Extremes (2010) 13:241–267
DOI 10.1007/s10687-010-0101-y

Detecting change in UK extreme precipitation using results from the climate*prediction*.net BBC climate change experiment

Hayley J. Fowler · Daniel Cooley · Stephan R. Sain · Milo Thurston

Received: 16 April 2009 / Revised: 9 December 2009 /
Accepted: 7 January 2010 / Published online: 4 February 2010

Abstract We investigate a question posed by policy makers, namely, "when will changes in extreme precipitation due to climate change be detectable?" To answer this question we use climate*prediction*.net (CPDN) model simulations from the BBC Climate Change Experiment (CCE) over the UK. These provide us with the unique opportunity to compare 1-day extreme precipitation generated from climate altered by observed forcings (time period 1920–2000) and the SRES A1B emissions scenario (time period 2000–2080) (the Scenario) to extreme precipitation generated by a constant climate for year 1920 (the Control) for the HadCM3L General Circulation Model (GCM). We fit non-stationary Generalized Extreme Value (GEV) models to the Scenario output and compare these to stationary GEV models fit to the parallel Control. We define the time of detectable change as the time at which we would reject a hypothesis at the $\alpha = 0.05$ significance level that the 20-year return level of the two runs is equal. We find that the time of detectable change depends on the season, with most model runs indicating that change to winter extreme precipitation may be detectable by the year 2010, and that change to summer extreme precipitation is not detectable by 2080. We also investigate which climate model parameters affect the weight of the tail of the precipitation distribution and which affect the time of

H. J. Fowler (✉)
School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK
e-mail: h.j.fowler@ncl.ac.uk

D. Cooley
Department of Statistics, Colorado State University, Fort Collins, CO, USA

S. R. Sain
Institute for Mathematics Applied to Geoscience, National Center for Atmospheric Research, Boulder, CO, USA

M. Thurston
Oxford e-Research Centre, Oxford University, Oxford, UK

detectable change for the winter season. We find that two climate model parameters have an important effect on the tail weight, and two others seem to affect the time of detection. Importantly, we find that climate model simulated extreme precipitation has a fundamentally different behavior to observations, perhaps due to the negative estimate of the GEV shape parameter, unlike observations which produce a slightly positive (∼0.0–0.2) estimate.

Keywords Extreme precipitation · Detection · Climate change · Climate*prediction*.net · Parameters · Generalized extreme value

AMS 2000 Subject Classifications 62-statistics

1 Introduction

In extreme value analyses, the primary questions researchers usually attempt to answer are to assess probabilities of rare events, to estimate tail quantiles such as r-year return levels, and to provide measures of uncertainty associated with these estimates. Although still related to extreme values and rare events, the primary questions we address in this study are quite different in nature. The first question was brought to us by policy makers; it is: "Under climate change and based on climate model output, when will a change in extreme precipitation be detectable?" Detection is the process of demonstrating that climate has changed in some defined statistical sense, without providing the reason(s) for the change (Hegerl et al. 2007). A second question that arose from undertaking this study comes from the perspective of climate modeling; it is: "Which climate model parameters affect the behavior of extreme precipitation in climate model output?" To our knowledge, this is the first time such questions have been addressed using a rigorous statistical approach to the modeling and analysis of extreme precipitation using the Generalized Extreme Value distribution (GEV).

1.1 Extremes and climate change

In a future subject to climate change, it is likely that changes to climatic and hydrological extremes will have the greatest impact on human society (Tebaldi et al. 2006). For this reason, there is great interest in assessing how climate change has already affected and will continue to affect the characteristics of extreme weather events. In the northern Hemisphere, many studies have noted upward trends in observations of both mean precipitation and high quantiles of precipitation (e.g. Alexander et al. 2006; Meehl et al. 2005; Trenberth et al. 2007; Fowler and Kilsby 2003a, b) and climate models suggest that these trends will continue under enhanced greenhouse conditions (Tebaldi et al. 2006; Meehl et al. 2005). However, many questions still remain about changes in extreme weather at more local scales and the causes of observed changes. Despite the steady accumulation of evidence of human influence on the hydrosphere (Tett et al. 2007; Hegerl et al. 2007), attribution of precipitation trends to human influence is not yet possible below the global scale (Lambert et al.

2004; Zhang et al. 2007). Attribution is the process of establishing the most likely cause(s) of detected changes at a defined level of statistical confidence (Hegerl et al. 2007). However, it is likely that changes to extreme precipitation will be detectable at smaller spatial scales and, perhaps, within the next 50 years (e.g. for UK observations, Fowler and Wilby 2010). Indeed, changes in moderately extreme precipitation events are, in theory, more robustly detectable than changes in mean precipitation (Frei and Schär 2001) because as precipitation increases (under the greater water holding capacity of a warmer atmosphere) a greater proportion of rainfall is expected to fall as heavy events (Hegerl et al. 2004, 2006; Katz 1999). This may increase the signal to noise ratio and enable more robust detection of changes in extreme rainfall.

Because it is impossible to collect observations for future climate conditions, much of what is known about the Earth's potential response to altered atmospheric conditions comes from climate model output. Global climate modeling is accomplished using General Circulation Models (GCMs), which are numerical models which represent the known physical processes in the atmosphere, ocean, cryosphere and land surface. They are generally run for decades of simulated time and the output includes fields of temperature, precipitation, barometric pressure, humidity and numerous other weather traits every few hours (of simulated time) and at every location using a three dimensional grid over the globe. Climate model output is deterministic in that, if given the same initial conditions, the model will output the exact same measurements. However, because of the chaotic nature of the models and their sensitivity to initial conditions, statistical models can be used to analyze the output in much the same way as they are used to analyze observed weather data.

GCM output is relatively coarse in resolution (e.g. the Hadley Centre's HadCM3 model is resolved at a spatial resolution of 2.5° latitude by 3.75° longitude) and GCMs are thus unable to resolve significant sub-grid scale features (such as topography) and processes (such as those related to clouds). These processes are instead approximated or "parameterized," where their known properties are averaged over the larger scale grid-boxes. Indeed, many uncertainties must be considered within the climate modeling process, including grid resolution, process parameterization, model structure and emissions scenario (e.g. Giorgi and Francisco 2000; Covey et al. 2003) and, for this reason, GCMs may simulate quite different responses, simply because of the way specific processes and feedbacks are modeled. One of our questions of interest is how different GCM parameterizations can affect the nature of extreme precipitation in climate model output.

There is some question as to whether GCM output should be representative of extreme weather phenomena. Due to their coarse resolution, GCMs would not be expected to represent extreme precipitation events with the same intensity and frequency as observations (Kiktev et al. 2003; Räisänen and Joelsson 2001). Wilson and Toumi (2005) derived a simple expression for precipitation as the product of advected mass, specific humidity and precipitation efficiency. The authors show that the tail of the distribution of the product of these three random variables will have a stretched exponential form with a shape parameter of two thirds, leading to estimates having an apparent heavy tail. More importantly, they argued that this shape parameter is unlikely to change under climate change (even if the scale parameter does), as it is invariant temporally with latitude.

Many studies have examined GCM output to investigate changes in tail behavior; examining the future projections from GCMs as indicative of what we may expect from future precipitation extremes (e.g. Tebaldi et al. 2006). Difficulties in comparing observed and modeled extreme precipitation result from both a lack of robustness in the chosen metric due to the infrequent nature of extreme events (the need for enough data—both observed and modeled—to provide a stable estimate of their frequency and intensity) and from the different scales upon which observations and modeled output work—point observations versus coarse areal model output (Osborn and Hulme 1997; Kharin and Zwiers 2005). Therefore, patterns of modeled changes do not necessarily match observed changes, although qualitative similarities have been found in some studies (e.g. Semenov and Bengtsson 2002; Groisman et al. 2005; Tebaldi et al. 2006). Part of this difference is expected since most GCMs do not simulate small-scale (<100 km) variations in precipitation intensity, as occurs with convective storms particularly during summer. However, when GCMs are compared with a reanalysis product (ERA40: a gridded data set at a similar spatial scale to GCMs and representing the state of the Earth's atmosphere, incorporating observations and global climate model output), they are found to reproduce observed precipitation extremes reasonably well over North America (Kharin et al. 2007). Therefore changes projected by GCMs are thought to be fairly robust, even if the model scale implies that they do not exactly match observed extremes. Nevertheless, it is unclear whether the tail behavior of observed precipitation will be well-represented by climate models. In fact, one of the aims of this work is to assess which climate model parameters directly affect the tail behavior of the GCM precipitation output.

Multiple-model or perturbed-physics ensembles offer the best way to explore these uncertainties. A perturbed-physics ensemble is an experiment where a single climate model is run using various parameterizations whereas a multiple-model ensemble is an experiment where multiple climate models with different structures and parameterizations are run. Climate models are extremely computationally expensive to run. Most GCM simulations can only be run at institutions with supercomputing capabilities and a single run can take weeks or months to complete. Therefore, multiple-model or perturbed-physics experiments are generally very limited in scope, and few studies have used multiple climate model outputs to explore change to extreme precipitation (e.g. Fowler and Ekström 2009). This is despite the fact that combining models through a multi-model ensemble generally increases the skill, reliability and consistency of predictions (Tebaldi and Knutti 2007) and allows assessment of the uncertainties within the climate modeling process.

1.2 The climate*prediction*.net (CPDN) perturbed-physics experiment

One of the largest perturbed-physics ensembles is maintained by the climate*prediction*.net (CPDN) project. The CPDN experiment was set up in 2001 (Allen 1999) to investigate the approximations in initial conditions, model parameterizations and forcings that have to be made in state-of-the-art climate models. Rather than running a model on a supercomputer, the CPDN project is a large on-line experiment

that utilizes distributed computing—that is, various ensemble members are run on people's home computers when they are idle (Christensen et al. 2005), producing thousands of climate model runs. In this work, we analyze model output from 304 climate model runs with different parameterizations.

