

# The Value of High-Resolution Met Office Regional Climate Models in the Simulation of Multihourly Precipitation Extremes

STEVEN C. CHAN\*

*School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, United Kingdom*

ELIZABETH J. KENDON

*Met Office Hadley Centre, Exeter, United Kingdom*

HAYLEY J. FOWLER AND STEPHEN BLENKINSOP

*School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, United Kingdom*

NIGEL M. ROBERTS

*Met Office Reading, Reading, United Kingdom*

CHRISTOPHER A. T. FERRO

*College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom*

(Manuscript received 18 November 2013, in final form 7 February 2014)

## ABSTRACT

Extreme value theory is used as a diagnostic for two high-resolution (12-km parameterized convection and 1.5-km explicit convection) Met Office regional climate model (RCM) simulations. On subdaily time scales, the 12-km simulation has weaker June–August (JJA) short-return-period return levels than the 1.5-km RCM, yet the 12-km RCM has overly large high return levels. Comparisons with observations indicate that the 1.5-km RCM is more successful than the 12-km RCM in representing (multi)hourly JJA very extreme events. As accumulation periods increase toward daily time scales, the erroneous 12-km precipitation extremes become more comparable with the observations and the 1.5-km RCM. The 12-km RCM fails to capture the observed low sensitivity of the growth rate to accumulation period changes, which is successfully captured by the 1.5-km RCM. Both simulations have comparable December–February (DJF) extremes, but the DJF extremes are generally weaker than in JJA at daily or shorter time scales. Case studies indicate that “gridpoint storms” are one of the causes of unrealistic very extreme events in the 12-km RCM. Caution is needed in interpreting the realism of 12-km RCM JJA extremes, including short-return-period events, which have return values closer to observations. There is clear evidence that the 1.5-km RCM has a higher degree of realism than the 12-km RCM in the simulation of JJA extremes.

---

 Denotes Open Access content.

---

\* Visiting scientist at the Met Office Hadley Centre, Exeter, United Kingdom.

---

*Corresponding author address:* Steven Chan, Met Office, FitzRoy Road, Exeter, EX1 3PB, United Kingdom.  
E-mail: steven.chan@metoffice.gov.uk

## 1. Introduction

Precipitation extremes have a large impact on society through floods and droughts, infrastructure damage, and even human casualties. Hence, understanding and quantifying their magnitude and frequency for the present and how they may change in the future is of great importance. For the United Kingdom, engineers and hydrologists have often relied on the Flood Estimation Handbook (FEH) (Reed 1999), which uses statistical methods to estimate

return levels and growth curves from historical observations. Improving the understanding of current and future extreme precipitation risks is one of the central goals of the U.K. Natural Environment Research Council (NERC) Changing Water Cycle (CWC) program of which the Convective Extremes (CONVEX) project is a part (NERC 2013).

Numerous studies exist on model-simulated (multi) daily extremes. Fowler and Ekström (2009), Hanel et al. (2009), and Hanel and Buishand (2010) have found that (multi)daily precipitation extremes are well simulated by various 25- to 50-km resolution regional climate models (RCMs) for all seasons except summer. This is facilitated by the widespread availability of quality-controlled daily precipitation datasets, such as the Met Office (UKMO) National Climatic Information Centre (NCIC) daily gridded precipitation (Perry et al. 2009), which is available for the United Kingdom at a high resolution (5 km) for 50+ years. As a consequence, UK daily precipitation extremes have been well characterized (Fowler and Kilsby 2003a,b; Maraun et al. 2009, 2011; Atyeo and Walshaw 2012; Jones et al. 2013).

However, to date, few studies have examined subdaily precipitation extremes because of both sparse subdaily observations and the inability of coarse-resolution dynamical models to reliably simulate subdaily precipitation. For the United Kingdom, hourly radar (“Radarnet”) precipitation measurements (Golding 1998; Harrison et al. 2000) are only available as a post-2003 gridded data product. The UKMO and the England and Wales Environment Agency (EA) also maintain tipping-bucket rain gauge networks, both of which are used in the FEH hourly extreme assessments. Some hourly gauges have measurements for 20–30 years, but hourly station availability is an order of magnitude less than for daily stations (Faulkner 1999). From the few studies of subdaily precipitation that have been, Madsen et al. (2009) found an intensification of Danish subdaily extreme precipitation in recent decades; Willems (2000) examined Belgian precipitation extremes across a range of accumulation periods (minutes to multiday) and found that extreme behavior is seasonal and weather system dependent; and Overeem et al. (2009) showed that radar observations produce subhourly to multiday precipitation extreme estimates that are comparable with gauges, which supports the use of radar precipitation for extreme studies.

Unlike observations, dynamical computer models of weather and climate can generate as much hourly data as the modeler can afford. The models are powerful tools with which to assess current and future extreme precipitation if the simulated extremes are deemed to be sufficiently accurate. However, there are a number of reasons to question the quality of precipitation from

coarse-resolution dynamical models, especially on hourly time scales. There can be large mean biases (Kjellström et al. 2010), poor timing and durations (Brockhaus et al. 2008), and incorrect spatial distributions (Gregersen et al. 2013). One modeling study, that of Hanel and Buishand (2010), assessed hourly extremes simulated by the Ensemble-Based Predictions of Climate Changes and Their Impacts (ENSEMBLES) project 25-km parameterized-convection RCMs (Hewitt and Griggs 2004) over the Netherlands with the help of radar data, and found that the model-simulated generalized extreme-value distribution shape (location) parameters are too high (too low), leading to weaker short-return-period hourly extremes but overly intense long-return-period hourly extremes.

Hourly extremes (especially summer ones) are dominated by convective precipitation. Convective precipitation tends to be both short-duration (1–4 h) and spatially localized. An example of a short-duration convective extreme is the 2004 Boscastle flood (Burt 2005; Golding et al. 2005). Long-duration (from 12 h to 3 day) extremes tend to have larger spatial scales and are often associated with synoptic weather systems; a good example of such an event is the 2007 U.K. summer floods (Blackburn et al. 2008). The nature of the event has important consequences for the extent to which we would expect it to be captured by a dynamical model grid box average or a point observing station. Convective events (with horizontal scales smaller than 10 km) have scales that are smaller than typical RCM grid box sizes. Synoptic systems have spatial scales  $O(10^3)$  km, and precipitation associated with synoptic systems has spatial scales larger than typical RCM grid box sizes.

A major contributor to the error in current global and regional models is the convective parameterization (CP) scheme, which aims to represent the effects of convection on the grid scale but does not capture the dynamics of individual storms. Very high-resolution models (with order of 1-km grid spacings) are able to represent convection explicitly without the need for a parameterization scheme (Hohenegger et al. 2008; Kendon et al. 2012). Such kilometer-scale models are not truly cloud resolving and are often termed “convective-permitting,” as larger storms and mesoscale organizations are permitted but convective plumes and smaller showers are not resolved. The use of convective-permitting models is now common practice in short-range numerical weather forecasts (e.g., Roberts and Lean 2008), where they have been shown to provide a significantly improved representation of topographically enhanced and convective precipitation compared to coarser-resolution models (Lean et al. 2008).

The UKMO has recently finished multiyear present-climate integrations of high-resolution 1.5-km RCM

simulations over the southern United Kingdom (SUK) driven by a coarser (12 km) RCM from the same Unified Model suite (Kendon et al. 2012). Aside from grid spacing, a difference between the two simulations is that the 12-km RCM uses convective parameterization to represent the effects of unresolved convection. The 1.5-km RCM does not use CP, but instead it has explicit convection. Kendon et al. (2012) focused on the intensity, duration, and extent of hourly precipitation. They found that the 1.5-km RCM generates peak precipitation intensities that are too high, but it has a superior representation of the diurnal cycle, structure, and the duration of precipitation. The differences in intensities and duration suggest that the (multi)hourly extremes of the two models could also be very different.

Here we seek to diagnose the differences between the 12- and 1.5-km RCMs with the help of extreme value theory (EVT). We examine the extent to which the very high-resolution (1.5 km) explicit convection RCM gives an improved representation of hourly extremes over a coarser (12 km) parameterized convection RCM, and hence evaluate the 1.5-km RCM value for providing future projections.

The main issues that we wish to address here are the following.

- Does the 1.5-km RCM have different extreme behavior for precipitation than the 12-km RCM?
- What can be done to assess model reliability for hourly extremes with limited U.K. hourly observations? Are the observations adequate for use in extreme value analysis and model extreme evaluations?
- Are the physical representations of the 1.5- and 12-km RCM precipitation extremes “realistic” (i.e., physically and meteorologically plausible) in comparison to observations?

This study seeks to extend the work of Kendon et al. (2012) with specific focus on extreme events. This paper is outlined as follows: Section 2 describes the model and observational data used; section 3 overviews our methodologies; section 4 presents the results from the extreme value analysis; section 5 presents a physical interpretation of the extremes; and this paper concludes in section 6.

## 2. Models and observations

This paper uses the same two UKMO high-resolution (12 and 1.5 km) RCM simulations as in Kendon et al. (2012) and Chan et al. (2013). For observations, we have used the gauge-based UKMO daily gridded values between 1990 and 2008 (UK5) (Perry and Hollis 2005; Perry et al. 2009), the radar-based UKMO hourly gridded values between 2003 and 2010 (Golding 1998; Harrison

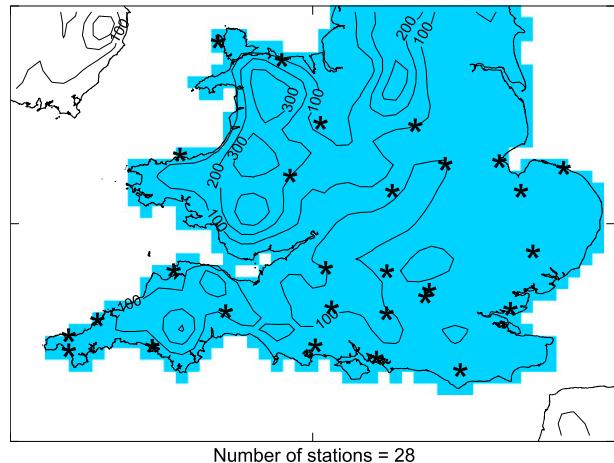


FIG. 1. The southern United Kingdom domain used in this study with the 12-km RCM surface geopotential height contoured. The approximate locations of the UKMO hourly gauges are marked with asterisks. This figure is adapted from Chan et al. (2013).

et al. 2000), and the UKMO hourly station values between 1992 and 2010 that are archived within the UKMO Integrated Data Archive System (MIDAS) (Sunter 2012). For brevity, we only discuss the key RCM differences and the MIDAS database here.

### a. High-resolution RCMs

The two RCM simulations are fully described in Kendon et al. (2012). They are as follows:

- a 12-km European simulation driven by the ERA-Interim reanalysis (Dee et al. 2011) and
- a 1.5-km southern U.K. simulation driven by the 12-km RCM.<sup>1</sup>

The RCM simulations are carried out for the years 1990 to 2008, and our southern United Kingdom (SUK) domain is shown in Fig. 1. A key difference between the two simulations is that the 1.5-km simulation does not use a convective parameterization. Daily and hourly precipitation totals are available from both simulations. For comparison, the 1.5-km RCM data are spatially averaged on to a 12-km grid for analysis. For the 12-km RCM, daily totals of “convective” (parameterized) and “large-scale” precipitation are also available.

### b. Observations

UKMO hourly gauge measurements from tipping-bucket gauges that measure with 0.2-mm increments are archived within the MIDAS observational database. We have selected 28 SUK hourly stations that have been in

<sup>1</sup> 1.5-km interior, variable spacing (up to 4 km) near the lateral boundaries.

nearly continuous operation since 1992 (Fig. 1). The data have been quality controlled, and the checks include unsupervised checks against unusual values and supervised checks at the station and the data center (UKMO 2013).

Analysis here uses precipitation accumulations from 1 h to 5 days. Comparisons with daily UK5 values are only made for 1-day+ accumulations. Multihour accumulations are computed as hourly running totals. To be consistent with UK5, all (multi)daily accumulations are computed as 0900–0900 UTC totals.

