Climate Dynamics manuscript No.
(will be inserted by the editor)

Does increasing spatial resolution improve the simulation of United Kingdom daily precipitation in a regional climate model?

Steven C. Chan · Elizabeth J. Kendon · Hayley J. Fowler · Stephen Blenkinsop · Christopher A. T. Ferro · David B. Stephenson

Received: date / Accepted: date

Abstract Three different resolution (50, 12, and 1.5 km) regional climate model simulations are compared in terms of their ability to simulate moderate and high daily precipitation events over the southern United Kingdom. Among the three simulations, the convection-permitting 1.5-km simulation is carried out without convective parametrisation. As in previous studies, increasing resolution (especially from 50 km to 12 km) is found to improve the representation of orographic precipitation. The 50-km simulation underestimates mean precipitation over the mountainous region of Wales, and event intensity tends to be too weak; this bias is reduced in both the 12-km and 1.5-km simulations for both summer and winter. In south-east England lowlands where summer extremes are mostly convective, increasing resolution does not necessary lead to an improvement in the simulation. For the 12-km simulation, simulated daily extreme events are overly intense. Even though the average intensity of summer daily extremes is improved in the 1.5-km

This research is supported by United Kingdom NERC Changing Water Cycle programme grant NE/I006680/1

Steven C. Chan
School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK.
(*visiting scientist at the Met Office Hadley Centre)
Tel.: +44-1392-884802
Fax: +44-1392-885681
E-mail: steven.chan@metoffice.gov.uk

Elizabeth J. Kendon
Met Office Hadley Centre, Exeter, UK.

Hayley J. Fowler
School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK.

Stephen Blenkinsop
School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK.

Christopher A. T. Ferro
College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.

David B. Stephenson
College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK.
Met Office Hadley Centre, Exeter, UK.

simulation, the 1.5-km simulation has a poorer mean bias with too many events exceeding high thresholds. Spatial density and clustering of summer extremes in the south-east England are poorly simulated in both the 12-km and 1.5-km simulations. In general, we have not found any clear evidence to show that the 1.5-km simulation is superior to the 12-km simulation, or vice versa at the daily level.

Keywords High resolution models · Dynamical downscaling · Hydroclimate · Precipitation

1 Introduction

With ever increasing computing power, dynamical climate model simulations can be performed at unprecedented high resolutions. There are many apparent benefits to high resolutions - most notably in the better representation of coastlines and topography. However, many atmospheric processes remain unresolved and require parametrisation, for example convection and cloud systems (Arakawa, 2004). Parametrisation schemes, in particular the cumulus convection scheme (Molinari and Dudek, 1992; Hohenegger et al, 2008), are often designed for coarser resolutions and may become less valid at increasingly high resolution. Molinari and Dudek (1992) argue that assumptions for traditional convective schemes begin to break down at horizontal resolutions of about 50 km. Therefore, an improvement in the representation of atmospheric processes by solely increasing resolution is far from certain.

Despite the high importance of accurate precipitation projections (in terms of social and economic impact), precipitation is among the most challenging climate variables to model, as precipitation is dependent on the representation of a wide range of processes. Precipitation can be caused by local convective instability, forced ascent near elevation changes, and synoptic variability ("weather patterns"), and relies on a number of model parametrisation schemes (e.g. convection, land-surface, boundary-layer, and cloud micro-physics schemes) for its representation in dynamical models. The modelled behaviour of these processes is likely to respond differently to resolution changes.

The varied geography of Britain leads to significant precipitation variations in a relatively small area (Wigley et al, 1984). Due to the different precipitation processes (orographic, convective, and synoptic-scale depressions) that are involved, one may expect a range of model sensitivities to horizontal resolution across the UK. Thus the UK provides a good test platform for assessing the precipitation sensitivity to model resolution.

There have been many studies testing model sensitivity to resolution changes considering various atmospheric and hydroclimate fields, with changes in mean bias and variance often used as the metric to assess model skill. Giorgi and Marinucci (1996) suggest that changes in topographic representation with resolution are the main cause of model sensitivity. Antic et al (2006) further argue that such sensitivity tends to lead to an improvement in the simulated climate. However, Laprise et al (2008) shows that while downscaling to higher resolution does tend to increase spatial variations (i.e. high spatial resolution fine features), the spatial variance increases are sensitive to domain size and do not necessarily improve deterministic skill (skill in simulating specific events) even if the climate representation is improved (overall statistics from the accumulation of many events).

Increase of inter-annual variability with increasing horizontal resolution from 50 km to 25 km is noted in Rauscher et al (2010); the same study also finds that resolution increases lead to a reduction of the ratio of convective to total precipitation and an improvement in seasonal precipitation over topographically complex regions. For Britain, the 25-km simulations showed a marked improvement in the spatial patterns of JJA precipitation.

The above work tends to focus on regional simulations that are of relatively coarse resolution (25+km). Molinari and Dudek (1992) argue that some convective processes begin to be partially resolved by the model at resolutions of about $20-25$km. Over the UK, however, grid scales of $\approx$ 1km are needed to give a satisfactory representation of the majority of convection (Roberts and Lean, 2008), although even at this scale not all convection is fully resolved. Such 'convection-permitting' resolutions are now commonly used for short-term weather forecasting, and these have shown considerable improvements in the representation of convective and topographically enhanced precipitation (Roberts and Lean, 2008; Roberts et al, 2009). There are relatively few studies applying such high resolutions for longer-term climate simulations, and such studies tend to be limited to a small domain or a given season (Hohenegger et al, 2008; Wakazuki et al, 2008; Knote et al, 2010). In this study, we assess the first 17-year extended length climate simulation at convection permitting scales over a region of the UK.

High resolution climate models typically span a limited area, and are forced at the lateral boundaries by reanalyses or a coarser-resolution global climate model (GCM). The regional climate/mesoscale model (RCM) develops its own local climatology in the interior of the domain, conditional on these lateral boundary conditions (LBCs). In a one-way nesting approach, which is typically used, the RCM does not feed information back to the driving model, with the assumption that the regional model does not diverge strongly from the driving model in terms of its representation of the large-scale conditions. The higher resolution RCM only aims at adding information to what is not resolved by the driving model (Jones et al, 1995, 1997). An obvious alternative to the one-way nesting approach is the more computational expensive two-way nesting, in which the higher resolution - smaller domain model interacts with the lower resolution - larger domain model. Similar to two-way nesting, variable resolution stretched grid models can be employed, where the modeller uses higher horizontal resolution in regions of interest (Déqué and Piedelievre, 1995).

The question that we seek to answer in this paper is "Does increasing model resolution lead to a better representation of the character of intense precipitation events?" While changes of variance and means are important, changes in the frequency and intensity of precipitation extremes are of equal concern as well, due to their relevance to floods and droughts. Model representation of event frequency, intensity, and extremes is far more important than the mean in a social and economic context (Meehl et al, 2000). Low probability "tail" extreme events are rare such that their contributions to the climatological mean are comparatively small. However, they may contribute strongly to inter-annual variations and anomalies for specific years. The social impact of such events also depends on local geography, antecedent hydroclimate conditions (i.e. soil moisture and groundwater levels) and mitigation measures (i.e. flood and drought management). Furthermore, changes in the mean do not always reflect changes in such rare "tail" events (Allen and Ingram, 2002; Allan et al, 2010). Over the UK, regional climate models have been

shown to project increases in the magnitude of extreme rainfall events (Fowler et al, 2007; Fowler and Ekström, 2009). Characterising the sensitivity of extreme rainfall events to model resolution changes is thus important.

In this study, we seek to understand the sensitivity of simulated extreme rainfall to horizontal resolution using high resolution regional climate simulations carried out with the Met Office Unified Model. In particular, we compare 50-, 12-, and 1.5-km simulations. The 1.5-km simulation corresponds to the first extended climate simulation at convection-permitting scales over a region of the UK (southern UK, hereby SUK). We focus on precipitation events over SUK using the following metrics:

- Extreme event occurrences - in both space and time
- Intensities of such events
- Spatial organization and density (as in the expected number of events per grid box, see Appendix) of such events

We do not limit ourselves here to only rare events (such as daily precipitation heavier than $50+$ mm/day), and also consider events that are more "moderate" ($20+$ mm/day). This study complements Kendon et al (in press), which analysed the same set of simulations. In summary, Kendon et al (in press) have found that the 12-km RCM precipitation tends to have lighter, more widespread and persistent precipitation relative to the 1.5-km RCM. The 1.5-km RCM is also found to have a better diurnal cycle. Here we focus only on daily precipitation (which is in contrast with the hourly precipitation focus in Kendon et al (in press)). The analyses here uses different metrics, and are seasonally and regionally stratified. Data from an additional 50-km RCM is also included.

This paper is divided into eight sections. In section 2, we present an overview of the modelling system and observational data that we have used. We then present our analysis methodologies in section 3. In sections 4 to 7, we compare the differences between the model simulations and observations with the use of different metrics. Finally, we conclude and discuss our results in section 8. There is also an appendix that discuss our methodologies in more detail.

2 Regional modelling system and observations used

Our simulations follow Kendon et al (in press) with the addition of a 50-km simulation. Here we analyse data for the period 1991 to 2007 from the three different resolution (50km, 12km, and 1.5km) RCM simulations. All three simulations are different configurations of the Met Office Unified Model. An overview of the three simulations is presented in Table 1.

2.1 50-km and 12-km HadGEM3-RA

The 50-km and 12-km simulations are limited area versions of the non-hydrostatic Hadley Centre Global Environmental Model version 3 (HadGEM3-RA) (Walters et al, 2011). The model dynamical core uses a semi-implicit semi-Lagrangian scheme to solve the non-hydrostatic and compressible dynamical equations (Davies et al, 2005). The model uses a staggered Arakawa-C horizontal grid (Arakawa and

Lamb, 1977), and has 63 Charney-Phillips terrain-following hybrid vertical levels (Charney and Phillips, 1953). The higher resolution 12-km simulation uses a shorter time-step than the 50-km simulation (see Table 1).

Both simulations use the Met Office mass-flux CMODS 4A convection scheme (Gregory and Rowntree, 1990), the updated Wilson and Ballard (1999) cloud microphysics scheme for large-scale precipitation (LSP) without prognostic rain, and the Wilson et al (2008) PC2 prognostic cloud scheme. For the land surface, the Met Office Surface Exchange Scheme 2.2 (MOSES2) (Essery et al, 2001) is employed. For sea surface temperatures (SSTs), we have prescribed observed high-resolution 0.25° daily SSTs (Reynolds et al, 2007).

Both simulations are forced by the ERA-Interim reanalysis (Dee et al, 2011) at the lateral boundaries. The ERA-Interim has a T255 ($\approx 0.75°$ or $\approx 80\texttt{km}$) spatial resolution and 60 hybrid vertical levels. Temporally, the (re-gridded) reanalysis is prescribed to the lateral boundaries every 6 hours. An illustration of the simulation domain is shown in Fig. 1a. The simulation domain covers Europe and parts of North Africa with the United Kingdom about a third of the way from the western boundary.

2.2 1.5-km RCM

The 1.5-km regional climate model uses a modified version of the non-hydrostatic Met Office operational UK variable-resolution model (UKV). Like the HadGEM3-RA, UKV is also one of the configurations of the Met Office Unified Model. It has a grid spacing of 1.5km in the interior with a transition to 4-km at the edges. This gives a $1:3$ downscaling ratio near the boundaries. This variable resolution rim reduces instabilities near the lateral boundaries. The operational 1km UKV has been shown to improve UK orographic and convective precipitation relative to coarser 12km simulations (Roberts et al, 2009; Roberts and Lean, 2008; Lean et al, 2008).