So far, most of the evaluation of CPDN output has concentrated on understanding the effects of parameter variation in the climate model on climate sensitivity—the global mean temperature response to a doubling of atmospheric CO_2. Most of these effects have been found to be due to a small subset of the parameters mostly concerned with cloud dynamics: notably, the entrainment coefficient in clouds (entcoef), the rate at which convective clouds mix with the surrounding air, was associated with 30% of the variation in climate sensitivity by Knight et al. (2007). Stainforth et al. (2005) additionally found relationships between the critical relative humidity (rhcrit) and the ice fall speed (vf1) and climate sensitivity.

In this paper we use General Circulation Model (GCM) outputs from the CPDN BBC Climate Change Experiment (CCE), to examine when formal detection of change in extreme precipitation may be possible within the UK. The CPDN BBC CCE has produced pairs of GCM runs under transient forcing for the 1920–2000 and 2000–2080 time periods (Frame et al. 2009)—transient forcing means that a different forcing is applied for every year of the simulation. These are collectively known as the Scenario run from 1920 to 2080. The 1920–2000 run is based on observed forcings, whereas the 2000–2080 run is based on the SRES A1B emissions scenario (a mid-range scenario where atmospheric CO_2 reaches $\sim$720 ppm by 2100; Nakicenovic et al. 2000). In addition a Control of the same length (160 years) was run for each GCM ensemble member, using the same initial conditions and parameter values. The Control simulations correspond to an unforced or stationary climate in the year 1920. This was so that each physically distinct model in the experiment could be checked for spurious model drifts (Frame et al. 2009) and also to provide an unforced pre-industrial climate against which change can be detected. Each 'data set' therefore consists of a pair of 160-year climate model runs: the Control and the Scenario, where each pair has a different combination of the 34 climate model parameters which are systematically varied (Table 1). See Frame et al. (2009) and the "Appendix" for further details.

Due to the magnitude of the experiment, only a limited amount of output can be retained for each climate model run. For most of the globe, output is averaged over the so-called Giorgi regions (sub-continental scale regions for which monthly summary data from the included model grid cells are retained; Giorgi and Francisco 2000). Grid cell level data is only retained for eight cells that lie over the UK at the monthly time scale. As our interest lies in understanding extreme precipitation we use the monthly maximum data. However, the maximum precipitation output for the Giorgi regions is further complicated, as the maximum measurement at each grid cell for each month is then averaged over the spatial domain of the region. It is questionable whether these data should follow an extreme value distribution, so here we limit our focus to the monthly maximum data for the eight UK grid cells (Table 2). Rather than analyze the monthly maximum daily precipitation amounts, we analyze the seasonal maxima, increasing our block size to roughly 90 days. This also allows for easy comparison with previous UK observational and modeling studies

Table 1 The parameters varied in the CPDN BBC CCE and the levels used for each in the one-way ANOVA analysis

Parameter name	Explanation	Levels
alpham	Albedo at melting point of ice	3
anthsca	Sulfur cycle: scaling factor for emission from anthropogenic sulfate aerosols	5
cloudtau	Time a circulating air parcel remains in a cloud	3
ct	Accretion constant	3
cw_land	Precipitation threshold over land	3
cw_sea	Precipitation threshold over sea	3
dtheta	IC ensemble: initial condition parameter	10
dtice	Temperature range of ice albedo variation	3
eacf	Empirically adjusted cloud fraction	3
entcoef	Entrainment coefficient	4
file_flux	Ocean: heat and salinity flux adjustment file particular to ocean spin up	10
file_nick	Ocean: heat and salinity flux adjustment file particular to atmospheric physics configuration	90
file_ocean	Ocean start file	10
file_solar	Solar_v01	9
file_volcanic	Volcanic forcing scenario	50
haney	Ocean: haney heat forcing coefficient	2
haneysfact	Ocean: haney salinity forcing factor	2
i_cnv_ice_lw	Type for convective ice	2
i_cnv_ice_sw	Type for convective water	2
i_st_ice_lw	Type for stratiform ice	2
i_st_ice_sw	Type for stratiform water	2
ice_size	Ice size in radiation	3
isopyc	Ocean: isopycnal diffusion of tracer at surface	3
l0	Sulfate mass scavenging parameter L0	3
l1	Sulfate mass scavenging parameter L1	3
mllam	Ocean: wind mixing energy scaling factor	2
num_star	Threshold for condensation onto accumulation mode particles	3
rhcrit	Critical relative humidity	3
so2_high_level	Sulfur cycle: model level for SO2 (high level) emissions	3
vdiffdepth	Ocean: increase of background vertical mixing of tracer with depth	3
vdiffsurf	Ocean: background vertical mixing of tracer (diffusion) at surface	3
vertvisc	Ocean: background vertical mixing of momentum (viscosity)	2
vf1	Ice fall speed	3
volsca	Sulphur cycle: scaling factor for emission from natural (volcanic) emissions	3

Table 2 The eight UK grid cells used in the study and descriptions of their location and latitude–longitude

Name	Description	Latitude (min)	Latitude (max)	Longitude (min)	Longitude (max)
ukhl	Highlands	56.25	58.75	−5.625	−1.875
ukni	N. Ireland	53.75	56.25	−9.375	−5.625
ukne	Borders	53.75	56.25	−5.625	−1.875
eire	Ireland	51.25	53.75	−9.375	−5.625
ukwm	Wales and Midlands	51.25	53.75	−5.625	−1.875
ukla	London and East Anglia	51.25	53.75	−1.875	1.875
ukcw	Cornwall	48.75	51.25	−5.625	−1.875
ukkt	Kent	48.75	51.25	−1.875	1.875

(e.g. Osborn et al. 2000; Fowler and Kilsby 2003b; Fowler et al. 2005, 2007; Ekström et al. 2005; Fowler and Ekström 2009). We analyze each season separately due to the inherent seasonal effects of weather data. The spatial extent of each 'data set' or model pair thus represents one grid cell of the eight GCM grid cells evaluated over the UK. As 304 climate model pairs were used, in total, 9,728 (304 model pairs × 8 grid cells × 4 seasons) combinations were analyzed.

In Section 2, we examine the structure of the statistical model needed to describe changing transient extreme precipitation through the fitting of a GEV distribution with time-varying parameters. In Section 3, we examine the changes in extreme precipitation projected for 2000–2080 by the perturbed-physics climate model ensemble. In Section 4, we then develop a methodology for estimating detection times for changes in seasonal precipitation extremes and apply this to a large ensemble of climate model pairs, much larger than has previously been examined in this regard. In Section 5, we examine the dependence of climate model simulations of extreme precipitation on parameter variations, the nature of this dependence and discuss these results in comparison to other studies. In Section 6 we then conclude the paper.

2 Fitting a statistical model to describe transient extreme precipitation

The climate*prediction*.net experiment retains only summary information from each climate model run including mean precipitation, mean temperature, maximum daily precipitation (for each month), and others, recorded for each month for each year of the simulation. As the only extreme precipitation data we have access to are the maximum daily values for each month, it is natural to perform a block maximum analysis and fit the generalized extreme value (GEV) distribution. Here, we fit using the seasonal daily maximum value. That is, for each season (winter = December, January, February; spring = March, April, May; summer = June, July, August; fall = September, October, November) of each year, we retain the maximum daily precipitation value. There are several excellent sources on extreme value theory and fitting block maximum data including Beirlant et al. (2004), Coles (2001), and deHaan and Ferreira (2006). The GEV is described by three parameters: a location

parameter μ, a scale parameter $\sigma > 0$, and a shape parameter ξ, and has a cumulative distribution function (cdf) given by:

$$P(X \leq x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \text{where} \quad 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0 \quad (1)$$

Since the Control run has an unforced climatology (that of 1920), we assume that the climate is stationary and thus choose a time-invariant model for this data. As the Scenario run is forced by an observed climatology up to 2000 and by the SRES A1B emissions scenario after 2000, we investigate models in which the GEV parameters change with time and allow for a change in behavior in the year 2000.

Fitting extreme value models with time-varying parameters has become a well-accepted practice, and Chapter 7 of Beirlant et al. (2004) and Chapter 6 of Coles (2001) both give excellent overviews. Commonly, a simple parametric form is given to the time varying parameter, for instance, the location parameter may be modeled as $\mu_t = at + b$. In recent work some authors (e.g. Chavez-Demoulin and Davison 2001) have advocated a non-parametric approach which allows for more flexibility in modeling parameter behavior through time. Here, we restrict our attention to simple parametric models as we must have an automated model selection procedure due to the number of data sets to which we fit models.

To simplify notation, we will assume that we are analyzing data from a particular climate model pair, season, and grid cell. Let t denote the number of years since 1920. We denote the random variable corresponding to the annual maximum for the Control run for year t as $X_{c,t}$, and denote the observation as $x_{c,t}$. The corresponding random variable for the Scenario run is denoted $X_{s,t}$.

Our goal is to find a statistical model which will capture the difference in behavior between the Scenario and Control runs across the different simulation model runs. A priori, we believe that the parameter that is most likely to change between these runs is the location parameter, μ. We proceed using a model selection process that increasingly adds complexity to this parameter. The statistical model selection model selection exercise is applied to all 9,728 datasets.

Table 3 Models tested in the statistical model selection procedure

Model 1	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$
Model 2	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_s, \sigma_c, \xi_c)$
Model 3	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_c + at, \sigma_c, \xi_c)$
Model 4	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_s + at, \sigma_c, \xi_c)$
Model 5	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_c + at + bt\, I_{\{t>2{,}000\}}, \sigma_c, \xi_c)$
Model 6	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu_s + at + bt\, I_{\{t>2{,}000\}}, \sigma_c, \xi_c)$
Model 7	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu^*, \sigma_s, \xi_c)$
Model 8	$X_{c,t} \sim \text{GEV}(\mu_c, \sigma_c, \xi_c)$	$X_{s,t} \sim \text{GEV}(\mu^*, \sigma_s, \xi_s)$

Here, t denotes number of years since 1920, and μ^*denotes the best fitting of statistical models 1–6

The tested statistical models are given in Table 3. Model 1 implies there is no change between the Control and Scenario runs, and model 2 summarizes the change as a level shift. Models 3 and 4 both allow the Scenario to have a linear trend in μ. Model 3's trend begins at the same level as the Control (as both the Control and Scenario begin with the same climatology), and model 4's linear trend has no restriction on the intercept. Models 5 and 6 allow for a change in behavior of μ at the year 2000 when the Scenario changes from observed to hypothesized forcings.

