Climate Dynamics (2006) 26: 489–511
DOI 10.1007/s00382-005-0097-1

David N. Barnett · Simon J. Brown · James M. Murphy
David M. H. Sexton · Mark J. Webb

Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations

Received: 13 January 2005 / Accepted: 14 November 2005 / Published online: 19 January 2006

Abstract We discuss equilibrium changes in daily extreme surface air temperature and precipitation events in response to doubled atmospheric CO_2, simulated in an ensemble of 53 versions of HadSM3, consisting of the HadAM3 atmospheric general circulation model (GCM) coupled to a mixed layer ocean. By virtue of its size and design, the ensemble, which samples uncertainty arising from the parameterisation of atmospheric physical processes and the effects of natural variability, provides a first opportunity to quantify the robustness of predictions of changes in extremes obtained from GCM simulations. Changes in extremes are quantified by calculating the frequency of exceedance of a fixed threshold in the $2 \times CO_2$ simulation relative to the $1 \times CO_2$ simulation. The ensemble-mean value of this relative frequency provides a best estimate of the expected change while the range of values across the ensemble provides a measure of the associated uncertainty. For example, when the extreme threshold is defined as the 99th percentile of the $1 \times CO_2$ distribution, the global-mean ensemble-mean relative frequency of extremely warm days is found to be 20 in January, and 28 in July, implying that events occurring on one day per hundred under present day conditions would typically occur on 20–30 days per hundred under $2 \times CO_2$ conditons. However the ensemble range in the relative frequency is of similar magnitude to the ensemble-mean value, indicating considerable uncertainty in the magnitude of the increase. The relative frequencies in response to doubled CO_2 become smaller as the threshold used to define the extreme event is reduced. For one variable (July maximum daily temperature) we investigate this simulated variation with threshold, showing that it can be quite well reproduced by assuming the response to doubling CO_2 to be characterised simply as a uniform shift of a Gaussian distribution. Nevertheless, doubling CO_2 does lead to changes in the shape of the daily distributions for both temperature and precipitation, but the effect of these changes on the relative frequency of extreme events is generally larger for precipitation. For example, around one-fifth of the globe exhibits ensemble-mean decreases in time-averaged precipitation accompanied by increases in the frequency of extremely wet days. The ensemble range of changes in precipitation extremes (relative to the ensemble mean of the changes) is typically larger than for temperature extremes, indicating greater uncertainty in the precipitation changes. In the global average, extremely wet days are predicted to become twice as common under $2 \times CO_2$ conditions. We also consider changes in extreme seasons, finding that simulated increases in the frequency of extremely warm or wet seasons under $2 \times CO_2$ are almost everywhere greater than the corresponding increase in daily extremes. The smaller increases in the frequency of daily extremes is explained by the influence of day-to-day weather variability which inflates the variance of daily distributions compared to their seasonal counterparts.

Keywords Climate change · Extreme events · Uncertainty · Ensemble · General circulation model · Parametrisation · Perturbed physics

D. N. Barnett · S. J. Brown · J. M. Murphy (✉)
D. M. H. Sexton · M. J. Webb
Met Office, Hadley Centre for Climate Prediction and Research, FitzRoy Road, Exeter, Devon, EX1 3PB UK
E-mail: james.murphy@metoffice.gov.uk
Tel.: +44-1392-886658
Fax: +44-1344-854898

1 Introduction

One of the major aims of climate research is to investigate the effects on climate, at various spatial and temporal scales, of anthropogenic emissions into the atmosphere. Of particular interest are climatic extremes, which have potentially larger impacts on population and the environment than changes in mean climate (e.g. Karl et al. 1997; Meehl et al. 2000a; Easterling et al. 2000). Previous analyses of changes in extremes from general

circulation model (GCM) simulations have been reviewed by a number of authors, including Easterling et al. (2000), Meehl et al. (2000b), Cubasch et al. (2001) and Giorgi et al. (2001). These studies generally indicate, at least in a globally-averaged sense, that the occurrence of extreme events in climatic elements such as surface temperature, precipitation and extratropical storm intensity (e.g. Gregory and Mitchell 1995; Carnell and Senior 1998; Senior et al. 2002) may become more frequent in the future. These conclusions are confirmed by results from the latest generation of coupled ocean-atmosphere GCM simulations (Tebaldi et al. 2006; Meehl et al. 2005; Weisheimer and Palmer 2005), which show widespread increases during the 21st century in several indicators of high temperature and precipitation events. Stott et al. (2004) quantified the relative contributions to southern European summer temperatures of changes in natural and anthropogenic forcings since the industrial revolution, finding that human activity has at least doubled the risk of a heatwave of the magnitude observed in 2003.

Any prediction of climate change carries an associated uncertainty, which arises from three fundamental sources: the processes represented in the climate model, the effects of natural variability and the future emissions or concentrations of forcing agents (CO_2, CH_4, sulphate aerosols, etc.). In much of the work carried out previously (e.g. Hennessy et al. 1997; Giorgi and Fracisco 2000; Cubasch et al. 2001; Palmer and Räisänen 2002; Tebaldi et al. 2005; Furrer et al. 2006), uncertainty has been estimated by considering the spread in the changes simulated by a small ensemble of different GCMs, which use the same scenario of forcing agents but differ from each other in their initial conditions, structural set-up (e.g. the choice of grid resolution and of which climate processes to include), and in the manner in which processes are parameterised. Thus the ensemble spread in a given prediction will be derived from many sources, whose individual contributions can not easily be determined. Other studies (e.g. Cubasch et al. 1994; Kharin and Zwiers 2000; Meehl and Tebaldi 2004) have used small ensembles of a particular GCM started from different initial conditions, which, for a given forcing scenario, encompass only the effects of natural variability. Indeed, many of the studies looking at extreme events have considered predictions derived from only a single run of a global and/or regional climate model, for a given CO_2 scenario (e.g. Durman et al. 2001; Jones and Reid 2001; Senior et al. 2002; Christensen and Christensen 2003; Huntingford et al. 2003), with no accompanying estimate of uncertainty.

The Hadley Centre quantifying uncertainty in model predictions (QUMP) project aims to produce probabilistic climate predictions through the production of large ensembles of a particular GCM, generated using a systematic methodology (see Murphy et al. 2004) and taking into account uncertainty from as many sources as possible. Since extreme events are rare by definition, and GCM simulations of them are dependent on the uncertainties outlined above, it is particularly important to use large ensembles designed to sample those uncertainties in order to obtain robust estimates of changes. Reliable estimates of changes in extremes are needed to inform risk assessments of the impacts of climate change and decisions concerning the corresponding adaptation options (Visser et al. 2000; Jones 2000; Pittock et al 2001; Reilly et al. 2001; Zwiers 2002).

In this paper, we take a first step in this direction by investigating equilibrium changes in the frequency of extreme events in response to a doubling of atmospheric CO_2. Changes are obtained from an ensemble of 53 GCM simulations which samples uncertainties arising both from natural variability and from the physical parameterisations of atmospheric processes. The climate variables analysed are screen temperature and precipitation, over both daily and seasonal timescales, with a view to identifying changes which are statistically robust across the ensemble. We also address the variation of changes in extreme event frequency with the threshold used to define the event and with timescale (daily or seasonal), and relate these variations to basic properties of the statistical distributions of the data. Detailed consideration of the physical mechanisms which drive the changes in extremes is reserved for future work. Analysis of the relative contributions of the different sources of uncertainty (i.e. perturbed physics and natural variability), as presented by Murphy et al. (2004) for changes in long-term mean fields, will also be reserved for a future paper.

2 Model and experiments

We use the HadAM3 atmospheric GCM (Pope et al. 2000), which includes a comprehensive range of physical processes. It therefore simulates many of the most uncertain climate change feedbacks, such as those associated with cloud (see e.g. Williams et al. 2003) and water vapour (Stocker et al. 2001), but does not include, for instance, an interactive carbon cycle or atmospheric chemistry scheme. HadAM3 uses a regular latitude/longitude grid with a horizontal resolution of 2.5°×3.75°, and with 19 levels on hybrid vertical coordinates (Simmons and Burridge 1981). These are terrain-following sigma (i.e. pressure/surface pressure) coordinates in the bottom four levels, pressure coordinates in the top three levels and a linear combination at intermediate levels. The timestep is 30 min.

We examine the equilibrium response to a doubling of CO_2, using 53 versions of the HadSM3 climate model, consisting of HadAM3 coupled to a mixed layer ('slab') ocean of depth 50 m. HadSM3 contains no explicit ocean dynamics, so ocean heat transport is prescribed as a heat convergence, which varies with position and season, but not from year to year (Hewitt and Mitchell 1997). Heat convergence values are diagnosed from a calibration experiment in which sea-surface temperatures (SSTs) are reset to seasonally-varying climatological values at each timestep.

Any climate model must include equations, expressed in terms of gridbox variables, to represent the effects of processes which are not explicitly resolved on the model grid. These equations contain parameters which cannot be specified precisely from theoretical considerations or observations, and are therefore subject to an uncertainty range. The parameter settings in the 'standard' version of the model were chosen to produce a realistic simulation of present-day climate when coupled to dynamic ocean models (Gordon et al. 2000). However there exist alternative settings, or combinations of settings, which are equally justifiable, and which would yield a range of simulations of present and future climate likely to be different from, but just as plausible as, those of the standard version.

We set out to design an ensemble of model versions which would sample the range of possible outcomes implied by these parameter uncertainties. A subset of 29 parameters thought likely to cover the major process uncertainties were identified by experts. A few parameters consisted of logical switches controlling activation of an additional process, or a choice between two alternative methods of representing a process. The remainder consisted of coefficients or thresholds subject to an uncertainty range. For these the experts were asked to specify plausible minimum, intermediate and maximum values. The values used in the standard version were often, but by no means always, the intermediate one (see Tables 1, 2 for details). Members of a perturbed physics ensemble were created by altering a single parameter relative to the standard version. This was achieved either by altering a logical switch, or setting a continuously-variable parameter to its minimum, intermediate or maximum value. The 53 member ensemble consisted of the standard version plus 52 members in which each parameter setting (on/off or min/intermediate/max) not covered by the standard version was sampled once. The ensemble thus represents a set of plausible model versions which explores the parameter space associated with physical processes represented in HadAM3. Murphy et al. (2004) provide key results from the ensemble relevant to the simulation of time-averaged climate change.

The perturbed physics ensemble provides a systematic basis for quantifying uncertainties in the equilibrium response to doubled CO_2 simulated by GCMs. However there are several caveats which imply that the uncertainty ranges derived from the ensemble should be interpreted as a lower limit, rather than a best estimate:

1. The ensemble does not fully sample parameter space: we only perturb a subset of the uncertain parameters (albeit a subset thought to address the main process uncertainties) and those we do perturb are only changed one at a time. These limitations imply that the distribution of outcomes is likely to be too heavily weighted towards the prediction of the standard model version. The importance of sampling multiple parameter perturbations has already been demonstrated in the context of global changes by results from the *climateprediction.net* project (Stainforth et al. 2005). We have addressed this by producing a new and larger HadSM3 perturbed physics ensemble (Webb et al. 2006) whose members contain multiple parameter perturbations and sample a range of values for continuous parameters, rather than just the values shown in Table 1. The simulation of extreme events in this new ensemble will be presented in a future paper.
2. We consider parameter uncertainties, but neglect two other potentially important types of modelling uncertainty. The first arises from the existence of alternative choices for basic structural aspects of the GCM, such as grid resolution or the fundamental physical assumptions used in a parameterisation scheme. Secondly the parameterisation equations should (but typically do not) contain a stochastic element, because the effects of sub-grid-scale processes on the gridbox model variables cannot be fully determined from values of the latter (Palmer 2001).
3. We do not account for uncertainties associated with ocean circulation feedbacks, which are capable of modifying the atmospheric response at a regional level (Boer and Yu 2003), or for biogeochemical feedbacks capable of altering the radiative forcing of climate change for a given profile of emissions (e.g. Cox et al. 2000). Also, the experimental methodology ensures that sea surface temperatures in the $1 \times CO_2$ simulations are close to observed values, so we do not fully sample uncertainties associated with systematic biases in SST (Murphy et al. 2004).

Each perturbed physics ensemble member consists of one $1 \times CO_2$ (present-day) run and one in which atmospheric CO_2 was doubled. Both were run to equilibrium, such that the global mean temperature showed a 20-year trend which was no greater than a threshold defined as the value exceeded by only 5% of 20-year sections of an extended 600 year simulation made using the standard model version. In each case, the equilibrium 20-year period (50 years in the case of the high cloud-fraction ensemble member [11] (Table 1), which showed marked decadal variability) was the section of the run analysed in this work.