Radar has deficiencies. In particular, attenuation can lead to an underestimate of high intensities (Harrison et al. 2000). Gauges observe events that they can sample with higher reliability, but cannot observe events that do not occur over them, and localized maxima are often missed. Orographic precipitation is systematically underestimated as most gauges are located in valleys. Instrumental errors also give systemic negative bias for high intensities (Legates and Willmott 1990; Molini et al. 2005). All observations used here are quality controlled. Substantial differences between different observational datasets exist, and the differences can effect model bias evaluation (Sunyer et al. 2013).

### c. Data corrections

Hourly precipitation observations are limited, and it is desirable to utilize all available data. This means there is a need to accept incomplete observations that cover somewhat different time periods (i.e., radar and gauge observations do not fully overlap with the model simulation period). Data period sensitivity tests show that our results are not sensitive to the period choice (not shown).

Missing values pose a difficulty in computing multihourly accumulations. Corrections are made to the radar and hourly gauge measurements to account for missing values as follows. For radar, we discard all subdaily values that encounter missing values, while daily totals with  $U$  hours missing are corrected with a multiplier as long as  $U < 12$ :

$$P_{\text{RADAR}}^* = \frac{24}{24 - U} \sum_{\text{defined}} P_{\text{RADAR}}. \quad (1)$$

Otherwise, the radar daily totals are set to missing.

Multihourly gauge accumulations that have missing or quality-control-fail values are corrected by the accumulation-period median insertion. If more than a quarter of the hourly values are missing, then the multihourly accumulations are set to missing (i.e., up to 3 h of missing data are tolerated for a 12-h total). This technique equates to simple persistence forecasting, that is, predicting unknown local hourly values from recent values.

All gridded data (models and observations) are gridded to 12-km grid boxes by water-conserving spatial

averaging before analysis to prevent intensity distortions. All analysis excludes nonland grid boxes.

### 3. Statistical methods

We have adopted peaks-over-threshold (PoT) extreme value theory (Coles 2001) to analyze our model and observational data. PoT is widely used and produces comparable results with competing methodologies (Madsen et al. 1997; Martins and Stedinger 2000, 2001). PoT has been used to model hydrological extremes—recently by Re and Barros (2009), Tomassini and Jacob (2009), and Acero et al. (2011)—and has been applied to gridded datasets (Coelho et al. 2008; Overeem et al. 2009). PoT characterizes both the extreme recurrence rate (defined as the annual frequency of exceedances above a threshold,  $\lambda$ ) and the probability distribution of the excesses above a specified threshold with the generalized Pareto (GP) distribution. The GP distribution is defined by three parameters: threshold ( $t$ ; the “extreme” threshold), scale ( $\sigma$ ; analogous to the standard deviation), and shape ( $\xi$ ; analogous to the skewness, the curvature of return levels):

$$z(n | t, \sigma, \xi) = \begin{cases} t + \frac{\sigma}{\xi} [(\lambda n)^\xi - 1], & \xi \neq 0 \\ t + \sigma \ln[\lambda n], & \xi = 0, \end{cases} \quad (2)$$

in which  $n$  and  $z$  are the return period and return level respectively. High quantiles (large  $n$  return levels) are sensitive to both parameters with higher scale and shape parameters giving higher return periods. The GP parameters can be constrained with Bayesian prior distributions (Martins and Stedinger 2001). In the present study, we do not constrain any fit parameters. Our goal is to diagnose possible model errors, and it would be inappropriate to assume that we know a priori what are the probable values of the fitted parameters for our model data. We would also note that no spatial or regional pooling (Hosking and Wallis 1993) is used in the present analysis.

The generalized Pareto distribution requires data samples to exceed some “reasonably high” event threshold. There is no general method to determine the threshold, but ad hoc methods exist (Ribatet 2006). We use the 95th percentile of wet values ( $\geq 0.1$  mm) as in Tomassini and Jacob (2009) and Acero et al. (2011). Data samples may not be independent and identically distributed (i.i.d.) without declustering. Nonindependence is accounted by automatic declustering (Ferro and Segers 2003), but we depart from the original methodology by imposing a minimum 1-day declustering time. L-moments are used to estimate the GP parameters (Hosking 1990). Standard errors of the GP parameters are estimated by refitting with bootstrapping postdeclustered data samples 1000 times.

To test goodness of fit, we employ the Anderson–Darling test (ADt) (Anderson and Darling 1952; Stephens 1977; Laio 2004). The null hypothesis is that the data are drawn from a GP distribution with unknown parameters and is rejected when the test statistic is larger than critical values that are estimated by Monte Carlo simulations (Lilliefors 1967; Waller et al. 2003).

The very extreme events [say  $z(100)$ ] are more important than less extreme “uncommon” events [say  $z(2)$ ] in terms of their social and economic impact. The hydrological community uses nondimensional “growth rates” and “curves” as a common standard to characterize the difference between uncommon and extreme events. The growth rate, or curve,  $G(n)$  (Reed 1999) is defined as the multiple increase of the  $n$ th year return level over an index extreme value:

$$G(n) = \frac{z(n)}{z(i)}, \quad (3)$$

in which  $z(i)$  is some standard return level. That is usually chosen to be  $z(2)$ , and is often referred as the annual maximum rainfall median (RMED) (Reed 1999). Note that  $G(n)$  is independent of any return period correction as long as the correction is a linear multiplier that is independent of the return level. Areal reduction factors (ARF) (de Michele et al. 2001; Kjeldsen 2007), which are used to upscale point-estimated return levels to areal return levels, are such linear multipliers that are assumed to be functions of accumulation period and catchment area only. The independence from ARF makes  $G(n)$  an attractive diagnostic to accommodate hourly gauge data. The present analysis focuses on June–August (JJA) and December–February (DJF) only, in order to gain a cleaner division between summer convective and winter stratiform precipitation.

#### 4. GP distributions of model and observed precipitation extremes

In this section, we present the estimations of the generalized Pareto parameters together with return levels and growth curves. It is impractical to examine the estimates at each grid point or gauge, so some diagnostics are presented as a spatial average.

##### a. Spatial GP structure

The spatial maps of PoT parameters are shown as

- 1-h extremes: JJA (Fig. 2), DJF (Fig. 3), and
- 1-day extremes: JJA (Fig. 4), DJF (Fig. 5).

The spatially averaged PoT parameters [ $E(\sigma)$ ,  $E(\xi)$ ,  $E(t)$ ,  $E(\lambda)$ ] and GP standard errors [ $E(\Delta\sigma)$  and  $E(\Delta\xi)$ ]

are above each panel. The parameters estimated from the hourly gauges are shown in Table 1. Radar-estimated fitted parameters appear to display more spatial variability than the model- and gridded-gauge estimates. This is unsurprising as radar data are shorter.

For 1-h JJA extremes (Fig. 2) both 12- and 1.5-km simulations have comparable spatially averaged event occurrence rates (Figs. 2d,h), but higher event frequencies over the orography are evident in the 12-km RCM. Larger discrepancies are found for the other three parameters. When compared with the 1.5-km RCM, the 12-km simulation tends to have lower thresholds and scale parameters, but higher shape parameters. The threshold differences between the model simulations are closely related to the differences in their mean precipitation intensities, in which the 1.5-km RCM precipitation is more intense (Kendon et al. 2012). The differences in the extreme probability distributions are diagnosed by the differences in scale and shape parameters. In particular, higher shape parameters indicate that extreme intensities increase rapidly with rarity. The largest event frequencies in both models are concentrated over orography, which is not surprising as precipitation occurs more often there. For the 1.5-km RCM, higher thresholds are evident over orography and the southern coast.

The spatially averaged radar-estimated shape parameters are closer to the 1.5-km RCM estimates, although with a high standard error. While the higher standard error may be related to lower radar sample sizes (radar has 8 years of data compared to 19 years of RCM data), some radar grid points have shape parameters that are much higher than the average of the whole domain ( $\xi \geq 0.5$ ). There is an apparent concentration of large shape parameters around the Thames Estuary, the Wash, and East Anglia. The high shape parameters around the Wash are at least partially due to localized “radar observed” extremes in August 2003 and 2004 that are not corroborated by the UK5 observations (not shown).

Model comparisons with the radar estimates show that the 12-km (1.5 km) simulation has lower (higher) scale parameters and thresholds, indicating negative (positive) biases in short return levels. Both radar and the 1.5-km RCM have considerably lower shape parameters (“thinner tailed”) than the 12-km RCM (“heavy tailed”). The higher scale parameters and thresholds in the 1.5-km RCM are consistent with its tendency to have precipitation that is too intense (Kendon et al. 2012). Along the southern English coast, the 1.5-km RCM simulates lines of high threshold and scale parameter. This is also where positive biases in JJA precipitation and heavy precipitation frequency are observed (Chan et al. 2013). The number of rejected fits for JJA in all of the gridded data is comparable to the number that is expected by

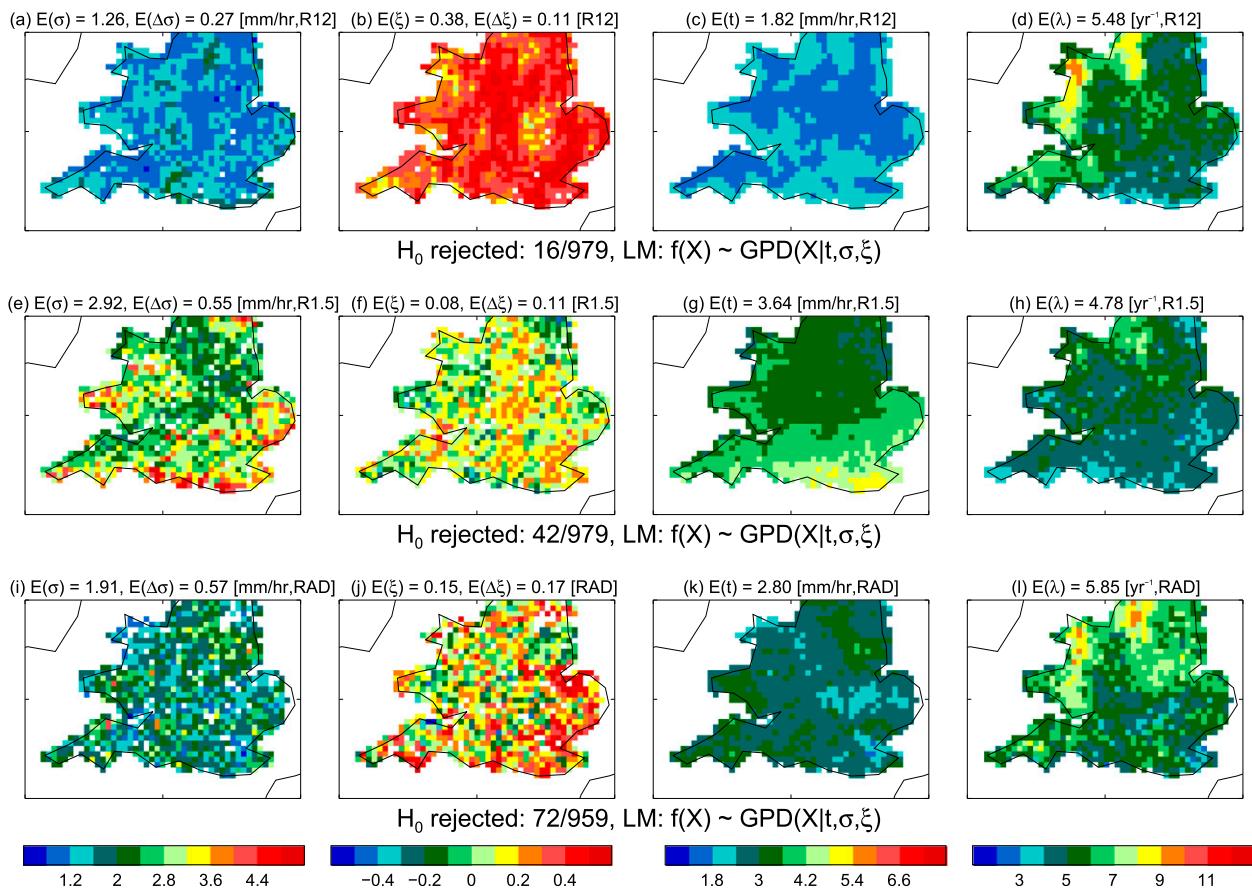


FIG. 2. PoT fit parameters for (top) 1990–2008 JJA 12-km RCM (R12), (middle) 1990–2008 JJA 1.5-km RCM (R1.5), and (bottom) 2003–10 Radarnet (RAD) 1-h precipitation accumulations: (a),(e),(i) scale ( $\sigma$ ,  $\text{mm h}^{-1}$ ), (b),(f),(j) shape ( $\xi$ ), (c),(g),(k) 95th percentile threshold ( $t$ ,  $\text{mm h}^{-1}$ ), and (d),(h),(l) declustered events per year ( $\lambda$ ,  $\text{yr}^{-1}$ ). Rejected GP fits are blanked in the scale and shape panels. The number of rejected fits, the spatially averaged fit parameters [ $E(\sigma)$ ,  $E(\xi)$ ,  $E(t)$ ,  $E(\lambda)$ ], and the spatially averaged standard errors of  $\sigma$  and  $\xi$  [ $E(\Delta\sigma)$ ,  $E(\Delta\xi)$ ] are also shown.

chance (a 5% test implies a 5% chance of rejection if the PoT model is correct).