Both the Control and Scenario data sets are fit simultaneously via maximum likelihood as some of the models in Table 3 share parameter values between the Control and Scenario models. The various models are then compared using the *AICc* criterion. The *AICc* is a standard tool for model selection based on the likelihood value and which penalizes for the number of model parameters, where a lower score implies a better fit (Burnham and Anderson 2002). It is given by the relation:

$$AICc = 2\log(L) + 2k + 2k\frac{(k+1)}{(n-k-1)}, \tag{2}$$

where L denotes the likelihood, k the number of parameters, and n the number of observations. While our belief is that the location parameter will capture much of the difference between the Control and Scenario runs, we also test to see if changes can be detected in the scale and shape parameters. Rather than test all possible combinations of μ, σ, and ξ we simply take the best fitting of models 1 through 6 and see if allowing σ or both σ and ξ to change improves the fit (models 7 and 8 in Table 3).

Two examples of this model selection procedure are shown in Fig. 1. The left panel of Fig. 1 shows the selection for model pair 1, grid cell 1 and season 1 (winter) and the right panel shows the selection for model pair 1, grid cell 8 and season 1 (winter). Using the *AICc* criterion, model 5 is chosen to be the best at grid cell 1 but

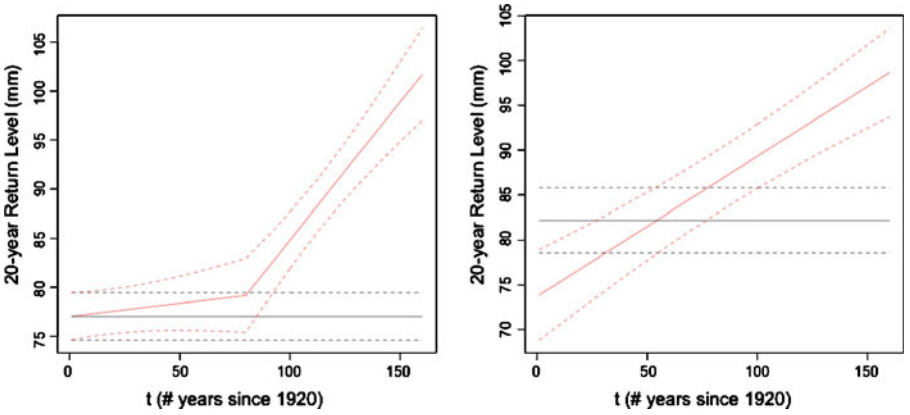

Fig. 1 Return level estimates (*solid*) and 95% confidence intervals estimated via the delta method (*dashed*) of statistical model 5 for the Control (*black*) and Scenario (*red*) for the simulation of climate model pair 1, grid cell 1, and season 1 (*left*) and for statistical model 4 for the Control (*black*) and Scenario (*red*) for the simulation of climate model pair 1, grid cell 8, and season 1 (*right*).

Table 4 Results of the model selection exercise for simulation climate model pair 1, grid cell 1, and season 1 for which model 5 was found to be best (left) and for simulation climate model pair 1, grid cell 8, and season 1 for which model 4 was found to be best (right)

Model pair 1	Season 1	Grid cell 1		Model pair 1	Season 1	Grid cell 8
	−2*llh	k	AICc	−2*llh	k	AICc
Model 1	2,547.778	3	2,553.854	2,637.648	3	2,643.724
Model 2	2,529.329	4	2,537.456	2,633.726	4	2,641.853
Model 3	2,480.919	4	2,489.046	2,604.304	4	2,612.431
Model 4	2,473.932	5	2,484.123	2,591.575	5	2,601.766
Model 5	2,463.743	5	2,473.935	2,595.880	5	2,606.071
Model 6	2,463.721	6	2,475.990	2,590.719	6	2,602.988
Model 7	2,462.069	6	2,474.337	2,589.580	6	2,601.848
Model 8	2,461.709	7	2,476.068	2,587.848	7	2,602.207

In the table, llh stands for log-likelihood and k is the number of parameters

model 4 is deemed best at grid cell 8. Figure 1 shows the point estimate and 95% confidence intervals for the return levels as a function of year and Table 4 details the results of the model selection exercise for these two examples. The 95% confidence intervals are estimated via the delta method (Oehlert 1992).

Results from the model experiment are summarized in Figs. 2, 3, and 4 for the 9,728 different model pairs. Figure 2 shows the histogram of the model (1–6)

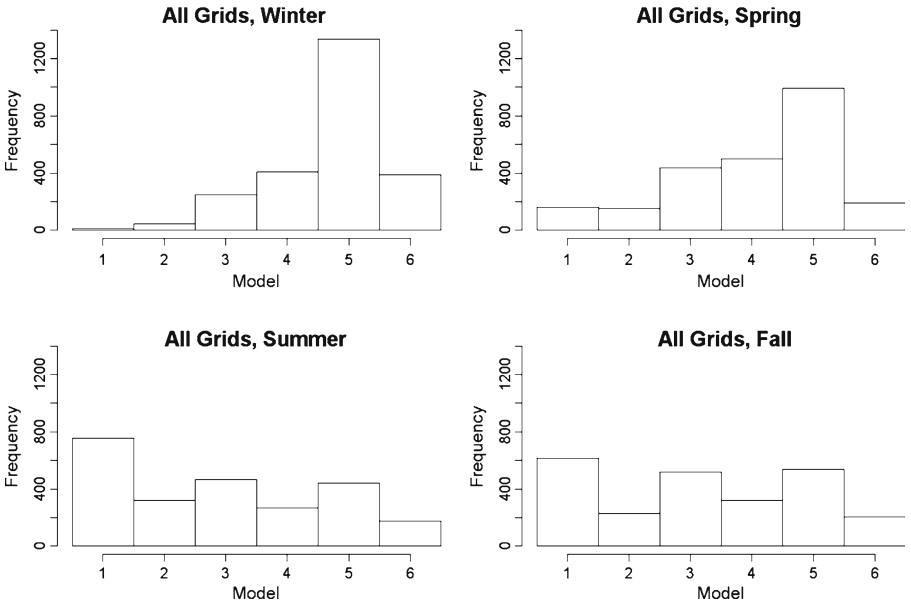

Fig. 2 Histogram of selected statistical model 1–6 for the four seasons combined for all grid cells.

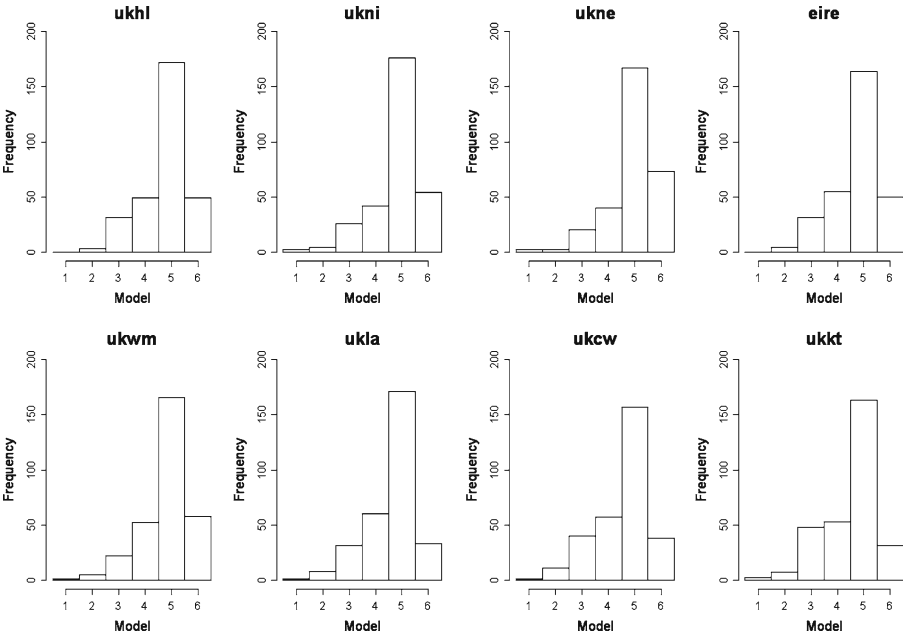

Fig. 3 Histogram of selected statistical model 1–6 for each grid cell (region) for the winter season.

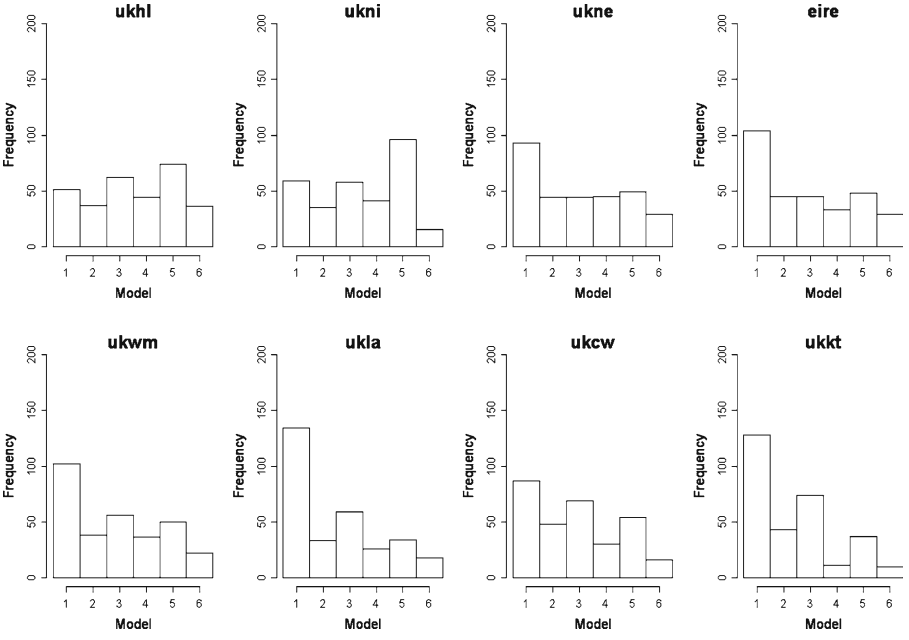

Fig. 4 Histogram of selected statistical model 1–6 for each grid cell (region) for the summer season.

chosen as best by the *AICc* criterion. This only selects among the 6 different statistical models for the location parameter μ, holding σ and ξ constant across the Control and Scenario runs. Model 5 was most often chosen as the best model in both the winter and spring seasons and this agrees with what we know about how the data were simulated, with a common starting point. There is an apparent different behavior between the seasons, with winter and spring tending to choose hinge-type models (5 or 6), while for summer and fall a common model for the Control and Scenario runs is more likely to be selected. Thus, there is less evidence of a change in extreme precipitation in future climate projections over the UK for summer and fall than in spring and winter.