The majority of the model physics in the 1.5-km simulation is the same as in the 12-km and 50-km simulations, but there are some important differences. Similar to the 50-km and 12-km HadGEM3-RA simulations, the 1.5-km simulation shares the same dynamical core (Davies et al, 2005). The same land surface (Essery et al, 2001) scheme and prescribed SSTs are used in all three simulations. Unlike the coarser simulations, however, the 1.5-km simulation uses no convective parametrisation, nor the prognostic cloud scheme (see Kendon et al (in press) for details). Time stepping and the number of vertical levels also differ (see Table 1). The 1.5-km simulation uses the (Wilson and Ballard, 1999) cloud microphysics scheme with prognostic rain. The Smagorinsky-Lilly model (Smagorinsky, 1963; Lilly, 1962) is used to represent sub-grid turbulent diffusion.

The simulation is driven by the 12-km RCM simulation, with no feedback from the 1.5-km simulation back to the 12-km simulation ('one-way nesting'; see section 1). Unlike the two coarser resolution simulations, the 1.5-km RCM domain is limited to SUK (see Fig. 1b) where convective events are observed to be the most common in Britain. Due to the small size of the 1.5-km domain, we expect strong similarities in the regional atmospheric conditions between the 1.5-km and 12-km simulations over the SUK. For inter-comparisons, we upscale the 1.5-km simulation results to the 12-km and 50-km scale.

2.3 Observations

The National Climate Information Centre daily UK gridded precipitation (Perry et al, 2009) is used as a reference to compare all three model simulations. The daily gridded dataset begins in 1958 and ends at the present day, and here we use data from 1991 to 2007. Between 1991 and 2007, the dataset has used approximately 2500-3500 surface gauge observations that are scattered over England, Wales, Scotland, and Northern Ireland. Approximately two-thirds of these gauges are in SUK.

Quality control is performed through computerized and manual comparisons of individual daily station values against the daily all-station average and daily values from nearby stations. Any stations that have failed quality control are excluded from the computation of the gridded values. The gridding of the gauge data to a $5\text{km} \times 5\text{km}$ grid uses a cubic inverse-distance weighting interpolation using stations within 50km radius of the grid box.

There are three notable issues for such a dataset:

- Values are undefined over water;
- Station gauge observations can only sample events that occur over the gauges themselves, and may not sample specific localized events;
- Gauges are often located in valleys, and that leads to an underestimation of precipitation in the vicinity of high topography;

To address the first issue, we restrict all our comparisons to land points only. The second and third issues are fundamental limitations to rain gauge data - one can only detect local events if they are sampled by the gauges. The under-sampling of convective and orographic precipitation extremes will cause the area averaged gridded values to be less than the true area-averaged value. Ensor and Robeson (2008) show that gridded gauge precipitation produces reasonable annual precipitation estimates, but selectively degrades the representations of high and low precipitation events.

We intend to investigate this problem in further detail in the near future with other observation datasets (see section 8). We expect the problem is more likely to affect rare (once every few years) localized extremes. Such extremes are more common in JJA when localized convection is the most common. During DJF, the under-sampling is expected to be lesser of a problem as extremes are more associated with large-scale precipitation.

3 Methodologies

In order to compare between model simulations and daily observations, we have re-gridded our observations and model simulations to the 12-km and 50-km grid:

- The 1.5-km simulation is upscaled to both 12-km and 50-km scales when compared against the 12-km and 50-km simulations;
- The 12-km simulation is upscaled to the 50-km scale when compared against the 50-km simulation;
- The 5-km gridded observation dataset is upscaled to the 12-km and 50-km grids.

Since we are only interested in days that have (significant) precipitation events, we only include days with events exceeding given minimum thresholds. The examined thresholds are: 1.0, 5.0, 10.0, 20.0, 30.0, 40.0, and 50.0mm/day. We also estimate spatial scales and the clustering of precipitation events with the Ripley K-function (Ripley, 1977). A description of the Ripley K-function can be found in the appendix.

Both parametric and non-parametric statistical significance tests are used in our analysis. For basic comparisons between climatological seasonal means, we have employed the Student-T test at the 5% level. Since there are 17 years of data, the degrees of freedom for the Student-T test are $dof = 17 - 1 = 16$.

For the comparisons between event intensities, we have used a 1000 member bootstrap (Efron and Tibshirani, 1993; Wilks, 1997), and test at the 10% significance level. We define a precipitation subset (P') in which a daily threshold ($p_{\mathrm{THRESHOLD}}$) has been exceeded (Equation 1). The average event intensity ($\langle P' \rangle$) is defined to be the expected daily intensity within the subset (Equation 2):

$$P' = \{P \geqslant p_{\mathrm{THRESHOLD}}\} \tag{1}$$

$$\langle P' \rangle = \frac{\sum_{n=1}^{N(p_{\mathrm{THRESHOLD}})} P'_n}{N} \tag{2}$$

$$N(p_c) = \left|P'\right| \tag{3}$$

In which P is the set of all non-zero precipitation values. N is the number of elements in subset P'. Both P' and N are functions of $p_{\mathrm{THRESHOLD}}$. P can be a set that is formed all values for all grid points (as in Figs. 6 and 7) or at each individual grid point (as in Figs. 4 and 5 in which we have denoted event counts (per year) at each grid point as n).

$\langle P' \rangle$ can be computed with the original dataset and a bootstrap. The bootstrap re-sampling is performed in 3-month seasonal blocks for each year (e.g. 1991 JJA, 1992 JJA, ..., 2007 JJA). We randomly select 17 seasons (out of the total of 17 years) with replacement, such that some years may be represented more than once and some not at all. We re-sample in seasonal blocks to account for possible auto-correlation. Wilks (1997) suggests that the block length can be estimated through independent sample number estimation assuming the process is a 1st-order autoregressive process. Such estimation is difficult practically as the auto-correlation of daily precipitation is caused by a number of mechanisms which act at a range of time scales: from 1-5 day synoptic variations to soil-precipitation feedbacks that operate over time scales of weeks and months. A seasonal block assumes that the precipitation intensities from the same season of the previous year to be independent of the precipitation intensities of the present season. Generally, long block sizes lead to Type-II errors - not enough null hypothesis rejections (i.e. significance tests favour higher probabilities for null hypothesis for non-difference between models and observations) (Wilks, 1997).

The re-sampling procedure is repeated 1000 times to produce 1000 simulated 17-season datasets. The original dataset is one out of $\binom{2n-1}{n}$ (for $n = 17$, $\binom{2n-1}{n} \sim 10^9$) possible outcomes from the re-sampling. For each bootstrap, we compute the intensity differences between the two compared datasets. The null hypothesis is that the differences are zero. We estimate the top and bottom 5% percentile (corresponding to a two-tail 10% significance test) of the 1000 differences from

the bootstrap, and check if the top (or bottom) 95% differences have same signed differences. If the sign is the same, we reject the null hypothesis.

4 Simulated climatologies

We first examine the difference in precipitation climatologies between the observations and models. This also serves as an introduction to the UK climate for unfamiliar readers. Shown in Fig. 2 and Fig.3 is the observed 3-month seasonal mean precipitation (and the three model-simulated fractional departures from the observed values) at both the 50-km and 12-km scale.

For all seasons, observed precipitation is highest over the Welsh mountains and south-west England. Lower amounts are observed in the lowlands to the east. Therefore for further analysis, we divide our SUK domain into two sample regions using the "Tees-Exe Line"[1] (see Fig. 1b). This line separates the meteorologically wetter and topographically higher north west (NW) and the meteorologically drier and topographically lower south east (SE). The models (at all three resolutions) tend to have negative (positive) precipitation bias over the NW (SE). However, the biases show seasonal variations, which are similar for all three simulations.

The Tees-Exe Line also separates the east which is subject to relatively more intense local extremes than the west. To the east of the line, the 100-year-return level event is on average $2.5 - 3.2$ times more intense than the 2-year-return level event[2], while the same ratio is lower $(2.0 - 2.5)$ to the west of the line (Faulkner, 1999).

The observations show that SON (MAM) is the wettest (driest) three-month period for SUK. JJA and DJF, which we will examine in detail here, have precipitation amounts in between SON and MAM. JJA and DJF are chosen for thorough analysis as they represent two different precipitation regimes: primarily convective rain concentrated to the east of the Tees-Exe Line during JJA, and frontal precipitation concentrated to the west during DJF (Maraun et al, 2009). Over highland regions, DJF precipitation is higher than JJA. Over the lowlands, the highest JJA precipitation values are lower than the highest DJF precipitation values over Southern England; however, over eastern England and East Anglia, JJA is wetter than DJF. For the models, the NW dry bias is largest during SON and DJF when highland precipitation is higher, and the SE wet bias is largest during MAM when lowland precipitation is highest.

By examining Fig. 2 and Fig. 3, it appears that increasing the model resolution has a positive impact on orographic precipitation in the NW. When resolution is increased from 50-km to 12-km, the (negative) bias over Wales is reduced. Minor reductions of positive bias are noticeable in the SE. Even though the 1.5-km simulation is driven by the 12-km simulation, the patterns of their bias differ significantly - the 1.5-km simulation is notably wetter than the 12-km simulation. Positive biases over the SE are higher in the 1.5-km simulation. The 1.5-km simulation does show (negative) bias reduction for orographic precipitation over Wales for SON and MAM. This improvement in orographic precipitation is consistent with what has been found in previous studies (see section 1).

[1] The line joining the mouths of the River Exe and the River Tees in UK

[2] Often called the "growth rate" in hydrology.

For the SE, there appears to be little value in going to a higher model resolution; in fact, results here in terms of seasonal mean precipitation show that the resolution increase has a negative impact on the bias. These results are consistent with Kendon et al (in press), and the excessive precipitation in the 1.5-km model is thought to be due to the inherent under-resolving of convection at the 1.5-km scale. However, we will show that resolution increases lead to changes in other important precipitation statistics.

5 Localized event frequencies

The climatological mean (and model biases in the mean) does not convey any information about the frequency and intensity of events. To begin our discussion, we have plotted the observed and simulated annual June-November (JJA+SON) and December-May (DJF+MAM) occurrences of precipitation events exceeding 20mm/day (Fig. 4) and 50mm/day (Fig. 5). We have plotted frequencies on both the 12-km and 50-km grid. The use of half year divisions is based on the similarities of occurrence frequencies between JJA and SON and between DJF and MAM (not shown). We expect the frequencies of these types of events to increase with decreasing grid size (evident when comparing panels a against d, and c against e in both figures). At a coarse grid size (i.e. the 50-km grid scale), area averaging favours events that are widespread as localized convective events are filtered out by area averaging.

Both observations and models indicate that one may expect between 1-10 20+ mm/day events per year (JJA+SON and DJF+MAM) at the 12-km and 50-km grid scale with the highest frequencies over the western part of our domain. On the east side of the Tees-Exe Line, event frequencies at both thresholds are higher in JJA+SON than in DJF+MAM, but such seasonal variations are not as evident over Wales and south-western England. All models captured the higher frequency of event occurrences observed over Wales, the east-west gradient of the frequency, and the seasonal variations in the south-east.

Frequencies for 20 + mm/day events are higher everywhere for the 1.5-km (12-km) simulation in SUK when compared with the 12-km (50-km) simulation. For the 1.5-km simulation, the increase in the number of events is most evident along the southern and south-eastern England coast in both JJA+SON and DJF+MAM, and the increased frequencies are higher than the observed estimates. The increase of 20 + mm/day event frequencies in the 12-km simulation (when compared with the 50-km simulation) occurs both to the west and the east of the Tees-Exe Line. The 12-km simulation is also superior to the 50-km simulation in terms of capturing the high frequencies over the Welsh highlands.

