightly on the return period. HIRHAM5 is the only RCM with a relative bias of less than 10% for all months and regions but one. For most of the other RCMs this is only true for Northern Ireland. In general the months with the lowest relative biases are the fall and spring months, since an annual cycle in the bias exists with over-representation of winter return levels and under-representation of summer return levels.

[88] A significant shortcoming is the misrepresentation of the amplitude of the annual cycle relative to the annual mean. The RCMs fail in representing its anomaly patterns across all return periods. Especially regions with a strong peak of the annual cycle in summer, such as East Anglia, are simulated with almost no annual cycle by some RCMs. Despite problems in representing the strength of the annual cycle the timing of its maximum value is simulated rather well. The spatial pattern of the peak time of the annual cycle, with peaks in the late summer in the east of the UK and late fall or early winter in the west, is well represented in almost all models and across almost all return periods.

[89] Some characteristics, such as the peak time of the annual cycle, have a temporal accuracy of one month and show a homogeneous spatial structure on larger scales. These characteristics might be simulated robustly enough to serve for regional climate change studies. Characteristics with finer spatial structures, such as the relative amplitude of the annual cycle, are rather poorly represented. Hence the simulated relation between heavy precipitation intensities occurring in different seasons cannot be interpreted in a meaningful way.

[90] Consequently simulations with impact models which strongly depend on the explicit annual cycle of heavy precipitation should be carefully considered.

[91] We also observe the well-known findings that the noise on grid point scales can be too large to extract meaningful information, [e.g., *Kendon et al.*, 2008; *Maraun et al.*, 2010]. Therefore, instead of studying a single grid point

Table A1. Pattern Bias of GEV Parameters and Annual Return Levels[a]

	GEV Parameter							Return Period in Years				
RCM	μ_0	$\phi(\mu)*$	$A(\mu)$	σ_0	$\phi(\sigma)*$	$A(\sigma)$	$\xi*$	2	10	20	50	100
UKMO	1	0	1	1	0	1	0	1	1	1	1	1
EOBS	***0.91***	**−0.01**	0.9	0.85	0.04	0.88	−0.01897	0.86	0.84	***0.83***	***0.825***	***0.82***
RACMO2	0.94	−0.27	0.84	0.88	−0.07	0.99	−0.01439	0.89	0.87	0.87	0.86	0.86
HADRM3Q0	0.95	−0.27	0.71	0.98	−0.09	1.11	0.02529	0.97	**1**	**1.01**	1.03	1.05
HADRM3Q3	0.92	***−0.71***	0.97	0.94	***−0.58***	1.07	0.01892	0.94	0.96	0.97	0.98	**1**
HADRM3Q16	***0.91***	−0.29	***0.64***	0.90	−0.34	**1**	−0.01019	0.89	0.88	0.87	0.87	0.87
HIRHAM5	**1**	−0.28	**0.98**	1.02	−0.24	1.27	**−0.00064**	1.01	1.01	1.02	1.02	1.03
HIRHAM	0.97	−0.32	1.03	0.97	**−0.02**	1.26	***0.03293***	**1**	1.03	1.04	1.068	1.09
RPNGEMLAM	1	−0.09	0.78	0.95	0.11	1.02	−0.02844	0.94	0.92	0.91	0.9	0.89
CLM	0.94	**−0.01**	0.75	0.93	0.08	1.1	0.01868	0.94	0.96	0.97	0.98	**1**
REMO	0.95	0.35	0.89	0.92	0.12	1.22	0.01676	0.95	0.96	0.96	0.98	0.99
SMHIRCA	0.94	−0.02	0.67	***0.8***	0.24	0.75	0.00578	***0.85***	***0.83***	***0.83***	***0.83***	***0.83***
C4IRCA3	0.98	−0.13	0.8	0.84	0.05	0.86	0.00721	0.9	0.88	0.88	0.89	0.89
PROMES	0.96	−0.29	1.11	**1.01**	−0.24	***1.44***	−0.00189	0.99	1.01	**1.01**	1.02	1.03
ALADIN	0.94	−0.57	0.77	0.9	−0.33	0.89	−0.02768	0.89	0.87	0.86	0.9	0.84
REGCM3	1.01	−0.54	0.65	0.93	−0.17	1.08	−0.00128	0.95	0.94	0.94	0.94	0.94