A different picture is found for 1-h DJF GP distributions (Fig. 3). The large JJA shape parameters in the 12-km RCM are no longer evident in DJF. Both models and the radar data suggest lower shape parameters in DJF, and that is corroborated by the hourly gauge estimates (Table 1)—suggesting DJF has considerably lower large return levels. Thresholds are lower for DJF except for the 12-km RCM. For the radar and 1.5-km RCM, this is consistent with the lower intensities expected from nonconvective winter precipitation. Overall, the estimates indicate lower hourly extremes for winter precipitation. The 1.5-km RCM also has lower event frequencies over orography relative to radar and the 12-km RCM.

The U.K. DJF precipitation is predominately nonconvective. A much larger fraction of the 12-km model-simulated precipitation comes from the resolved dynamics and large-scale frontal ascent. Simulated precipitation is

less dependent on the convective parameterization. This hints that the CP may play a central role in the “fat tail” shape parameter biases seen for the 12-km-RCM JJA extremes.

Fits with DJF 1-h totals show a higher number of rejected fits than for JJA for both models and for radar gridded values. For the 12-km RCM, about 25% of the grid points have fits rejected, and the rejections appear to be concentrated over orography. For radar and the 1.5-km RCM, the number of rejections is still twice the number expected by chance at approximately 10%. Unlike the 12-km RCM, both radar and the 1.5-km RCM have no apparent concentration of rejections over orography.

The lower threshold, shape parameter, and goodness of fit for 1-h DJF precipitation are consistent with meteorological understanding. Winter and orographic precipitation are predominately stratiform, which is less intense on hourly time scales; such lower totals may not

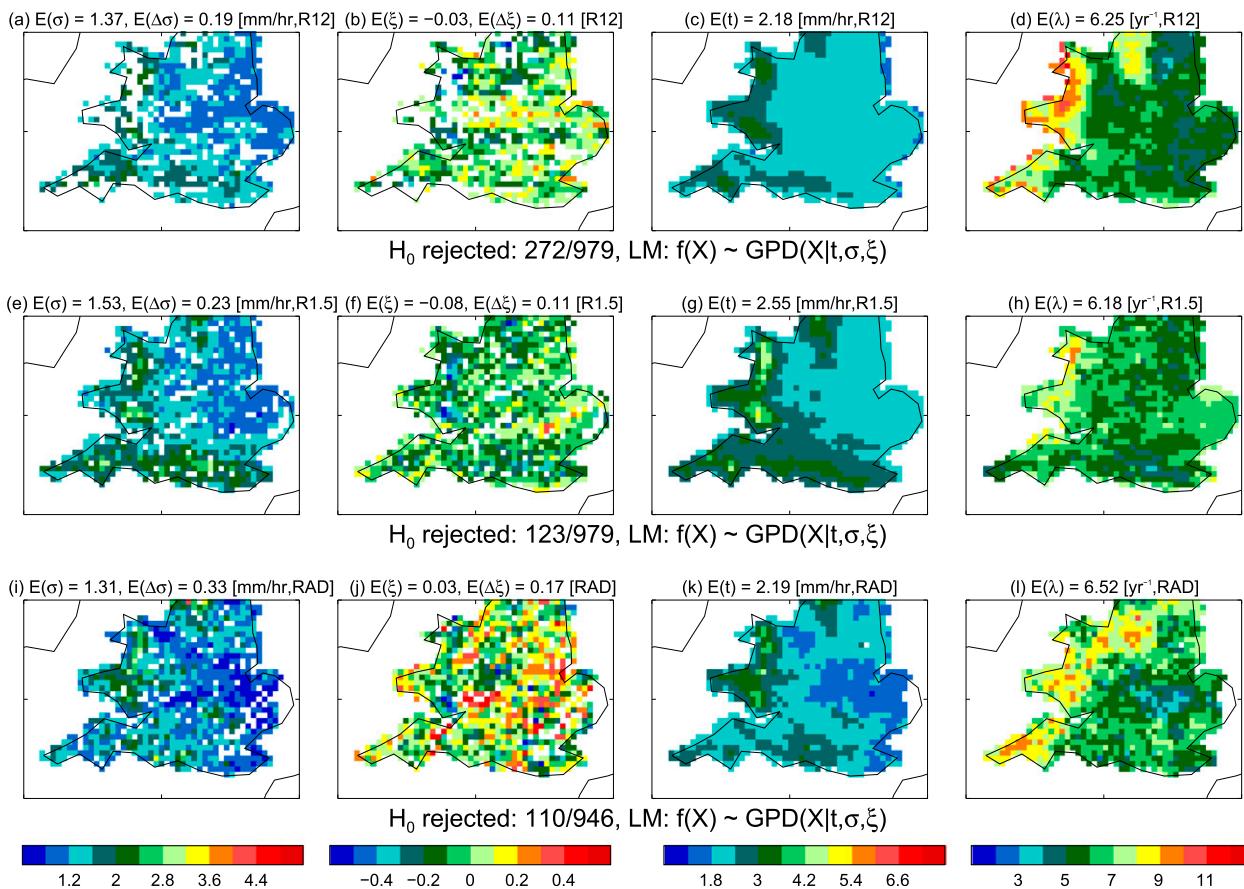


FIG. 3. As in Fig. 2, but for DJF 1-h precipitation accumulations.

be high enough to be considered “extreme,” and thus GP may less well describe their distribution.

Moving into 1-day JJA extremes (Fig. 4), the shape parameters ( $\xi \approx 0.1$ ) for the 1.5-km RCM and observations are found to be insensitive to the accumulation period change. This insensitivity is also noted in Hanel and Buishand (2010). Unlike the other datasets, the 12-km RCM shape parameter decreases when the accumulation period is extended to 1 day. The decreased 12-km RCM shape parameter is still higher than for both observations and the 1.5-km RCM, and the large shape parameter values are concentrated in the southeastern part of the model domain.

Similarly to 1-h extremes, the 1.5-km RCM has higher scale parameters and thresholds than both observations and the 12-km RCM, indicating a systematic positive bias in heavy precipitation. While the 12-km RCM tends to have higher event frequencies and lower thresholds (gives negative bias to short return levels), it has more accurate values for the scale parameter than the 1.5-km RCM. Large 1.5-km-RCM southern coastal scale parameters and thresholds are also seen for the daily estimates.

Radar and UK5 estimates are generally comparable. Both have similar spatial patterns in thresholds and event frequencies. The orographic enhancement is well captured by both. There is a tendency for the radar to give a high shape parameter over the southeast, which is not seen in the UK5 data.

All models and observations show the event frequencies are down by 50%–75% for 1-day extremes compared to 1-h extremes in JJA, but the spatial patterns are similar—the largest event frequencies are concentrated over orography in the western half of the domain. Event frequencies reflect the number of wet values, plus the effects of declustering.

Goodness of fit for 1-day DJF extremes (Fig. 5) is much better than for the 1-h DJF fits. The number of rejections is now comparable to that expected by chance, and is consistent with the physical understanding that winter extremes tend to occur at longer accumulations. A higher threshold may be needed for the 1-h DJF values to yield a good goodness of fit. Similar to the 1-h DJF estimates, the spatial average of the observation-fitted shape parameters is close to zero (or slightly negative for the

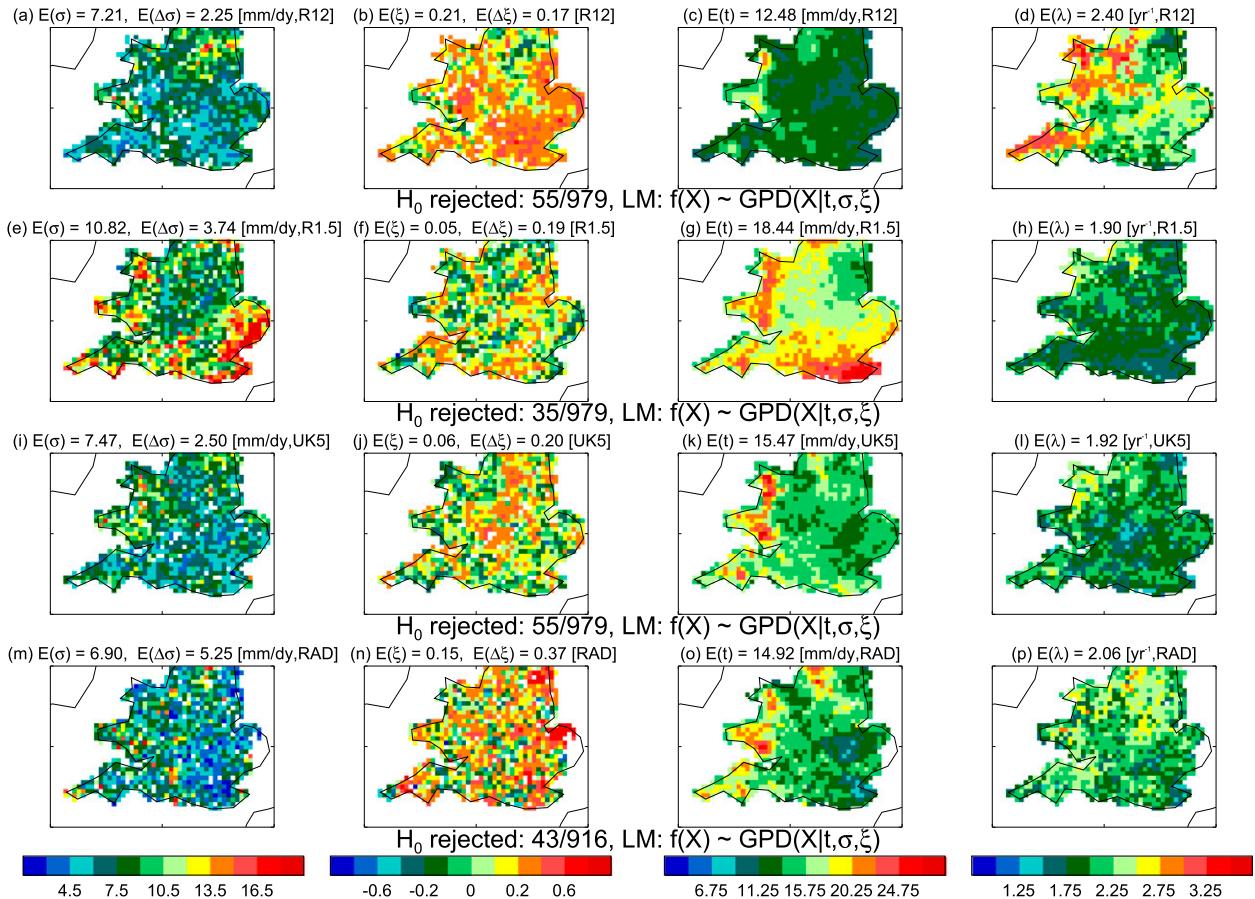


FIG. 4. As in Fig. 2, but for JJA 1-day precipitation accumulations. Both UK5 (middle bottom) and Radarnet (bottom) are shown. The units for threshold ( $t$ ) and the scale parameter ( $\sigma$ ) are millimeters per day.

gauges), and this is well simulated by both the 12- and the 1.5-km RCMs. Like the JJA 1-day extremes, there is a concentration of larger event frequencies, thresholds, and scale parameters over orography as one may expect from orographic precipitation.