As one moves to the 50 + mm/day threshold, event occurrences decrease (Fig. 5). Typically there are no more than 1-2 events per year (JJA+SON and DJF+MAM) at any grid point for the 17 years of analysed data (sometimes just once within all 17 years of data; the 0.025 events/year contour in Fig. 5 is chosen based on $\frac{1}{17} \times \frac{1}{2} \approx 0.025$). Similar to the 20+mm/day threshold, observations show an east-west gradient in occurrence number with the highest frequencies observed over the Welsh highlands. In the 50-km simulation, there are too few 50 + mm/day events to discern such a gradient. The southern / south-eastern coast event increase that is evident at the 20 + mm/day threshold within the 1.5-km simulation is also evi-

dent at the 50 + mm/day threshold for both JJA+SON and DJF+MAM, leading to an increased bias with respect to observations.

Given the above analysis, there is a clear improvement in the model simulation of the occurrence of heavy precipitation when model resolution is increased from 50-km to 12-km in both the orographic regime to the west and lowland regime to the east. However, there is little value of increasing the resolution from 12-km to 1.5-km as biases worsen. However, Figs.4 and 5 convey no information about spatial structures and average intensities of 20+mm/day and 50+mm/day events.

6 Excess intensity and event-based statistics

In this section, we examine the average intensity ($\langle P' \rangle$, see Equation 3) of events exceeding various thresholds from 1mm/day to 50mm/day. The results are presented in Fig. 6 and Fig. 7 for JJA and DJF respectively. As in previous figures, results are presented with data that are upscaled to 50-km (left panels) and 12-km (right panels). We have plotted up to the 50mm/day threshold due to the rarity of 50 + mm/day events in the 50-km simulation (as indicated in panel c of both figures).

6.1 JJA

At the 12-km grid scale, in both the SE and NW subregions, the 12-km simulation simulates precipitation intensities that are $10\% - 20\%$ higher than observations for thresholds above 30mm/day, and the differences are statistically significant at the 10% level. The 1.5-km simulation simulates a lower and closer-to-observation intensity for the same 30 + mm/day thresholds. When the comparisons are made at the 50-km scale, both the 12-km and 1.5-km simulations show SE intensities that are 5-10% higher than observations.

For lower thresholds (below 10mm/day, where all events above this threshold are included), only the 1.5-km simulation has higher ($\approx 20\%$ for the SE) average intensity than the observations. That is true for both the NW and SE subregions. The other two (12-km and 50-km) simulations have either intensities that are lower (NW) or that are not statistically different (SE) from observations.

The 50-km model simulation underestimates event intensities. Over the NW, this underestimation is significant for events exceeding thresholds up to 30mm/day, and becomes insignificant at higher thresholds. By contrast over the SE, this underestimation only becomes significant at higher thresholds (40mm/day and 50mm/day).

6.2 DJF

DJF precipitation intensities are better simulated by the 12-km and 1.5-km simulations than by the 50-km simulation, with the negative intensity bias in the NW reduced or eliminated entirely in the higher resolution simulations. The 50-km simulation has consistently lower intensities in the NW than observations; differences at all but one threshold (40mm/day) are statistically significant at the 10% level.

For the 12-km simulation, intensities of higher threshold NW events are well simulated at both the 12-km and 50-km scale, and are not statistically different from observations at any threshold above 10mm/day. Unlike the 12-km simulation, the 1.5-km simulation tends to show positive biases of about $5\% - 10\%$ across multiple thresholds when the data is upscaled to the 12-km scale. At the 50-km scale, the 1.5-km simulation has positive biases in two (20mm/day and 40mm/day) out of seven thresholds.

At lower thresholds (below $10 +$ mm/day), all three simulations have lower NW intensities than observed at the 50-km grid scale. The biases are reduced for the 1.5-km and 12-km simulations if the comparisons are made at the 12-km scale. Since the differences disappear or are reduced at higher thresholds (above 10mm/day), this shows that all simulations have deficiencies in the simulation of moderate events ($1 - 10$mm/day).

SE precipitation intensities that are simulated by the 50-km and 12-km simulations are not statistically different from observations. The 1.5-km simulation simulates higher intensities in the SE at both 12-km and 50-km grid scale. In Fig. 3, it is notable that the 1.5-km simulation also has the highest SE DJF positive bias among the three simulations.

6.3 Event occurrences as a function of intensity threshold

Total event numbers (across all grid boxes and days) are shown in the lower panels (c,d) in both Fig.6 and Fig. 7. The number of events decreases with increasing intensity threshold, as would be expected, and the decrease rates are highest for the 50-km simulation. The number of NW and SE 1+mm/day events is comparable for all model resolutions and observations, but the number of $40 +$ mm/day events in the 50-km simulation is up to an order of magnitude less than the two other simulations and observations.

Panels Fig. 6 c,d and Fig. 7 c,d are integrated measures of event occurrence in both time and space. In order to partition out temporal occurrences, in Table 2, we present the number of days which have at least one event greater than the threshold at the 12-km scale. For the 50-km scale (not shown), higher (40mm/day and 50mm/day) threshold events are lacking especially for the 50-km simulation.

For JJA (Fig. 6 c,d), the total number of events in the 50-km simulation and the observations are consistently lower than in the 12-km and 1.5-km simulations for thresholds above 40mm/day. However, we expect the estimated observed counts to be lower than the true value (see section 2.3), so the positive event number biases in the 12-km and 1.5-km simulations may be less than shown. The 50-km simulation itself has less counts than the observations; using the same argument as above, the underestimation by the 50-km simulation may actually be higher due to under-sampling by the observations. A comparison between Table 2 and Fig. 6 indicates that in the 1.5-km simulation the excessive number of events is partially due to the large number of days having at least one event somewhere in the SUK domain. For the 12-km simulation, results are more curious as the number of days having at least one event is less than in the observations until the $50 +$ mm/day threshold. We shall show that the spatial density of events is the cause of the discrepancy.

For DJF, the 1.5-km and 12-km simulations are more realistic than the 50-km simulation in simulating the number of events over the mountainous NW, especially with thresholds of $30+$ mm/day or higher. At the 12-km scale, the 1.5-km simulation has $2-5$ times more SE events than the observations for the $30+$ mm/day and $40+$ mm/day thresholds. It is worth noting that such heavier SE DJF precipitation events are rare when compared with JJA, and the number of DJF events for the whole domain (SE + NW) is mostly attributable to the NW. All three model simulations are able to capture that NW-SE difference.

6.4 Section summary

Here are some of the key results from this section:

- Both the 12-km and 1.5-km simulations and observations have higher precipitation intensities and event numbers than the 50-km simulation in both JJA and DJF;
- The 1.5-km simulation has more realistic intensities for JJA intense precipitation events than the 12-km simulation;
- The 12-km simulation has more realistic intensities for DJF intense precipitation events than the 1.5-km simulation;
- The 1.5-km simulation has a larger number of events than the other two model simulations and observations across many thresholds;
- For most thresholds above 20mm/day, the 12-km (1.5-km) simulation has fewer (more) days in JJA with at least one precipitation event in the SUK domain compared to observations.

Our results indicate that the 50-km simulation performs least well in comparison to the higher resolution simulations in simulating intense daily precipitation. However, a mixed picture is shown between the two higher resolution simulations.

7 Spatial density and clustering

We have presented the temporal occurrences of intense precipitation events; and in this section, we examine the spatial characteristics of intense daily precipitation events. Clustered local precipitation events pose a larger risk than scattered events in triggering flooding. Thus the realistic simulation of the spatial characteristics of precipitation is essential for using models as a flood risk assessment tool.

Using the NW and SE regional division, we compute the average spatial density (Tables 3 and 4)[3], the number of near neighbours, and spatial clustering. The *spatial density* is the total number of precipitating grid points with the threshold exceeded divided by the total number of defined grid points (see Equation 5 in Appendix). The spatial density can then be averaged across all days that have at least 2 grid points that have exceeded the threshold (see Equation 17 in Appendix).

Clustering is a localized (an area subset of the whole domain) density enhancement (excess density above the average density of the whole domain). Density and

[3] Only 1.5-km model simulated SE DJF $50+$ mm/day event statistics are shown due to the lack of events in the other datasets.

clustering are two different concepts; clustering may exist in the absence of "high" densities. A schematic that illustrates the differences between density and clustering is shown in Fig. 8. As in density, we present the time average values only (see Equations 18 and 19 in Appendix).

Clustering is assessed by calculating the area normalized Ripley K-function (also known as the Besag L-function; see appendix) of precipitation events exceeding two specific thresholds (20 + mm/day and 50 + mm/day). High spatial density values may arise from many individual disorganized single grid-point showers (with low clustering) or many clustered "precipitation blobs" (non-zero Besag L-function). The results are presented in Fig. 9 and Fig. 10 for JJA, and Fig. 11 and Fig. 12 for DJF. Shown are:

- The estimated number of near neighbours[4] and Besag L-function assuming events are not clustered (left panels, dash line)
- The actual number of near neighbours (left panels, solid line) and Besag L-function as observed in the gridded observations (right panels), 12-km and 1.5-km simulations

In a nutshell, the right panel (Besag L-function) measures the excess of clustered points (left panels, solid lines) above the background (left panels, dashed lines).

Here we limit our comparisons to only the 12-km horizontal scale. This is because coastlines and island geography become too coarse at the 50-km scale, and the poor simulation of rare high-intensity events by the 50-km model. For example, East Anglia and Cornwall-Devon are only 1-3 grid points wide at the 50-km scale (see Fig. 2).

7.1 JJA

In general, the clustering and spatial density are much better simulated in the NW than in the SE by both simulations. For the NW, both models are reasonably successful in simulating the average clustering at both thresholds (Figs. 9 a and 10 a, dashed lines). Naturally, clustering over the NW is tied with orography, and a better simulated clustering in the NW is not surprising. The 20 + mm/day events are also better handled than the 50 + mm/day events.

By comparison with the NW, both models are more challenged to simulate the spatial density and clustering over the SE. Figs. 9 and 10 (lower panel c, dashed lines) show that both the 12-km and 1.5-km simulations tend to simulate precipitation that is spatially too dense at both thresholds in the SE. Both models are more successful in reproducing the observed clustering at the 20 + mm/day threshold (Fig. 9 d) than at the 50 + mm/day (Fig. 10 d). For the 50 + mm/day threshold, there is severe underestimation of clustering in the SE, but the same is not evident at the 20 + mm/day threshold. In general over the SE, the 12-km simulation gives a better representation of the average spatial density than the 1.5-km simulation, whilst the 1.5-km simulation gives a better (but still poor) representation of clustering for the 50 + mm/day threshold.

[4] The average spatial event density multiplied by the surface area

The excessive density in the 12-km simulation over the SE at the 20 + mm/day threshold compensates for the reduced number of days of event occurrence - leading to an overall comparable total number of events between the 12-km simulation and observations (Fig. 6 d). For 50 + mm/day events, both too many days with events and the excessive spatial density contribute to the excessive number of events in the 12-km and 1.5-km simulations.

We note that observed clustering peaks at around 4-10 grid points (for both thresholds) in the SE. This is consistent with length scales of organized mesoscale convective events ($\approx 10^1 - 10^2$km).

7.2 DJF

Relative to JJA, both models generally show much better skill in simulating the spatial densities and clustering in DJF. For the DJF 20 + mm/day threshold (Fig. 11), the SE clustering (panel d) in the 12-km and 1.5-km simulations is similar, with both simulations tend to underestimate observed clustering at "large" ($r \geq 10$ grid points) radii. The observed spatial density of 20+mm/day events is in between the values simulated by the 12-km and 1.5-km simulations in both NW and SE (panels a and c, dashed lines; Tables 3 and 4 column c). This is different to the situation in JJA, where both simulations have SE spatial densities that are too high relative to observations.

For 50 + mm/day events, the simulated clustering in the NW (Fig. 12b) is higher for both simulations (with the 1.5-km simulation closer to observations). This is in contrast with JJA (Fig.10 b) when the simulated clustering is lower than observed. Both simulations simulate spatial densities that are higher than observed (Fig. 12 a, dashed lines; Table 3 column d).