[a]The biases to the reference data (UKMO) are calculated for each grid point and then averaged over all grid points. The relative biases are (as a ratio) unitless; the absolute biases, marked with an asterisk, are given in radians (ϕ) or unitless (ξ). In each column the best RCM (boldface only) and the worst RCM (boldfaced italics) are highlighted. In the row for E-OBS the characteristics which are represented as good as/better than the best RCM are set in boldface and the characteristics which are represented as worse as/worse than the worst RCM are set in italic.

separately, one should make use of a large ensemble of RCMs. Results from a single RCM should be based on a set of grid points to avoid noise on grid point scales. Either way a bias of the RCMs remains. Further averaging over larger regions only seemingly reduces the overall bias by averaging out compensating biases.

[92] Apart from the validation of the RCMs, we also investigate a second observation-based data set, the E-OBS data set. It fails to represent the intensity of the monthly return levels, but is able to represent the spatial patterns of the phase and the relative amplitude of their annual cycle. The exact values of the phase and relative amplitude of the annual cycle of heavy precipitation are well represented as well. In this regard it could be used as a reference data set for validation studies. However, it should not be used for validating the amount of heavy precipitation or the absolute amplitude of its annual cycle.

Appendix A: GEV Parameters and Annual Return Level

A1. GEV Parameters

[93] The GEV maximum likelihood estimates for both observation-based data sets show similar spatial patterns but differ in magnitude, regional smoothness and spatial extension of the regional characteristics (see auxiliary material).[1]

[94] For all data sets the spatial patterns in the offset of the location, μ, and scale, σ, parameters show an east-west gradient with smaller values in the east. By the RCMs the offset of μ is the best represented of all characteristics in terms of pattern bias and pattern accordance. The spatial variation of the offset of σ is very divergently represented, mostly underestimated. The amplitude of μ as well as the phase of μ are diversely represented ($0.5 \le pcor(A(\mu)) \le 0.92$ and $0.2 \le$

[1]Auxiliary materials are available in the HTML. doi:10.1029/2012JD017828.

$pcor(\phi(\sigma)) \le 0.8$, and overall bias in Table A1). The amplitude of σ is rather poorly represented: the centered pattern root mean squared distance varies between 1 and 2.1. The phase of σ is represented fairly well by half of the RCMs. The shape parameter, ξ, which is difficult to estimate, exhibits the largest differences between the data sets (see auxiliary material). In observations, the values of ξ are positive in the eastern parts and in Northern Ireland, zero (or slightly negative) along the south coasts of Wales and England and the west coast of Scotland.

[95] As we have seen in section 4.2, the pattern accordance of the peak time of the annual cycle of the monthly return levels is worse represented at small return periods than at large return periods. However, the pattern accordance of the phase of the location parameter is better represented than the accordance of the phase of the scale parameter (not shown).

[96] The RCM HADRM3Q3, which seriously misrepresents the annual cycle in the eastern parts of the UK, does not have a strong annual cycle in the scale parameter in these parts of the UK (auxiliary material). The pattern bias of the phase of both location and scale parameter is greatest for HADRM3Q3 as well (Table A1), resulting in the deficits shown in previous sections.

A2. Annual Return Level

[97] Table A1 shows the pattern bias for five return periods: 2,10,20,50,100 years. The overall biases are small, less than 20% for all return periods and all RCMs. HADRM3Q0 and HADRM3Q3 show very little biases in the annual return levels. Surprisingly the second observation-based data set, E-OBS, shows largest bias in the annual return levels.

[98] The pattern accordance is best for HIRHAM5 (rp2,10) and RACMO2 (rp20-100) and worst for REMO (rp2,10) and HADRM3Q3 (rp20-100) (not shown).