Similar to the 1-day JJA estimates, both radar and UK5 have similar spatial patterns in the threshold and frequency of events. Again, radar estimates are showing high shape parameters in the southeast.

#### b. Return levels as functions of return periods

In Fig. 6, we compare the JJA and DJF return level estimates from the model and observational data. Since UK5 are daily, return levels are only shown for 1- and 5-day accumulations. The hourly gauge return levels are not shown to avoid the usage of ARF. Generally speaking, return levels are higher in JJA than in DJF for both the models and observations and across all compared accumulation periods.

For JJA, the 12-km RCM appears to simulate 5–10-yr return levels better than the 1.5-km RCM across a range

of accumulation periods (from 3- to 12-h totals); however, the good estimates there are a consequence of underestimating shorter return periods and overestimating longer return periods. For 1-h accumulations,  $z(2)$  values are overestimated (underestimated) by the 1.5-km (12-km) RCM. For the 1.5-km RCM,  $z(2)$  values are overestimated for all accumulation periods; this is in contrast with the 12-km RCM in which  $z(2)$  biases decrease with increasing accumulation period. Both RCMs have comparable 20–50-yr return levels—higher than both gridded observation estimates. For JJA, the 12-km RCM 1-h to 6-h accumulations return levels exceed the 1.5-km RCM return levels for 40–60-yr and longer return periods. For 6- and 12-h accumulations, both model-simulated  $z(100)$  values are nearly two times higher than the radar estimates. In summary, the 1.5-km RCM shows consistent positive return level biases across all examined accumulation periods, but the 12-km RCM biases depend on the accumulation period.

A key difference between the 12- and 1.5-km RCM return levels is in the gradient of  $z(n)$ . Owing to the

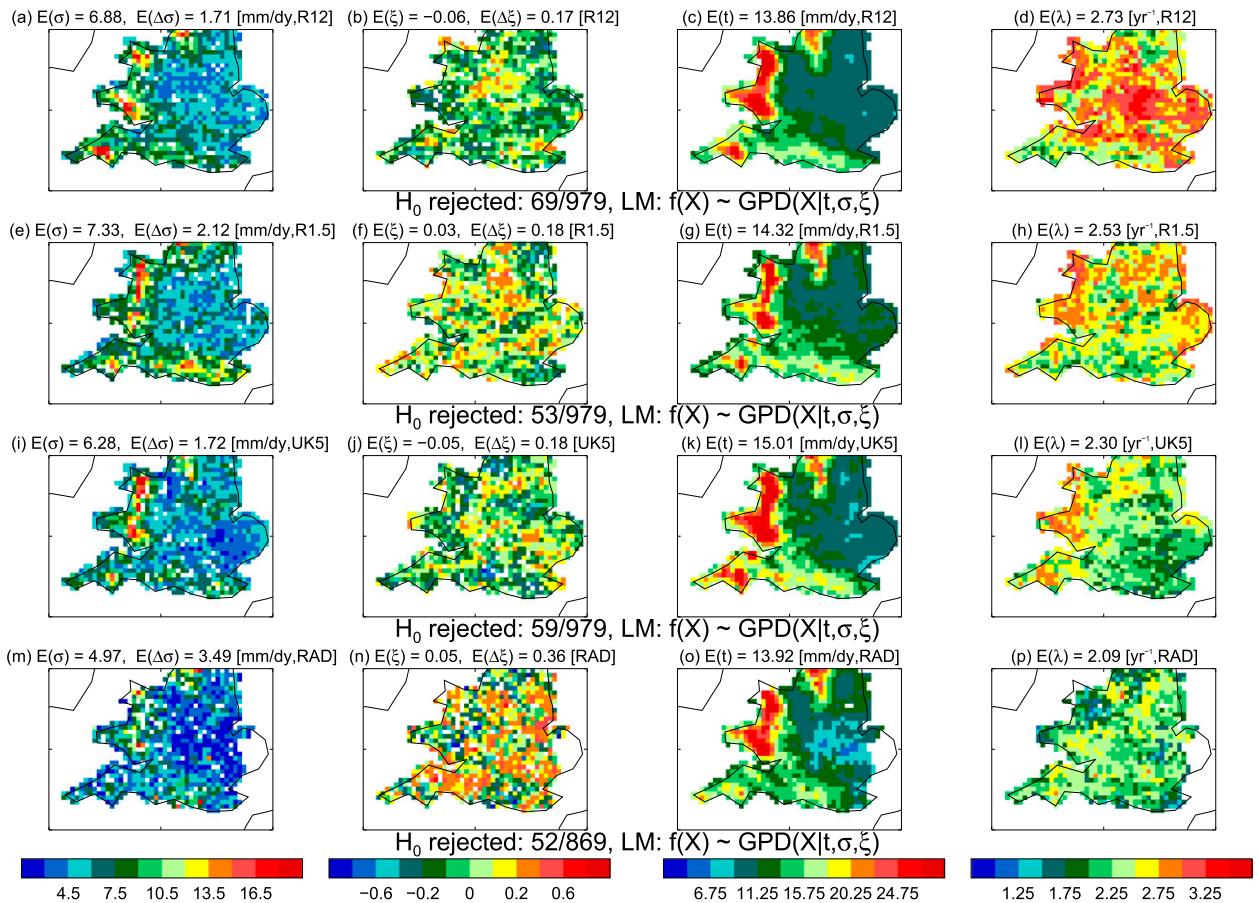


FIG. 5. As in Fig. 4, but for DJF 1-day precipitation accumulations.

positive biases in the shape parameter, the 12-km RCM return levels are growing increasingly fast as log-return periods ( $\log_{10}n$ ) increase, and eventually surpasses other datasets at a long enough return period. Near-zero shape parameters for the observations and the 1.5-km

RCM give quasi-linear increases of return levels with  $\log_{10}n$ . The 1.5-km RCM positive biases in threshold and the scale parameter mean its return levels are shifted upward and have a quasi-linear slope somewhat steeper than the radar.

TABLE 1. JJA and DJF 25th and 75th quartiles of PoT parameters for the ensemble of analyzed Met Office hourly gauges for various accumulation periods: parentheses indicate negative values. All precipitation valued figures are quoted in millimeters per its accumulation period [i.e., 12-h accumulation figures will be presented as  $\text{mm} (12 \text{ h})^{-1}$ ].

	Met Office hourly gauges accumulation period					
	1 h	3 h	6 h	12 h	1 day	5 days
JJA						
$\sigma$ (mm)	2.24–3.18	3.87–4.79	4.67–6.54	5.84–7.96	6.05–9.94	7.45–10.52
$\xi$	0.03–0.14	(0.03)–0.16	(0.07)–0.16	(0.09)–0.18	(0.08)–0.18	0.02–0.36
$t$ (mm)	3.2–3.6	7.2–8.0	10.6–12.2	14.8–16.8	18.8–22.2	29.7–34.4
$\lambda$ (yr <sup>-1</sup> )	4.11–5.26	2.95–3.42	2.21–2.42	1.58–1.84	1.16–1.47	0.79–1.00
DJF						
$\sigma$ (mm)	1.45–1.80	2.42–2.92	3.62–4.37	3.96–5.40	5.07–6.78	5.48–9.20
$\xi$	(0.23)–(0.11)	(0.19)–(0.02)	(0.22)–(0.02)	(0.14)–0.04	(0.16)–0.06	(0.19)–0.03
$t$ (mm)	2.0–2.6	5.0–6.4	7.8–9.8	10.6–13.8	13.4–18.5	21.8–35.3
$\lambda$ (yr <sup>-1</sup> )	5.53–7.05	4.00–5.05	3.00–3.79	2.32–3.08	1.89–2.26	1.17–1.41

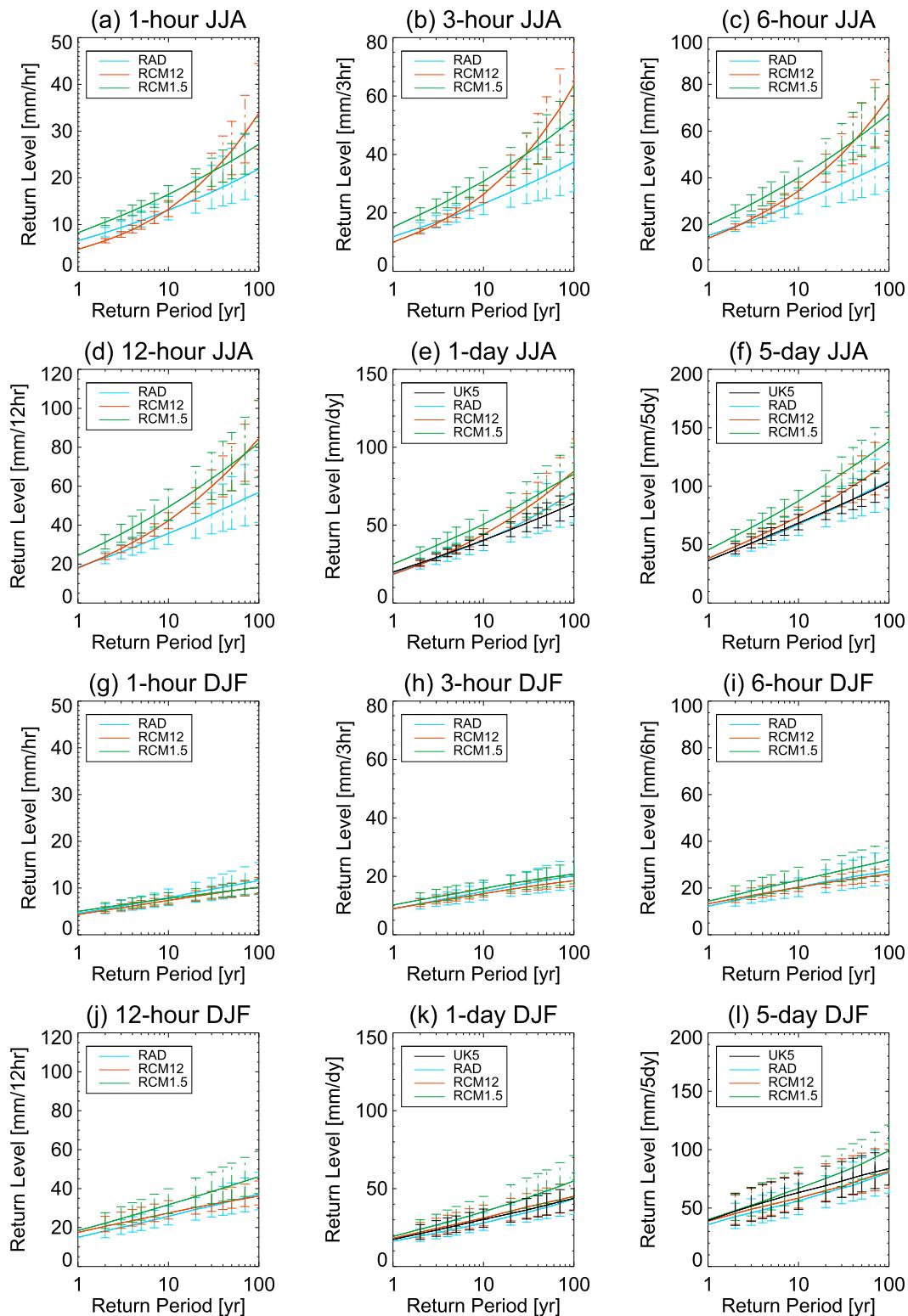


FIG. 6. PoT-estimated return levels  $[z(n)]$  as a function of return period  $(n)$  for the 12-km RCM (orange), 1.5-km RCM (green), daily gauges (black), and radar (blue) plotted on linear-log graphs. Return periods are computed at each grid box, and the spatial median (central value) and interquartile range (error bars) are shown. The different panels represent different accumulations (1, 3, 6, and 12 h; 1 and 5 days) and seasons: (a)–(f) JJA and (g)–(l) DJF.