We have not shown the clustering and near neighbour number counts for DJF SE 50 + mm/day events for 2 out of the 3 datasets. There are only 1 and 3 DJF SE 50 + mm/day events between 1991 and 2007 that have spatial scales more than 2 or more grid points for the observations and 12-km model simulation respectively. The difference in the number of events can be seen in Fig. 5 j-l. The probability of having a 50+mm/day event in the SE during DJF+MAM is highest for the 1.5-km simulation with most events concentrated along the southern coast of England. In summary, the lack of DJF SE 50 + mm/day events in the 12-km simulation and observations is consistent with the event frequency differences between the two models and the observations.

7.3 Section summary

The above results reveal a few important points:

- With the exception of the 1.5-km model simulated NW DJF 20 + mm/day spatial density, all examined model simulated spatial densities are either higher or comparable to observations.
- Despite the higher spatial density, both simulations tend to underestimate spatial clustering of 50 + mm/day events over the SE in JJA.

Similar to the results for precipitation intensities, there are no clear improvements to precipitation clustering and spatial density from increasing the model resolution from 12 km to 1.5 km. The simulation of the SE during JJA has been especially challenging for both simulations - overall spatial density is overestimated for both thresholds, and spatial clustering (organization) of daily extremes (50+mm/day) is poor. Due to the nature of the observations, the observed clustering may be underestimated, and this suggests model discrepancies may be larger. We note the above picture may change if different accumulation periods are examined (hourly or multi-hourly), and is suggested by the results in Roberts and Lean (2008).

8 Discussions and Conclusions

We have presented a number of event-based metrics in this paper. These have been chosen to gauge differences in the model simulations across different resolutions. Key results include:

- Increasing model resolution from 50 km to 12 km is beneficial to the simulation of DJF orographic precipitation. The 50-km simulation underestimates the occurrence and intensity of heavy precipitation and has a negative mean precipitation bias over orography. This is reduced in the 12-km simulation.
- Seasonal biases in precipitation totals increase when resolution is increased from 12 km to 1.5 km.
- Moderate-to-heavy daily precipitation occurs too often in the 1.5-km simulation - especially in SE England. The average intensity of the JJA (DJF) daily extremes is better simulated by the 1.5-km (12-km) simulation. Both the 12-km and 1.5-km simulations have too many days with extreme (50 + mm/day) JJA precipitation.
- Both 12-km and 1.5-km simulations overestimate JJA spatial density of events over the SE for two thresholds examined here (20+mm/day and 50+mm/day). On top of the overestimation, there is clear deficiency for both simulations in capturing the appropriate spatial clustering for SE JJA 50 + mm/day events.

In general, there are some improvements in simulating daily intense precipitation when model resolution is increased from 50 km to 12 km. When resolution is further increased to 1.5-km, there is no further clear cut improvement. Comparisons between the 12-km and 1.5-km simulation give a mixed picture: better JJA daily extreme intensity in the 1.5-km simulation versus smaller seasonal biases in the 12-km simulation in multiple seasons. We acknowledge that there are many ways to compare model simulations, and one of the objectives of this study is trying different ways to do the comparison.

Improved representation of precipitation in DJF over orography between the 50-km and 12-km is a result consistent with other similar studies (see section 1). Our results indicate that boreal summer (JJA) precipitation away from orography may or may not have benefited with increasing resolution depending on the metric used.

The summer of SE UK represents the convective precipitation regime over UK. The 1.5-km simulation shows an improvement over the 12-km simulation in the simulation of average intensities of high threshold JJA events, but such events are

too common in the the 1.5-km model (event day counts are too high, Table 2). Spatial clustering and density are deficient in both simulations especially at the high thresholds.

Daily precipitation extremes are linked to multi-hour slowly moving (frontal or convective) precipitating systems. Short duration and/or fast moving precipitation systems do not lead to extreme daily accumulations (but they may be linked to extremes at shorter time scales). The above argument suggests that a different result may be obtained if the analysis is performed for accumulations over shorter periods. The above also highlights that a good simulation of precipitation duration, spatial organization and temporal evolution of precipitation systems is essential to give reasonable extremes in the model climate.

The 12-km simulation has fewer JJA heavy precipitation days, and the 12-km heavy events have a more "outbreak" nature than the 1.5-km simulation. The 12-km simulation has less days with moderate-high precipitation. When moderate-high daily precipitation events are triggered, they become overly intense (daily totals are too high) and widespread (spatial density of grid points above a moderate/high daily threshold is too high).

In JJA, for more moderate thresholds (20+mm/day), the 12-km model appears to have the best simulated intensity. However, the 12-km simulation has nearly 20-25% less days with any such events. Therefore, it is hard to conclude if the 12-km simulation is any better than the 1.5-km simulation in simulating more moderate daily events.

For the metrics that we have compared, the 12-km simulation appears to outperform the 1.5-km simulation in DJF. On the whole, boreal winter (DJF) heavy precipitation is easier to simulate than boreal summer (JJA) precipitation. Simulated intensities, spatial densities and clustering are all better simulated in DJF. Our results indicate that there is no benefit in increasing model resolution higher than 12 km for DJF for the used models within the context of regional climate modelling. Similar results have been demonstrated for other regions of the globe (Hong and Leetmaa, 1999). This is due to the dynamical processes that drive winter precipitation events - caused by fronts and synoptic depressions with scales of the order of $10^2 - 10^3$km. That is 2-3 orders of magnitude larger than 1.5 km.

An important question remains unanswered - the value of increasing model resolution above 12-km for JJA precipitation. Results presented here for daily precipitation are mixed, but Kendon et al (in press) have shown clear improvements when the same data is examined at the sub-daily time scale in metrics such as diurnal variability, duration, and spatial extent. Many of the remaining biases at 1.5 km may be explained by the fact that the 1.5-km simulation is "convection permitting"; i.e. even at 1.5 km, convection is still under-resolved.

Our results indicate that the 1.5-km model simulates too many heavy precipitation events in JJA (too many event counts, spatial density that is too high). This is a result consistent with Lean et al (2008). With convection under-resolved, the explicitly-resolved convective motion and vertical mass flux become too intense. That is because cloud-scale up- and down-drafts are still under-resolved where grid box vertical motion is either all up or all down. This is in contrast with CP-enabled model simulations where vertical convective motion is parametrised under the presumption that cloud-scale convective motions (mixed between up and downward motion) are not resolved, and the thermodynamical and dynamical consequence of the unresolved convective motions are estimated and feedbacks to

683 the larger model resolved scale. The use of Smagorinsky-Lilly turbulent diffusion
684 (Smagorinsky, 1963; Lilly, 1962) has alleviated problems with convective showers
685 being too intense in the 1.5-km simulation. However, this is still an active research
686 area in "convective permitting" models. Another option is to increase resolution
687 yet further (e.g. to $\sim 10^2$m); this would lead to considerable increases in compu-
688 tational cost, and also the suitability of other model physics would then need to
689 be re-examined, especially the representation of the boundary layer and the land
690 surface scheme.

691 We have not discussed the sensitivity of our results to domain size and bound-
692 ary changes. Such sensitivity is well discussed in the literature (Jones et al, 1995;
693 Seth and Giorgi, 1998; Leduc and Laprise, 2009); that is large (small) domains give
694 the model greater (less) freedom to develop its own features. The current work can-
695 not explore such sensitivity, and running many high resolution RCM simulations
696 are computationally expensive. The 1.5-km RCM domain is $\sim$ 1000km wide - small
697 in comparison with the domain sensitivity study by Leduc and Laprise (2009). It is
698 intriguing to ask how such domain size sensitivity manifests itself in high resolution
699 RCM simulations. The last question should be explored in the future.

700 Gridded gauge observations may underestimate clustering, variance, and inten-
701 sity for specific events, particularly in summer when extremes are more localised.
702 In order to estimate the degree of underestimation of clustering and intensity of
703 local extremes in the observations, one requires independent high-resolution esti-
704 mates of daily/hourly precipitation. We intend to revisit this problem again in the
705 near future using radar and hourly gauge data (Golding, 1998).

706 We have not discussed inter-annual variability in this paper. Out of the 17
707 years that we have analysed (1991-2007), 2007 was a major flood year in the SUK
708 (Blackburn et al, 2008). The models do not capture the heavy rain in central and
709 western UK that was seen in observations (not shown). The 2007 UK floods were
710 caused by two synoptic events, whose tracks and precipitation patterns are not well
711 captured by the simulations despite the use of reanalysis data as lateral boundary
712 conditions. We note that, however, reanalysis information is only fed in at the
713 edge of the European domain, and one should not expect exact agreement in the
714 positioning of events over the UK - not only between models and observations,
715 but also between models themselves.

716 Any type of model projection for future climate change needs to be carefully
717 interpreted in the context of an understanding of model strengths and limitations.
718 Even with increases in model complexity (such as increasing the horizontal reso-
719 lution as shown here), many limitations can still exist in their simulations for the
720 current climate.

721 Appendix: The Ripley K function

722 The Ripley K-function (Ripley, 1977) is a measure of spatial clustering which com-
723 pares the number of near neighbours with the average spatial density of the whole
724 region of interest (Fig.3). Given any (time-varying) map (e.g. gridded precipita-
725 tion), one marks all the events with ones (1) and non-events with zeroes (0). We
726 denote that map (effectively a matrix/vector) with I:

$$I(x,y,t) = \begin{cases} 1, & \text{if } P \geq P_c \\ 0, & \text{if } P < P_c \end{cases} \tag{4}$$

, where P_c is the threshold. The average spatial density of the events is simply defined as:

$$\varrho(t) = \frac{N(t)}{A} \tag{5}$$

In which, N is the total number of events, and A is the area of the map (e.g. the number of grid points). By definition, $\varrho(t) \leq 1$ ($\varrho(t) = 1$ implies events are occurring at every single grid point). The average number of near neighbours is a function of distance (or area which is proportional to the distance squared), and is given as:

$$V(r,t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}(-(\delta_{ij}-1))I(d_{ij} \leq r,t) \tag{6}$$

Index i represents the summation over all existing events, and index j represents the other events. d_{ij} is the distance between them:

$$d_{ij} = |\mathbf{x}_i - \mathbf{x}_j| \tag{7}$$

We have used the Kronecker delta function:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{8}$$

Therefore, $-(\delta_{ij}-1))$ denotes the self-exclusion during near neighbour counting. If there is spatial clustering of events, the number of near neighbours ($V(r,t)$) to any existing event is higher than the value expected by computing the average background density:

$$V(r,t) > \varrho(t)\pi r^2 - 1 \tag{9}$$

The Ripley K-function is defined as the number of near neighbours divided by the average density:

$$K(r,t) = \frac{V(r,t)}{\varrho(t)} \tag{10}$$

The Ripley K-function has the dimensions of area (radius squared), and is the non-clustered area (grid boxes) needed to match the number of events as observed in a localized clustered area. By definition, it is not defined if there is only one event (i.e. there are 0 near neighbours). If spatial density is perfectly uniform, then the Ripley K-function is exactly the geometric area of a circle with radius r. If events occur as Poisson processes in space (i.e. each events occurring independently with each other), the Ripley K-function is approximately but not exactly the same circle geometric area. If events are clustered, then the Ripley K-function at a given radius exceeds the geometric area of the circle given with the same radius:

$$K(r,t)\begin{cases}\approx \pi r^2, & \text{if events are not clustered}\\ > \pi r^2, & \text{if events are clustered}\end{cases} \tag{11}$$

Simulation study by Ripley (1979) show that the spatial Poisson process null hypothesis can be rejected at the 0.05 and 0.01 level if the observed supermum (maximum) Besag L-function exceeds a total area (A) dependent threshold:

$$\sup_r(L(r)) \geq 1.42\frac{\sqrt{A}}{N}, p = 0.05 \tag{12}$$

$$\sup_r(L(r)) \geq 1.68\frac{\sqrt{A}}{N}, p = 0.01 \tag{13}$$

In the present analysis, all individual K-functions that the null hypothesis cannot be rejected at the 0.05 level are excluded from the time average.