[99] The RCMs which represent the annual 2-year return levels best, in terms of relative bias or *pcrmsd* values

(Table A1), are not the same RCMs which represent the annual cycle best, in terms of relative amplitude or phase.

References

Barrow, E., and M. Hulme (1997), Describing the surface climate of the British Isles, in *Climates of the British Isles*, edited by M. Hulme and E. Barrow, chap. 3, pp. 33–62, Routledge, London.

Beniston, M., et al. (2007), Future extreme events in European climate: An exploration of regional climate model projections, *Clim. Change*, *81*(S1), 71–95, doi:10.1007/s10584-006-9226-z.

Böhm, U., M. Kücken, W. Ahrens, A. Block, D. Hauffe, K. Keuler, B. Rockel, and A. Will (2006), CLM-the climate version of LM: Brief description and long-term applications, *COSMO Newsl.*, *6*, 225–235.

Brown, S. J., J. Caesar, and C. A. T. Ferro (2008), Global changes in extreme daily temperature since 1950, *J. Geophys. Res.*, *113*, D05115, doi:10.1029/2006JD008091.

Buonomo, E., R. Jones, C. Huntingford, and J. Hannaford (2007), On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations, *Q. J. R. Meteorol. Soc.*, *133*(622), 65–81, doi:10.1002/qj.13.

Christensen, J. H., and O. B. Christensen (2003), Climate modelling: Severe summertime flooding in Europe, *Nature*, *421*(6925), 805–806, doi:10.1038/421805a.

Christensen, J. H., B. Machenhauer, R. G. Jones, C. Schär, P. M. Ruti, M. Castro, and G. Visconti (1997), Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions, *Clim. Dyn.*, *13*(7–8), 489–506, doi:10.1007/s003820050178.

Christensen, J., O. Christensen, P. Lopez, E. Van Meijgaard, and M. Botzet (1996), The HIRHAM4 Regional Atmospheric Climate Model, *Sci. Rep. 96–4*, Dan. Meteorol. Inst., Copenhagen.

Christensen, J., et al. (2007), Regional climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 849–940, Cambridge Univ. Press, Cambridge, U. K.

Christensen, J., E. Kjellström, F. Giorgi, G. Lenderink, and M. Rummukainen (2010), Weight assignment in regional climate models, *Clim. Res.*, *44*(2–3), 179–194, doi:10.3354/cr00916.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

Coles, S., J. Heffernan, and J. Tawn (1999), Dependence measures for extreme value analyses, *Extremes*, *2*(4), 339–365, doi:10.1023/A:1009963131610.

Collins, M., B. B. B. Booth, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb (2006), Towards quantifying uncertainty in transient climate change, *Clim. Dyn.*, *27*(2–3), 127–147, doi:10.1007/s00382-006-0121-0.

Eastoe, E. (2007), Statistical models for dependent and non-stationary extreme events., PhD thesis, Lancaster Univ., Lancaster, U. K.

Fawcett, L., and D. Walshaw (2007), Improved estimation for temporally clustered extremes, *Environmetrics*, *18*(2), 173–188, doi:10.1002/env.810.

Fowler, H. J., and M. Ekström (2009), Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, *Int. J. Climatol.*, *29*(3), 385–416, doi:10.1002/joc.1827.

Fowler, H. J., and C. G. Kilsby (2003), A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000, *Int. J. Climatol.*, *23*(11), 1313–1334, doi:10.1002/joc.943.

Fowler, H. J., M. Ekström, C. G. Kilsby, and P. Jones (2005), New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate, *J. Hydrol.*, *300*(1–4), 212–233, doi:10.1016/j.jhydrol.2004.06.017.

Fowler, H. J., M. Ekström, S. Blenkinsop, and A. P. Smith (2007), Estimating change in extreme European precipitation using a multimodel ensemble, *J. Geophys. Res.*, *112*, D18104, doi:10.1029/2007JD008619.