Some of the differences in the JJA model return levels are not apparent in DJF. For this season (DJF), the 1.5-km RCM has consistently higher return levels than the 12-km RCM and the gridded observations for accumulation periods longer than 6 h. Unlike JJA, the 12-km RCM return levels no longer surpass the 1.5-km RCM at long return periods. Since the DJF shape parameters and thresholds for 6-h+ accumulations are comparable across the three datasets, the higher 1.5-km RCM  $z(100)$  come from the higher scale parameters and thresholds.

Despite possible large radar measurement errors, the return levels estimated by the radar data are comparable with UK5 for 1-day accumulations. Obviously, this agreement cannot be verified for subdaily precipitation. Even though individual estimates of radar data may be unreliable, useful information can be extracted as a spatial average. This supports the idea that radar data are usable and useful if spatial averaging and pooling are used as in [Hanel and Buishand \(2010\)](#) and [Overeem et al. \(2009\)](#).

### c. Growth curves

The growth curves are now examined, allowing the inclusion of the point gauge estimates with the gridded data estimates. The JJA and DJF growth curves for different accumulation periods are shown in [Figs. 7 and 8](#).

The JJA growth curves are steeper than DJF across all observational and model estimates. The hourly-gauge-estimated 100-yr JJA growth rates are about 2.0–3.0, in agreement with the radar data. Daily accumulations show a similar agreement between the gauges and UK5 estimates. The lower DJF estimates ( $\sim 1.5$ ) are also consistent between radar and hourly gauge estimates, that is, despite the radar data having a much shorter record length. For short (1 h) accumulations, the radar estimates are higher than the gauge estimates, and both RCM estimates are in between the two. Results here give further support to the use of radar data for the understanding of observed subdaily extremes as in [Hanel and Buishand \(2010\)](#).

Radar northwest – southeast differences (not shown) suggest growth curves are steeper in the southeast than in the northwest part of the domain for 1-h to 1-day totals, for both JJA and DJF (see [Faulkner 1999](#)). The same is suggested in the model simulations; however, it is not evident in the hourly gauge observations. Gridded gauge observations (see [Figs. 4 and 5](#)) hint that some parts of orography are associated with lower shape parameters (and hence lower growth rates); however, the gauges tend to be placed in valleys, and may undersample the orographic maximums.

The 12-km RCM JJA growth curve is higher than the observations and the 1.5-km RCM, and the differences

here are consistent with the fitted shape parameter differences. For the 12-km RCM,  $z(100)$  is nearly 4–6 times higher than the  $z(2)$ . The differences between the 12-km RCM and the other datasets are gradually reduced as accumulation periods lengthen, and the differences are not evident for 5-day accumulations.

Changing the accumulation period has little impact on both JJA and DJF growth curves for all compared datasets except for the 12-km RCM. This is illustrated in [Fig. 9](#), where  $G(100)$  is shown as a function of accumulation period. For the observations, this is a remarkable result because it suggests that the normalized probability distribution of extreme precipitation totals is invariant with accumulation period, and this fact is well captured by the 1.5-km RCM. However, this is not true for the 12-km RCM, where the JJA extreme probability distribution is sensitive to the accumulation period.

The above problems are not evident in DJF when 1-h precipitation growth rates are well simulated by both 12- and 1.5-km RCMs. This is further corroborated by the insensitivity of the shape parameter to accumulation period in DJF ([Figs. 3 and 5](#)).

The FEH ([Faulkner 1999](#)) shows higher 100-yr growth rates for 1-h extremes than the 1-day extremes, in contradiction to our results above. However, an invariance with accumulation period is found in the analysis by [Madsen et al. \(2009\)](#). The reason behind the differences is not clear—possibly related to differences in analysis strategy, or the length of data analyzed. The cause of these differences would require further investigation.

## 5. The physical behavior of the model extremes

In the previous section, we have shown that the 12-km RCM tends to produce too intense long-return period extremes in summer. Here we seek an explanation by examining why extreme events appear too large within the 12-km RCM. It is not possible to examine each single climate model-simulated extreme event. For case comparisons, we focus only on the very tail of the model extremes. We will take advantage of the fact that the 12-km RCM data are used to drive the 1.5-km RCM, which means some events are common.

An examination of the largest events in both simulations show two types of events:

- “stationary” grid-point storms in the 12-km RCM, and
- mesoscale convective systems (MCS).

Raw data checks indicate that 7 out of 10 of the largest 12-km RCM events are localized gridpoint storms (not shown). Highly localized events with high hourly totals exceeding  $\sim 100 \text{ mm h}^{-1}$  over a few hours are not unprecedented in the United Kingdom ([Burt 2005](#); [Golding](#)

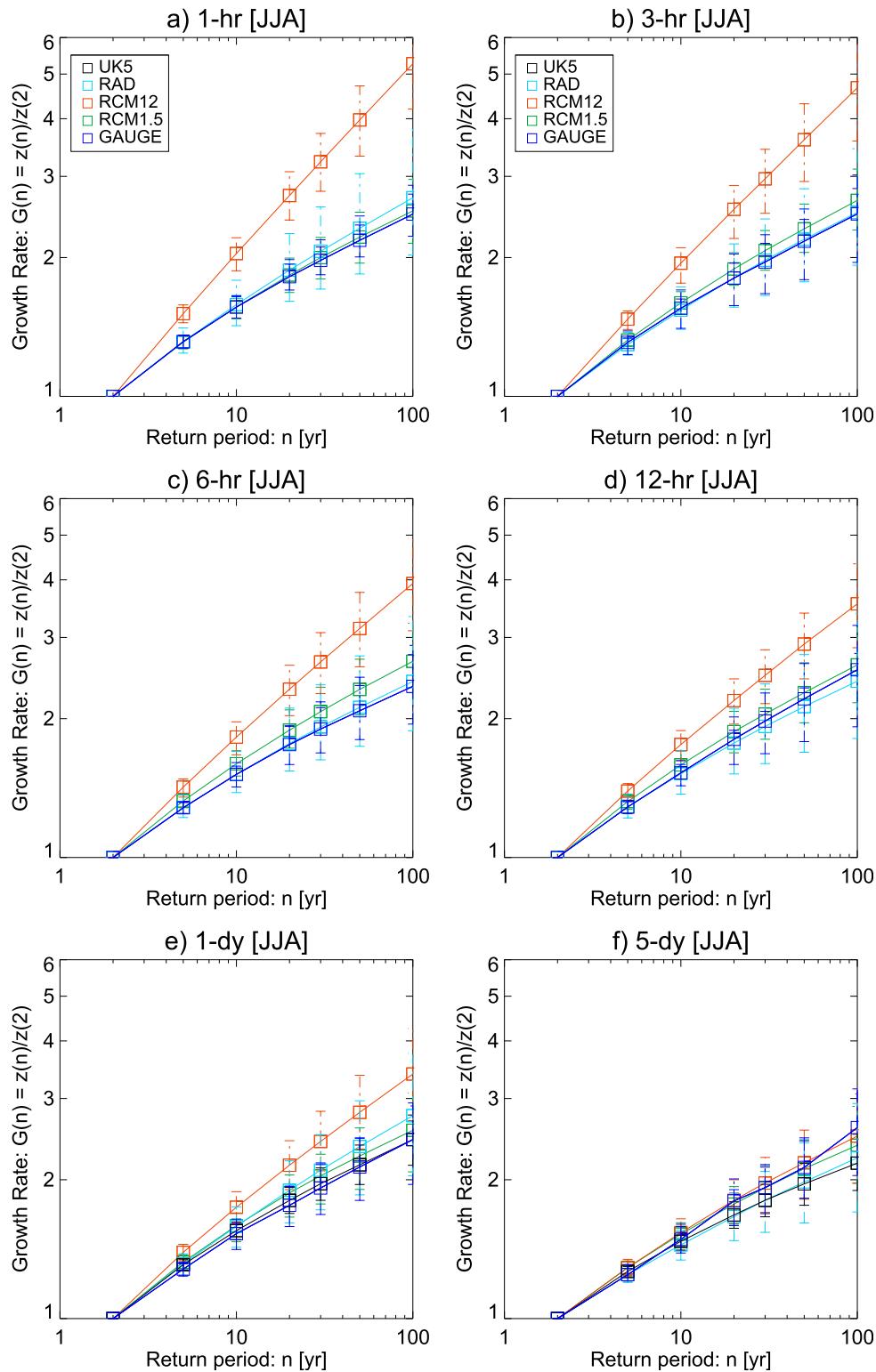


FIG. 7. JJA PoT-estimated growth curves [ $y$  axis:  $G(n) = z(n)/z(2)$ ] as a function of return period [ $x$  axis:  $n$  (yr)]. Different accumulation periods are shown in different panels. Line color indicates the different dataset: orange for 1990–2008 12-km RCM, green for 1990–2008 1.5-km RCM, black for 1990–2008 gridded daily gauges, dark blue for 1992–2010 station hourly gauges, and light blue for 2003–10 radar. The central value and error bars indicate the spatial median and interquartile range.

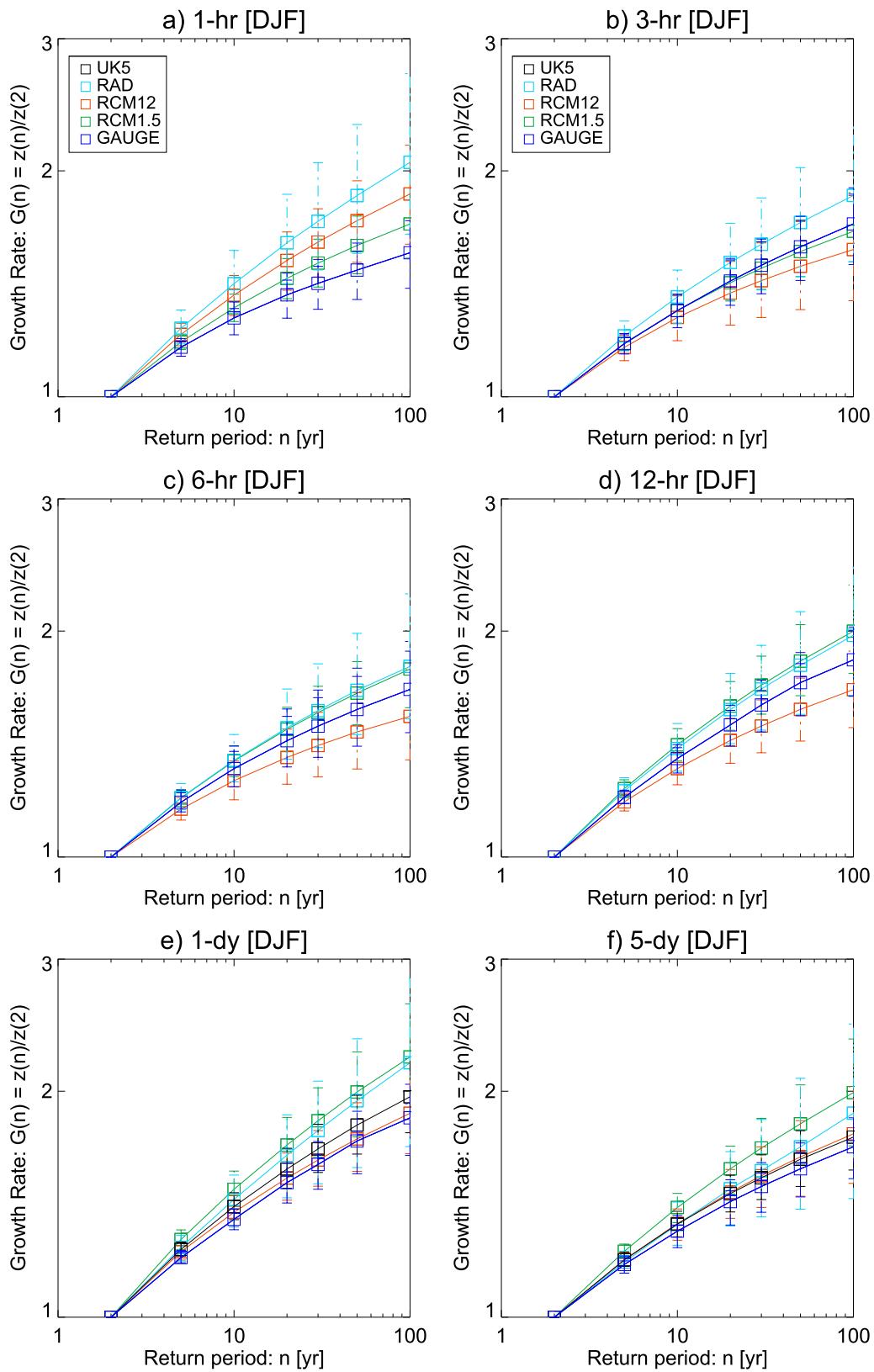


FIG. 8. As in Fig. 7, but for DJF. Note: the y-axis scale has been reduced.