It was also examined whether the chosen model was improved by allowing σ and ξ to vary. In winter, 50% of model 7 and model 8 had a lower *AICc* score than the selected model (1–6) which only modeled a difference in the μ parameter. The results for the other three seasons were: spring (36%), summer (39%), fall (39%). To keep things simple in this initial investigation, we therefore restrict our attention to the first six statistical models in the subsequent results.

Figure 3 shows the selected model (1–6) for each of the eight grid cells in the winter season. Likewise, Fig. 4 shows the selected model (1–6) for each grid cell in the summer season. There appears to be more consistent behavior between the grid cells for the winter season than for the summer season, with very few selections of either model 1 or 2 suggesting that change in extreme precipitation is projected for the Scenario run. The overwhelming selection of model 5 or 6 for the winter season (Fig. 3) for all grid cells also suggests that the trend in the μ parameter changes in the year 2000 and implies that extreme precipitation may increase at a greater rate after 2000. Figure 4 suggests that there is less consistent behaviour among the eight UK grid cells in summer with the majority of selections choosing model 1 or 2 and thus suggesting no change in extreme precipitation in the Scenario run, although in the ukhl and ukni regions (UK Highlands and Northern Ireland) there is a shift towards the selection of model 5 or 6.

3 Projected changes in extreme precipitation

Before estimating the detection times, it is useful to consider the changes in seasonal extreme precipitation projected by the CPDN ensemble as larger changes, in general, may imply earlier detection times (Fowler and Wilby 2010). The goal of this section, therefore, is to simply summarize the projected changes in seasonal extreme precipitation across the 304-member CPDN ensemble.

For each ensemble member, the best GEV model fitted in Section 2 to the Scenario climate model run was used to provide point estimates for the 20-year return level at 2020, 2050 and 2080. This estimate was, in each case, then divided by the point estimate for the 20-year return level at 1975 (approximating the baseline from 1961 to 1990) from the same GEV model fit to provide an estimate of the percent change by 2020, 2050 and 2080 respectively. This was performed across the 304-member CPDN 'Scenario' model ensemble and is presented as box and whisker

plots in Figs. 5, 6, 7 and 8 which show the mean and uncertainty in the projected changes to the 20-year return level of 1-day extreme precipitation totals in winter, spring, summer and fall respectively for each of the eight UK grid cells (regions). The 20-year return level was chosen arbitrarily as an example; previous work using regional climate model projections has projected even larger changes at higher return levels (e.g. Fowler and Ekström 2009) although as higher return level estimates have higher sampling uncertainty, so detection may be more difficult.

Figure 5 shows that winter 1-day extreme precipitation is projected to increase across the UK. The median change by 2020, 2050 and 2080 from the 1961 to 1990 baseline is projected to be an increase of around 5%, 10% and 20% respectively. Although there is large uncertainty surrounding these projections, particularly by 2080, the majority of the CPDN ensemble members project substantial increases in 1-day extreme winter precipitation over the next 70 years.

In spring, the majority of CPDN ensemble members again project increases in the 20-year return level of 1-day extreme precipitation (Fig. 6). However, the projected changes are not as large as for winter, with median changes of less than 20% in all UK grid cells projected by 2080.

In summer and fall, a similar pattern of change emerges for the UK, with small increases projected for northern and western grid cells (UK Highlands and Northern Ireland), no change projected for north-east and north-west England (ukne, nkwm

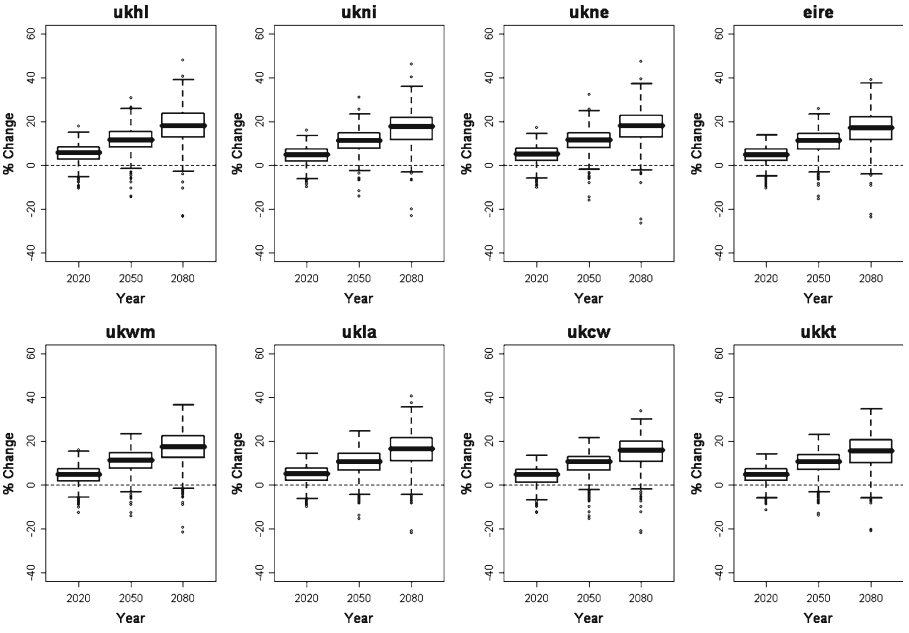

Fig. 5 Box and whisker plots showing the mean and uncertainty in percentage changes to the 20-year return level of 1-day extreme precipitation by 2020, 2050 and 2080 from the 1961 to 1990 baseline projected by the CPDN BBC CCE for winter.

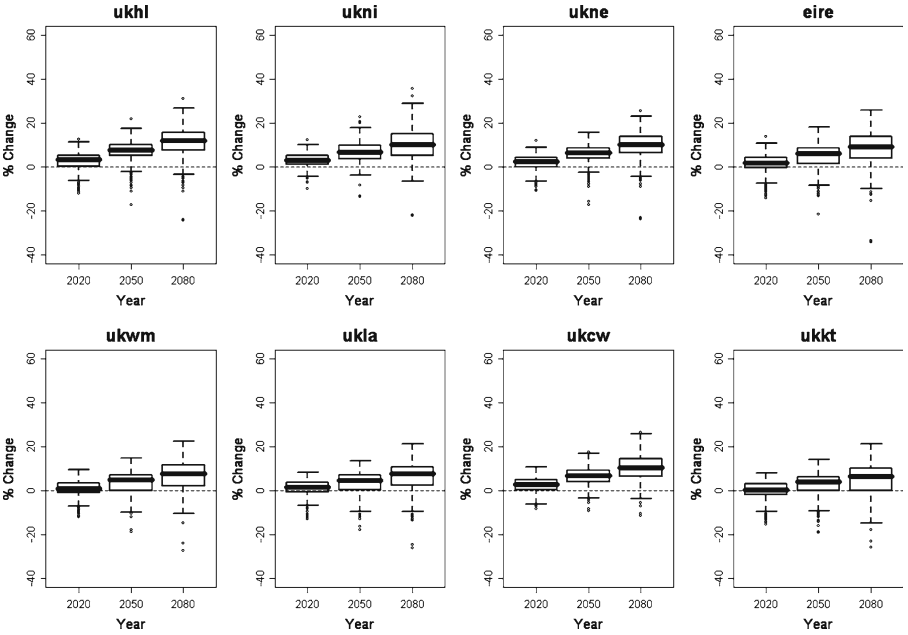

Fig. 6 As in Fig. 5 but for spring.

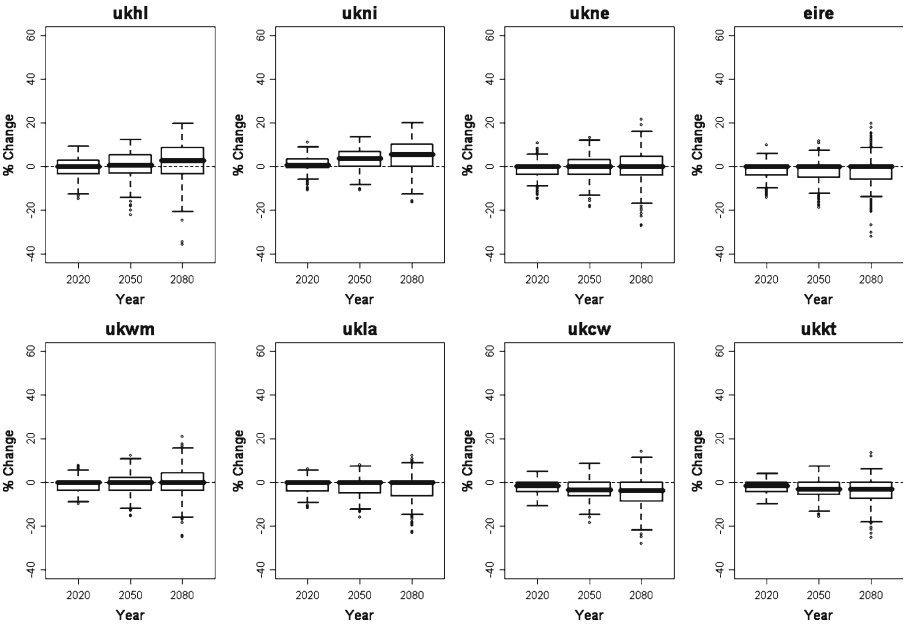

Fig. 7 As in Fig. 5 but for summer.