The above assumes that event sampling is not limited by domain specifications. In our datasets, we have undefined points because of:

- No observations outside of model/observation domain;
- Observations over water are undefined, and model non-land points are masked out.

This leads to under-sampling as there are unobserved events over the undefined area. Therefore, a correction factor (w) (Ripley, 1977) should be used on V(r). We denote the corrected V(r) with a "hat".

$$\hat{V}(r,t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w|\delta_{ij} - 1|I(d_{ij} \leq r, t) \tag{14}$$

$$\hat{K}(r,t) = \frac{\hat{V}(r,t)}{\varrho(t)} \tag{15}$$

We have used an area based correction by Besag (1977) (discussed in the postscript of the original Ripley paper) due to its easy implementation with complex coastlines (problem degenerates to counting undefined grid boxes). Given a circle with radius r with only an area $\hat{A}$ within the defined domain (over land and within the SUK domain), the correction factor is defined as:

$$w(r, x_0, y_0) = \frac{\pi r^2}{\hat{A}(r, x_0, y_0)} \tag{16}$$

Functions V and K are generally time dependent (maps of daily precipitation). We compute daily ϱ, V, and K values, and present only their time-averaged values.

$$\langle\hat{\varrho}\rangle = \frac{1}{T}\sum_{t=1}^{T}\varrho(t) \tag{17}$$

$$\left\langle\hat{V}\right\rangle(r) = \frac{1}{T}\sum_{t=1}^{T}\hat{V}(r,t) \tag{18}$$

$$\left\langle \hat{K} \right\rangle (r) = \frac{1}{T} \sum_{t=1}^{T} \hat{K}(r,t) \tag{19}$$

If events are not clustered, $K(A)$ is a linear function ($K(A = \pi r^2) \approx \varrho A$). Since the area ($A$) is known for any given radius (r), K-function is often plotted as a square root (the Besag L-function) (Besag, 1977).

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r \tag{20}$$

Clustering is largest where $L(r)$ is largest, and unclustered data will have $L(r) \approx 0$. Note that both the K- and L-function are normalized in a way that they do not favour higher average spatial density. Both functions only measure *the inflation of local density due to event clustering*.

Since zonal (δx) and meridional (δy) grid point distances can only take on whole number values (0, 1, 2, 3, ...; "quantized grid space"), and can be diagonal. r is defined as:

$$r(\delta x, \delta y) = \sqrt{\delta x^2 + \delta y^2} \tag{21}$$

Note that $L(r)$ and $K(r)$ are functions of r. We have assumed the clustering and density to be isotropic (independent of direction), which is not true in general for precipitation (such as frontal and orographic precipitation). The sampled region is also assumed to be uniform. That is the same as saying the mechanisms behind rainfall within each region is assumed to be the same everywhere. We mimic that by sampling only the north-western orographic or south-eastern convective rain regions. That is, of course, only an approximation; non-uniformity clearly exists within each of the regions - such as non-uniform topography, changing land surface types, and irregular coastlines. A perfect stationary region is impossible to obtain, and we approximate that by slicing our domain with the Tees-Exe Line.

List of Figures

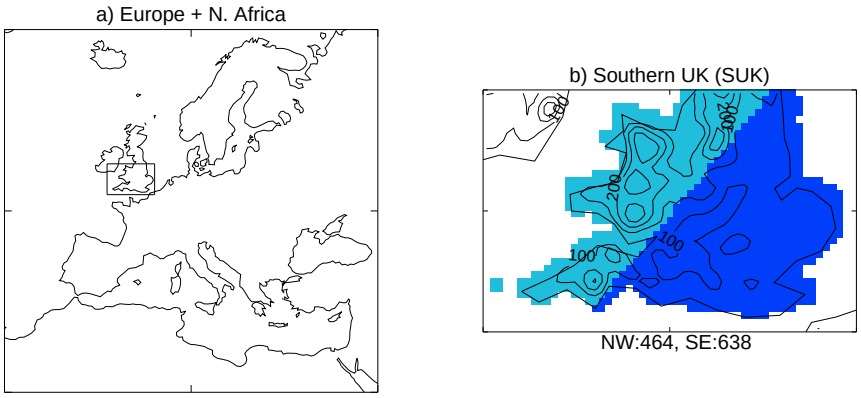

Fig. 1: The Europe/North African (left) and southern UK (SUK) (right) domain. The inner domain is marked as a square on the left panel. Surface height (m, in the 12-km simulation) for the SUK domain is contoured with a 100-metre interval. For SUK, the north west (south east) sub-domain is coloured in light (dark) blue. There are a total of 464 (638) grid points in the north-west (south-east) sub-domain.

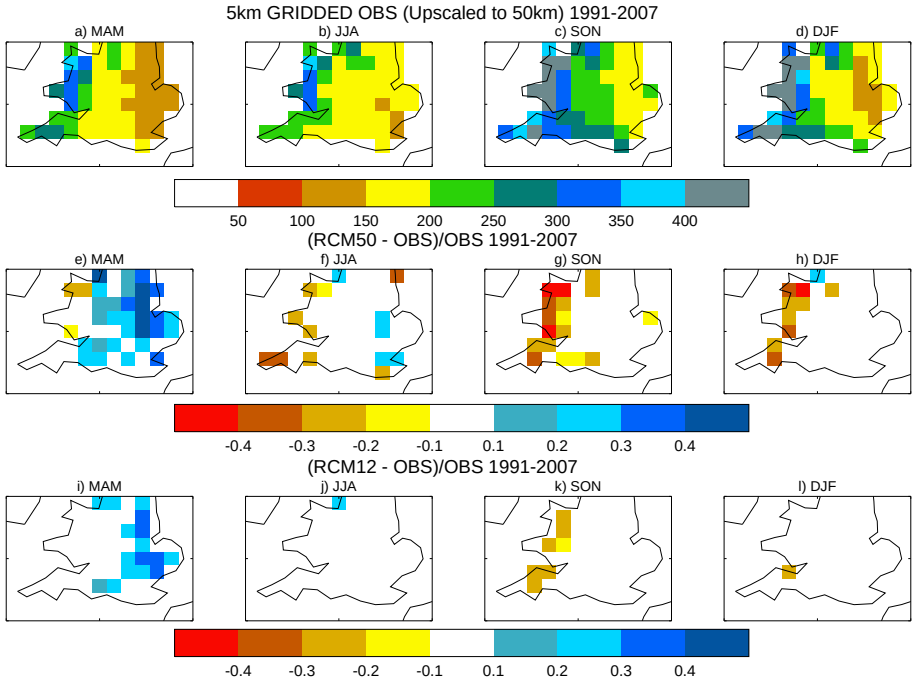

Fig. 2: Observed 1991-2007 climatological daily precipitation (mm/day) at the 50-km scale (panels a-d) for different 3-month periods: March-April-May (MAM), June-July-August (JJA), September-October-November (SON), and December-January-February (DJF). The 50-km and 12-km (upscaled to 50-km) simulation's fractional departure from the observed values are shown in panels e to l. Only differences that are significant at the 5% level and larger than ± 0.1 are shown.

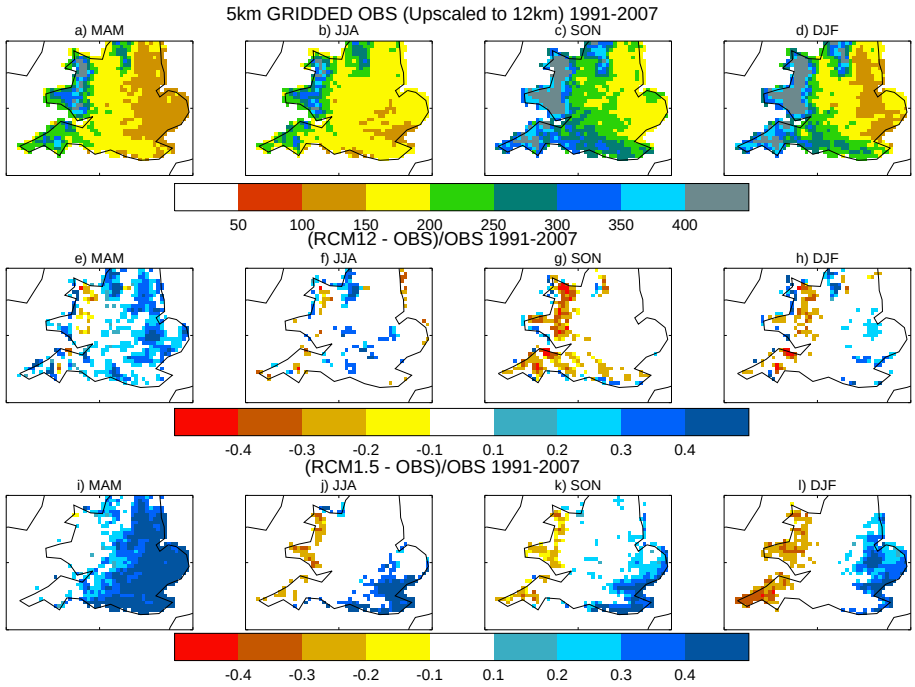

Fig. 3: Same as in Fig. 2, but at the 12-km scale and the comparisons are made against the 12-km and 1.5-km simulations.