If the ensemble members are to produce plausible estimates of climate change, a necessary (though not sufficient) condition is that they should realistically simulate the observed present-day climate, since a GCM exhibiting regional biases will be likely to simulate climate feedbacks in the wrong location (Mitchell et al. 1987), or of the wrong strength (Hewitt et al. 2001). Figure 1 shows the ensemble-mean bias in simulated 20 year mean 1.5 m air temperature (hereafter screen temperature) and precipitation for June–July–August (JJA) and December–January–February (DJF). The biases are calculated relative to observed climatologies of screen temperature for 1961–1990 (New et al. 1999) and ocean and land precipitation for 1979–1999 (Xie

Table 1 The parameter values used in producing the different ensemble members, grouped by parameterisation scheme

Parameter/scheme	Lower	Low	Standard [1]	High	Higher
Large scale cloud					
RH_{crit}		0.6[2]	0.7	0.9[3]	
V_{fl} (ms^{-1})		0.5[4]	1.0	2.0[5]	
C_w land ($kg\ m^{-3}$)		1.0×10^{-4} [6]	2.0×10^{-4}	2.0×10^{-3}[7]	
C_w sea ($kg\ m^{-3}$)		2.0×10^{-5} [6]	5.0×10^{-5}	5.0×10^{-4} [7]	
C_t (s^{-1})		5.0×10^{-5} [8]	1.0×10^{-4}	4.0×10^{-4} [9]	
Cloud frac. at gridbox satn. (BL/free trop.)			0.5/0.5	0.7/0.6[10]	0.8/0.65[11]
Flow dependent RH_{crit} scheme			Off	On[12]	
Vertical gradient cloud area scheme			Off	On[13]	
Convection					
Entrainment rate coefficient		0.6[14]	3.0	9.0[15]	
Timescale for destruction of CAPE (h)[a]		(1)[16]	(2)[17]	(4)[18]	
Convective updraught factor[b]		(0.1)[19]	(1.0)[20]		
Convective anvil shape factor[b]			(1)	(2)[21]	(3)[22]
Radiation					
Ice particle size (μm)		25[23]	30	40[24]	
Non-spherical ice particles			Off	On[25]	
SW water vapour continuum absorption			Off	On[26]	
Sulphur cycle			Off	On[27]	
Sea ice					
Sea ice albedo:					
Albedo at 0°C	0.65[28]	0.57[29]	0.50		
Albedo at T_{cold}	0.80[28]	0.80[29]	0.80		
T_{cold} (°C)	−2[28]	−5[29]	−10		
Ocean-ice diffusion coefficient ($m^2\ s^{-1}$)	2.50×10^{-5} [30]	1.00×10^{-4}[31]	3.75×10^{-4}		
Boundary layer and surface processes					
Boundary layer flux profile parameter,G_0		5[32]	10	20[33]	
Asymptotic neutr. Mixing length parameter, λ		0.05[34]	0.15	0.50[35]	
Charnock constant			0.012	0.016[36]	0.020[37]
Free convective roughness length over sea (m)		2.0×10^{-4} [38]	1.3×10^{-3}	5.0×10^{-3} [39]	
Roughness length z_0(m):					
Dense evergreen needleleaf forest		0.50[40]	0.78[41]	2.00[42]	
Dense deciduous needleaf forest		0.50[40]	0.78[41]	2.00[42]	
Dense deciduous broadleaf forest		0.50[40]	0.70[41]	2.00[42]	
Equatorial rainforest			1.05[40]	2.10[41]	2.90[42]
No. soil lev. Access. For t/piration (forest/grass)	2/1[43]	3/2[44]	4/3		
Surface-cancopy decoupling scheme			Off	On [45]	
Stomatal conductance response to ΔCO_2		Off[46]	On		
Dynamics					
Order of dynamical diffusion operator		4[47]	6		
Dynamical diffusion e-folding time (h)		6[48]	12	24[49]	
Surface gravity wave constant	1.00×10^4	1.50×10^4 [51]	2.00×10^4		
Trapped lee wave constant ($m^{-3/2}$)	1.50×10^5 [50]	2.25×10^5 [51]	3.00×10^5		
Gravity wave drag start level			3	4[52]	5[53]

The ensemble member number is given in square brackets, and shows the parameter value setting by which the corresponding ensemble member differs from the standard version. Note that, in cases where the standard parameter setting falls at the extreme upper (lower) end of the estimated uncertainty range, two lower (higher) perturbation values were selected so as to encompass this range

[a]The timescale must be set when CAPE closure, rather than the standard buoyancy closure, is used in the convection scheme

[b]When the convective anvil scheme is used, a logical flag is set to exclude convective precipitation from the cloud water path. This flag is not set in the standard version

and Arkin 1998), omitting areas of the ocean where the observed dataset is not reliable. Screen temperature biases are generally smaller than the ensemble-mean response to a doubling of CO_2. Precipitation biases exhibit a more complex pattern, particularly in tropical regions where small shifts in the location of features such as the intertropical convergence zone can cause large regional errors.

Murphy et al. (2004) report a quantitative verification of the quality of the perturbed physics ensemble $1 \times CO_2$ simulations, using a climate prediction index (CPI) obtained from normalised errors in multiannual mean fields averaged over a wide range of climate observables. In principle, a measure of reliability such as the CPI can be used to weight the contributions of ensemble members when forming ensemble means, ensemble variances or probability distributions from the simulated changes (Giorgi and Mearns 2002; Murphy et al. 2004; Tebaldi et al. 2005). Here, however, the simulations of all ensemble members are given equal weight, since further work is needed to establish the suitability of the CPI, or an alternative metric, for measuring the reliability of simulated changes in extreme events at a regional level. All of the ensemble members produce reasonable simulations of screen temperature or precipitation according to the CPI (Murphy et al. 2004), although there is significant varia-

Table 2 Description of the function of the parameters, the settings of which are varied between ensemble members

Parameter/scheme	Description/process affected
Large scale cloud	
RH_{crit}	Threshold of relative humidity for cloud formation
V_{f1}	Ice particle fall speed
C_w	Cloud droplet to rain conversion threshold
C_t	Cloud droplet to rain conversion rate
Cloud fraction at gridbox saturation	Cloud cover calculation
Flow-dependent RH_{crit} scheme	Parameterisation of RH_{crit} in terms of local variance of cloud water
Vertical gradient cloud area scheme	Accounts for effect of gridbox vertical cloud water gradient on cloud cover calculation
Convection	
Entrainment rate coefficient	Rate of mixing between environmental air and convective plume
Timescale for destruction of CAPE	Intensity of convective mass flux
Convective updraught factor	Fraction of convective cloud in which updraught occurs
Convective anvil shape factor	Shape of convective cloud
Radiation	
Ice particle size (μm)	Effective radius of cloud ice spheres
Non-spherical ice particles	Accounts for effect of non-spherical ice particles
SW water vapour continuum absorption	Accounts for short-wave absorption due to the self-broadened continuum of water vapour
Sulphur cycle	Includes interactive calculation of sulphate aerosol loadings, accounting for sources, transport, physical removal and chemistry
Sea ice	
Sea ice albedo	Dependence of sea ice albedo on temperature
Ocean-ice diffusion coefficient	Ocean to ice heat transfer
Boundary layer and surface processes	
Boundary layer flux profile parameter,G_0	Functions used to determine stability dependence of turbulent mixing coefficients
Asymptotic neutral mixing length parameter, λ	Calculation of turbulent mixing coefficients
Charnock constant	Roughness lengths and surface fluxes over sea
Free convective roughness length over sea	Surface fluxes over tropical oceans
Roughness length z_0	Surface fluxes over areas containing forest
No. soil levels accessible for transpiration	Root depths
Surface-cancopy decoupling scheme	Accounts for effect of vegetation canopy on surface energy balance
Stomatal conductance response to ΔCO_2	Variance of stomatal conductance with carbon dioxide concentration
Dynamics	
Order of dynamical diffusion operator	Spatial scale of diffusive dampling of heat, momentum and moisture
Dynamical diffusion e-folding time	Diffusion coefficients for heat, momentum and moisture
Surface gravity wave constant	Magnitude of hydrostatic gravity wave stress
Trapped lee wave constant	Magnitude of non-hydrostatic gravity wave stress
Gravity wave drag start level	Lowest model level at which drag is applied

tion in regional biases (not shown), some of which arises from natural variability as well as from the impact of the parameter perturbations. However certain errors, such as a cold bias over northern continents in DJF, are common to all the ensemble members, whereas this is not the case in ensembles constructed from GCMs developed at different modelling centres (Lambert and Boer 2001). This reinforces the point made above that the perturbed physics ensemble only addresses a subset of possible modelling uncertainties (see Sect. 6 for further discussion).

3 Extreme daily events

3.1 Method

We analyse extreme daily events in January and July, considering maximum and minimum daily screen temperature and accumulated daily precipitation, which are among the most directly relevant variables in terms of their impacts on population. For each ensemble member, the 20-year equilibrium period yields 600 days of data, for both the $1 \times CO_2$ and $2 \times CO_2$ runs. At each gridbox, and for each ensemble member, a frequency distribution was produced for each climate variable, under both $1 \times CO_2$ and $2 \times CO_2$ conditions. An extreme event in a given ensemble member was defined as one which exceeds the 99th percentile threshold (i.e. occurs on less than 1% of days) under $1 \times CO_2$ conditions. Use of a particular percentile to define an extreme event threshold is well established in previous work (e.g. Jones et al. 1999; Durman et al. 2001; Bonsal et al. 2001; Horton et al. 2001; Senior et al. 2002; Frich et al. 2002; Kiktev et al. 2003). In the case of precipitation, we emphasise that the 99th percentile for a given ensemble member was calculated from the full 600-day-time series containing both wet and dry days, rather than over wet

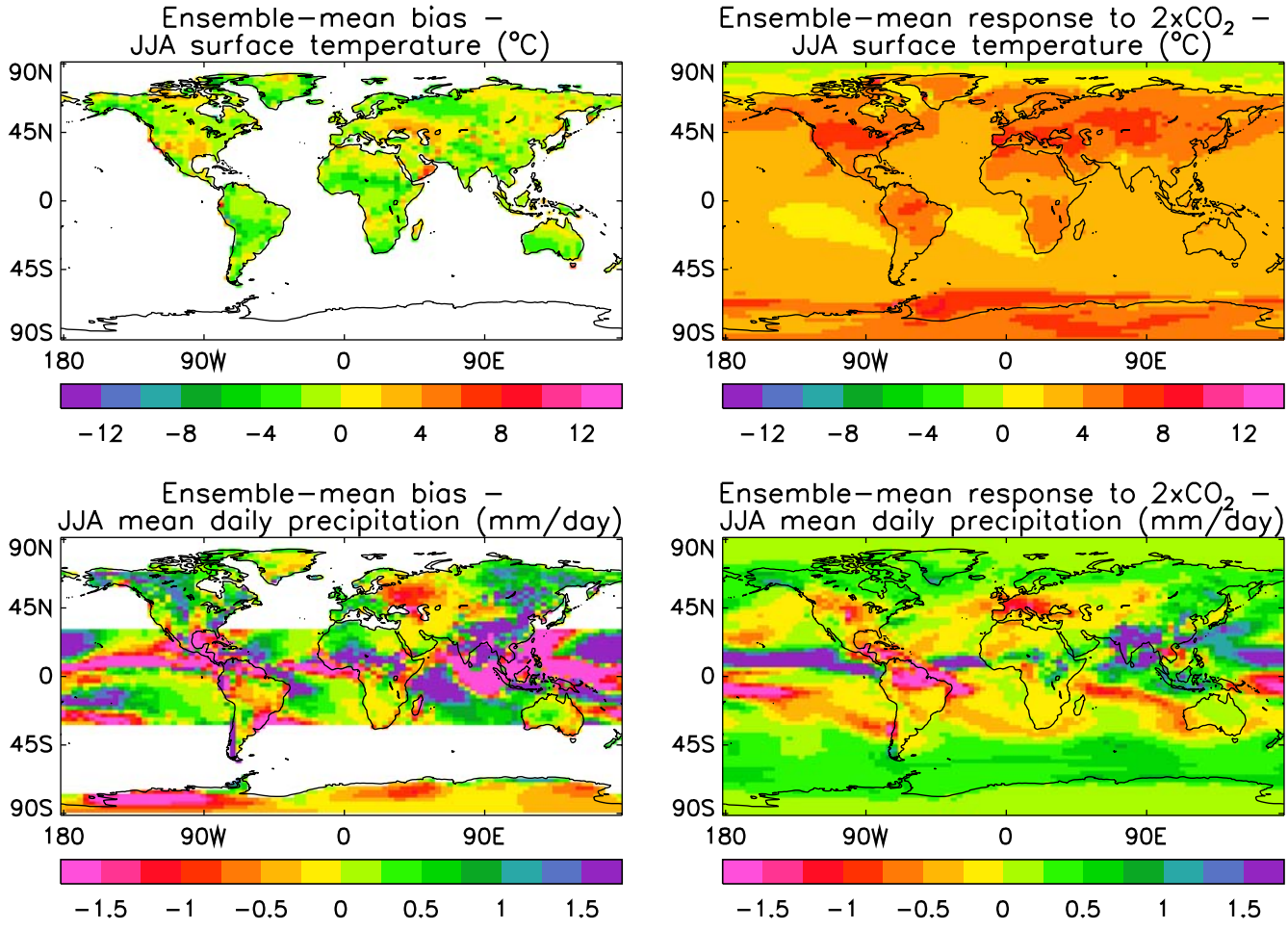

Fig. 1 Ensemble-mean bias of the models with respect to observations (*left*), and the ensemble-mean equilibrium response to a doubling of CO_2 (*right*), for 20 year averages of JJA screen temperature (*top*) and precipitation (*bottom*)

days alone. The frequency of events exceeding this threshold under $2 \times CO_2$ conditions was then calculated, and divided by the frequency of exceedance under $1 \times CO_2$ conditions to give a relative frequency of extreme events (RFEE). We use the abbreviation RFEE throughout the paper when referring to the equilibrium response of extreme events to a doubling of CO_2. The ensemble average of RFEE, and the ensemble range (defined below), are calculated for screen temperature and precipitation at each gridbox, and are interpreted as the best estimate of the expected change and its associated uncertainty. For the sake of brevity we plot results only for July, supporting these with tables which include statistical summaries of key results for both January and July.