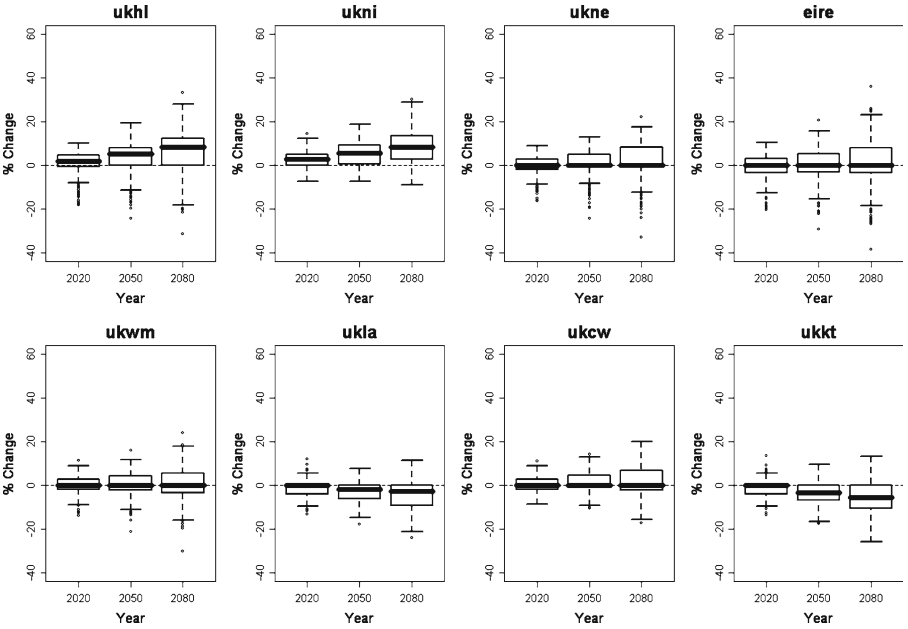

Fig. 8 As in Fig. 5 but for fall.

and ukla) and small decreases projected for southern grid cells such as Cornwall and Kent (Figs. 7 and 8 respectively). Median projected increases in the 20-year return level of 1-day extreme precipitation by 2080 in northern and western grid cells is around 10%, whereas the projected decreases in southern grid cells are of a similar magnitude.

4 Estimating time of a detectable difference in extreme precipitation

We define a detectable increase in extreme precipitation as the point when we would reject (at the $\alpha = 0.05$ level) the null hypothesis that the 20-year return levels from the two runs are equal in favor of the alternative hypothesis that the 20-year return level from the Scenario run is greater than that from the Control run, thus detecting a change from the background climatology pre-1920. As we are estimating the parameters for the Control run and Scenario runs simultaneously, we can estimate the information matrix, and calculate the associated error of our estimate for the difference in return levels via the delta method. Note that to define detectability we could have chosen other return levels but chose to only use one return level as an example. In other work (Fowler and Wilby 2010) it has been noted that as the return level increases the time to detection also increases.

Figure 9 shows two examples of this process. In this figure the 90% confidence interval is shown because of correspondence to the one-sided test of significance

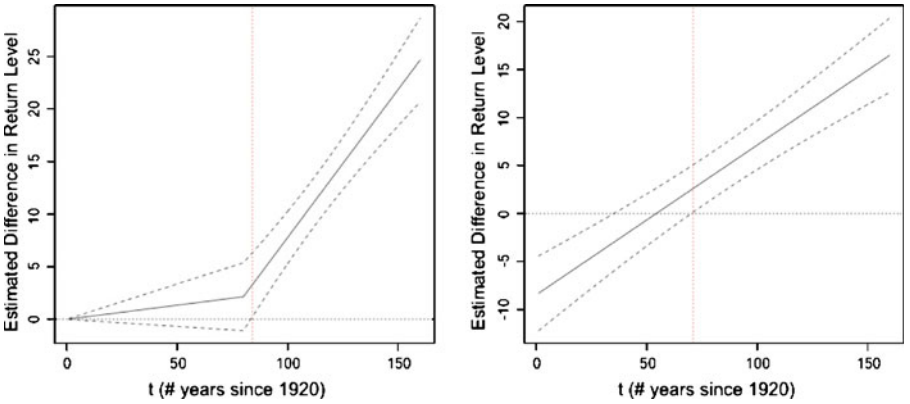

Fig. 9 Illustration of our definition of detected difference. Plot shows the estimated difference in return levels and 90% confidence intervals. Climate model pair 1, grid cell 1, season 1 (*left*) has a detected difference at 84 years. The confidence bands narrow to zero when t equals zero as the Control and Scenario runs in statistical model 5 have the same parameters when t = 0. Climate model pair 1, grid cell 8, season 1 (*right*) was fit best with statistical model 4 and has a detected difference at at 71 years.

at the 5% level. For the first example, model 5 was chosen to be the best model. According to model 5, the Control run and the Scenario run have the same parameters at the start, thus the difference in return levels is zero initially. As time progresses the Scenario run's model parameters change, and the uncertainty increases. A detectable difference was found at year 84 (2004). In the second example, model 4 was chosen to be best. Model 4 is not a hinge model, and when fit to this data, it found that the Scenario run had lower initial return levels than the Control run. By year 71 (1991), it was found that the Scenario run had a significantly higher 20-year return level.

Figure 10 summarizes the results of this exercise across model pairs and grid cells for the winter, spring, summer and fall seasons using empirical cumulative distribution functions (cdfs) of the time of detected difference. That is, letting t denote year and D_i denote the time of the detected difference for model i as defined above, then we plot $n^{-1} \sum_{t=1}^{n} I\,(D_t > t)$ verses t where I denotes the indicator function. The cdfs do not reach a height of 1 as in no season do all model pairs detect a difference in extreme precipitation in the 160-year simulation run. For the winter season, more than 50% of the climate model pairs found a detectable difference by ∼90 years (2010), although ∼18% of the data sets showed a detectable difference at year 1 (1920: caused by level shifts in the fitting of statistical models 2 or 4), and ∼8% of the datasets did not detect a difference in return levels over the 160 year model run. For the spring season, more than 50% of the climate model pairs found a detectable difference by ∼110 years (2030), although ∼22% of the data sets showed a detectable difference at year 1 (1920: caused by level shifts in the fitting of statistical models 2 or 4), and ∼20% of the datasets did not detect a difference in return levels over the 160 year model run. However, for the summer and fall seasons respectively, only ∼30% and 40% of the data sets showed a detectable difference over the 160-year model run. This difference in behavior between the summer/fall and winter/spring

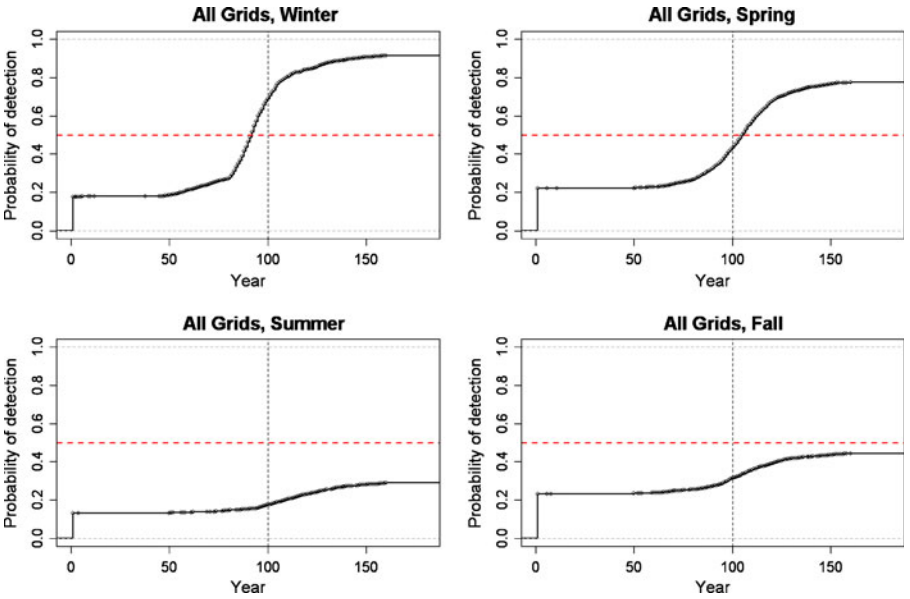

Fig. 10 Empirical cumulative distribution functions (cdfs) of the time of detected difference for the winter, spring, summer and fall seasons. The *vertical black dotted line* indicates the year 100 (real year 2020) and the *horizontal red dashed line* the point beyond which the probability of detection is more likely than not. The cdf does not reach 1 in any season as there are some climate model pairs for which no difference was detected over the course of the model run. Equally, for some climate model pairs a statistical model was chosen which produced a difference at year 1.

seasons is not unexpected, especially given the known difficulties that climate models have in resolving convective precipitation.

5 Climate model parameter effect on the shape parameter and time of detection

One of the primary goals of CPDN is to understand the effect that different climate model parameters have on the simulated climate and, as such, the CPDN ensemble uses a parameter sampling strategy that chooses one of a small number of possible values for each parameter detailed in Table 1. In terms of describing extreme precipitation, a parameter of primary interest is the GEV shape parameter, ξ, as it controls the weight of the tail. To our knowledge, it is not well understood which climate model parameters could possibly affect ξ.

To attempt to answer this question, we ran a sequence of simple one-way analysis of variance (ANOVA) tests for each of the 34 different model parameters found in Table 1. The analysis was run only on the Control climate model runs as we wish to limit our focus to the effect of the model parameters on the simulation of extreme precipitation rather than possible effects of the different forcing scenarios. In each ANOVA, ξ as estimated by the 2,432 different simulations serves as our variable and

the different parameter settings serve as treatment. Normality and equality of variance of the ξ's were checked for a handful of the model parameters, and the standard ANOVA assumptions could not be rejected. Regardless, the ANOVA analysis here is done only as a data mining exercise rather than as a traditional hypothesis test, and the p-values simply tell us which climate model parameters are worthy of further investigation rather than as a strict assessment of statistical significance.