Fowler, H. J., D. Cooley, S. R. Sain, and M. Thurston (2010), Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment, *Extremes*, *13*(2), 241–267, doi:10.1007/s10687-010-0101-y.

Frei, C. (2003), Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, *J. Geophys. Res.*, *108*(D3), 4124, doi:10.1029/2002JD002287.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2006), Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *J. Geophys. Res.*, *111*, D06105, doi:10.1029/2005JD005965.

Giorgi, F., and L. O. Mearns (1999), Introduction to special section: Regional climate modeling revisited, *J. Geophys. Res.*, *104*(D6), 6335–6352, doi:10.1029/98JD02072.

Gregory, J. M., P. D. Jones, and T. M. L. Wigley (2007), Precipitation in Britain: An analysis of area-average data updated to 1989, *Int. J. Climatol.*, *11*(3), 331–345, doi:10.1002/joc.3370110308.

Haugen, J., and H. Haakenstad (2006), Validation of HIRHAM version 2 with 50 km and 25 km resolution, *RegClim Gen. Tech. Rep. 9*, Norw. Meteorol. Inst., Oslo.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.

Hofstra, N., M. Haylock, M. New, and P. D. Jones (2009), Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res.*, *114*, D21101, doi:10.1029/2009JD011799.

Intergovernmental Panel on Climate Change (IPCC) (2012), Summary for policymakers, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, edited by C. Field et al., pp. 1–19, Cambridge Univ. Press, Cambridge, U. K.

Jacob, D. (2001), A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin, *Meteorol. Atmos. Phys.*, *77*(1–4), 61–73, doi:10.1007/s007030170017.

Kallache, M., M. Vrac, P. Naveau, and P.-A. Michelangeli (2011), Nonstationary probabilistic downscaling of extreme precipitation, *J. Geophys. Res.*, *116*, D05113, doi:10.1029/2010JD014892.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*(8–12), 1287–1304, doi:10.1016/S0309-1708(02)00056-8.

Katz, R. W., G. S. Brush, and M. B. Parlange (2005), Statistics of extremes: Modeling ecological disturbances, *Ecology*, *86*(5), 1124–1134, doi:10.1890/04-0606.

Kendon, E. J., D. P. Rowell, R. G. Jones, and E. Buonomo (2008), Robustness of future changes in local precipitation extremes, *J. Clim.*, *21*(17), 4280–4297, doi:10.1175/2008JCLI2082.1.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007), Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *J. Clim.*, *20*(8), 1419–1444, doi:10.1175/JCLI4066.1.

Kjellström, E., L. Bärring, S. Gollvik, U. Hansson, C. Jones, and P. Samuelsson (2005), A 140-year simulation of European climate with the new version of the Rossby Centre regional atmospheric climate model (RCA3), report, SMHI, Norrköping, Sweden.

Lemon, J. (2010), Plotrix: A package in the red light district of R, *R News*, *6*(4), 8–12.

Maraun, D., H. W. Rust, and T. J. Osborn (2009a), The annual cycle of heavy precipitation across the United Kingdom: A model based on extreme value statistics, *Int. J. Climatol.*, *29*(12), 1731–1744, doi:10.1002/joc.1811.

Maraun, D., T. J. Osborn, and H. W. Rust (2009b), The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships, *Clim. Dyn.*, *36*(1–2), 261–275, doi:10.1007/s00382-009-0710-9.

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, *48*, RG3003, doi:10.1029/2009RG000314.

Maraun, D., T. J. Osborn, and H. W. Rust (2012), The influence of synoptic airflow on UK daily precipitation extremes. Part II: Regional climate model and E-OBS data validation, *Clim. Dyn.*, *39*, 287–301, doi:10.1007/s00382-011-1176-0.

Meehl, G., et al. (2007), Global climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. SM10.1–SM10.8, Cambridge Univ. Press, Cambridge, U. K.

Pal, J. S., et al. (2007), Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET, *Bull. Am. Meteorol. Soc.*, *88*(9), 1395–1409, doi:10.1175/BAMS-88-9-1395.