In what follows we define the range of an ensemble distribution of values as the interval between its 10th and 90th percentiles, identified by ranking the 53 values of the relevant variable in ascending order and calculating values midway between values 5 and 6 (approximately the 10th percentile) and midway between values 48 and 49 (approximately the 90th percentile). We remind the reader that differences between the simulations of alternative ensemble members arise both from the effects of parameter perturbations and from sampling differences associated with natural climate variability.

3.2 Results

The results for a single gridbox, centred on the south-east of England (at 52.5°N, 0°E), are presented first, to illustrate the information available from the ensemble, before considering results for other regions. Figure 2 shows the frequency distributions for July daily maximum screen temperature, for the SE England gridbox, for all 53 ensemble members. The values of the extreme event thresholds for each ensemble member are shown in the first column of Table 3. The second column gives the corresponding relative frequency of extreme events (RFEE). The results show that extreme daily temperatures become more frequent in the $2 \times CO_2$ simulation relative to the $1 \times CO_2$ simulation in all but one ensemble member. The ensemble-mean RFEE is 14.1, indicating that hot summer temperatures expected once every hundred days under $1 \times CO_2$ conditions would

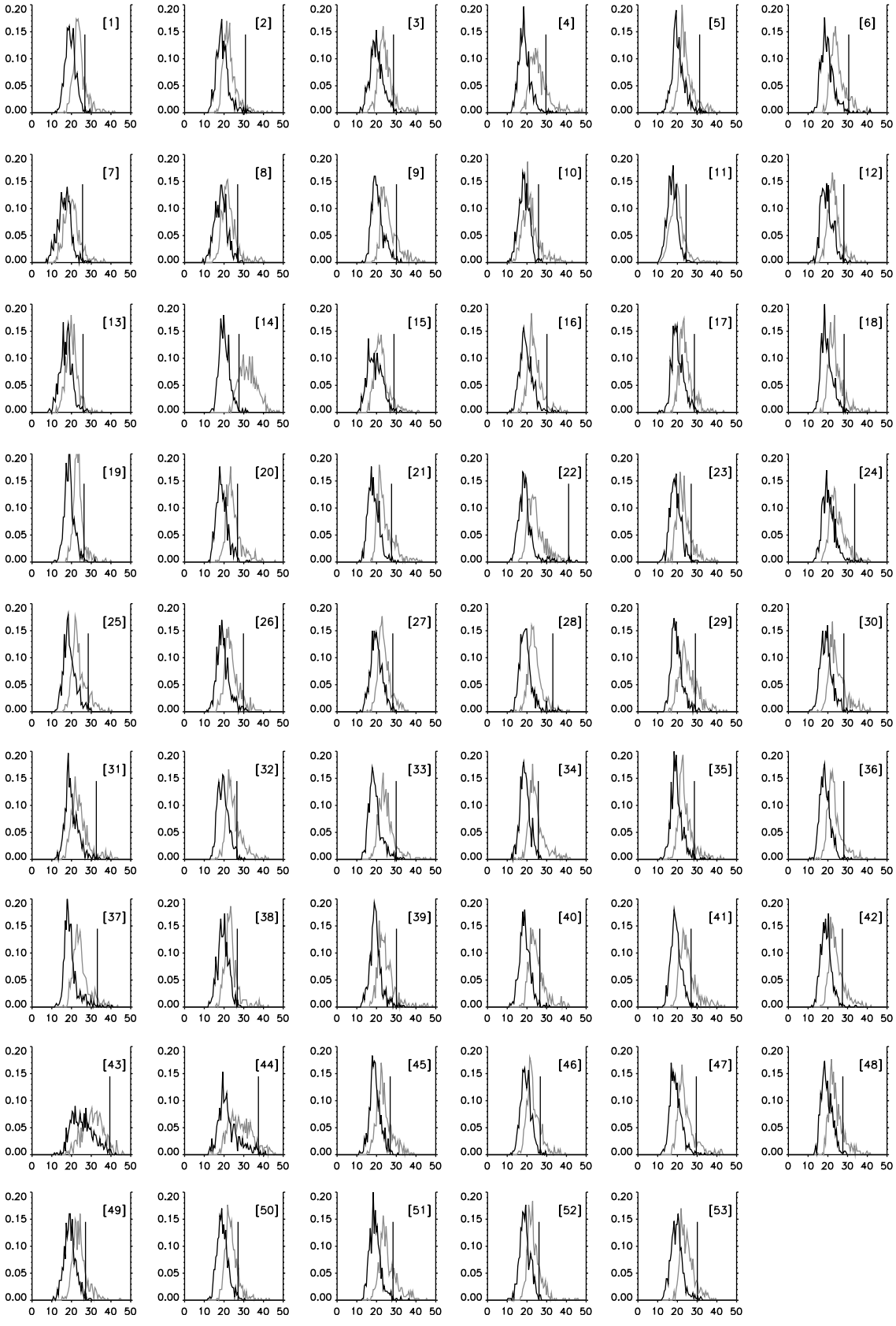

Fig. 2 Relative frequencies for each ensemble member of values of July maximum daily screen temperature (shown in °C along the abscissa), under 1 × CO_2 (*black*) and 2 × CO_2 (*grey*) conditions, for the SE England gridbox. Ensemble members are numbered in square brackets (see Table 1), and in each case the extreme event threshold, defined as the 99th percentile threshold under 1 × CO_2 conditions, is shown by a *solid vertical line*

Table 3 Values for each ensemble member (see Table 1) of the extreme event threshold (EET) for the SE England gridbox, defined as the 99th percentile of the distribution of daily values simulated under $1 \times CO_2$ conditions (1st percentile threshold in the case of minimum screen temperature in January), and the relative frequency of extreme events (RFEE) in the $2 \times CO_2$ simulation of each ensemble member, with respect to the $1 \times CO_2$ EET

Ens.memb.	Max daily screen temp(°C)				Min daily screen temp (°C)				Daily precip. (mm/day)			
	July		January		July		January		July		January	
	EET	RFEE	EET	RFEE	EET	RFEE	EET	RFEE	EET	RFEE	EET	RFEE
1	26.8	16.17	12.1	19.67	17.5	14.87	−11.1	0.00	15.5	0.75	13.5	2.17
2	30.9	4.67	10.8	25.77	17.3	10.33	−14.9	0.00	14.0	0.67	13.6	4.10
3	28.6	11.50	11.2	30.67	18.1	16.50	−12.6	0.00	15.0	0.17	14.2	2.43
4	29.6	18.33	11.5	35.83	16.8	38.44	−11.9	0.00	12.6	0.47	14.5	4.67
5	31.3	7.33	11.1	25.83	18.4	13.67	−12.6	0.00	16.3	0.33	14.3	2.39
6	30.6	10.33	11.0	34.73	19.4	8.83	−6.6	0.00	13.5	0.92	12.0	5.17
7	25.6	8.17	11.3	16.17	16.5	7.91	−12.3	0.00	17.0	0.33	17.0	0.67
8	26.9	9.83	11.3	22.05	17.7	10.46	−13.0	0.00	15.0	0.83	14.5	1.83
9	30.1	18.00	12.0	22.19	17.7	28.33	−12.4	0.00	14.0	0.83	12.8	3.22
10	25.9	15.08	11.9	10.33	16.5	16.15	−11.9	0.00	14.7	0.00	14.5	2.25
11	24.4	10.03	11.7	12.96	16.6	8.47	−11.4	0.00	15.8	1.43	14.0	2.07
12	28.1	9.00	11.4	21.08	16.9	18.67	−14.9	0.00	17.3	0.50	15.0	1.67
13	25.9	3.92	11.7	17.55	16.8	6.89	−8.3	0.00	12.3	2.22	16.5	1.83
14	27.6	84.33	12.0	55.47	15.5	84.67	−12.1	0.00	13.0	0.17	13.0	4.83
15	28.9	6.83	12.1	22.67	18.4	10.22	−11.4	0.00	16.0	1.17	16.5	2.67
16	30.1	5.83	11.0	22.25	18.2	12.37	−8.1	0.00	16.0	0.33	14.5	2.08
17	28.6	11.17	11.5	37.00	17.5	21.06	−9.4	0.00	13.3	0.94	12.3	4.17
18	28.3	11.17	11.6	32.12	17.9	12.20	−13.3	0.00	13.5	1.00	13.0	1.83
19	26.4	10.50	11.1	26.33	17.6	13.33	−10.9	0.00	15.2	0.63	13.7	2.94
20	26.9	20.08	11.9	22.94	17.9	15.77	−10.4	0.00	14.0	1.83	14.7	2.22
21	27.6	19.50	12.1	25.60	18.1	10.50	−10.6	0.00	14.0	1.83	14.7	2.22
22	41.1	0.00	11.6	25.67	22.1	0.83	−9.1	0.00	15.3	1.12	13.0	5.00
23	27.0	14.62	11.5	23.14	18.1	8.67	−10.1	0.00	13.0	1.83	16.0	1.17
24	33.6	3.83	11.5	23.31	18.5	15.04	−11.6	0.00	14.0	0.50	15.0	2.33
25	28.5	12.58	11.6	17.50	19.4	3.67	−14.6	0.00	14.0	0.33	15.7	2.67
26	29.9	8.00	11.6	16.67	18.3	8.10	−11.4	0.00	13.5	0.50	14.5	2.83
27	28.2	10.46	12.5	14.58	18.7	6.46	−13.4	0.00	14.0	1.50	17.0	1.17
28	33.1	2.33	11.5	25.71	18.4	11.92	−7.4	0.00	17.0	0.33	14.7	2.39
29	29.1	20.00	11.5	28.57	18.8	11.56	−8.6	0.00	14.3	1.50	16.0	1.17
30	28.1	17.17	11.3	24.04	18.1	17.50	−10.1	0.00	14.7	1.22	13.8	3.75
31	32.6	6.50	11.5	22.54	19.6	7.00	−13.4	0.00	16.0	0.33	15.0	2.17
32	26.5	26.81	12.1	23.67	17.5	17.58	−9.4	0.00	14.7	0.17	14.5	2.17
33	29.9	10.11	11.5	26.31	17.4	23.08	−13.9	0.00	16.0	0.50	16.0	1.33
34	25.7	28.75	11.5	18.24	16.8	30.77	−11.1	0.00	15.7	0.89	13.5	3.08
35	28.6	10.50	12.7	23.70	17.9	9.67	−12.4	0.00	16.0	0.50	12.5	4.00
36	28.1	11.50	10.8	24.44	18.1	9.33	−16.4	0.00	11.9	1.18	13.0	3.83
37	33.1	5.33	11.6	21.17	18.9	12.11	−10.00	0.00	12.8	1.25	14.3	3.39
38	26.7	12.17	12.1	17.50	16.5	25.19	−13.4	0.00	13.7	1.94	16.3	1.29
39	30.1	7.83	11.5	24.27	18.0	13.58	−7.6	0.00	14.0	1.00	16.50	0.67
40	26.5	25.33	11.8	25.28	18.4	13.33	−12.9	0.00	14.3	0.83	14.7	3.06
41	26.9	25.22	11.3	28.33	18.6	13.00	−7.9	0.00	16.5	0.33	12.5	4.75
42	27.4	18.00	12.7	15.37	16.9	22.00	−9.9	0.00	19.0	0.17	16.5	1.42
43	39.4	5.50	11.3	26.11	20.1	15.33	−10.9	0.00	7.8	1.00	14.0	2.33
44	37.4	8.56	11.9	26.17	20.1	15.50	−9.8	0.00	12.0	0.00	13.5	4.25
45	26.9	19.11	11.7	18.75	17.5	19.67	−12.9	0.00	15.0	0.83	16.0	2.33
46	26.8	15.06	12.1	16.67	17.5	11.37	−11.1	0.00	15.5	0.75	13.5	2.50
47	29.6	12.17	11.4	23.50	17.9	17.25	−10.9	0.00	14.0	0.83	13.5	3.08
48	27.6	16.00	11.3	28.00	17.3	17.97	−7.4	0.00	13.5	0.92	12.7	3.61
49	27.1	16.33	12.1	17.33	17.1	23.17	−10.4	0.00	15.0	0.50	15.0	2.50
50	27.1	13.33	11.8	22.94	18.1	9.00	−9.9	0.00	14.5	0.50	12.0	6.33
51	28.4	22.83	12.2	14.14	18.3	20.20	−9.1	0.00	13.3	1.06	13.6	3.13
52	26.1	17.83	11.6	25.00	17.1	17.05	−10.9	0.00	15.9	0.87	16.5	1.25
53	30.1	9.67	11.9	26.80	17.4	25.67	−9.3	0.00	16.0	0.83	13.0	3.00
Mean	29.0	14.06	11.6	23.67	17.9	16.08	−11.1	0.00	14.5	0.79	14.4	2.75
Range	7.1	19.03	1.0	15.63	2.7	17.97	5.4	0.00	3.7	1.33	3.9	3.50

The figures for daily maximum screen temperatures in July are plotted as frequency distributions in Fig. 2. The bottom rows show the mean and range of the ensemble of values. The range is defined as the interval between the 10th and 90th percentiles of the ensemble distribution of values