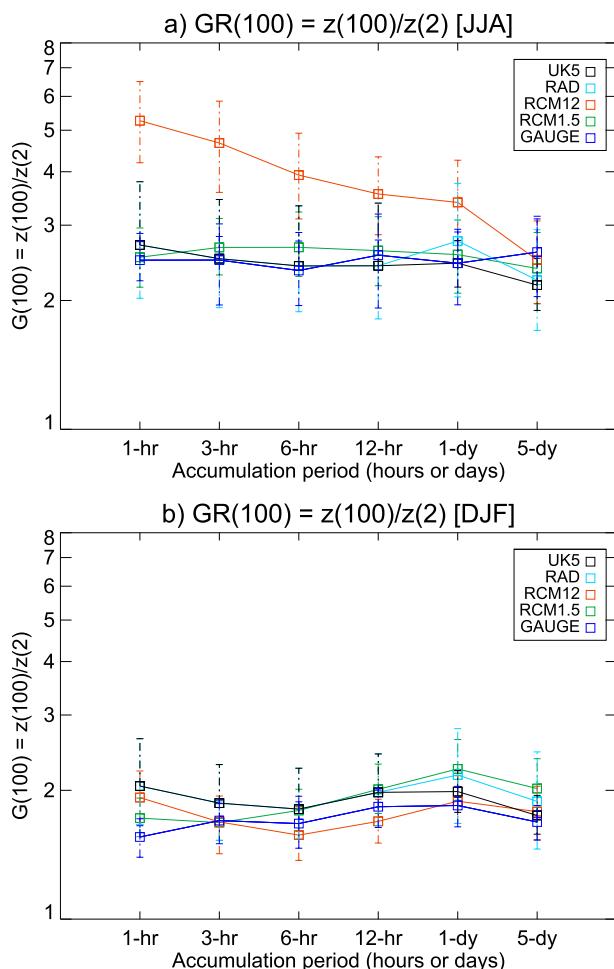


FIG. 9. Variation of the  $G(100)$  as a function of accumulation period for (a) JJA and (b) DJF. Different lines represent different datasets (see legend), and the central values (error bars) are the spatial median (25th and 75th spatial quartiles).

et al. 2005). Such events are not “common,” and typically have areas less than the grid box area of the 12-km RCM. The single largest 1-h accumulation within the 12-km RCM has peak intensity of  $93.3 \text{ mm h}^{-1}$ . Figure 10 shows the daily totals for this event, and its partitioning between CP parameterized and explicitly resolved precipitation. This localized event maintains  $50+ \text{ mm h}^{-1}$  intensity for 3 h (not shown) and has a high “large-scale” fraction within an area in which most of the rain comes from the CP.

Using the 12-km RCM estimates above, this event would have been a 1000-yr event for 1-h totals, which sounds plausible enough. However, if this event were to take place in the 1.5-km RCM or radar, it would correspond to a 1 000 000-yr event. The 5-h totals from Boscastle were 75–140 mm spanning over approximately a  $50 \text{ km}^2$  area, and localized  $200+ \text{ mm h}^{-1}$  instantaneous

rates were measured (Golding et al. 2005). Such spatial scales are smaller than the 12-km RCM is expected to resolve. The 12-km RCM event has a 5-h total of 280+ mm (not shown) in a single grid box. A Boscastle-like event in the 12-km RCM would have considerably lower totals after spatial averaging.

While exact correspondences between the 12- and 1.5-km RCMs are not expected, the large-scale conditions of both are similar. Examination of the 1.5-km RCM (not shown) indicates that there is precipitation for this event over approximately the same area at much lower intensity (up to  $\sim 16 \text{ mm h}^{-1}$ ).

The maximum hourly intensity of this event in the 12-km RCM is 1.8 times the largest hourly intensity ( $\sim 52 \text{ mm h}^{-1}$ ; averaged to a 12-km grid box) occurring at any time in the 1.5-km RCM (not shown). The 12-km RCM cannot be expected to have the correct dynamical structure because the storm is underresolved. The existence of both explicit and parameterized precipitation in the same area is a “red flag” (Molinari and Dudek 1992). Such a large amount of explicit or “large-scale” precipitation at a single grid point indicates gridpoint saturation. CP (including Gregory and Rowntree 1990) assumes that convective clouds have areas much smaller than the grid box, and the convective equilibrium state is restored at the subgrid level. These equilibrium assumptions are invalid for “big” convection, and CP alone cannot restore the equilibrium.

Such gridpoint events have detrimental impacts on the probability distribution of extremes for the 12-km RCM. Apart from tail fattening, it explains the 12-km shape parameter dependence on accumulation period, as such gridpoint storms are less important when accumulation period lengthens. The largest 1.5-km RCM events do not appear as such gridpoint events, but are more like traveling MCSs.

All of the examined MCS-like events are captured by both models. An example of such an event is shown in Figs. 11 and 12. The event corresponds to the single largest and fifth largest 1-h intensity for the 1.5- and the 12-km RCM, respectively. The 1.5-km RCM representation of this event has much more moderate precipitation than the 12-km RCM. The 12-km RCM peak hourly intensity is about 1.6 times higher than the 1.5-km RCM peak hourly intensity.

Unlike the gridpoint storms, this event appears more realistic from a meteorological perspective—a traveling system with temporal continuity that undergoes life cycle changes. The large hourly accumulations ( $\sim 70 \text{ mm h}^{-1}$ ) in the 12-km RCM appear near the core of the simulated MCS and are larger than the single largest intensity for the 1.5-km RCM. The core is likely to be underresolved by the 12-km grid boxes (even if the MCS is large), which

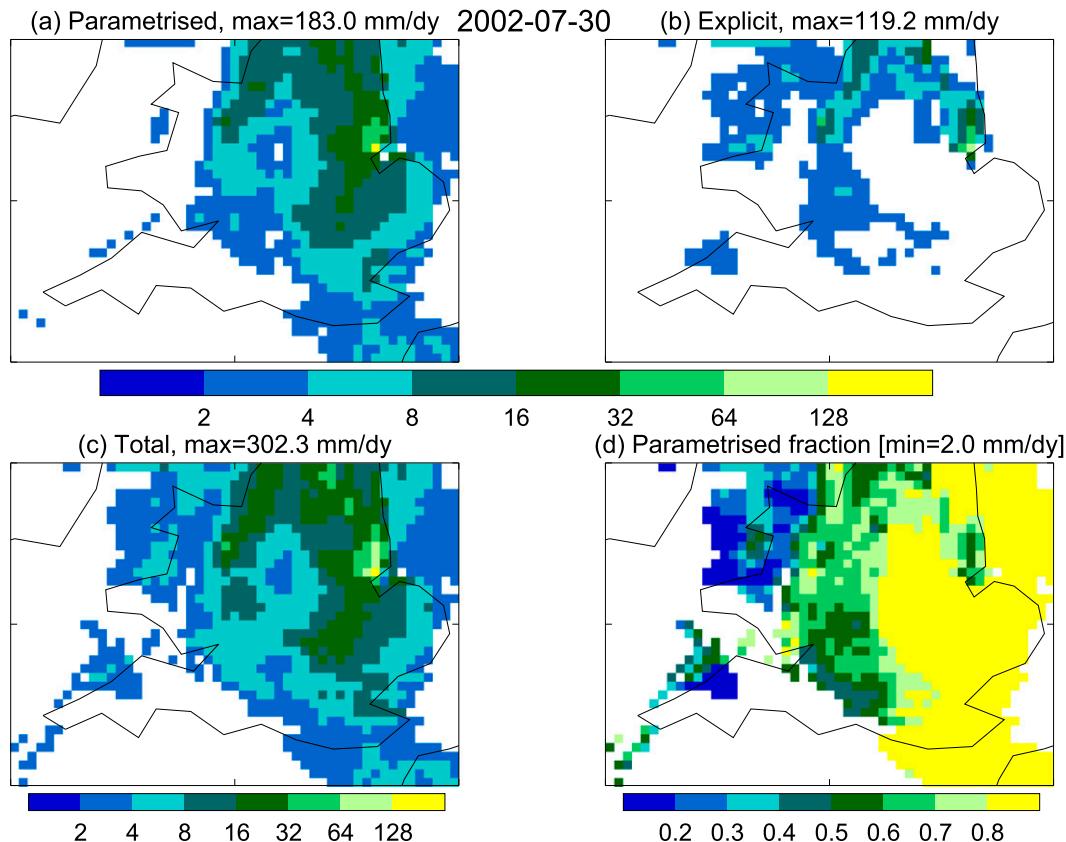


FIG. 10. (a) CP parameterized (daily total), (b) explicit large-scale (daily total), (c) total of (a) and (b), and (d) parameterized fraction of a "gridpoint storm" (model day 30 Jul 2002) event in the 12-km RCM. The maximum simulated grid box hourly (parameterized plus explicit) intensity is  $93.3 \text{ mm h}^{-1}$ .

means that there is still aliasing of the 12-km-model large-scale precipitation.

## 6. Discussion and conclusions

We seek to understand observed and model-simulated southern U.K. precipitation extremes across a range of accumulation periods, and ask whether the two selected UKMO high-resolution models (12- and 1.5-km RCMs) are able to realistically simulate precipitation extremes. We also want to know whether there is added value in the convective-permitting 1.5-km RCM in the simulation of extremes. Our answer appears to be yes to the second question—the 1.5-km convective-permitting RCM has more realistic very large extremes than the 12-km RCM, despite the 1.5-km RCM having a positive bias overall. The JJA 1.5-km RCM 1-h return level biases against the radar are approximately 25%–30%, which are generally in line with the intensity biases found in Kendon et al. (2012).

We have shown evidence that JJA extremes in the 1.5-km RCM are not only different, but have a much

higher degree of realism than those in the 12-km RCM. The high growth rates in the 12-km RCM at the subdaily time scale are closely related to high shape parameters in the peaks-over-threshold estimates to extreme intensities. The high shape parameters lead to a heavy tail in the 12-km RCM extremes and rapid non-log-linear return level increases, and are at least partially caused by events that appear meteorologically and statistically implausible. When accumulation periods are extended toward 1 day or longer, the 12-km RCM simulated JJA extremes generally improve as the impact from subdaily grid point storms is reduced. By 5-day accumulations, the JJA growth curve differences become negligible between the 12- and 1.5-km RCM.