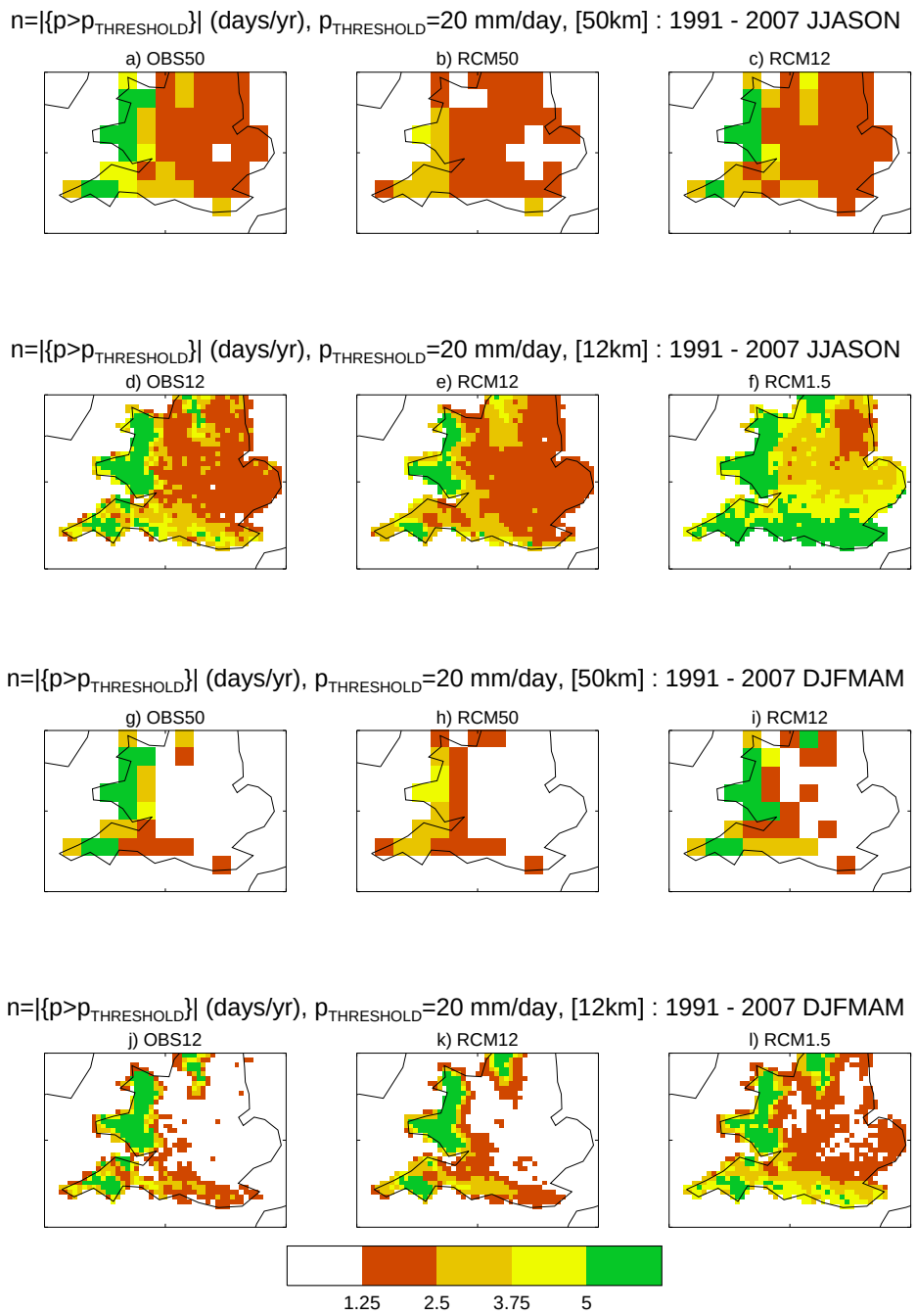

Fig. 4: Average frequencies (days/year) that daily precipitation surpasses the 20 mm/day threshold during JJA+SON (upper half, a-f) and DJF+MAM (lower half, g-l) between 1991 and 2007 at each grid point. Panels a-c (JJA+SON) and g-i (DJF+MAM) show the frequencies for observations, 50-km and 12-km simulations at the 50-km scale with observations and 12-km daily amounts pre-upscaled to 50-km scale. Panels d-f (JJA+SON) and j-l (DJF+MAM) show the same at the 12-km scale for observations, 12-km and 1.5-km simulations.

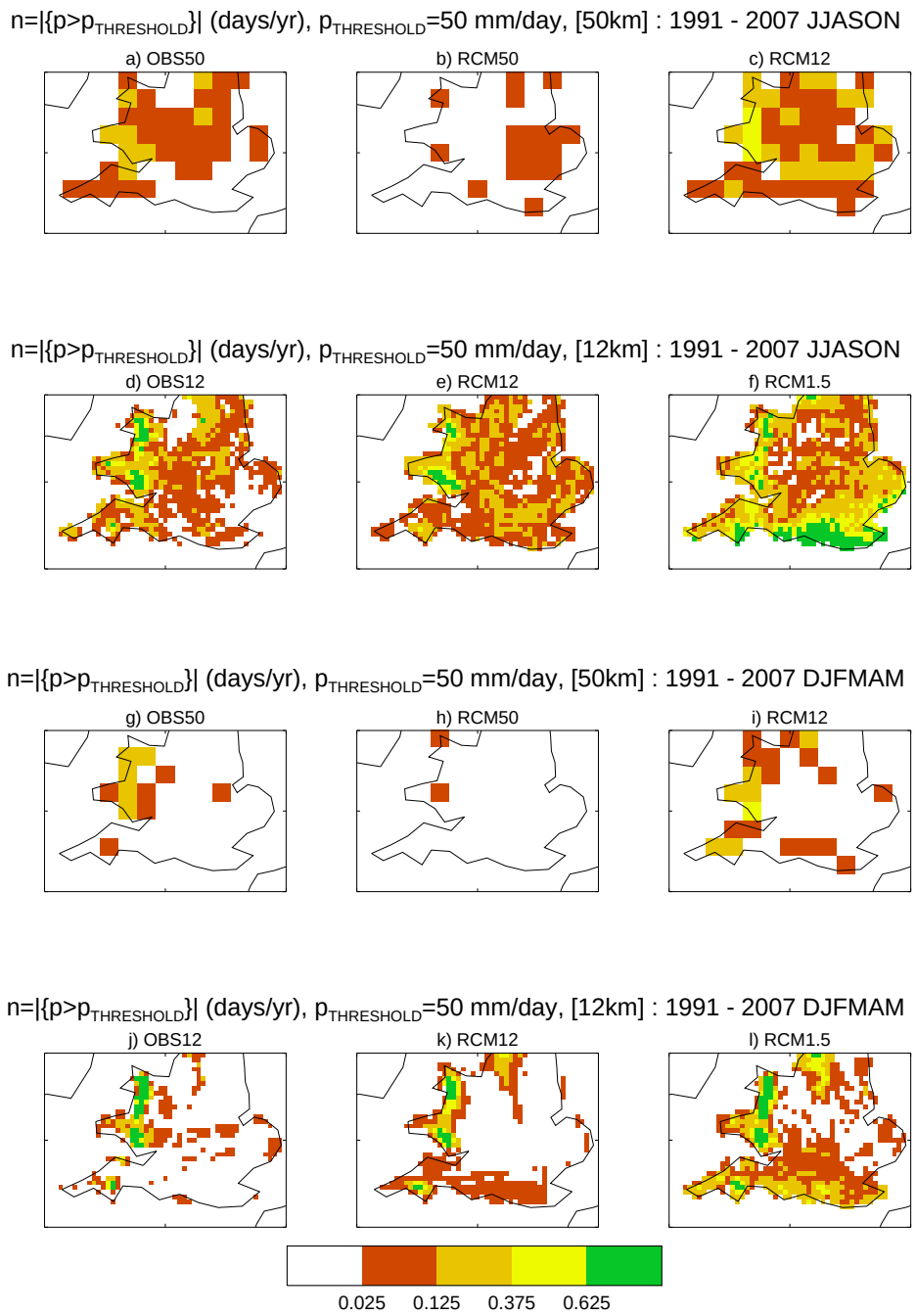

Fig. 5: Same as Fig. 4, but with a 50mm/day threshold.

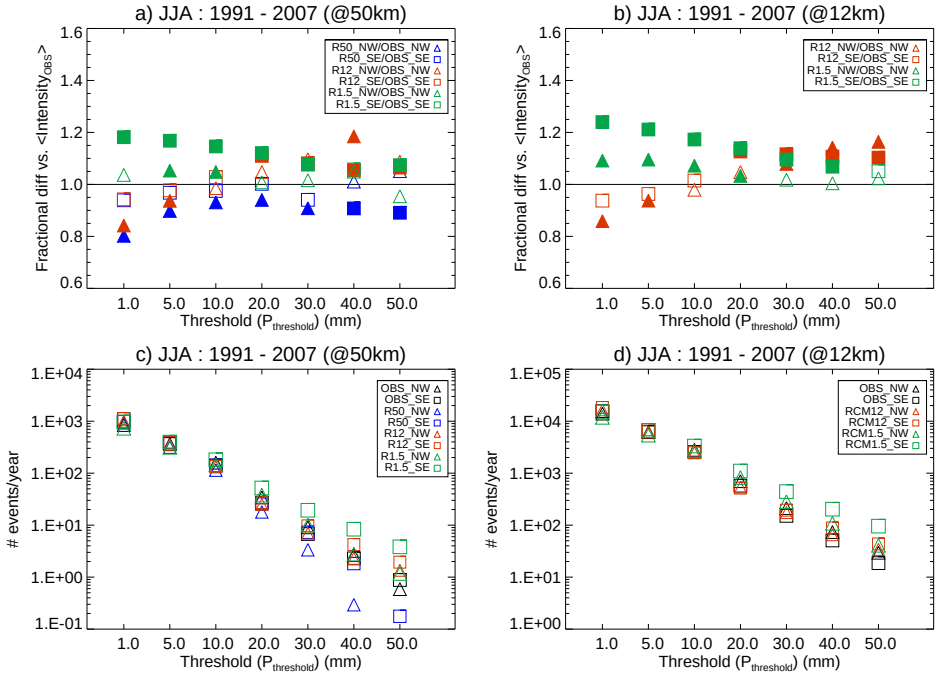

Fig. 6: The fractional differences in average intensity relative to observations ($\frac{\langle P'_{\mathrm{MODEL}} \rangle}{\langle P'_{\mathrm{OBS}} \rangle}$, upper panels) and average event counts per year ($\frac{N_{\mathrm{MODEL}}}{\text{Number of years}}$ and $\frac{N_{\mathrm{OBS}}}{\text{Number of years}}$, lower panels) for precipitation exceeding a range of thresholds, for all JJAs between 1991 and 2007. For the upper panels, filled symbol indicates the fractional differences are significant at the 10% level. No symbols are drawn if there are zero samples from either the model or observations (see lower panels c and d). For average event counts, y-axis is plotted with a logarithmic scale.

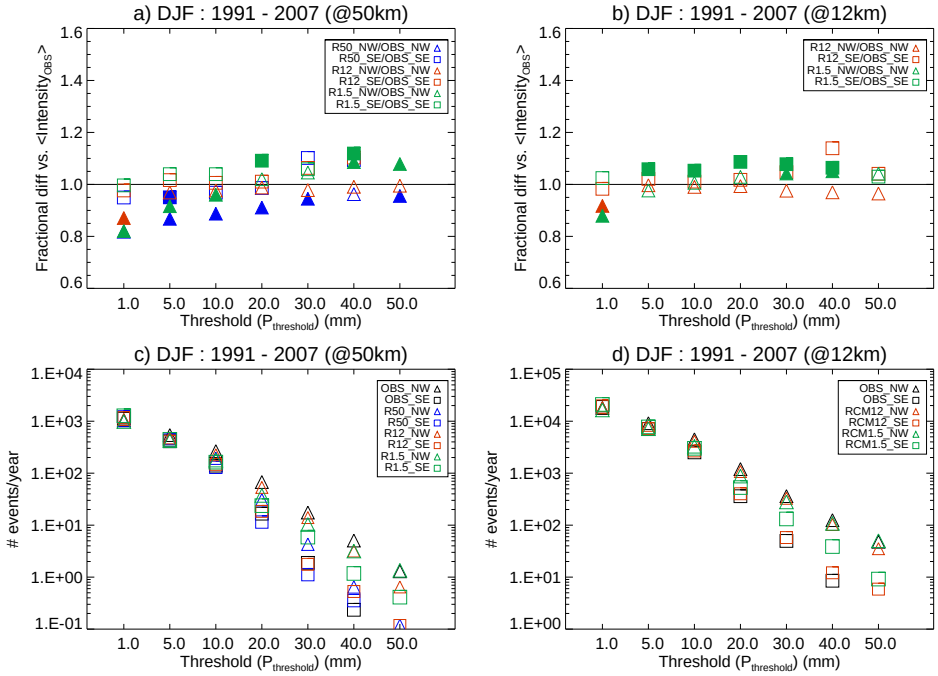

Fig. 7: Same as in Fig. 6, but for all DJFs between 1991 and 2007.