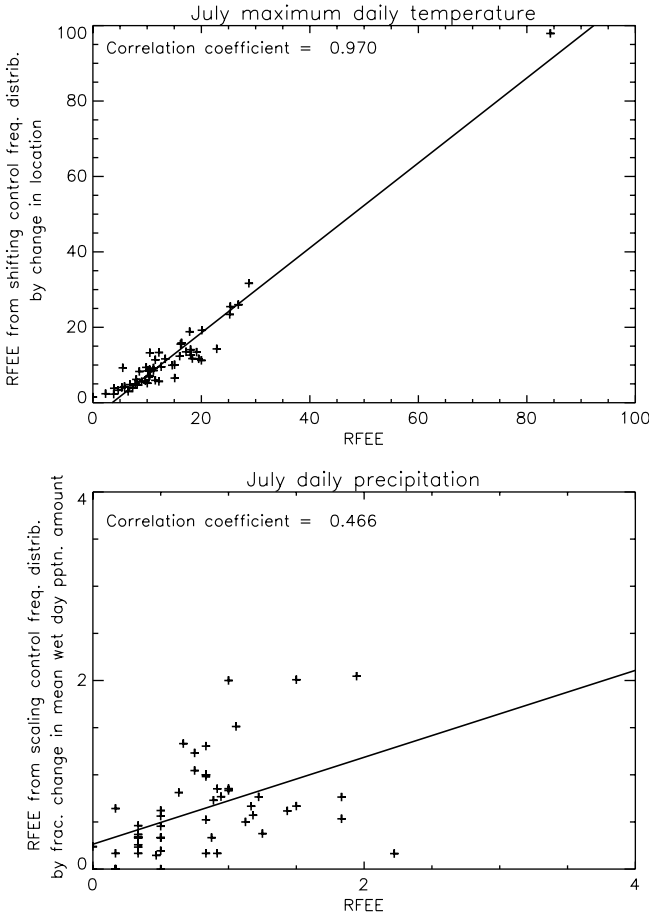

Fig. 3 (*Top*) The relationship between simulated RFEE for July maximum daily screen temperature and an empirical estimate of RFEE obtained by shifting the $1 \times CO_2$ simulation frequency distribution by the change ($2 \times CO_2$ minus $1 \times CO_2$) in the time averages (means) of the daily distributions, assuming no change in the shape of the distribution simulated in the $1 \times CO_2$ run. *Bottom* The relationship between the simulated RFEE for July daily precipitation and an empirical estimate of RFEE obtained by scaling the precipitation amounts on wet days (defined as a day on which precipitation exceeds 0.1 mm) in the $1 \times CO_2$ run by a uniform factor which ensures that the mean wet day precipitation amount is then equal to that simulated in the $2 \times CO_2$ run. The empirical calculation assumes no change in wet day frequency on doubling CO_2. Both panels show results from all 53 ensemble members for the SE England gridbox

typically be expected to occur about once every seven days under $2 \times CO_2$ conditions. However the ensemble range of RFEE values is 19, indicating substantial uncertainty in the magnitude of the simulated increase. There is a very strong correlation between the RFEE and that which would be predicted simply by shifting the $1 \times CO_2$ frequency distribution by the change in the time-averaged value between the $1 \times CO_2$ and $2 \times CO_2$ cases, but with no change in shape (see Fig. 3). Thus ensemble members with a larger time-averaged warming generally show a larger RFEE, with changes in the shape of the frequency distributions being of lesser importance in explaining the RFEE. Nevertheless, in many cases there are significant changes in the shape of the distribution in response to doubled CO_2 (Fig. 2), with increases in standard deviation and positive skewness being common. In fact, the ensemble-mean skewness of the July maximum daily temperature distributions increases from 0.75 in the $1 \times CO_2$ cases to 1.25 in the $2 \times CO_2$ simulations, extending the 'hot tails' of the distributions and thus increasing the RFEE. This change (i.e. a disproportionate increase in the frequency of the hottest summer days) is accompanied by an ensemble-mean decrease of 25% in the soil moisture content. The simulation of reductions in soil moisture is consistent with many previous studies (for example Wetherald and Manabe 1999; Cubasch et al. 2001) which found that reductions in summertime soil moisture levels in mid-latitude continental regions reduced the extent to which increases in screen temperature can be limited by increases in evaporation. In HadCM3, the configuration of the present model in which the atmosphere is coupled to a dynamical ocean, Brabson et al. (2005) find that simulated reductions in future soil moisture over the UK are an important driver of increases in extreme summer temperatures and the risk of extrended hot spells. Note, however, that the magnitude of soil moisture reductions, and hence their impact on the temperature distributions, is uncertain due to the influence of several different physical mechanisms (Rowell and Jones 2006).

Table 3 also shows the ensemble statistics for the other daily climate variables under consideration, for the SE England gridbox. January and July maximum screen temperatures and July minimum screen temperatures all exhibit large increases in ensemble-mean extreme event frequency in response to doubled CO_2. The ensemble range of RFEE varies between from 15.6 to 19.0 for these variables, indicating significant uncertainty in the expected change. Note that ensemble member 14 gives much larger values of RFEE for daily temperatures than the other members do. This simulation (which uses a low value for the convective entrainment rate coefficient) has a climate sensitivity of 7.0°C, well outside the envelope of values (2.2–4.1°C) encompassing the other ensemble members. Our choice of the 10th–90th percentile interval to characterise the ensemble range of RFEE values ensures that the range is not unduly sensitive to this outlier experiment, which gives a $1 \times CO_2$ simulation of lower quality than the other ensemble members (its value of the CPI is 8.3, compared with a range of 5.3–6.8 for the other 52 model versions). Despite the uncertainty in the response, RFEE invariably exceeds unity for each ensemble member, showing that a prediction of an increase in extremely warm winter and summer days and summer nights is highly robust according to the ensemble.

In the case of winter minimum temperatures, changes in the frequency of cold events at middle and high latitudes are at least as important to society as changes in the frequency of warm events. For example, Meehl et al. (2004) reported large reductions in the frequency of frost days over northern Europe during the twenty-first century in an ensemble of four simulations of the PCM

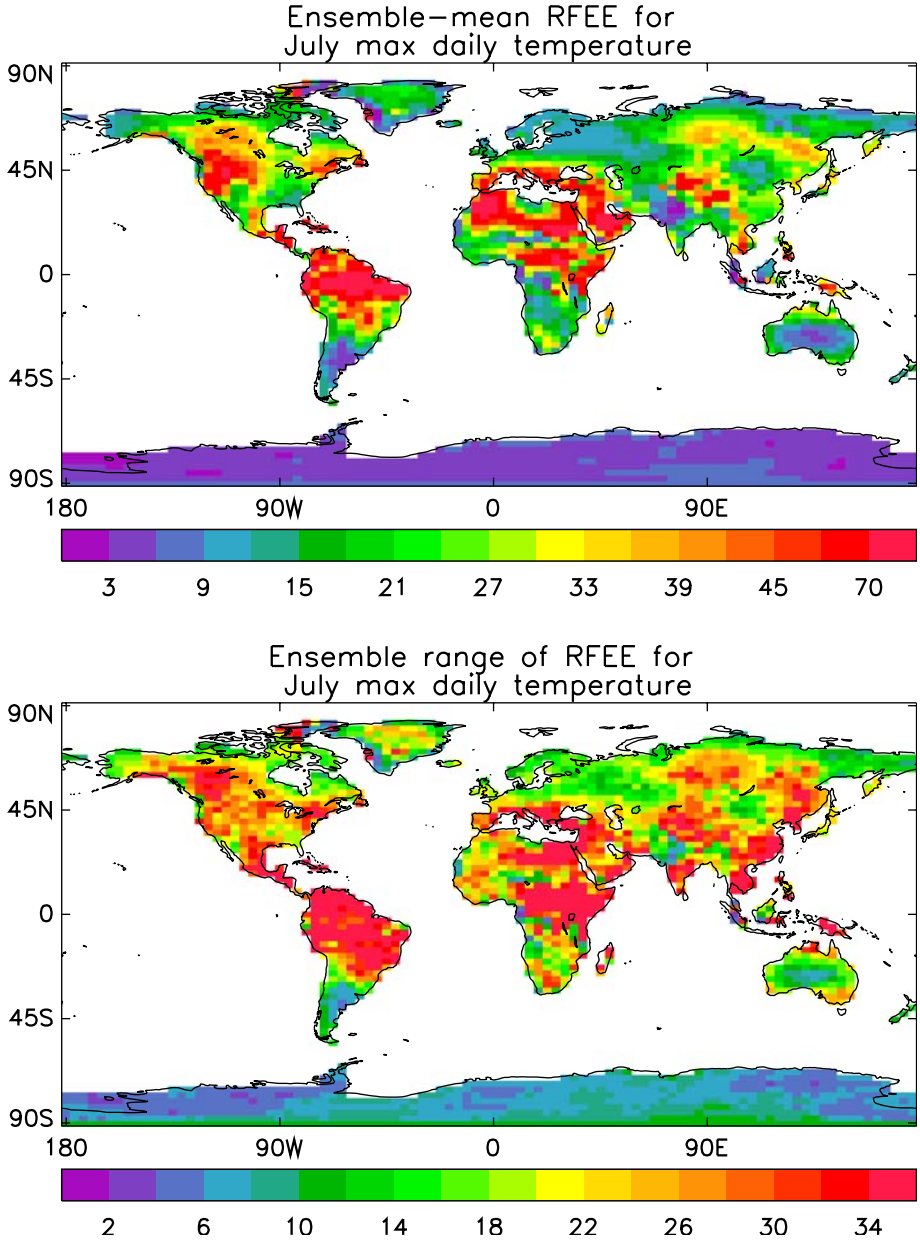

Fig. 4 Maps of the ensemble-mean relative frequency of extreme events (RFEE) simulated in response to doubled CO_2 (*top*) and ensemble range of the RFEE (*bottom*), for July maximum daily screen temperature. The range is defined as the interval between the 10th and 90th percentiles of the ensemble distribution of RFEE values

climate model driven by a business-as-usual scenario of changes in greenhouse gas concentrations. Here we consider January minimum temperatures, defining the extreme event threshold for a given ensemble member as the first percentile of the daily frequency distribution from the $1 \times CO_2$ run. An extreme cold night is one on which the temperature falls below this threshold (typically around -10°C). According to this definition the occurrence of extreme cold nights over SE England is completely eradicated in the $2 \times CO_2$ climate in all 53 ensemble members (Table 3). Increases in extreme warm days and decreases in cold nights under increasing CO_2 are a consistent prediction of previous studies (e.g. Cubasch et al. 2001). However, the present results provide the first quantitative estimate of the associated uncertainty.

An increase in extremely wet January days is obtained (see Table 3), consistent with results from regional climate model integrations (Jones and Reid 2001). The prediction of an increase is robust, in that all but two ensemble members simulate RFEE > 1. However there is substantial uncertainty in the magnitude of the increase, given that the ensemble range in RFEE comfortably exceeds the ensemble-mean of (RFEE-1), the change in event frequency. For July, an ensemble-mean decrease in extremely wet days is predicted. However, the magnitude of the decrease is much smaller than the ensemble range, and 13 ensemble members actually simulate an increase. It is therefore less clear whether to expect an increase or decrease in the RFEE of July precipitation for SE England under $2 \times CO_2$ conditions. In both January and July the magnitude of the ensem-