Although the 1.5-km RCM JJA GP shape parameters are in line with observations, its thresholds and scale parameters are too high. This reflects the tendency for the 1.5-km RCM precipitation intensity to be positively biased (Kendon et al. 2012) and return level curves to be shifted upward with steeper linear slopes. This is likely to be partly a consequence of convection still being underresolved at 1.5-km grid spacing. For long return

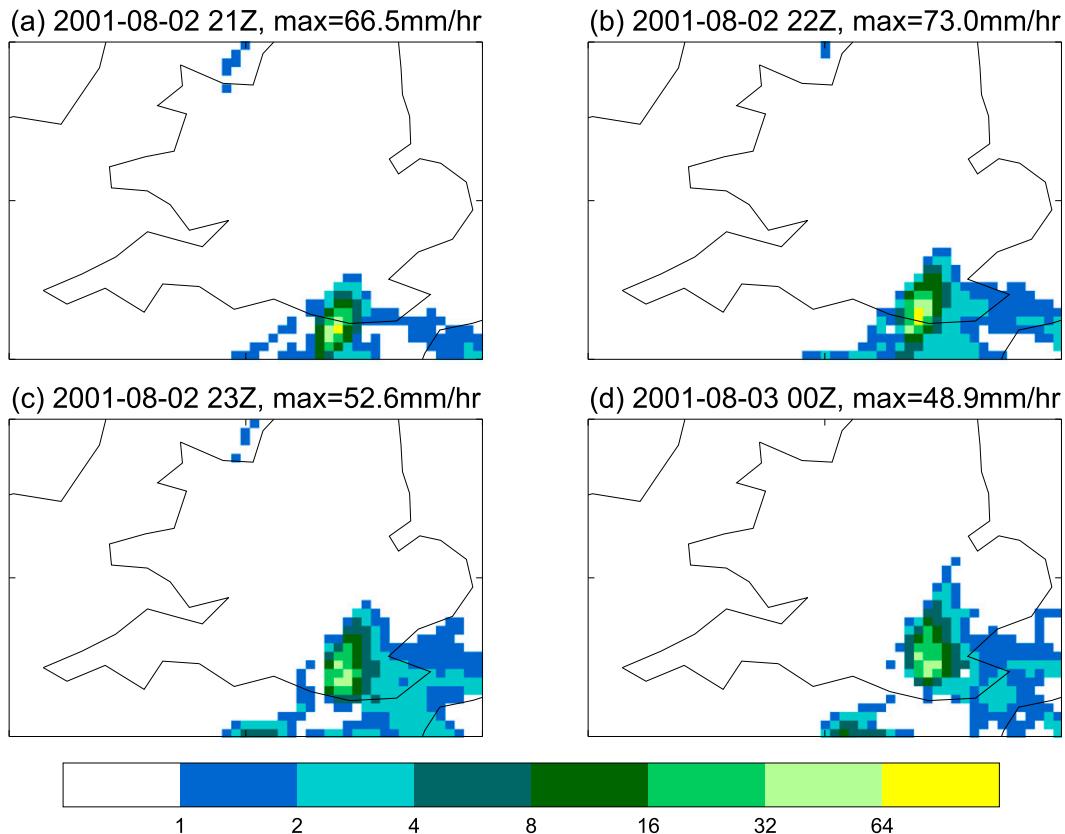


FIG. 11. Hourly progression of a mesoscale convective system (MCS) extreme event within the 12-km RCM.

periods, good tail representation and the curvature of return levels becomes more important than lower mean and linear slope biases.

DJF extremes are much better simulated by both models. Observed hourly and daily DJF extremes and their growth rates are lower than the JJA growth rates across a range of accumulation periods, and this is well captured by both simulations. The 12-km RCM appears sufficient to simulate DJF extremes, and the advantage of 1.5-km RCM for DJF is low; yet this is overshadowed by the fact that U.K. JJA extremes are more intense.

Growth curves allow direct comparisons between station and gridded data. That is because growth curves are independent of areal reduction factors and mean biases. Our results indicate that the observed JJA growth rates are insensitive to accumulation periods that are 1 h or longer in contrast with the FEH estimates, which give higher growth rates for 1-h extremes than 1-day extremes (Faulkner 1999). The hourly extreme estimates in FEH are less robust than the daily extreme estimates. The number of hourly precipitation measurements by station years (stations  $\times$  years) is about  $1/50$  of the number of daily precipitation measurements (Faulkner 1999).

The growth curve differences may be caused by an underestimation of weaker extremes or an overestimation of extremes with longer return periods. With the exception of 1-h totals, the 12-km RCM simulates the RMED [ $z(2)$ ] reasonably well across a range of accumulation periods. This is in contrast with the 1.5-km RCM, which has higher RMED levels than the observations. The high JJA goodness of fit indicates the suitability of PoT for our analysis. We conclude that the growth curve differences are a reflection of the deficiencies of the 12-km RCM in simulating the higher JJA extremes.

Climate model biases can be caused by the symmetric (scale parameter) and asymmetric (shape parameter) stretching, or the shifting of the probability distributions. The biases in the probability distribution of simulated extremes and nonextremes may also differ. As discussed in Kendon et al. (2012) and Chan et al. (2013), the 1.5-km RCM produces daily/hourly precipitation intensities that are higher than observed. Based on the results here, these positive biases appear to be linked to the shifts in the mean—and are likely to be related to the 1.5-km RCM underresolving convection (Kendon et al. 2012). The biases in the extremes within the 12-km RCM have a different character, with skewness being different

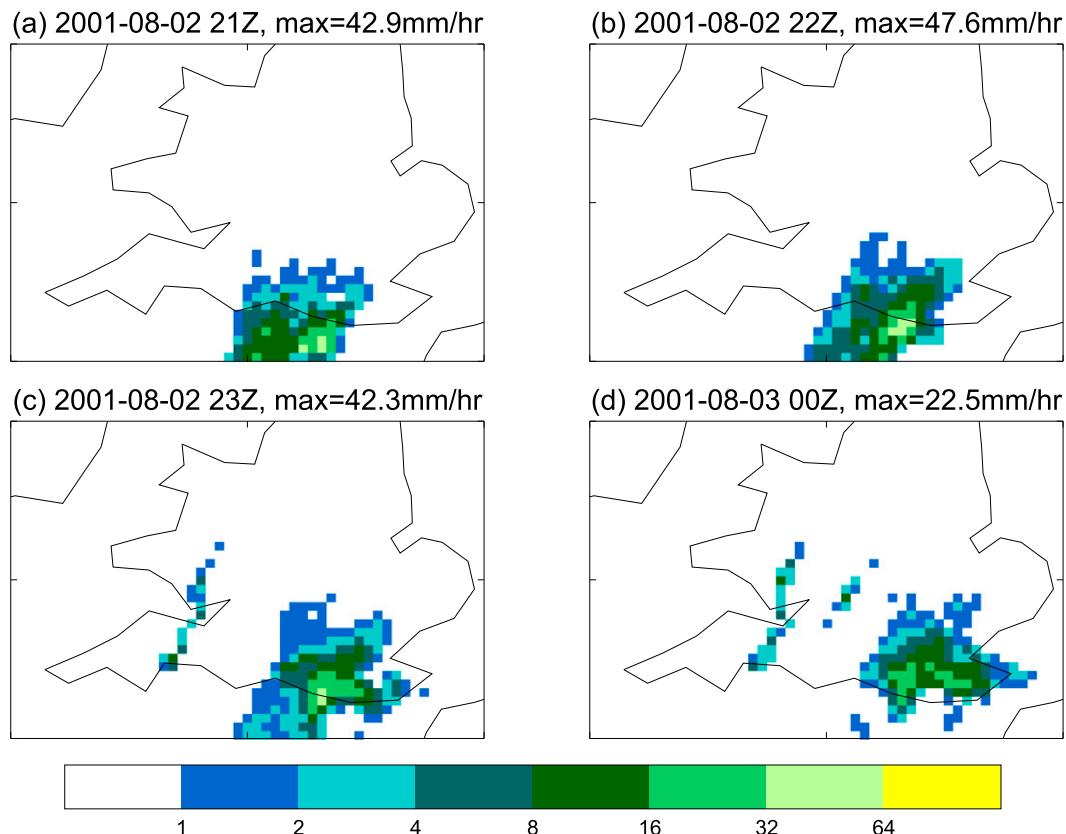


FIG. 12. As in Fig. 11, but for the 1.5-km RCM simulation of the same event.

compared to the 1.5-km RCM and observations; this skewness bias can be linked to 12-km RCM very different and unphysical representation of convection when explicitly resolved convection becomes as vigorous as parameterized convection.

Radar-estimated precipitation intensities are often considered to be of lower quality than gauge data, and its flaws are well documented (Harrison et al. 2000). Indeed, some results here appear to be affected by questionable radar-estimated precipitation. In particular, there are spurious large radar values near the Wash and Thames Estuary. More generally over the SUK, we find that radar-estimated precipitation gives good estimates of growth curves and return levels that are comparable with station estimates. Our results indicate that useful information can be derived from the use of radar-estimated precipitation. However, we note that due to radar attenuation and gauge undersampling of local maxima (which particularly affects their representation of convective storms and orographic precipitation), both datasets tend to underestimate high-intensity precipitation. This suggests that the 1.5-km RCM constant positive return-level biases are overstated.

Our results show positive benefits of the explicit convection 1.5-km simulations for summer. Even though the

mean biases may appear higher (Chan et al. 2013), the 1.5-km convective-permitting simulation is more physically realistic than the 12-km RCM. The 1.5-km RCM is able to explicitly represent the dynamical structure and life cycle of convective storms (albeit not perfectly). The lower mean biases for the 12-km RCM are in part caused by compensation between too much light precipitation and excessively high extreme intensities.

In the case for the 12-km RCM, one would expect the precipitation intensity to be sensitive to the convective parameterization and its internal parameters. The shape parameter is known to be sensitive to the model's CP entrainment coefficient (Fowler et al. 2010). However, caution should be exercised in tuning the CP parameters to improve the simulation of extremes, as many CPs are simply not designed for 12-km grid spacing (Arakawa 2004).

Gridpoint storms, which lead to erroneous extremes in the 12-km RCM, are a consequence of the fact that CP assumptions are not valid for 12-km grid boxes. The 12-km RCM operates within the "gray zone" horizontal grid spacing [a 5–50-km range is quoted by Molinari and Dudek (1992)]—the resolution is insufficient to resolve all convective spatial scales. CP is needed to represent

unresolved convection, but the CP equilibrium principle is no longer valid. CP introduces artificial and unnatural separation of cloud processes and scales (Arakawa 2004). Convection cannot be properly represented either explicitly or parameterized, and this can lead to undesired interaction and creates an arbitrary partition between what is deemed “resolved” and what is not. We note that, while the gray-zone problem does not disappear as grid spacing approaches the kilometer scale, the need for CP becomes less acute. Some state-of-the-art gray-zone-resolution dynamical models are now addressing problematic assumptions in older-generation CPs (Gerard et al. 2009), and problems with gray-zone convection may be somewhat alleviated in the future. Williamson (2013) has found that runaway storms can be contained by having a CP closure time scale comparable to the model time step. While it is desirable to remove gridpoint storms from the 12-km model and therefore reduce the skewness bias in the probability distribution, such improvements will not provide a model capable of representing short-duration extremes as CPs are not designed to have that capability.

As part of the CONVEX program, future-climate GCM-driven 12- and 1.5-km RCM simulations are being carried out by the UKMO. Growth curve analysis will be carried out with these simulations to understand future changes of extreme precipitation, particularly for sub-daily extremes.

From the stakeholder perspective, realistic “very extreme” events (with long return periods) are far more important than “less extreme” events. The unrealistic 12-km RCM very extreme events are a major concern for its use in providing extreme precipitation guidance. While this paper has touched on the meteorology of extremes, the triggers of model extremes have not been examined. A full understanding of the unusual model-simulated meteorological events will require non-supervised detection and categorizing of extremes and their convective environment, and existing tools from mesoscale meteorology and cloud dynamics may be employed for such work (Tsakraklides and Evans 2003; Baldwin et al. 2005; Davis et al. 2006; Roberts and Lean 2008; Hanley et al. 2014).