Each season's data was analyzed separately, and we performed the analysis across the eight grid cells, assuming that the effect of altering the climate model parameter did not interact with location. For each ANOVA, there were 2,432 estimated shape parameters unequally divided into approximately three or four treatment levels for each climate model parameter (see last column of Table 1). Because the sample size was huge, the ANOVA could detect very small differences in the means of the treatment groups and assess them as significant. Most of the ANOVA analyses returned a p-value of less than 0.01. To better assess the *importance* of the differences in treatment means, we employ the statistic ω^2, which is an effect size measure for one-way ANOVA (Olejnik and Algina 2003). Like r^2 in a regression analysis, ω^2 can be interpreted as the percentage of the total variability explained by the treatment effect. The definition is given by:

$$\omega^2 = \left(\sigma^2_{total} - \sigma^2_{error}\right) \Big/ \sigma^2_{error} \tag{3}$$

where $\sigma^2_{error} = E\left(Y_{ij} - \mu\right)^2$ and $\sigma^2_{error} = E\left(Y_{ij} - \mu_i\right)^2$, and the common estimator is: $\widehat{\omega^2}$ = (ssEffect − dfEffect × msError) / (ssTotal + msError) (Hays 1973, page 484). Here ssEffect is the sum-of-squares due to the treatment effect (i.e. between-treatment variance), dfEffect is the degrees of freedom associated with the treatment effect, the treatment effect, msError = ssError/dfError which are all associated with the within-treatment variance, and ssTotal is the total sums-of-squares. These quantities can all be found in a standard ANOVA table.

Despite having low p-values indicating significant differences, the importance ($\widehat{\omega^2}$) associated with these differences for most climate model parameters was also found to be small. For most climate model parameters, the difference in treatment means explained less than 2% or 3% of the total variability in the ξ's. However, two climate model parameters were found to explain a non-negligible amount of the variability in the GEV shape parameter estimates. Figure 11 shows box plots for the different treatments for the climate model parameters "entcoef" (entrainment coefficient) and "vf1" (ice fall speed) for the summer season. The respective ω^2 values for these two parameters were found to be 0.35 and 0.09. That is, these two climate model parameters (each when analyzed independently of all other parameters) explain respectively 35% and 9% of the total variability found in the estimates of the shape parameters. Entcoef has a positive relationship with the GEV shape parameter whereas vf1 has a negative relationship. Similar, but less dramatic relationships can be seen for these same two parameters for the spring and fall seasons, and the effect is minimal for entcoef and negligible for vf1 for the winter.

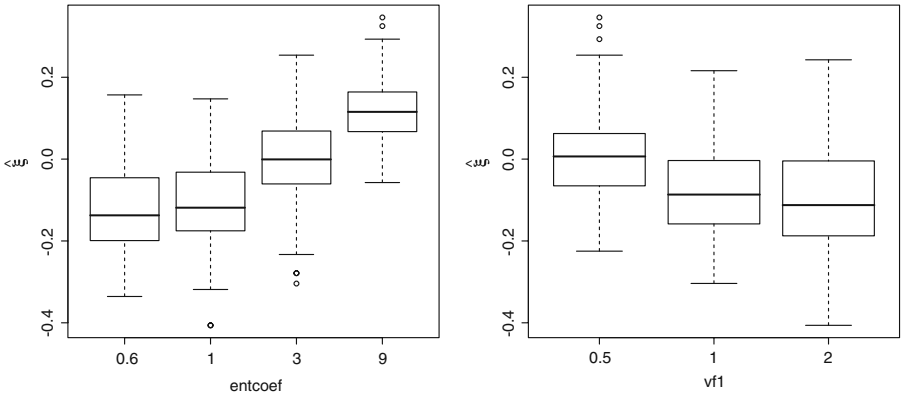

Fig. 11 Box plots showing different treatments of the climate model parameters "entcoef" (*left*) and "vf1" (*right*) and their effect on the estimated shape parameter for the summer season. These two variables were detected by the data mining exercise which employed a one-way ANOVA analysis.

The above results are interesting in that (to our knowledge) no one has before tied tail behavior associated with simulated extreme precipitation events to any particular climate model parameters although much is known about heavy precipitation behavior from observational and theoretical analyses (e.g. Wilson and Toumi 2005). Previous investigations into CPDN ensemble parameters have suggested that optimum values of two parameters—the entrainment coefficient (entcoef) and the ice fall speed (vf1)—show first order sensitivity to climate sensitivity—the mean global temperature response to a doubling of atmospheric CO_2 (Sanderson et al. 2008). Indeed, all CPDN investigations suggest that entcoef is dominant in establishing relative humidity profiles that lead to strongly different responses to greenhouse gas forcing (Sanderson et al. 2008; Stainforth et al. 2005; Sanderson and Piani 2007; Knight et al. 2007).

It is well known that the variation in humidity and atmospheric circulation are critical in determining the occurrence of heavy precipitation. However, Wilson and Toumi (2005) suggest that precipitation efficiency must play a critical role in the occurrence of heavy precipitation. Precipitation efficiency is controlled by cloud microphysics, entrainment, detrainment, the water holding capacity of the column, and large-scale divergence above the moisture level (Trenberth et al. 2003). The entrainment coefficient in HadCM3L affects how air is diluted in rising cumulus cloud columns and has a big impact on the global top-of-the-atmosphere energy budget as well as on climate sensitivity. Therefore, the value of entcoef partially controls the amount of convective activity (Gregory and Rowntree 1990) and it makes physical sense that it would affect intense summer precipitation, which is presumably mostly convective (Myles Allen, personal communication). For prediction of changes to heavy precipitation, Wilson and Toumi (2005) assert that predicting the future evolution of the precipitation efficiency is perhaps at least as important as predicting humidity changes. Our results suggest that entcoef also has an important effect on

heavy precipitation generation in climate models. However, parameterizing (choosing a particular value for) this variable will be challenging as it is poorly quantified even for our current climate (Wilson and Toumi 2005).

The ice fall speed in clouds, "vf1", has also been associated with a significant percentage of the variation in climate sensitivity and has a major impact on cloud cover and cloud optical properties (Myles Allen, personal communication); a large value of this parameter allows the fast fallout of cloud ice (Sanderson et al. 2008). Sanderson and Piani (2007) found that reducing the ice fall speed parameter resulted in increased long-wave clear sky and increased low-level layer clouds, allowing the air to remain moister. This 'moistening' effect causes the simulation of increases in extreme precipitation in the climate model output in the same way as the effect due to increased warming. In the future climate, warmer air will be able to hold more moisture generated by increased evaporation from warmer oceans. When this moister air moves over land, more intense precipitation is produced (Meehl et al. 2005). Nevertheless, it is interesting that for extreme precipitation, these climate model parameters change the distribution not only in terms of a possible location and/or scale shift, but also by changing the rate at which the tail decays.

In addition to the effects that parameters entcoef and vf1 have on the GEV shape parameter, Fig. 11 also gives interesting information about the values of the shape parameter estimates. It is clear from both box plots that many of the CPDN BBC CCE climate model runs produce a negative estimate of the shape parameter. Most studies of observed extreme precipitation yield shape parameter estimates that are slightly positive (usually in the range of about 0.0 to 0.2), which indicates heavy tailed-behavior (e.g. Fowler and Kilsby 2003a). It is accepted in the climate modeling community that climate models do not model convective precipitation well, and perhaps this causes the simulated extreme precipitation to have this fundamentally different behavior.

Returning to our initial idea of investigating the time of a detectable difference in extreme precipitation, we also investigated if any of the climate model parameters had an effect on when a change in extreme precipitation would be detectable. Here, of course, we analyze the model pair data sets (Control and and Scenario). We did not consider our response variable to be continuous as it is possible to never have a detectable change during the 160 years of the climate model run and also to have a detectable change at 1920 (year 1: where statistical model 2 or 4 is fitted). Instead, we perform a simple contingency table analysis. We bin the response variable into four categories: (1) detect at 1920 (year 1), (2) detect before 2000 (by year 80), (3) detect after 2000 (after year 80), and (4) no detection by 2080 (by year 160). We perform the analysis only for the winter season as this is when we see the greatest detection rate and also the greatest variability among the climate models for when a detectable change occurs.

We perform the standard chi-squared test for contingency tables and, because of the number of separate tests we run, reject the null hypothesis of independence only if the p-value associated with the test is less than 0.01. Only two climate model parameters are found to have p-values low enough to reject the null hypothesis. The parameter "ct", which describes the accretion constant, has a p-value of 1e-4, and the

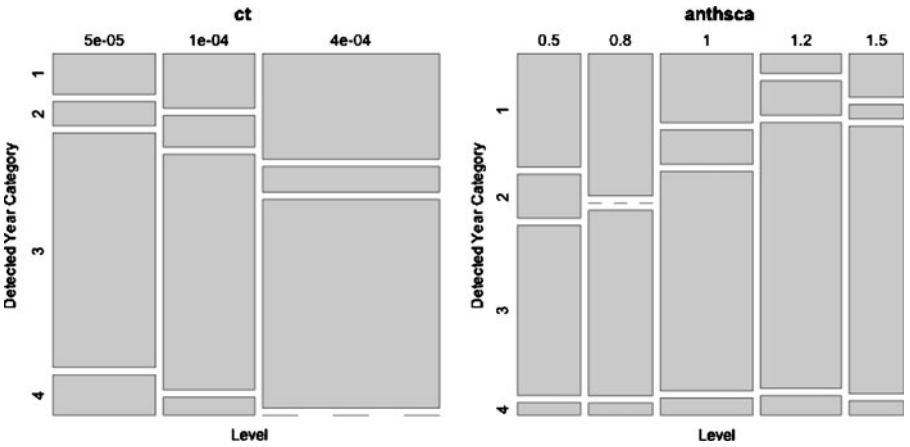

Fig. 12 Contingency table. Areas of each block correspond to the percentage of models that correspond to each category. The *horizontal direction* corresponds to the different levels of the climate model parameters "ct" (*left*) and "anthsca" (*right*). The *vertical direction* corresponds to the year that a change in extreme precipitation was detectable (1—detect at year 1, 2—in the time period 1920–2000, 3—in the time period 2000–2080, 4—no detect).

parameter "anthsca", which describes the scaling factor for emissions from anthropogenic sulfur aerosols, has a p-value of 2e-4. Figure 12 summarizes the data in the contingency tables for each of these significant climate model parameters. The plot for ct shows that an increase in this parameter corresponds to an earlier detectable change in extreme precipitation. The pattern for anthsca is a bit less evident, but it appears that increasing this parameter tends to result in a later detectable change.