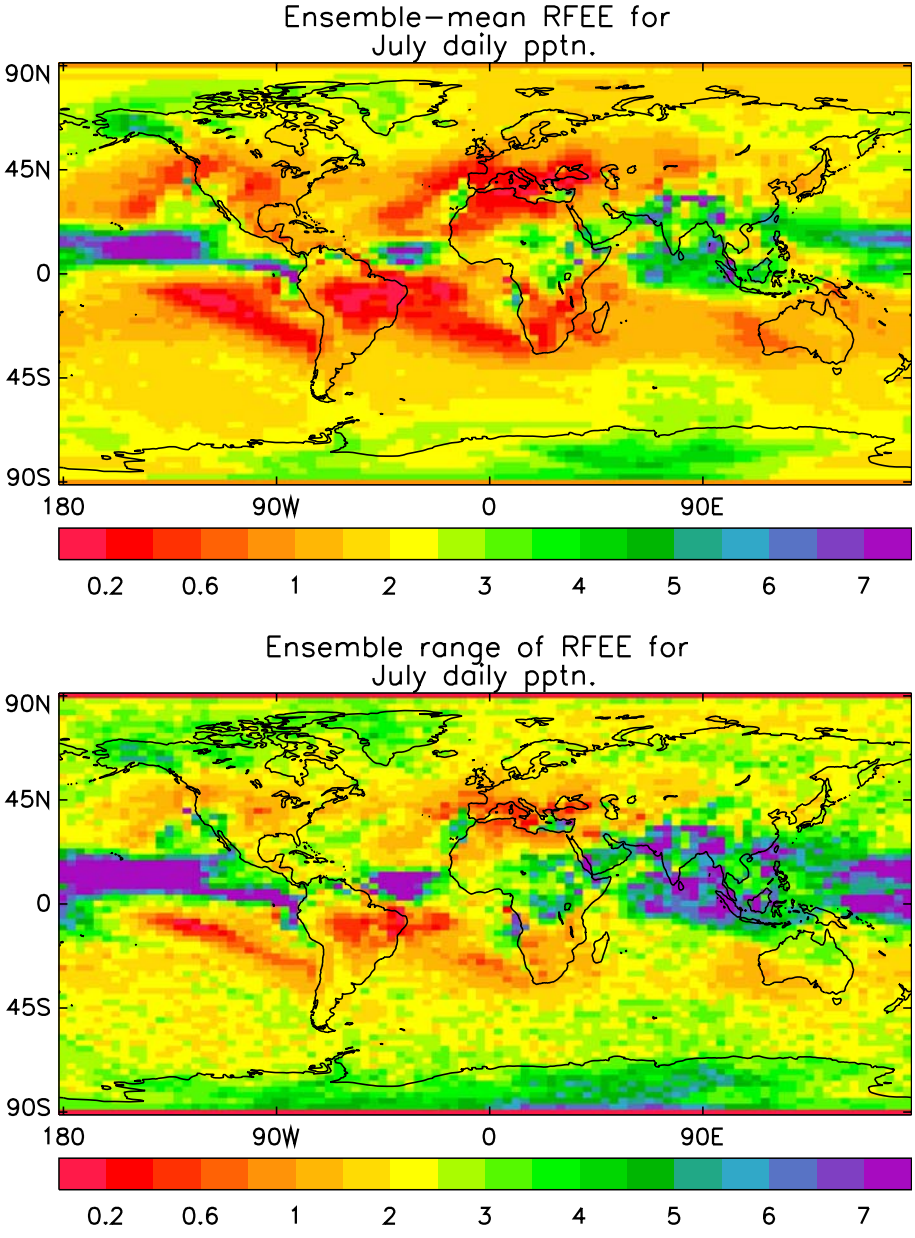

Fig. 5 As Fig. 4, but for July daily precipitation

ble-mean change is modest compared with the results for screen temperature. This is consistent with the results of Zwiers and Kharin (1998) and Kharin and Zwiers (2000), who found that the signal of change in precipitation extremes under $2 \times CO_2$ was typically weaker (compared with natural variability) than for screen temperature. Figure 3 shows the correlation between RFEE for daily precipitation over SE England in July and the RFEE which would be obtained by scaling all $1 \times CO_2$ wet day intensities so as to give a mean wet day intensity equal to that in the $2 \times CO_2$ case. The correlation of 0.47 is sufficiently low to indicate that changes in the shape of the daily distributions under $2 \times CO_2$ (e.g. arising from changes in the frequency of wet days) are much more important in explaining the RFEE changes for precipitation than in the case of screen temperature. This is discussed further below.

In Figs. 4, 5 and 6 we summarise the results of repeating the analysis of changes over SE England for the whole globe, for July. Table 4 summarises key statistics of the global changes, for both July and January. In all cases extreme events are defined as those exceeding the 99th percentile threshold. This is done to avoid confusion when interpreting plots showing results from both hemispheres for a given month, however note that this approach differs from the approach taken for the SE England analysis, where an extreme event for night minimum temperature in winter was defined as one falling below the first percentile.

Figure 4 shows maps of the ensemble mean and range of RFEE for maximum daily screen temperature. We do not plot results for ocean gridboxes, because our use of a slab ocean means that variability arising from the ocean circulation is not captured, hence the simulated fre-

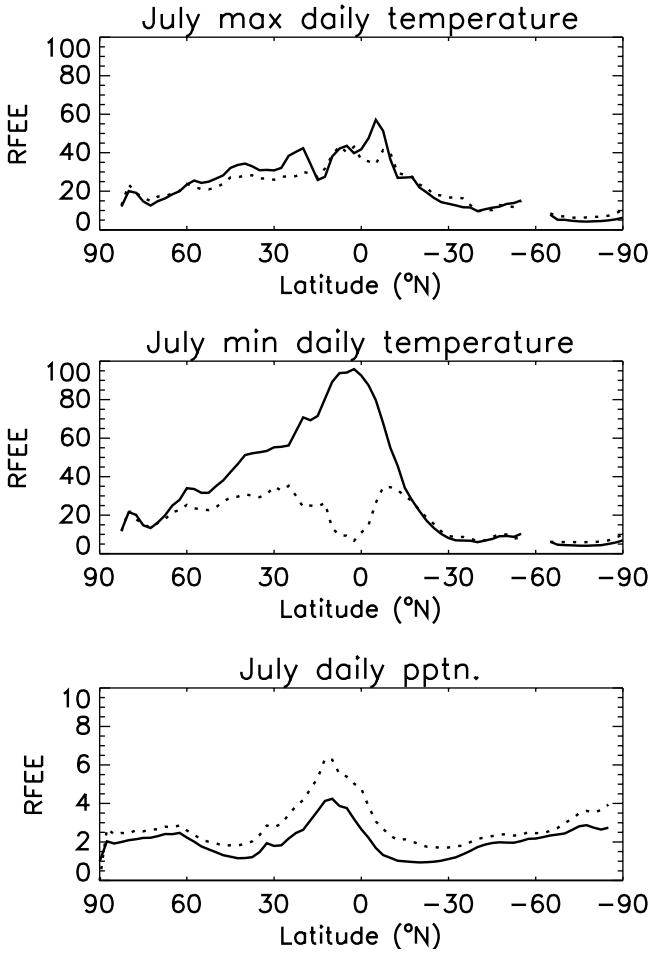

Fig. 6 Zonal averages of the ensemble-mean relative frequency of extreme events (RFEE) in the $2 \times CO_2$ simulations (*solid lines*) and the ensemble range of the RFEE (*dotted lines*) for (from *top to bottom*) maximum daily screen temperature, minimum daily screen temperature, and daily precipitation, in July. The global mean values of these variables are shown in Table 4

quency distributions are likely to be unrealistically narrow and the values of RFEE too high. Over land, extremely warm days are predicted to become much more frequent in a $2 \times CO_2$ climate, with global averages of the ensemble-mean RFEE of 19.5 (January) and 27.5 (July) (see Table 4). The largest values of RFEE (50 or greater) are simulated in western North America, northern South America (including Amazonia), central and northern Africa, the Middle East, southern Europe and central Asia. To place these values in context, an RFEE of 50 indicates that temperatures exceeding the threshold defining a one in a hundred event in the $1 \times CO_2$ simulations occur once every two days in the $2 \times CO_2$ simulations. This implies (on average) a change in time averaged temperature similar in magnitude to the difference between the temperatures associated with the 99th and 50th percentiles in the $1 \times CO_2$ simulation. The typical range in RFEE across the ensemble is around 20 (Table 4), with larger values tending to occur at locations where the ensemble-mean RFEE is larger than average (Fig. 4). Values of the ensemble-mean of (RFEE-1), the change in the frequency of extreme events, exceed the ensemble range over 39% of the land area in July, and 24% in January (Table 4), indicative of a projected increase in extreme maximum temperatures which is highly robust across the ensemble. Nevertheless, the ensemble range is large enough to imply substantial uncertainty in the response at most locations.

Ensemble-mean RFEE values for daily minimum screen temperature tend to be even larger than for maximum temperatures, with zonal average values approaching 100 in equatorial regions (Fig. 6), and global averages of 28.9 (January) and 43.7 (July) (Table 4). Once again these predictions are highly robust across the ensemble at many locations, with the ensemble-mean change in extreme event frequency exceeding the ensemble range over 68% of the land in July, and 41% in January (Table 4). The RFEE values for minimum screen temperature exceed those for maximum screen temperature by the greatest amount in the tropics, particularly just to the north of the equator in July and to the south in January.

Spatial variations in RFEE for maximum and minimum screen temperature, and differences between the changes for minimum and maximum temperature at a given location, could plausibly be related to a number of variables (Kharin and Zwiers 2000). These include soil

Table 4 Global averages of the ensemble-mean extreme (and unusual) event threshold (EET), the ensemble-mean RFEE (and RFUE), the ensemble range of the RFEE (and RFUE) and the proportion of the globe over which the change in extreme (unusual) event frequency (i.e. [ensemble-mean RFEE (or RFUE) - 1]) exceeds the ensemble range, for July and January maximum daily screen temperature, July and January minimum daily screen temperature and July and January daily precipitation

	July max T		Jan max T		July min T		Jan min T		July pptn		Jan pptn	
	Extr.	Unus.	Extr.	Unus.	Extr.	Unus.	Extr.	Unus.	Extr.	Unus.	Extr.	Unus.
Global mean of ens. mean (E/U) ET (°C or mm/day)	21.9	20.1	18.1	16.1	17.6	16.0	14.4	12.4	16.7	6.8	18.0	6.8
Global mean of ens. mean RF(E/U)E	27.5	5.7	19.5	4.8	43.7	6.7	28.9	5.4	2.0	1.2	2.1	1.2
Global mean of ensemble range	24.8	2.4	20.5	2.4	22.1	1.7	15.9	1.9	3.0	0.8	3.0	0.8
Percentage of area for which (RF(E/U)E − 1) > range	38.6	84.8	24.4	86.0	67.5	92.3	40.5	90.7	0.0	2.0	0.2	3.6

Global averages are calculated over land and sea points for precipitation, but only over land points for screen temperature. In each ensemble member, 'extreme' events are defined as those exceeding the 99th percentile (EET) of the distribution of daily values simulated in the $1 \times CO_2$ integration, and 'unusual' events as those exceeding the 90th percentile (UET)

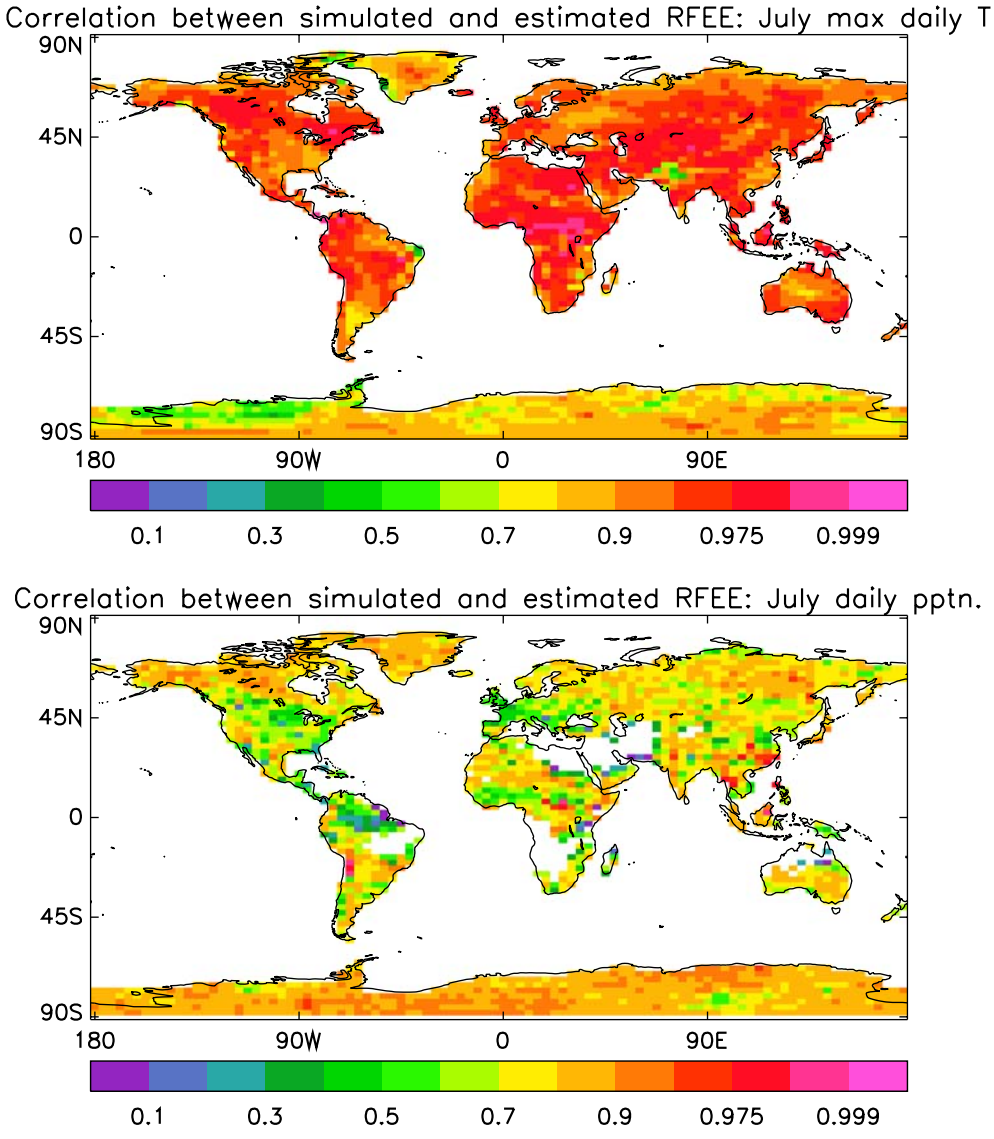

Fig. 7 *Top* Correlations across the 53 member ensemble between the simulated RFEE for July maximum daily screen temperature and an empirical estimate of RFEE obtained by uniformly shifting the $1 \times CO_2$ simulation of the daily frequency distribution by the change ($2 \times CO_2$ minus $1 \times CO_2$) in the mean of the distribution. *Bottom* Correlations across the ensemble between the simulated RFEE for July daily precipitation and an empirical estimate of RFEE obtained by scaling the precipitation amounts on wet days in the $1 \times CO_2$ run by a uniform factor which ensures that the mean wet day precipitation amount is then equal to that simulated in the $2 \times CO_2$ run. The empirical calculation assumes no change in wet day frequency on doubling CO_2