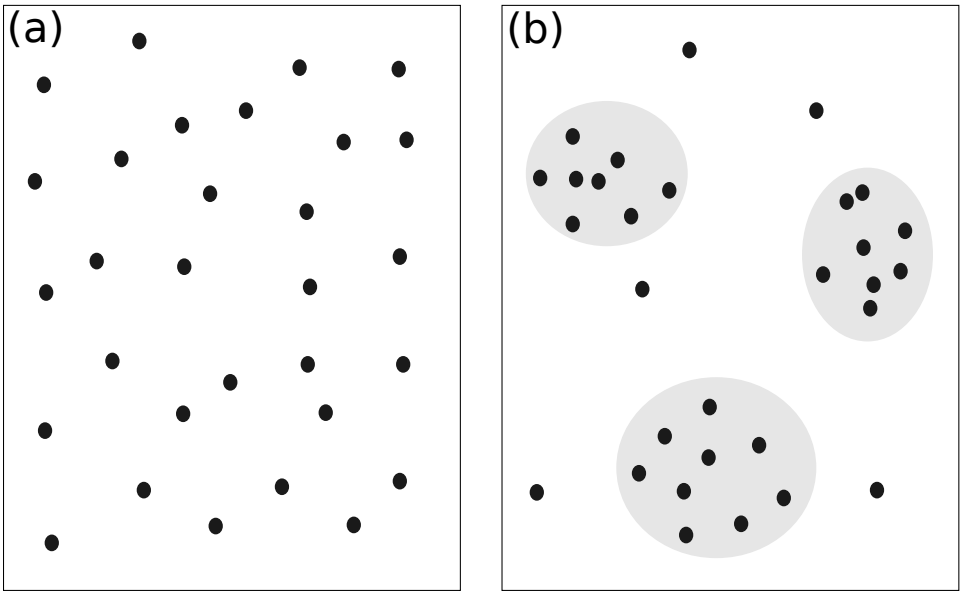

Fig. 8: Shown above is a schematic of (a) non-clustered and (b) clustered precipitation. The spatial density of (a) is the same as (b); there are thirty precipitating grid points (dark spots) enclosed within the same area. For (a) the non-clustered case, individual grid point "showers" are approximately spaced at regular spatial intervals. For (b) the clustered case, precipitation organize into clustered "blobs" (light grey circles). One would expect (gridded) precipitation to cluster in space across a of spatial scales. If horizontal resolution is small enough ($\approx$ 1km, clustering can be caused by convective clouds occupying more than one grid point. Precipitation and clouds are also clustered in the meso- (fronts, organized convective storms, orographic precipitation), synoptic- (mid-latitude and tropical low pressure systems), and planetary scales (ITCZ, monsoon systems).

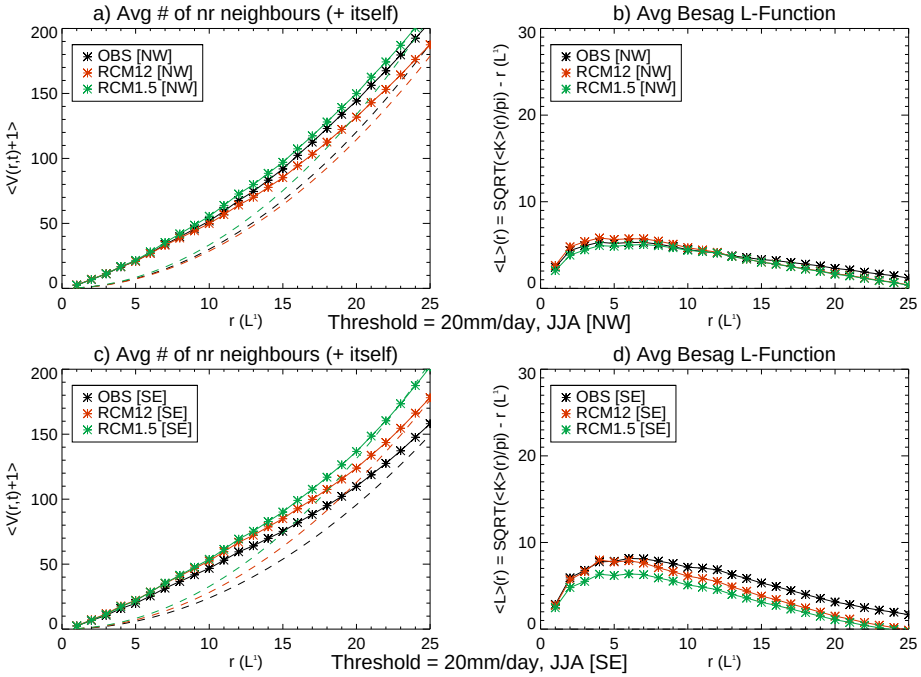

Fig. 9: The time average of three spatial descriptive statistic for JJA days with at least 2 grid points that has exceeded the 20mm/day threshold are shown. Days with clustering that cannot be rejected at the 5% level with the Poisson process null hypothesis are excluded. In the left panels (a and c), solid lines are the observed number of events (plus itself) from an existing event, and the dashed lines are the expected number of events (plus itself) if spatial density is uniform in space. In the right panels (b and d), the Besag L-function (local increase of spatial density due to spatial clustering) are plotted. Black, red, and green represent gridded observations, 12-km RCM, and 1.5-km RCM estimates respectively. The upper panels (a and b) are for the NW, and the lower panels (c and d) are for the SE.

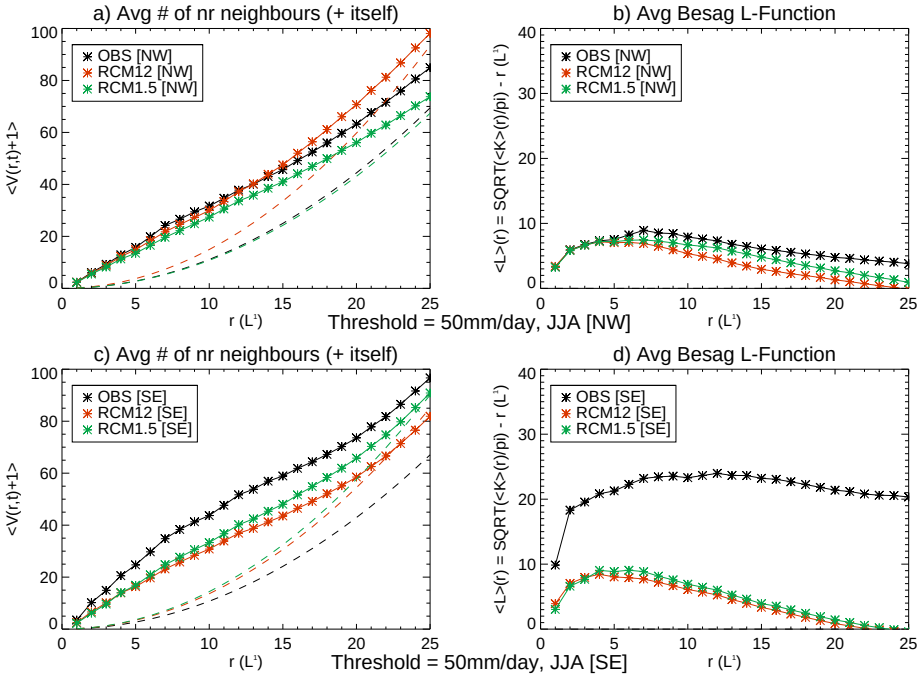

Fig. 10: Same as in Fig. 9, but the threshold is set to 50 mm/day.

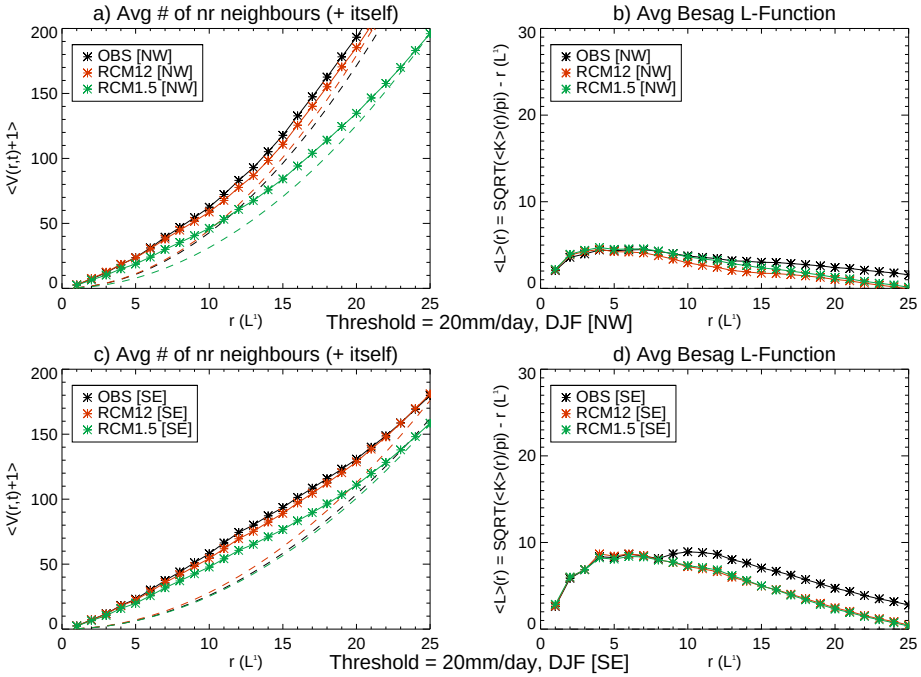

Fig. 11: Same as in Fig. 9 with the 20mm/day threshold, but for DJF instead of JJA.

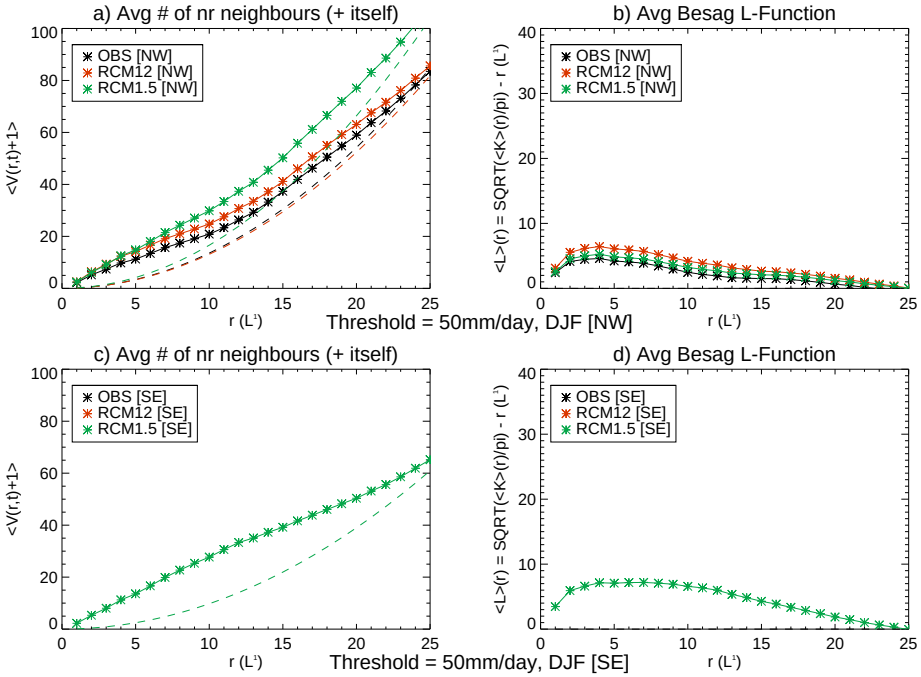

Fig. 12: Similar as in Fig. 10 with the 50mm/day threshold, but for DJF instead of JJA. For SE, only 1.5-km model simulated values are shown; there are only 1 and 3 DJF SE 50mm/day valid events (events that have more than 1 grid point) in the observations and the 12-km model simulations respectively. The 1.5-km simulation has 12 SE events.