6 Discussion and conclusions

In this paper we have investigated a methodology for estimating detection times for changes in seasonal precipitation extremes and applied this to CPDN BBC CCE outputs from the HadCM3L GCM. These parallel climate runs allow us to compare model output forced by observed climatology from 1920 to 2000 and the SRES A1B emissions scenario from 2000 to 2080 to model output with an unforced climatology for 1920. We have investigated the statistical models that are best applied to describe extreme precipitation from transient GCM runs in different seasons for grid cells over the UK and the changes projected by those models for 2020, 2050 and 2080. Additionally, we have investigated the climate model parameters that most affect the simulation of extreme precipitation.

We used 304 climate model pairs, providing 9,728 unique combinations for analysis. For winter and spring, the statistical model most frequently chosen by the *AICc* criterion suggested a change in the GEV location parameter at the year 2000, that is, when the forcings switch from observed to the SRES A1B emissions scenario. However, in summer and fall the statistical model most commonly chosen suggested that

there was no difference between the Control and Scenario, or that there was no trend in GEV location parameter within the Scenario. Although there was evidence of significant trend in the GEV scale and shape parameters in the winter season, there was less evidence in other seasons. Therefore, statistical models were restricted to fitting time-varying trends in the GEV location parameter only in this initial investigation.

The CPDN BBC CCE ensemble projects substantial changes to the 20-year return level of 1-day extreme precipitation during the next 70 years, with median changes in winter approximating increases of 5%, 10% and 20% for 2020, 2050 and 2080 respectively when compared to 1961–1990. These changes are comparable in magnitude to those projected for the UK by Regional Climate Models (RCMs) from the PRUDENCE ensemble (Fowler et al. 2007; Fowler and Ekström 2009). Increases in spring are projected to be lower than winter and projected changes in summer and fall range from small increases to small decreases depending on the grid cell; increases are commonly seen in northern grid cells and decreases in southern grid cells.

Detecting significant differences in extreme precipitation is related to the magnitude of the projected change as well as the statistical model best fitted to the Scenario run. For winter, a detectable difference from the 1920 background climatology at the 20-year return level is found by most models within the 160 year window and at ∼90 years (2010) for more than 50% of the models. However, the majority of the models did not detect a difference over the 160-year climate model run in summer and fall. Although we do not have good observations for many rain gauges back to 1920, other work using RCMs to determine when a climate change signal may be detectable in UK extreme precipitation suggests that, even within observational constraints, this may be achievable before 2050 (Fowler and Wilby 2010). Therefore, for some regions of the UK and particularly during the winter season, flood managers will soon be able to make adaptation decisions about these types of extreme event in the light of formally detected changes in flood risk, useful for the prioritization of spending on flood defence infrastructures. However, in other seasons, and particularly for summer flash flooding caused by convective rainfall events which are poorly simulated by both GCMs and RCMs (Fowler and Ekström 2009), adaptation decisions will need to be made before changes in flood risk are formally detected.

The second part to this study was to examine which parameters have an influence on the simulation of extreme precipitation in the HadCM3L GCM. Two effects were analyzed: the effect upon the GEV shape parameter using a simple one-way analysis of variance (ANOVA) test, and the effect upon the time of detection using a contingency table analysis. Two climate model parameters were found to have important effects on the nature of the summer season GEV shape parameter which governs tail behavior: “entcoef”, the entrainment coefficient, and “vf1”, the ice fall speed. Two other parameters were found to have a significant influence on the time of detectable change: “ct”, the accretion constant, and “anthsca”, the scaling factor for emissions from anthropogenic sulfur aerosols. Significantly, our results suggest that precipitation efficiency (through the entrainment coefficient) has an important effect on heavy precipitation generation in climate models as has been found for observations (Wilson and Toumi 2005) although climate model simulated extreme precipitation seems to have a fundamentally different behavior to observations, perhaps due to

the negative estimate of the GEV shape parameter, unlike observations which produce a slightly positive (∼0.0–0.2) estimate. Wilson and Toumi (2005) suggest that the stretched exponential shape of the tail of the observed rainfall distribution will be largely unaffected by climate change. However, the scale of the distribution will be determined by multiplicative changes in the magnitudes of vertical mass flux, specific humidity and the precipitation efficiency (Wilson and Toumi 2005). Correctly parameterizing the precipitation efficiency will be challenging, as it is poorly quantified even for our current climate (Wilson and Toumi 2005), but this will be necessary for good predictions to be made for changes to precipitation extremes using climate models.

One current caveat to this study is that all climate models within the CPDN BBC CCE ensemble are treated equally, even though we may not believe that they are all equally likely or valid. However, although it intuitively makes sense to trust, and thus weigh, the better models more, it is difficult to objectively quantify model skill and therefore derive model weights (Tebaldi and Knutti 2007). In further work we hope to produce a method of weighting better models by assessing not only their ability to simulate the properties of observed extreme precipitation (an application-specific weighting) but also other measures. Ultimately, it is likely that a multi-scale approach to weighting, assessing not only the simulation of synoptic-scale regional climate, but also the simulation of continental-scale and global modes of variability may be more appropriate for the weighting of climate models from very large ensemble simulations such as CPDN.

Acknowledgments This work was supported by a Visiting Scientist position at the National Center for Atmospheric Research, Boulder, Colorado held by Dr. Hayley Fowler from November 2007 to February 2008 and a NERC Postdoctoral Fellowship award to Dr. Hayley Fowler (2006–2009) NE/D009588/1. Dan Cooley's research is supported, in part, through NSF grant DMS-0905315. This research was also supported, in part, through a grant from the Weather and Climate Impact Assessment Science Program at the National Center for Atmospheric Research. The research of the third author was also supported, in part, through a grant from the National Science Foundation (ATM-0502977 and ATM-0534173). The National Center for Atmospheric Research is sponsored by the National Science Foundation. We would also like to thank all the researchers from the climateprediction.net team and the Oxford University Centre for the Environment who have helped us or given advice on this study: in particular, David Frame, Myles Allen, Ana Lopez, Milena Cuéllar and Mark New. We would like to thank the Coordinating Editor and three anonymous reviewers for comments that have substantially improved an early version of this paper.

Appendix: The CPDN BBC climate change experiment

The CPDN project comprises three separate experiments which each examine uncertainties in model initial conditions, model parameterizations and external forcings. The first experiment was used to characterize realistic ranges for model parameters using the 'slab ocean' version of the Hadley Centre's General Circulation Model (GCM), HadSM3, which was run for 15 years each for a 'calibration' phase, a 'control phase' where greenhouse gases are kept at pre-industrial levels, and a 'future' phase where CO_2 in the atmosphere is double that of the control phase (Stainforth et al. 2005; Piani et al. 2005). In the second and third experiments, CPDN

conducted a forcing ensemble designed to explore uncertainty in past (and future) forcings. This is an important source of uncertainty in climate change projections. Kiehl (2007) have shown that the Coupled Model Intercomparison Project (CMIP) models under-sample uncertainty in historical forcing (Frame et al. 2009) and Joshi and Gregory (2008) have suggested that the physical climate response can be quite different depending on the forcing.

The second and third CPDN experiments are collectively known as the "CPDN BBC Climate Change Experiment" (CCE) (Frame et al. 2009). These experiments used a reduced resolution ocean version of the atmosphere-ocean coupled HadCM3 GCM of the UK Meteorological Office Hadley Centre, HadCM3L (Jones and Palmer 1998). The atmosphere component of the model is at a resolution of 2.5 × 3.75° latitude–longitude and has 19 vertical levels (known as N48; comparable resolution to ∼T42). The ocean component of the model is run at a reduced resolution in comparison to the standard HadCM3 model; with a resolution of 2.5 × 3.75° latitude–longitude, 20 vertical levels and a 1 h time-step. For each ensemble member, there is a 'flux-readjustment' spin up of HadCM3L with a standard atmosphere using the 1880–1920 climatology. In addition, the BBC CCE version of HadCM3L includes a modification to the ocean bathymetry: Iceland was removed and the Denmark straits deepened (Jones 2003), improving the northward transport of heat in the coarse-resolution ocean; and includes the interactive sulfur cycle described by Ackerley et al. (2009a, b).

In the second experiment this was used to produce a 'transient hindcast ensemble' of 1920–2000 using historical forcings, exploring the stable parameterizations of the climate model identified in the first experiment (Frame et al. 2009). Four observed datasets were used to obtain a range of plausible solar forcings (see Frame et al. 2009 for details). As these all underestimate the trend in solar index, a fifth dataset was arbitrarily created by doubling the trend in solar index in one of the data sets. Volcanic forcings were also added using five datasets based on observations of volcanic aerosol in the stratosphere. The third experiment was then run to produce a 'transient prediction ensemble' of 2000–2080 using variable natural (solar and volcanic forcings) under the SRES A1B emissions scenario (Nakicenovic et al. 2000: a mid-range scenario) using the same model parameterizations. Future solar forcing used three scenarios: increasing at the same rate as over the past 80 years; decreasing at the same rate; no significant trend either way. Future volcanic forcing is simulated using ten possible scenarios based on sampling from historic datasets. A full description of the models used in the BBC CCE can be found in Frame et al. (2009).