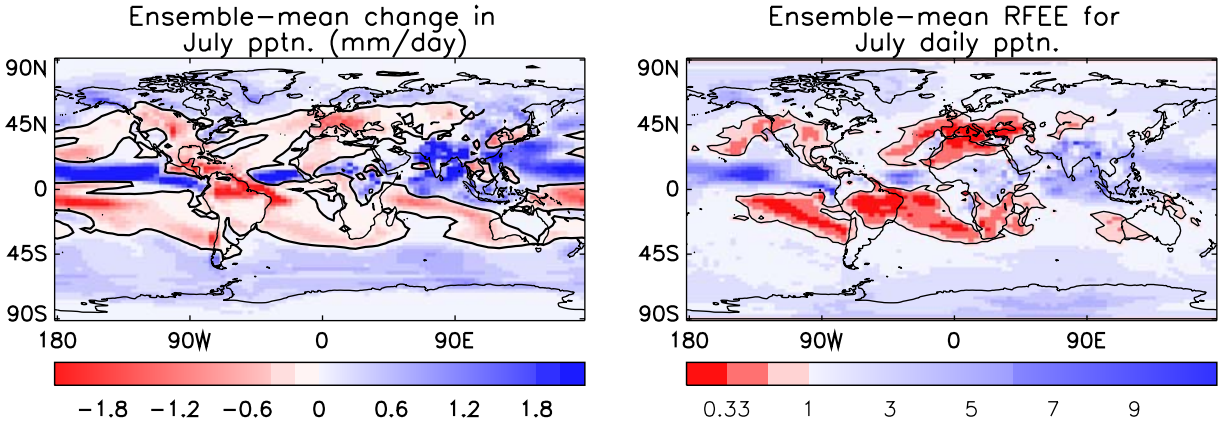

Fig. 8 A comparison between the ensemble-mean change in time-averaged precipitation (*left*) and the ensemble-mean relative frequency of extreme daily precipitation events (*right*), in response to doubled CO_2, for July. *Solid line* denotes zero change (*left*) and RFEE = 1 (*right*)

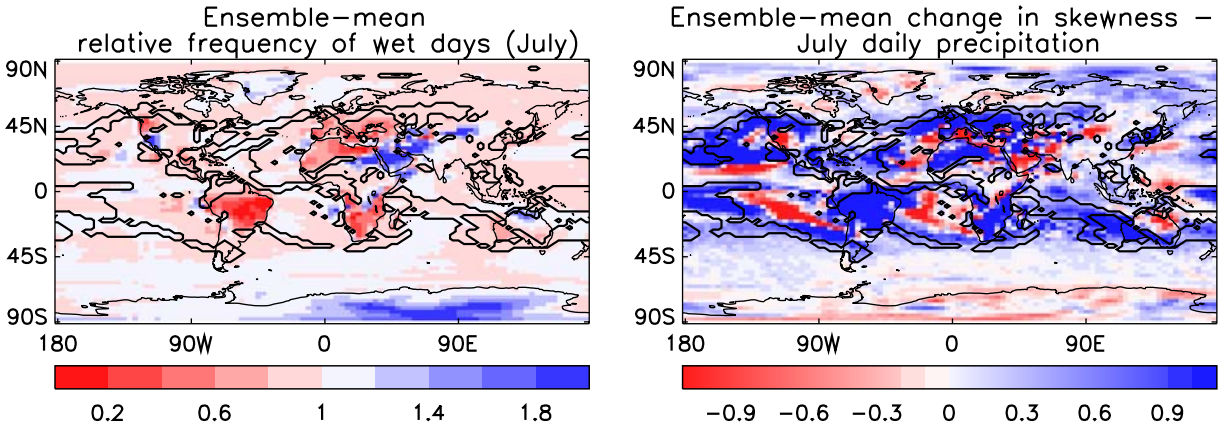

Fig. 9 Ensemble-mean relative frequencies of wet days (*left*), and the ensemble-mean change in the skewness of the daily precipitation frequency distributions (*right*), in response to doubled CO_2, for July, where a wet day is defined as one on which precipitation exceeds 0.1 mm. The solid contour encloses regions in which the ensemble-mean response to doubled CO_2 shows decreases in time-averaged precipitation accompanied by increases in daily extreme events (RFEE > 1, see Fig. 8)

moisture, precipitation, evaporation, cloud cover, snow and sea-ice cover, diurnal temperature range and near surface wind. Inspection of the spatial distributions of changes in these variables suggests that none of them can be singled out as possessing a dominant link with the changes in daily surface temperature extremes presented above. Hence a number of mechanisms and feedbacks are likely to be involved. These will be investigated in a future paper.

Simulated increases in the frequency of extremely wet days are much more modest than the corresponding results for screen temperature (Table 4 and Fig. 5 cf. Fig. 4). Nevertheless the results suggest a doubling of such events in the $2 \times CO_2$ simulations relative to the $1 \times CO_2$ simulations in the ensemble and global mean, providing strong support for results from previous experiments predicting an increased intensity of precipitation events in response to increased CO_2 (e.g. Cubasch et al. 2001; Jones and Reid 2001; Senior et al. 2002). Ensemble mean values of RFEE vary substantially with location, exceeding 5 in July in some equatorial regions and parts of Antarctica. Conversely, values are less than unity in many subtropical areas, notably over large parts of north-western Africa and Amazonia, southern Africa and the Mediterranean region. Note, however, that care must be taken in interpreting the significance of the results in locations where there is very little rainfall [e.g. Antarctica and the Mediterranean in July (see Fig. 5)], since longer simulations might be required to obtain robust estimates of an extreme threshold, and of changes in the frequency of exceedance in response to changes in radiative forcing. Over Europe, values of ensemble-mean RFEE are similar to those found from a pair of individual global and regional climate model simulations by Durman et al. (2001), who obtained typical values of around 2.5 in winter and 1.5 in summer.

The ensemble simulations reveal a large uncertainty in the expected changes in extremes in most regions. The ensemble-mean change in event frequency is much

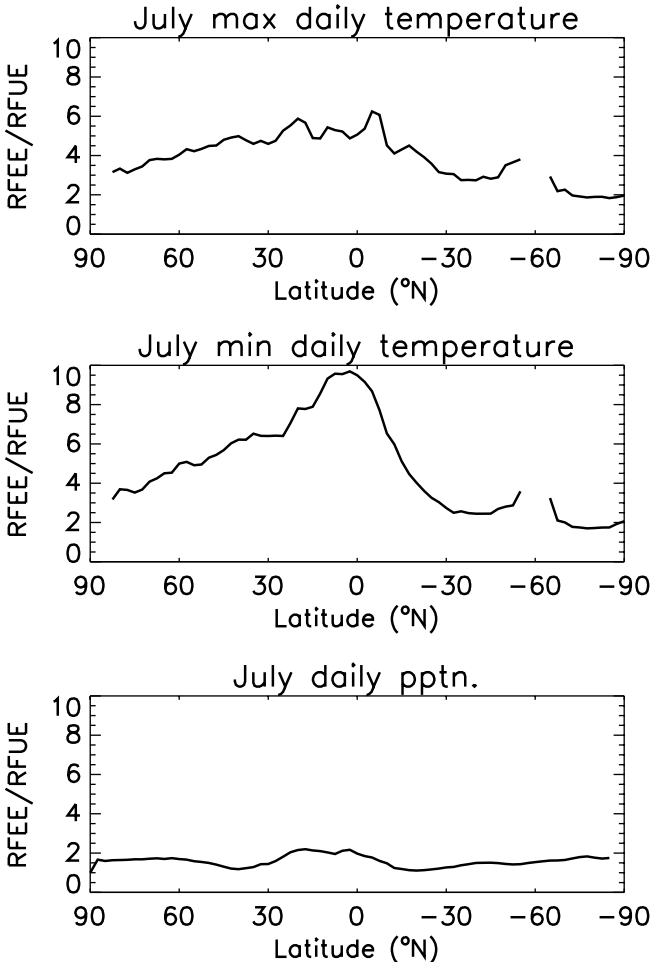

Fig. 10 Zonal averages of the ratio of the ensemble-mean RFEE to the ensemble-mean RFUE (with 'unusual' events defined with respect to the 90th (rather than the 99th) percentile threshold under $1 \times CO_2$ conditions) for July maximum daily screen temperature (*top*), minimum daily screen temperature (*middle*) and daily precipitation (*bottom*)

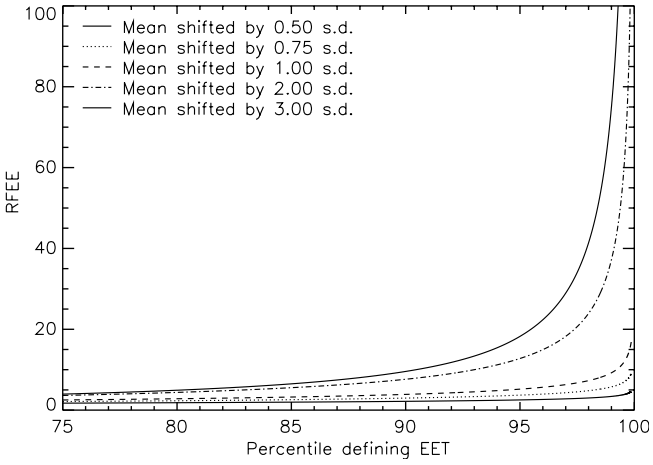

Fig. 11 The relationship between the relative frequency of 'extreme' events under $2\times CO_2$ (RFEE) and the percentile of the $1 \times CO_2$ frequency distribution used to define the 'extreme' event, for an idealised case where the $1 \times CO_2$ distribution is Gaussian and the response to doubled CO_2 is characterised by a shift (increase) in the mean of the distribution with no change in shape. Values are shown for a number of hypothetical shifts, expressed as the number of standard deviations of the $1 \times CO_2$ distribution

smaller than the ensemble range over large areas of the globe (Fig. 6). The ensemble mean change hardly ever exceeds the range (Table 4), so it is not generally possible to identify a change in the frequency of extreme precipitation at an individual location with a high degree of confidence. On the other hand, at most locations the possibility of large increases cannot be ruled out. For example, the global average of the 90th percentile of the ensemble of RFEE values is 3.6, in both January and July.

Figure 7 generalises the analysis of Fig. 3 to all land grid points, showing maps of the correlation across ensemble members between simulated RFEE and that obtained by uniformly shifting (in the case of maximum screen temperature) or scaling (in the case of wet day precipitation amounts) the daily frequency distributions from the $1 \times CO_2$ integrations, with no change in shape. For screen temperature the correlation is high everywhere (global mean value 0.93), suggesting that changes in the mean of the frequency distributions under $2 \times CO_2$ dominate in explaining the RFEE. For precipitation the correlations are lower, though the global mean value of 0.72 is significantly higher than was obtained for the SE England grid point in Fig. 3. Thus around half of the variance of RFEE can typically be explained by a uniform scaling of wet-day precipitation amounts, with the remaining variance attributable to changes in the shape of the daily distributions in response to doubled CO_2. Kharin and Zwiers (2005) obtain qualitatively consistent results from a different approach, based on fitting generalised extreme value distribution parameters to annual temperature and precipitation maxima in transient simulations of a coupled ocean-atmosphere GCM. They find that changes in return values for annual temperature extremes are, in most regions, explained principally by a uniform shift in the distributions, whereas changes in the width and shape of the distributions are more important in explaining the changes in precipitation extremes. Wehner (2004) finds that simulated changes in 20 year return period values for extreme daily precipitation are more strongly related to changes in the standard deviation of daily precipitation than to changes in the time average, based on analysis of transient climate change simulations of the NCAR parallel climate model.

The relationship between changes in time-averaged precipitation and RFEE in our ensemble is shown in Fig. 8. Regions over which ensemble- and time-averaged precipitation is predicted to decrease (41% of the globe in July) are much more widespread than areas over which extreme precipitation events become less common (i.e. ensemble mean RFEE < 1) (20% of the globe). Thus there are substantial areas (21% of the globe) over which time-averaged precipitation decreases under $2 \times CO_2$ conditions while the relative frequency of extreme precipitation events increases. These include parts of north-eastern Europe, central and eastern Asia, Canada, South America, western Africa and Australia. We find results of a similar nature in January (not shown), and indeed similar conclusions for the European region were drawn by Christensen and Christensen (2003), based on individual simulations of the response to SRES A2 and B2 scenario forcing for 2071–2100 relative to 1961–1990 using a high-resolution regional climate model. Our results have important implications for impact studies in the regions concerned, suggesting a trend towards fewer rain events, but an increase in the intensity of the heaviest events (see also Gregory and Mitchell 1995). This behaviour is consistent with the argument that extreme precipitation events will tend to increase more rapidly than time-averaged precipitation in a warming climate (Allen and Ingram 2002), since the frequency of intense events is likely to be controlled by the moisture-carrying capacity of the air, which increases exponentially with temperature, whereas time-averaged precipitation is likely to be controlled by radiative cooling of the troposphere, which increases approximately linearly with temperature. Results supporting this argument were also found by Zwiers and Kharin (1998) and Kharin and Zwiers (2000) using a different GCM to that used here. Emori and Brown (2005) analyse results from six GCMs, finding that globally averaged increases in extreme values of daily precipitation are about twice the increase in time averaged precipitation, and that in most regions the difference is explained mainly by the increase in moisture-holding capacity of the atmosphere, with changes in circulation playing only a secondary role.