List of Tables

Table 1: Description of model simulations: RCM horizontal resolution (δx), model used, time-step (δt), number of vertical levels (N_z), simulation domain, lateral boundary conditions, convective parametrisation

δx	Model	$\delta t, N_z$	Domain	LBC	CP
50km	HadGEM3-RA	720sec, 63	Eur.+N. Africa	ERA-Int.	CMODS
12km	HadGEM3-RA	300sec, 63	Eur.+N. Africa	ERA-Int.	CMODS
1.5km	modified UKV	50sec, 70	SUK	12-km RCM	No CP

Table 2: The number of JJA and DJF days (out of the total number of valid JJA and DJF days) where there is at least one daily-threshold excess event anywhere in domain at the 12-km scale

—	(a)	(b)	(c)
Threshold	Obs. (12-km)	12-km RCM	1.5-km RCM
20+ mm/day (JJA)	544 / 1564	405 / 1564	564 / 1564
30+ mm/day (JJA)	314 / 1564	215 / 1564	359 / 1564
40+ mm/day (JJA)	163 / 1564	148 / 1564	233 / 1564
50+ mm/day (JJA)	75 / 1564	97 / 1564	149 / 1564
20+ mm/day (DJF)	528 / 1534	429 / 1534	540 / 1534
30+ mm/day (DJF)	304 / 1534	218 / 1534	303 / 1534
40+ mm/day (DJF)	166 / 1534	98 / 1534	157 / 1534
50+ mm/day (DJF)	88 / 1534	56 / 1534	79 / 1534

Table 3: Average $20 + mm/day$ and $50 + mm/day$ event density (events per grid box) for JJA and DJF for the NW

—	(a)	(b)	(c)	(d)
Season	JJA	JJA	DJF	DJF
Threshold	$20 + mm/day$	$50 + mm/day$	$20 + mm/day$	$50 + mm/day$
OBS	0.096	0.035	0.136	0.043
12-km RCM	0.091	0.047	0.143	0.042
1.5-km RCM	0.106	0.034	0.100	0.053

Table 4: Average $20 + mm/day$ and $50 + mm/day$ event density (events per grid box) for JJA and DJF for the SE

—	(a)	(b)	(c)	(d)
Season	JJA	JJA	DJF	DJF
Threshold	$20 + mm/day$	$50 + mm/day$	$20 + mm/day$	$50 + mm/day$
OBS	0.076	0.034	0.083	—
12-km RCM	0.090	0.043	0.090	—
1.5-km RCM	0.104	0.046	0.080	0.031

Acknowledgements This research is part of the CONVEX project - a collaboration between Newcastle University, the Met Office, and the University of Exeter. CONVEX is supported by the United Kingdom NERC Changing Water Cycle programme (grant NE/I006680/1), and the presented model simulations are supported by the Met Office. The lead author is financially supported by Newcastle University, and is a visiting scientist at the Met Office Hadley Centre in Exeter, United Kingdom.

References

Allan RP, Soden BJ, John VO, Ingram W, Good P (2010) Current changes in tropical precipitation. Environ Res Lett 5(2):025,205, DOI 10.1088/1748-9326/5/2/025205

Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. Nature 419:224–232

Antic S, Laprise R, Denis B, de Elía R (2006) Testing the downscaling ability of a one-way nested regional climate model in regions of complex topography. Clim Dyn 26(2-3):305–325, DOI 10.1007/s00382-005-0046-z

Arakawa A (2004) The cumulus parameterization problem: Past, present, and future. J Climate 17(13):2493–2525, DOI 10.1175/1520-0442(2004)017⟨2493:RATCPP⟩2.0.CO;2

Arakawa A, Lamb VR (1977) Computational design of the basic dynamical processes of the UCLA general circulation model. Methods Comput Phys 17:173–265

Besag JE (1977) Comments on ripley's paper. J Roy Stat Soc 39(2):193–195

Blackburn M, Methven J, Roberts N (2008) Large-scale context for the uk floods in summer 2007. Weather 63(9):280–288, DOI 10.1002/wea.322

Charney JG, Phillips NA (1953) Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. J Meteorol 10:71–99, DOI 10.1175/1520-0469(1953)010⟨0071:NIOTQG⟩2.0.CO;2

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Stainforth A, White AA, Wood N (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q J R Meteorol Soc 131(608):1759–1782, DOI 10.1256/qj.04.101

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hölm EV, Isaksen L, Kallberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park PK, Peubey C, de Rosnay P, Tavolato C, Thêpaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597, DOI 10.1002/qj.828

Déqué M, Piedelievre JP (1995) High resolution climate simulation over Europe. Clim Dyn 11:321–339

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, vol 57. Chapman and Hall, New York

Ensor LA, Robeson SM (2008) Statistical characteristics of daily precipitation: Comparisons of gridded and point datasets. J Appl Meteorol Clim 47:2468–2476, DOI 10.1175/2008JAMC1757.1

Essery R, Best M, Cox P (2001) MOSES 2.2 technical documentation. Hadley Centre Technical Note 30, Hadley Centre, Met Office, Fitzroy Road, Exeter, UK

Faulkner D (1999) Flood estimation handbook. Volume 2. Rainfall frequency estimation. NERC Centre for Ecology and Hydrology

Fowler HJ, Ekström M (2009) Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. Int J Climatol 29:385–416

930 Fowler HJ, Ekstrom M, Blenkinsop S, Smith AP (2007) Estimating change in
931 extreme European precipitation using a multi-model ensemble. J Geophys Res
932 112(D18), art No. D18104
933 Giorgi F, Marinucci MR (1996) An investigation of the sensitivity of simulated
934 precipitation to the model resolution and its implication for climate studies. Mon
935 Weather Rev 124:148–166, DOI 10.1175/1520-0493(1996)124⟨0148:AIOTSO⟩2.
936 0.CO;2
937 Golding BW (1998) Nimrod: A system for generating automated very short range
938 forecasts. Meteorol Appl 5:1–16
939 Gregory D, Rowntree PR (1990) A mass-flux convection scheme with rep-
940 resentation of cloud ensemble characteristics and stability dependent clo-
941 sure. Mon Weather Rev 118:1483–1506, DOI 10.1175/1520-0493(1990)118⟨1483:
942 AMFCSW⟩2.0.CO;2
943 Hohenegger C, Brockhaus P, Schär C (2008) Towards climate simulations at cloud-
944 resolving scales. Meteorol Z 17(4):383–394, DOI 10.1127/0941-2948/2008/0303
945 Hong SY, Leetmaa A (1999) An evaluation of the ncep rsm for regional climate
946 modeling. J Climate 12(2):592–609, DOI 10.1175/1520-0442(1999)012⟨0592:
947 AEOTNR⟩2.0.CO;2
948 Jones RG, Murphy JM, Noguer M (1995) Simulation of climate change over Europe
949 using a nested regional–climate model. I: Assessment of control climate, includ-
950 ing sensitivity to location of lateral boundaries. Q J R Meteorol Soc 121:1413–
951 1449
952 Jones RG, Murphy JM, Noguer M, Keen AB (1997) Simulation of climate change
953 over Europe using a nested regional–climate model. II: Comparison of driving
954 and regional model responses to a doubling of carbon dioxide concentration. Q
955 J R Meteorol Soc 123:265–292
956 Kendon EJ, Roberts NM, Senior CA, Roberts MJ (in press) Realism of rain-
957 fall in a very high resolution regional climate model. J Climate DOI 10.1175/
958 JCLI-D-11-00562.1
959 Knote C, Heinemann G, Rockel B (2010) Changes in weather extremes: Assessment
960 of return values using high resolution climate simulations at convection-resolving
961 scale. Meteorol Z 19(1):11–23, DOI 10.1127/0941-2948/2010/0424
962 Laprise R, de Elía R, Caya D, Biner S, Lucas-Picher P, Diaconescu E,
963 Leduc M, Alexandru A, Separovic L (2008) Challenging some tenets of re-
964 gional climate modelling. Meteorol Atmos Phys 100(1-4):3–22, DOI 10.1007/
965 s00703-008-0292-9
966 Lean HW, Clark PA, Dixon M, Roberts NM, Fitch A, Forbes R, Halliwell C (2008)
967 Characteristics of high-resolution versions of the Met Office Unified Model for
968 forecasting convection over the United Kingdom. Mon Weather Rev 136:3408–
969 3424, DOI 10.1175/2008MWR2332.1
970 Leduc M, Laprise R (2009) Regional climate model sensitivity to domain size.
971 Clim Dyn 32(6):833–854, DOI 10.1007/s00382-008-0400-z
972 Lilly DK (1962) On the numerical simulation of buoyant convection. Tellus A
973 14:148–171
974 Maraun D, Rust HW, Osborn TJ (2009) The annual cycle of heavy precipitation
975 across the united kingdom: a model based on extreme value statistics. Int J
976 Climatol 29(12):1731–1744, DOI 10.1002/joc.1811
977 Meehl GA, Karl T, Easterling DR, Changnon S, Changnon D, Jr RP, Evans J,
978 Groisman PY, Knutson TR, Kunkel KE, Mearns LO, Parmesan C, Pulwarty

R, Root T, Sylves RT, Whetton P, Zwiers F (2000) An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. Bull Am Meteorol Soc 81(3):413–416, DOI 1520-0477(2000)081⟨0413:AITTIE⟩2.3.CO;2

Molinari J, Dudek M (1992) Parameterization of convective precipitation in mesoscale numerical models: A critical review. Mon Weather Rev 120(2):326–344, DOI 10.1175/1520-0493(1992)120⟨0326:POCPIM⟩2.0.CO;2

Perry M, Hollis D, Elms M (2009) The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK. Met Office National Climate Information Centre, FitzRoy Road, Exeter, Devon EX1 3PB, United Kingdom

Rauscher SA, Coppola E, Piani C, Giorgi F (2010) Resolution effects on regional climate model simulations of seasonal precipitation over Europe. Clim Dyn 35(4):685–711, DOI 10.1007/s00382-009-0607-7

Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution blended analyses for sea surface temperature. J Climate 20:5473–5496

Ripley BD (1977) Modelling spatial patterns. J Roy Stat Soc 39(2):172–212

Ripley BD (1979) Tests of 'randomness' for spatial point patterns. J Roy Stat Soc 41(3):368–374

Roberts NM, Lean HW (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon Weather Rev 136:78–97, DOI 10.1175/2007MWR2123.1

Roberts NM, Cole S, Forbes RM, Moore R, Boswell D (2009) Use of high-resolution NWP rainfall and river flow forecasts for advanced warning of the Carlisle flood, north-west England. Meteorol Appl 16(1):23–34, DOI 10.1002/met.94

Seth A, Giorgi F (1998) The effects of domain choice on summer precipitation simulation and sensitivity in a regional climate model. J Climate 11:2698–2712, DOI 10.1175/1520-0442(1998)011⟨2698:TEODCO⟩2.0.CO;2

Smagorinsky J (1963) General circulation experiments with the primitive equations, Part 1: The basic experiments. Mon Weather Rev 91:99–164

Wakazuki Y, Nakamura M, Kanada S, Muroi C (2008) Climatological reproducibility evaluation and future climate projection of extreme precipitation events in the Baiu season using a high-resolution non-hydrostatic RCM in comparison with an AGCM. J Meteorol Soc Jpn 86(6):951–967

Walters DN, Best MJ, Bushell AC, Copsey D, Edwards JM, Falloon PD, Harris CM, Lock AP, Manners JC, Morcrette CJ, Roberts MJ, Stratton RA, Webster S, Wilkinson JM, Willett MR, Boutle IA, Earnshaw PD, Hill PG, MacLachlan C, Martin GM, Moufouma-Okia W, Palmer MD, Petch JC, Rooney GG, Scaife AA, Williams KD (2011) The Met Office Unified Model global atmosphere 3.0/3.1 and JULES global land 3.0/3.1 configurations. Geosci Model Devel 4:919–941, DOI 10.5194/gmd-4-919-2011

Wigley TML, Lough JM, Jones PD (1984) Spatial patterns of precipitation in england and wales and a revised, homogeneous england and wales precipitation series. Int J Climatol 4(1):1–25, DOI 10.1002/joc.3370040102

Wilks DS (1997) Resampling hypothesis tests for autocorrelated fields. J Climate 10:65–82

Wilson DR, Ballard SP (1999) A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. Q J R Meteorol Soc 125:1607–1636

1027 Wilson DR, Bushell AC, Kerr-Munslow AM, Price JD, Morcrette CJ (2008) PC2:
1028 A prognostic cloud fraction and condensation scheme. I: Scheme description. Q
1029 J R Meteorol Soc 134:2093–2107, DOI 10.1002/qj.333