*Acknowledgments.* This research is part of the CONVEX project—a collaboration between Newcastle University, the UKMO, and the University of Exeter. CONVEX is supported by the United Kingdom NERC CWC programme (Grant NE/I006680/1), and the presented model simulations are supported by the UKMO. The lead author is financially supported by Newcastle University, and is a visiting scientist at the UKMO Hadley Centre in Exeter, United Kingdom. We thank

Simon Brown of UKMO for his scientific recommendations and Mathieu Ribatet of the University of Montpellier for his R technical assistance. We also thank the anonymous reviewers for their valuable comments that have greatly improved this paper. Large portions of the analysis were carried out with the FOSS software R.

## REFERENCES

- Acerro, F. J., J. A. García, and M. C. Gallego, 2011: Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *J. Climate*, **24**, 1089–1105, doi:10.1175/2010JCLI3627.1.
- Anderson, T. W., and D. A. Darling, 1952: Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Stat.*, **23**, 193–212, doi:10.1214/aoms/1177729437.
- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, doi:10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.
- Atyeo, J., and D. Walshaw, 2012: A region-based hierarchical model for extreme rainfall over the UK, incorporating spatial dependence and temporal trend. *Environmetrics*, **23**, 509–521, doi:10.1002/env.2155.
- Baldwin, M. E., J. S. Kain, and S. Lakshmiarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.*, **133**, 844–862, doi:10.1175/MWR2892.1.
- Blackburn, M., J. Methven, and N. Roberts, 2008: Large-scale context for the UK floods in summer 2007. *Weather*, **63**, 280–288, doi:10.1002/wea.322.
- Brockhaus, P., D. Lüthi, and C. Schär, 2008: Aspects of the diurnal cycle in a regional climate model. *Meteor. Z.*, **17**, 433–443, doi:10.1127/0941-2948/2008/0316.
- Burt, S., 2005: Cloudburst upon Hendraburnick Down: The Boscastle storm of 16 August 2004. *Weather*, **60**, 219–227, doi:10.1256/wea.26.05.
- Chan, S. C., E. J. Kendon, H. J. Fowler, S. Blenkinsop, C. A. T. Ferro, and D. B. Stephenson, 2013: Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? *Climate Dyn.*, **41**, 1475–1495, doi:10.1007/s00382-012-1568-9.
- Coelho, C. A. S., C. A. T. Ferro, D. B. Stephenson, and D. J. Steinskog, 2008: Methods for exploring spatial and temporal variability of extreme events in climate data. *J. Climate*, **21**, 2072–2092, doi:10.1175/2007JCLI1781.1.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp.
- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, doi:10.1175/MWR3145.1.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- de Michele, C., N. T. Kottegoda, and R. Rosso, 2001: The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resour. Res.*, **37**, 3247–3252, doi:10.1029/2001WR000346.
- Faulkner, D., 1999: *Rainfall Frequency Estimation*. Vol. 2, *Flood Estimation Handbook*, NERC Centre for Ecology and Hydrology, 110 pp.

- Ferro, C. A. T., and J. Segers, 2003: Inference for clusters of extreme values. *J. Roy. Stat. Soc.*, **65**, 545–556, doi:10.1111/1467-9868.00401.
- Fowler, H. J., and C. G. Kilsby, 2003a: Implications of changes in seasonal and annual extreme rainfall. *Geophys. Res. Lett.*, **30**, 1720, doi:10.1029/2003GL017327.
- , and —, 2003b: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int. J. Climatol.*, **23**, 1313–1334, doi:10.1002/joc.943.
- , and M. Ekström, 2009: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *Int. J. Climatol.*, **29**, 385–416, doi:10.1002/joc.1827.
- , 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**, 241–267, doi:10.1007/s10687-010-0101-y.
- Gerard, L., J.-M. Piriou, R. Brozkova, J.-F. Geleyn, and D. Banciu, 2009: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model. *Mon. Wea. Rev.*, **137**, 3960–3977, doi:10.1175/2009MWR2750.1.
- Golding, B. W., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteor. Appl.*, **5**, 1–16, doi:10.1017/S1350482798000577.
- , P. Clark, and B. May, 2005: The Boscastle flood: Meteorological analysis of the conditions leading to flooding on 16 August 2004. *Weather*, **60**, 230–235, doi:10.1256/wea.71.05.
- Gregersen, I. B., H. J. D. Sorup, H. Madsen, D. Rosbjerg, P. S. Mikkelsen, and K. Arnbjerg-Nielsen, 2013: Assessing future climatic changes of rainfall extremes at small spatio-temporal scales. *Climatic Change*, **118**, 783–797, doi:10.1007/s10584-012-0669-0.
- Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483–1506, doi:10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2.
- Hanel, M., and T. A. Buishand, 2010: On the value of hourly precipitation extremes in regional climate model simulations. *J. Hydrol.*, **393**, 265–273, doi:10.1016/j.jhydrol.2010.08.024.
- , —, and C. A. T. Ferro, 2009: A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *J. Geophys. Res.*, **114**, D15107, doi:10.1029/2009JD011712.
- Hanley, K. E., R. S. Plant, T. H. M. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2014: Mixing length controls on high resolution simulations of convective storms. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2356, in press.
- Harrison, D. L., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteor. Appl.*, **7**, 135–144, doi:10.1017/S1350482700001468.
- Hewitt, C. D., and D. J. Griggs, 2004: ENSEMBLES-based predictions of climate changes and their impacts. *Eos, Trans. Amer. Geophys. Union*, **85**, 566, doi:10.1029/2004EO520005.
- Hohenegger, C., P. Brockhaus, and C. Schär, 2008: Towards climate simulations at cloud-resolving scales. *Meteor. Z.*, **17**, 383–394, doi:10.1127/0941-2948/2008/0303.
- Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.*, **52**, 105–124.
- , and J. R. Wallis, 1993: Some statistics useful in regional frequency analysis. *Water Resour. Res.*, **29**, 271–281, doi:10.1029/92WR01980.
- Jones, M. R., H. J. Fowler, C. G. Kilsby, and S. Blenkinsop, 2013: An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *Int. J. Climatol.*, **33**, 1178–1194, doi:10.1002/joc.3503.
- Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high resolution regional climate model. *J. Climate*, **25**, 5791–5806, doi:10.1175/JCLI-D-11-00562.1.
- Kjeldsen, T. R., 2007: The revitalised FSR/FEH rainfall-runoff method. Flood Estimation Handbook Supplementary Rep., Vol. 1, NERC Centre for Ecology and Hydrology, 68 pp.
- Kjellström, E., F. Boberg, M. Castro, J. H. Christensen, G. Nikulin, and E. Sánchez, 2010: Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Res.*, **44**, 135–150, doi:10.3354/cr00932.
- Laio, F., 2004: Cramer–von Mises and Anderson–Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour. Res.*, **40**, W09308, doi:10.1029/2004WR003204.
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, **136**, 3408–3424, doi:10.1175/2008MWR2332.1.
- Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127, doi:10.1002/joc.3370100202.
- Lilliefors, H. W., 1967: On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Amer. Stat. Assoc.*, **62**, 399–402, doi:10.1080/01621459.1967.10482916.
- Madsen, H., P. F. Rasmussen, and D. Rosbjerg, 1997: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water Resour. Res.*, **33**, 747–757, doi:10.1029/96WR03848.
- , K. Arnbjerg-Nielsen, and P. S. Mikkelsen, 2009: Update of regional intensity–duration–frequency curves in Denmark: Tendency towards increased storm intensities. *Atmos. Res.*, **92**, 343–349, doi:10.1016/j.atmosres.2009.01.013.
- Maraun, D., H. W. Rust, and T. J. Osborn, 2009: The annual cycle of heavy precipitation across the United Kingdom: A model based on extreme value statistics. *Int. J. Climatol.*, **29**, 1731–1744, doi:10.1002/joc.1811.
- , T. J. Osborn, and H. W. Rust, 2011: The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships. *Climate Dyn.*, **36**, 261–275, doi:10.1007/s00382-009-0710-9.
- Martins, E. S., and J. R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.*, **36**, 737–744, doi:10.1029/1999WR900330.
- , and —, 2001: Generalized maximum likelihood Pareto–Poisson estimators for partial duration series. *Water Resour. Res.*, **37**, 2551–2557, doi:10.1029/2001WR000367.
- Molinari, J., and M. Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.*, **120**, 326–344, doi:10.1175/1520-0493(1992)120<0326:POCPIM>2.0.CO;2.
- Molini, A., L. G. Lanza, and P. La Barbera, 2005: The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrol. Processes*, **19**, 1073–1088, doi:10.1002/hyp.5646.

- NERC, cited 2013: Changing water cycle. Natural Environment Research Council. [Available online at <http://www.bgs.ac.uk/changingwatercycle/>.]
- Overeem, A., T. A. Buishand, and I. Holleman, 2009: Extreme rainfall analysis and estimation of depth–duration–frequency curves using weather radar. *Water Resour. Res.*, **45**, W10424, doi:10.1029/2009WR007869.
- Perry, M., and D. Hollis, 2005: The generation of monthly gridded datasets for a range of climatic variables over the UK. *Int. J. Climatol.*, **25**, 1041–1054, doi:10.1002/joc.1161.
- , —, and M. Elms, 2009: The generation of daily gridded datasets of temperature and rainfall for the UK. Met Office National Climate Information Centre Climate Memo. **24**, 7 pp. [Available online at [http://www.metoffice.gov.uk/climatechange/science/downloads/generation\\_of\\_daily\\_gridded\\_datasets.pdf](http://www.metoffice.gov.uk/climatechange/science/downloads/generation_of_daily_gridded_datasets.pdf).]
- Re, M., and V. R. Barros, 2009: Extreme rainfalls in se South America. *Climatic Change*, **96**, 119–136, doi:10.1007/s10584-009-9619-x.
- Reed, D., 1999: *Overview*. Vol. I, *Flood Estimation Handbook*, NERC Centre for Ecology and Hydrology, 108 pp.
- Ribatet, M. A., 2006: A user's guide to the PoT package (version 1.4). University of Quebec, 31 pp.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/2007MWR2123.1.
- Stephens, M. A., 1977: Goodness of fit for the extreme value distribution. *Biometrika*, **64**, 583–588, doi:10.1093/biomet/64.3.583.
- Sunter, M., 2012: UK hourly rainfall data. Met Office Integrated Data Archive System (MIDAS) data. [Available online at [http://badc.nerc.ac.uk/view/badc.nerc.ac.uk\\_\\_ATOM\\_\\_dataent\\_ukmo-midas](http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ukmo-midas).]
- Sunyer, M. A., H. J. D. Sorup, O. B. Christensen, H. Madsen, D. Rosbjerg, P. S. Mikkelsen, and K. Arnbjerg-Nielsen, 2013: On the importance of observational data properties when assessing regional climate model performance of extreme precipitation. *Hydrol. Earth Syst. Sci.*, **10**, 7003–7043, doi:10.5194/hessd-10-7003-2013.
- 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.
- Tsakraklides, G., and J. L. Evans, 2003: Global and regional diurnal variations of organized convection. *J. Climate*, **16**, 1562–1572, doi:10.1175/1520-0442-16.10.1562.
- UKMO, cited 2013: Met Office surface data users guide. [Available online at [http://badc.nerc.ac.uk/data/ukmo-midas/ukmo\\_guide.html](http://badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html).]
- Waller, L. A., D. Smith, J. E. Childs, and L. A. Real, 2003: Monte Carlo assessments of goodness-of-fit for ecological simulation models. *Ecol. Modell.*, **164**, 49–63, doi:10.1016/S0304-3800(03)00011-5.
- Willems, P., 2000: Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *J. Hydrol.*, **233**, 189–205, doi:10.1016/S0022-1694(00)00233-X.
- Williamson, D. L., 2013: The effect of time steps and time-scales on parametrization suites. *Quart. J. Roy. Meteor. Soc.*, **139**, 548–560, doi:10.1002/qj.1992.