Figure 9 examines the nature of changes in shape of the distributions of daily precipitation in response to doubled CO_2. There are reductions in wet day frequency over much of the globe (as expected given the discussion in the previous paragraph), including many areas where time-averaged precipitation increases. Reductions in wet

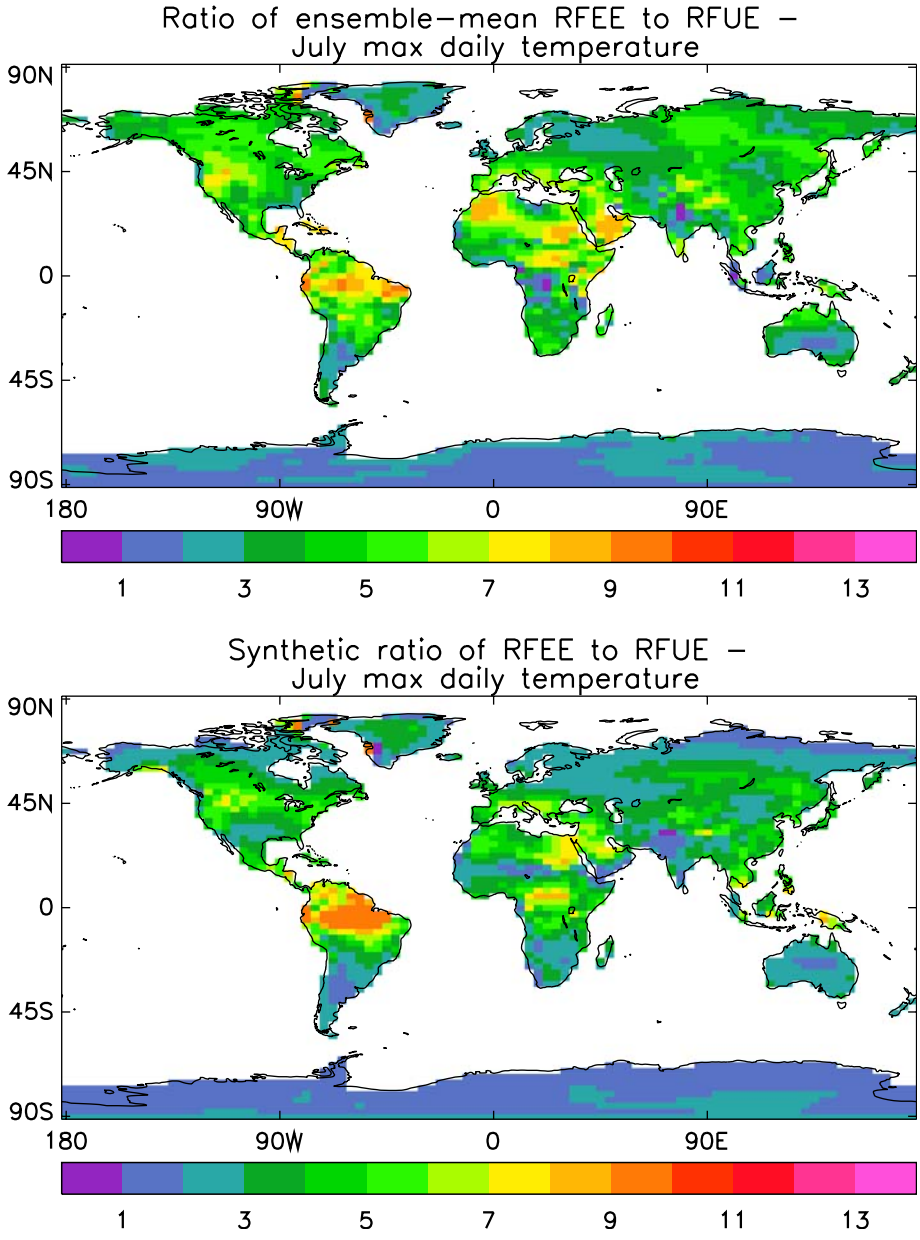

Fig. 12 *Top* Ratios of the ensemble-mean RFEE to the ensemble-mean RFUE for July maximum daily screen temperature (also shown as zonal averages in Fig. 10). *Bottom* Ratios of RFEE to RFUE which would occur if the model simulations replicated the relationship shown in Fig. 11. These are obtained from the ensemble-mean change in the time average of the daily frequency distributions in response to doubled CO_2 and the ensemble-mean standard deviation of the $1 \times CO_2$ run frequency distributions

day frequency are generally accompanied by increases in skewness and vice versa. We focus in particular on regions enclosed by the black contour, which constitute areas where the ensemble-mean response shows reductions in time-averaged precipitation accompanied by increases in RFEE. In these areas, increases in extreme precipitation frequency could theoretically be explained either by a reduction in wet day frequency accompanied by a uniform increase in wet day intensity, or by a reduction in both wet day frequency and average wet day intensity accompanied by an increase in skewness. The results consistently identify the latter as the correct explanation (the changes in average wet day intensity are not shown), indicating that the wet tails of the distributions become longer and flatter under $2 \times CO_2$ conditions.

In Fig. 10 and Table 4 we compare changes in the frequency of extreme events in response to doubled CO_2 with changes in the frequency of 'unusual' events (defined with respect to a 90th, rather than a 99th, percentile threshold). The results show, for all variables and latitude zones, that unusual events (UE) do not increase in frequency as much as extreme events. The contrast between the changes in extreme and unusual events is larger for screen temperature than precipitation, and is particularly large for minimum screen temperature in the tropics (Fig. 10). The ubiquitous nature of the result that extreme events increase more than unusual results, coupled with the fact that it is obtained from a large ensemble of GCM versions designed to sample a wide range of modelling uncertainties, suggests that it is likely to be a consistent feature of the response to increases in

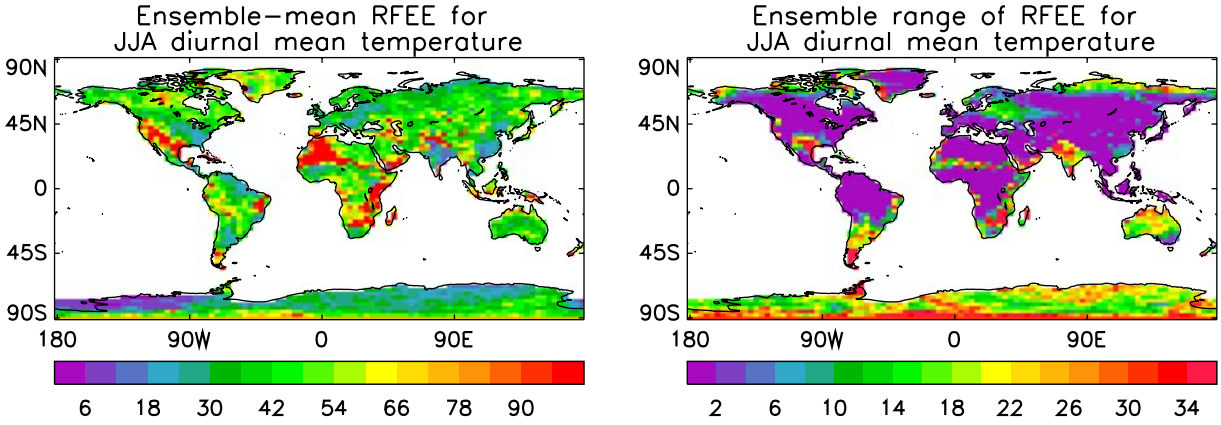

Fig. 13 Ensemble-mean relative frequency of extreme events under $2 \times CO_2$ (RFEE) (*left*) and ensemble range of RFEE (*right*) (global means of which are shown in Table 5), for seasonal values of diurnal mean screen temperature in June–July–August (JJA)

greenhouse gas forcing. The predicted increases in unusual temperature events, though smaller, tend to be more robust than the increases in extreme events, with the value of [ensemble-mean RFUE - 1] exceeding the ensemble range over 85% or more of the land area (Table 4).

Figure 11 shows how the relative frequency of an extreme event would vary as a function of the threshold used to define that event, given a uniform shift in the time average (mean) of an assumed Gaussian frequency distribution in response to doubled CO_2. For a given change in time average, larger values of RFEE are found as the threshold defining the event becomes more extreme, qualitatively consistent with the comparison between simulated extreme and unusual events described above. For screen temperature (though not precipitation), results presented earlier (Fig. 7) suggest that the uniform shift shown in Fig. 11 explains the simulated changes well, at least to first order. Accordingly, Fig. 12 compares simulated ratios of ensemble-mean RFEE to ensemble-mean RFUE for July maximum daily screen temperature against estimated ratios obtained from the relationship of Fig. 11, using the ensemble-mean change in the time average of the daily frequency distributions in response to doubled CO_2 and the ensemble-mean standard deviation of the $1 \times CO_2$ frequency distributions. The correspondence between the large-scale spatial variations in these two plots is good, implying that the difference between RFEE and RFUE is primarily controlled by simple geometric characteristics of the frequency distributions. Note, however, that there are regional differences between the two plots, indicating that in some areas changes in distributional shape, such as the increases in positive skewness and development of bimodality reported by Clark et al. (2006) in areas of restricted evaporation, do play a significant role in determining the changes in extreme event frequencies at different thresholds.

4 Extreme seasonal means

Some of the largest impacts of weather and climate extremes occur when anomalous conditions are sustained

Table 5 Global averages of the ensemble-mean extreme event threshold (EET), the ensemble-mean RFEE, the ensemble range of the RFEE (cf. Figs. 13 and 14) and the proportion of the globe over which the ensemble-mean of the change in extreme event frequency (RFEE − 1) exceeds the ensemble range, calculated from both seasonal and daily data, for screen temperature and precipitation in boreal summer and winter

	JJA mean T/ July max T		DJF mean T/ Jan max T		JJA mean pptn/July mean pptn		DJF mean pptn/Jan mean pptn	
	Seas.	Dail.	Seas.	Dail.	Seas.	Dail.	Seas.	Dail.
Global mean of ens. mean EET (°C or mm/day)	16.8	21.3	13.2	17.5	4.5	12.9	4.5	13.7
Global mean of ens. mean RFEE	56.6	15.6	59.0	11.7	4.8	1.6	5.3	1.7
Global mean of ensemble range of RFEE	10.7	11.3	23.8	9.8	8.4	1.9	8.4	1.8
Percentage of area for which (RFEE − 1) > range	98.2	58.0	98.8	47.6	4.6	0.4	5.5	1.1

Global averages are calculated over land and sea points for precipitation, but only over land points for screen temperature. Seasonal extremes are calculated from distributions of JJA or DJF precipitation or diurnal mean screen temperature, while daily extremes are calculated from distributions of July or January precipitation or daily maximum screen temperature. Here, daily extreme events were defined with respect to a 97.725th percentile threshold, which corresponds to two standard deviations above the mean assuming a Gaussian distribution. This gives a threshold consistent with that used for the seasonal data

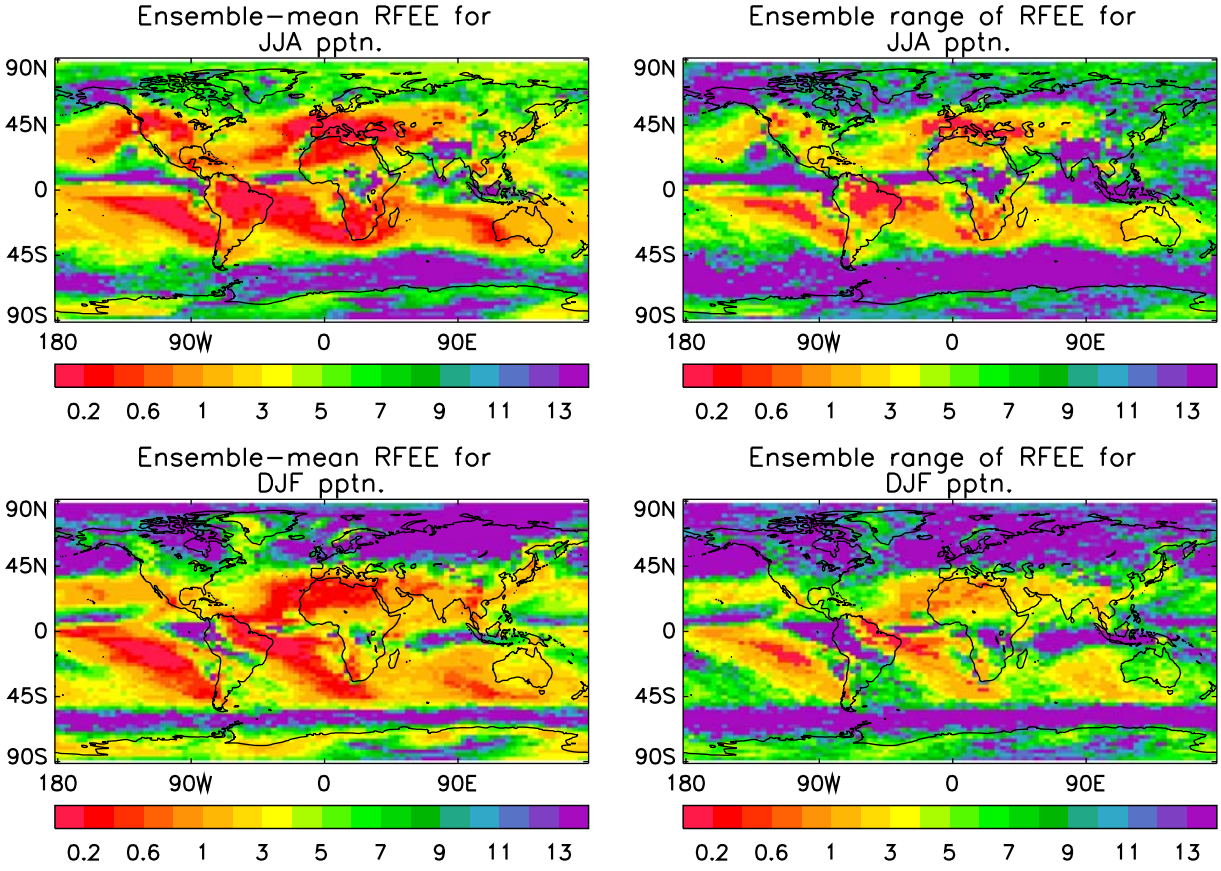

Fig. 14 As Fig. 13, but for JJA precipitation (*top*) and DJF precipitation (*bottom*)

over an extended period. Summer heatwaves (e.g. Karl and Knight 1997; Schär et al. 2004) can lead to a range of adverse economic and societal impacts, including increases in human mortality and the risk of wildfires, while unusually warm winter temperatures can have both positive impacts (such as reduced energy costs (Meehl et al. 2000a)) and negative impacts (such as species loss through disruption of seasonal population dynamics (Parmesan et al. 2000)). Periods of high precipitation sustained over several months are also important, playing a major role in determining regional flooding risks (e.g. Marsh 2001). Here we consider extreme values of seasonal averages of screen temperature and precipitation in order to characterise anthropogenic changes in sustained climate anomalies.