Parry, M., C. Rosenzweig, and M. Livermore (2005), Climate change, global food supply and risk of hunger, *Philos. Trans. R. Soc. B*, *360*, 2125–2138.

Perry, M., D. Hollis, and M. Elms (2009), The generation of daily gridded datasets of temperature and rainfall for the UK, *Clim. Memo. 24*, UK Met Off., Exeter, U. K.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.

Rivington, M., D. Miller, K. B. Matthews, G. Russell, G. Bellocchi, and K. Buchan (2008), Evaluating regional climate model estimates against site-specific observed data in the UK, *Clim. Change*, *88*(2), 157–185, doi:10.1007/s10584-007-9382-9.

Rosenzweig, C., A. Iglesias, X. Yang, P. R. Epstein, and E. Chivian (2001), Climate change and extreme weather events: Implications for food production, plant diseases, and pests, *Global Change Human Health*, *2*(2), 90–104, doi:10.1023/A:1015086831467.

Rust, H. W., D. Maraun, and T. J. Osborn (2009), Modelling seasonality in extreme precipitation, *Eur. Phys. J. Spec. Top.*, *174*(1), 99–111, doi:10.1140/epjst/e2009-01093-7.

Samuelsson, P., C. G. Jones, U. Willén, A. Ullerstig, S. Gollvik, U. Hansson, C. Jansson, E. Kjellström, G. Nikulin, and K. Wyser (2011), The Rossby Centre Regional Climate model RCA3: Model description and performance, *Tellus, Ser. A*, *63*(1), 4–23, doi:10.1111/j.1600-0870.2010.00478.x.

Sánchez, E., C. Gallardo, M. Gaertner, A. Arribas, and M. Castro (2004), Future climate extreme events in the Mediterranean simulated by a regional climate model: A first approach, *Global Planet. Change*, *44*(1–4), 163–180, doi:10.1016/j.gloplacha.2004.06.010.

Semenov, V., and L. Bengtsson (2002), Secular trends in daily precipitation characteristics: Greenhouse gas simulation with a coupled AOGCM, *Clim. Dyn.*, *19*(2), 123–140, doi:10.1007/s00382-001-0218-4.

Smith, R. L. (1989), Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone, *Stat. Sci.*, *4*(4), 367–377.

Smith, R. L. (1990), Regional estimation from spatially dependent data, technical report, 17 pp., Dep. of Stat., Univ. of N. C., Chapel Hill. [Available at http://www.stat.unc.edu/postscript/rs/regest.]

Stephenson, A. G. (2002), evd: Extreme value distributions, *R News*, *2*(2), 31–32.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.

Tomassini, L., and D. Jacob (2009), Spatial analysis of trends in extreme precipitation events in high-resolution climate model results and observations for Germany, *J. Geophys. Res.*, *114*, D12113, doi:10.1029/2008JD010652.

Toreti, A., E. Xoplaki, D. Maraun, F. G. Kuglitsch, H. Wanner, and J. Luterbacher (2010), Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns, *Nat. Hazards Earth Syst. Sci.*, *10*(5), 1037–1050, doi:10.5194/nhess-10-1037-2010.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*(612), 2961–3012, doi:10.1256/qj.04.176.

van der Linden, P., and J. Mitchell (Eds.) (2009), ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project, technical report, Met Off. Hadley Cent., Exeter, U. K.

van Meijgaard, E., L. H. van Ulft, W. J. van de Berg, F. C. Bosveld, B. van den Hurk, G. Lenderink, and A. P. Siebesma (2008), The KNMI regional atmospheric climate model RACMO version 2.1, *KNMI Publ. TR-302*, R. Neth. Meteorol. Inst., De Bilt, Netherlands.

Wigley, T. M. L., J. M. Lough, and P. D. Jones (1984), Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series, *J. Climatol.*, *4*(1), 1–25, doi:10.1002/joc.3370040102.

Wilks, D. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 304 pp., Academic, San Diego, Calif.