References

Ackerley, D., Highwood, E.J., Frame, D.J.: Changes in the global sulphate burden due to perturbations in global CO^2 concentrations. J. Clim. **22**, 5421–5432 (2009a)

Ackerley, D., Highwood, E.J., Frame, D.J.: Quantifying the effects of perturbing the physics of an interactive sulphur scheme using an ensemble of GCMs on the climateprediction.net platform. J. Geophys. Res. **114**, D01203 (2009b)

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Ambenje, P., Rupa Kumar, K., Revadekar, J.,

Griffiths, G.: Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res. **111**, D05109 (2006). doi:10.1029/2005JD006290
Allen, M.: Do-it-yourself climate prediction. Nature **401**, 642 (1999)
Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes, Theory and Applications. Wiley, West Sussex (2004)
Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: A Practical-Theoretic Approach, 2nd edition. Springer, London (2002)
Chavez-Demoulin, V., Davison, A.C.: Generalized Additive Models for Sample Extremes. J. Roy. Stat. Soc. C-App. **54**, 207–222 (2001)
Christensen, C., Aina, T., Stainforth, D.: The challenge of volunteer computing with lengthy climate modelling simulations. In: Proc. 1st IEEE Conference on Science and Grid Computing (2005)
Coles, S.G.: An Introduction to Statistical Modeling of Extreme Values. Springer, London (2001)
Covey, C., Achuta Rao, K.M., Cubasch, U., Jones, P., Lambert, S.J., Mann, M.E., Phillips, T.J., Taylor, K.E.: An overview of results from the Coupled Model Intercomparison Project. Global Planet. Change **37**, 103–133 (2003)
deHaan, L., Ferreira, A.: Extreme Value Theory. An Introduction. Springer, Heidelberg (2006)
Ekström, M., Fowler, H.J., Kilsby, C.G., Jones, P.D.: New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Future estimates and use in impact studies. J. Hydrol. **300**, 234–251 (2005)
Fowler, H.J., Ekström, M.: Multi-model ensemble estimates of climate change impacts on UK seasonal rainfall extremes. Int. J. Climatol. **29**, 385–416 (2009)
Fowler, H.J., Kilsby, C.G.: A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. Int. J. Climatol. **23**, 1313–1334 (2003a)
Fowler, H.J., Kilsby, C.G.: Implications of changes in seasonal and annual extreme rainfall. Geophys. Res. Lett. **30**, 1720 (2003b). doi:10.1029/2003GL017327
Fowler, H.J., Wilby, R.L.: Detecting changes in seasonal precipitation extremes using regional climate model projections: implications for managing fluvial flood risk. Water Resour. Res. (2010). doi:10.1029/2008WR007636
Fowler, H.J., Ekström, M., Kilsby, C.G., Jones, P.D.: New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. J. Hydrol. **300**, 212–233 (2005)
Fowler, H.J., Ekström, M., Blenkinsop, S., Smith, A.P.: Estimating change in extreme European precipitation using a multi-model ensemble. J. Geophys. Res. Atmos. **112**, D18104 (2007). doi:10.1029/2007JD008619
Frame, D.J., Aina, T., Christensen, C.M., Faull, N.E., Knight, S.H.E., Piani, C., Rosier, S.M., Yamazaki, K., Yamazaki, Y., Allen, M.R.: The climateprediction.net BBC climate change experiment: design of the coupled model ensemble. Philos. Trans. Roy. Soc. A **367**, 855–870 (2009)
Frei, C., Schär, C.: Detection probability of trends in rare events: theory and application to heavy precipitation in the Alpine region. J. Climate **14**, 1568–1584 (2001)
Giorgi, F., Francisco, R.: Evaluating uncertainties in the prediction of regional climate change. Geophys. Res. Lett. **27**, 1295–1298 (2000)
Gregory, D., Rowntree, P.: A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. Mon. Wea. Rev. **118**, 1438–1506 (1990)
Groisman, P.Ya., Knight, R.W., Easterling, D.R., Karl, T.R., Hegerl, G.C., Razuvaev, V.N.: Trends in intense precipitation in the climate record. J. Climate **18**, 1326–1350 (2005)
Hays, W.L.: Statistics for the Social Sciences, 2nd edn. Holt, Rinehart, and Winston, New York (1973)
Hegerl, G.C., Zwiers, F.W., Stott, P.A., Kharin, V.V.: Detectability of anthropogenic changes in annual temperature and precipitation extremes. J. Climate **17**, 3683–3700 (2004)
Hegerl, G.C., Karl, T.R., Allen, M., Bindoff, N.L., Gillett, N., Karoly, D., Zhang, X.B., Zwiers, F.: Climate change detection and attribution: beyond mean temperature signals. J. Climate **19**, 5058–5077 (2006)
Hegerl, G.C., Zwiers, F.W., Braconnot, P., Gillet, N.P., Luo, Y., Marengo, J.A., Nicholls, N., Penner, J.E., Stott, P.A.: Understanding and attributing climate change. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (eds.) Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge (2007)
Jones, C.: A fast ocean GCM without flux adjustments. J. Atmos. Oceanic Technol. **20**, 1857–1868 (2003). doi:10.1175/1520–0426(2003)020<1857:AFOGWF>2.0.CO;2

Jones, C.D., Palmer, J.R.: Spin up methods for HadCM3L. Technical report CRTN 84, Hadley Centre for Climate Prediction and Research (1998)
Joshi, M., Gregory, J.: The dependence of the land-sea contrast in surface climate response on the nature of the forcing. Geophys. Res. Lett. **35**, L24802 (2008)
Katz, R.W.: Extreme value theory for precipitation: sensitivity analysis for climate change. Adv. Water Resour. **23**, 133–139 (1999)
Kharin, V.V., Zwiers, F.W.: Estimating extremes in transient climate change simulations. J. Climate **18**, 1156–1173 (2005)
Kharin, V., Zwiers, F.W., Zhang, X., Hegerl, G.C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J. Climate **20**, 1419–1444 (2007)
Kiehl, J.T.: Twentieth century climate model response and climate sensitivity. Geophys. Res. Lett. **34**, L22710 (2007). doi:10.1029/2007GL031383
Kiktev, D., Sexton, D.M.H., Alexander, L., Folland, C.K.: Comparison of modeled and observed trends in indices of daily climate extremes. J. Climate **16**, 3560–3571 (2003)
Knight, C.G., Knight, S.H.E., Massey, N., Aina, T., Christensen, C., Frame, D.J., Kettleborough, J.A., Martin, A., Pascoe, S., Sanderson, B., Stainforth, D.A., Allen, M.R.: Association of parameter, software and hardware variation with large scale behavior across 57,000 climate models. Proc. Natl. Acad. Sci. USA **104**, 12259–12264 (2007)
Lambert, F.H., Stott, P.A., Allen, M.R., Palmer, M.A.: Detection and attribution of changes in 20th century land precipitation. Geophys. Res. Lett. **31**, L10203 (2004). doi:10.1029/2004GL019545
Meehl, G.A., Arblaster, J.M., Tebaldi, C.: Understanding future patterns of increased precipitation intensity in climate model simulations. Geophys. Res. Lett. **32**, L18719 (2005). doi:10.1029/2005GL023680
Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grübler, A., Jung, T.Y., Kram, T., La Rovere, E.L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Raihi, K., Roehrl, A., Rogner, H.-H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., Dadi, Z.: IPCC Special Report on Emissions Scenarios. Cambridge University Press, Cambridge (2000), 599 pp
Oehlert, G.W.: A note on the delta method. Am. Stat. **46**, 27–29 (1992)
Olejnik, S., Algina, J.: Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol. Methods **8**, 434–447 (2003)
Osborn, T.J., Hulme, M.: Development of a relationship between station and grid-box rain day frequencies for climate model evaluation. J. Climate **10**, 1885–1908 (1997)
Osborn, T.J., Hulme, M., Jones, P.D., Basnett, T.A.: Observed trends in the daily intensity of United Kingdom precipitation. Int. J. Climatol. **20**, 347–364 (2000)
Piani, C., Frame, D.J., Stainforth, D.A., Allen, M.R.: Constraints on climate change from a multi-thousand member ensemble of simulations. Geophys. Rev. Lett. **32**, L23825 (2005)
Räisänen, J., Joelsson, R.: Changes in average and extreme precipitation in two regional climate model experiments. Tellus **53A**, 547–566 (2001)
Sanderson, B., Piani, C.: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. Climate Dyn. **30**, 175–190 (2007)
Sanderson, B., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D.J., Ingram, W.J., Piani, C., Stainforth, D.A., Stone, D.A., Allen, M.R.: Constraints on model response to greenhouse gas forcing and the rose of sub-grid scale processes. J. Clim. **21**, 2384–2400 (2008)
Semenov, V.A., Bengtsson, L.: Secular trends in daily precipitation characteristics: greenhouse gas simulation with a coupled AOGCM. Climate Dyn. **19**, 123–140 (2002)
Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J., Allen, M.R.: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature **433**, 403–406 (2005)
Tebaldi, C., Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections. Philos. T. Roy. Soc. A **365**, 2053–2075 (2007)
Tebaldi, C., Hayhoe, K., Arblaster, J.M., Meehl, G.A.: Going to the extremes; an intercomparison of model-simulated historical and future changes in extreme events. Clim. Change **79**, 185–211 (2006)
Tett, S.F.B., Betts, R., Crowley, T.J., Gregory, J., Johns, T.C., Jones, A., Osborn, T.J., Ostrom, E., Roberts, D.L., Woodgate, M.J.: The impact of natural and anthropogenic forcings on climate and hydrology since 1550. Clim. Dynam. **28**, 3–34 (2007)
Trenberth, K.E., Dai, A.G., Rasmussen, R.M., Parsons, D.B.: The changing character of precipitation. Bull. Am. Meteorol. Soc. **84**, 1205–1217 (2003)

Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B., Zhai, P.: Observations: surface and atmospheric climate change. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (eds.) Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge (2007)

Wilson, P.R., Toumi, R.: A fundamental probability distribution for heavy rainfall. Geophys. Res. Lett. **32**, L14812 (2005). doi:10.1029/2005GL022465

Zhang, X., Zwiers, F.W., Hegerl, G.C., Lambert, F.G., Gillett, N.P., Solomon, S., Stott, P.A., Nozawa, T.: Detection of human influence on twentieth-century precipitation trends. Nature **448**, 461–465 (2007)