4.1 Method

We focus on JJA and DJF, calculating precipitation and diurnal mean screen temperature fields for each ensemble member ($1 \times CO_2$ and $2 \times CO_2$), for each of the 20 years in the equilibrium period. Then, for each ensemble member and at each gridbox, the extreme event threshold was defined to be two (interannual) standard deviations above the multiannual mean, calculated from the twenty $1 \times CO_2$ seasons. An extreme season is identified as one in which the seasonal mean exceeds this threshold. (Note that this approach necessarily differs from the method of defining the extreme event threshold as a fixed percentile, as in Sect. 3.1, since the sample size available from each $1 \times CO_2$ simulation (20) is too small to support the accurate identification of percentiles). Thus, for each gridbox of each member (i), the percentage of seasons exceeding the extreme event threshold, under both $1 \times CO_2$ and $2 \times CO_2$ conditions (C_i and D_i, respectively), was calculated.

Ideally, at each gridbox, the RFEE for ensemble member i should be calculated as $\mathrm{RFEE}_i = D_i/C_i$. For grid boxes where the distribution of seasonal means approximates a Gaussian distribution, we would expect an average value of about 0.5 for C_i, given samples of 20. This alerts us to the problem that the ratio D_i/C_i will be very sensitive to sampling uncertainties in C_i, and will be undefined at gridboxes where $C_i = 0$. Therefore, we adopt a more stable definition of

$$\mathrm{RFEE}_i = \frac{D_i}{\langle C \rangle} \tag{1}$$

where $\langle C \rangle$ denotes the ensemble mean of the C_i. In other words, we assume that the probability of exceeding the mean plus two standard deviation threshold in the $1 \times CO_2$ run is the same for all ensemble members.

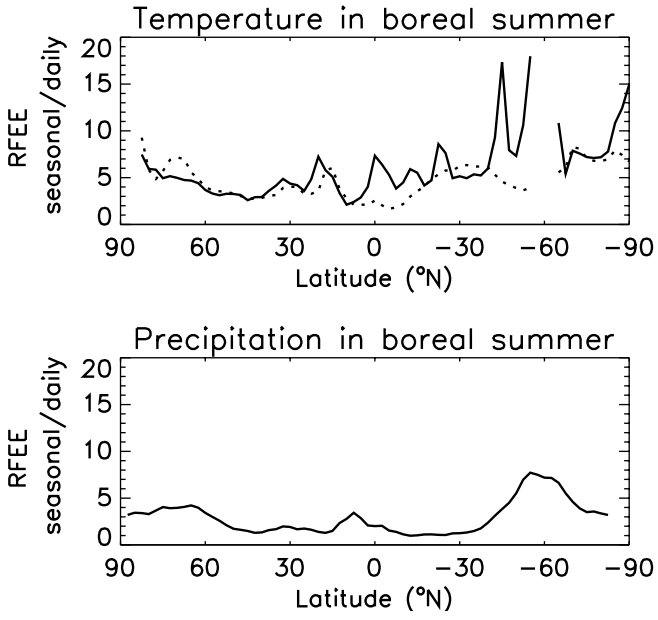

Fig. 15 Zonal averages of the ratio of the ensemble-mean RFEE calculated from seasonal data (Figs. 13, 14) to the ensemble-mean RFEE calculated from daily data, for screen temperature (*top*) and precipitation (*bottom*) in boreal summer. Daily extreme events were defined with respect to a 97.725th percentile threshold, which corresponds to two standard deviations above the mean assuming a Gaussian distribution. This gives a threshold consistent with that used for the seasonal data. Seasonal extremes are calculated from distributions of JJA diurnal mean screen temperature or precipitation, while daily extremes are calculated from distributions of July daily maximum screen temperature or precipitation. The *dashed curve in the top panel* shows ratios of seasonal RFEE to the daily RFEE which would occur if the model response to doubled CO_2 took the form of a uniform shift relative to an assumed Gaussian $1 \times CO_2$ distribution. These ratios are obtained from the ensemble-mean change in the time average of the (daily or seasonal) distributions in response to doubled CO_2 and the ensemble-mean standard deviation of the $1 \times CO_2$ run distributions, using the relationship shown in Fig. 11

This work follows on from the study by Palmer and Räisänen (2002), who used a multi-model ensemble of 19 coupled ocean-atmosphere GCMs to estimate the ensemble-mean relative frequency of extreme wet seasons under a doubling of atmospheric CO_2. The larger size of ensemble considered here allows us to quantify not only an ensemble mean, but also an ensemble range in the RFEE, giving an indication of the robustness of the results. We characterise the range of RFEE values as the interval between the 10th and 90th percentiles of the ensemble distribution, as in the previous section. We emphasise that sampling variability in D_i will contribute significantly to the range in RFEE, in addition to the effects of perturbing the physical parameterisations in HadSM3.

4.2 Results

Figure 13 shows maps of the ensemble mean and range of RFEE for seasonally averaged screen temperature in JJA (cf. the results of Fig. 4 for daily extremes). Extremely warm seasons are predicted to become very much more common under $2 \times CO_2$ conditions relative to $1 \times CO_2$ conditions, in all areas. The same is true for DJF events (not shown). The global mean relative frequency over all land points is around 60 in both JJA and DJF (see Table 5), indicating that seasonal temperatures two standard deviations above the long term average in the $1 \times CO_2$ simulations occur almost every year in the $2 \times CO_2$ simulations. These predictions are very robust across the ensemble, the ensemble range of the RFEE being generally very much smaller than the mean; in particular, the ensemble-mean of (RFEE- 1), the change in the frequency of extreme seasons, exceeds the range at almost all gridboxes (Table 5).

Figure 14 gives the ensemble mean and range of RFEE for seasonal precipitation in JJA, and also shows corresponding results for DJF. Extensive regions are found over which extremely wet seasons are predicted to become more common under doubled CO_2 conditions. Large values of RFEE occur at middle and high latitudes in the winter hemisphere, coincident with regions in which increases in time-averaged precipitation occur (note that the increases of (typically) 0.5 mm per day obtained in these regions, though smaller in absolute terms than the changes in some tropical regions (see Fig. 1), often represent large fractional increases of 30–50% or more, compared with the levels found in the $1 \times CO_2$ simulations). The detailed spatial variations of RFEE in the tropics are also closely related to changes in time-averaged precipitation (Fig. 14 cf. Fig. 1). Values of RFEE are generally below unity in the sub-tropics, again consistent with the changes in time-averaged precipitation. In general, the ensemble range of RFEE for seasonal precipitation is larger then the ensemble mean value (Table 5), implying that the simulated changes are less robust than for screen temperature, consistent with the results found for daily extremes (Sect. 3.2). In fact, the ensemble-mean of (RFEE-1), the change in the frequency of extreme seasons, only exceeds the ensemble range of RFEE over about 5% of the globe for precipitation. Nevertheless, our ensemble-mean values of RFEE for European winter precipitation are somewhat greater than those found by Palmer and Räisänen (2002), who considered the effects of a transient increase of CO_2 at the time of doubling, rather then the equilibrium changes considered here.

Extremely wet summers are predicted to become ten (or more) times more common over parts of the Asian monsoon region (Fig. 14), an area in which such events can have large impacts. Many previous studies have also predicted an increase in the intensity (e.g. Zwiers and Kharin 1998) or interannual variability (see e.g. Easterling et al. 2000) of the Asian monsoon under increased CO_2. The geographical distribution of our ensemble-mean changes is similar to the changes found by Bhaskaran and Mitchell (1998) using the HadCM2 version of the Met Office coupled ocean-atmosphere GCM. However, our values of RFEE over southern

Asia are typically about twice as large as those found by Palmer and Räisänen (2002). Extremely wet seasons are predicted to become much less common over parts of Amazonia and northern Africa, and over southern Europe in summer (Fig. 14). These same areas (and, in addition, Arabia in JJA and DJF, and India in DJF) also show an increase in the frequency of extremely dry seasons under $2 \times CO_2$ (defined using a threshold of two standard deviations below the mean).

In Fig. 15 we compare RFEE values for daily and seasonal time scales in boreal summer. Due to the sensitivity of RFEE to threshold (Fig. 11), it was necessary to recalculate the daily extremes using a 97.725th percentile threshold (which corresponds to 2 standard deviations above the mean in the case of a Gaussian distribution), to be consistent with the threshold used to calculate the seasonal extremes. The results show that the ensemble-mean RFEE for seasonal events is almost everywhere greater than for daily events (see also Table 5). This is because the standard deviations of the simulated frequency distributions in the $1 \times CO_2$ simulations (arising from natural variability) are much larger for daily fields than for seasonal fields, since the formation of seasonal averages removes much of the influence of sub-seasonal variability. Thus the time-averaged temperature change in response to doubled CO_2 typically corresponds to a shift of several standard deviations of the $1 \times CO_2$ simulation for seasonal data, whereas the shift is typically of the same order as the standard deviation for daily data (not shown). Provided the definitions of extreme event threshold are consistent, we would therefore expect to find larger values of RFEE for the seasonal data than for the daily data (Fig. 11). In fact, Fig. 15 shows that the differing values of daily and seasonal RFEE for screen temperature in boreal summer can be explained, to first order, by assuming the response to doubled CO_2 to be a uniform shift in the mean of an assumed Gaussian frequency distribution (see Fig. 11 and related discussion in Sect. 3.2).

5 Conclusions

An ensemble of 53 versions of the HadSM3 climate model was used to produce simulations of equilibrium climate changes in response to a doubling of atmospheric CO_2. The ensemble, produced by perturbing poorly constrained model parameters one at a time, samples uncertainties arising both from the parameterisation of surface and atmospheric physical processes and from natural variability (see Murphy et al. 2004). Large ensembles (see also Stainforth et al. 2005) are essential to support reliable quantification of the spread of outcomes consistent with the uncertainties explored, in turn allowing identification of features of the response robust to the uncertainties explored. We focus on the simulated changes in extreme climatic events at individual model grid points in the solstitial seasons.

Large increases in the frequency of extremely warm days over land are simulated in the $2 \times CO_2$ integrations relative to the $1 \times CO_2$ integrations. In January and July respectively, the occurrence of days on which maximum screen temperatures exceed the simulated present-day 99th percentile threshold become 20 and 28 times more frequent under $2 \times CO_2$ conditions, in the global mean. For minimum screen temperatures, the corresponding global-mean increases are even larger, amounting to 29 in January and 44 in July. The magnitudes of the increases in daily temperature extremes varies substantially with region. In July, for example, the largest increases are found over western parts of North America, the northern half of South America, and much of southern Europe, northern Africa and the Middle East. Predictions of an increase are robust across the ensemble in many regions (although the magnitude of the increase does vary substantially between different ensemble members), thus the results provide strong support for evidence of increases found by previous studies using individual simulations or small ensembles. The occurrence of extremely high daily precipitation totals is also predicted to increase over large areas of the globe under $2 \times CO_2$, though only by a factor of two in the global- and ensemble-mean. At individual model grid points, it is rarely possible to identify a robust change across the ensemble.

Changes in the shape of the daily data distributions are more important in accounting for the modelled values of the relative frequency of extreme events (RFEE) for precipitation than they are for screen temperature; in the ensemble mean, for a number of regions (accounting for around 20% of the globe), time-averaged precipitation is predicted to decrease under $2 \times CO_2$, while the relative frequency of extreme precipitation events increases (see also Christensen and Christensen 2003).These changes are manifested as a decrease in wet day frequency and an increase in the skewness of the distributions. Such climatic changes are likely to have adverse effects on vegetation and population in the affected regions, increasing the likelihood of both drought and flooding.

The RFEE is found to reduce as the threshold used to define the event reduces. For example, the global-mean ensemble-mean relative frequency for extremely warm days in July (defined with respect to the 99th percentile of the distribution of simulated present day values) is 27.5, compared to a value of 5.7 for unusual events defined according to the 90th percentile. However, the ensemble range of values is smaller for unusual events, so the fractional area of land points experiencing an increase which is robust across the ensemble is larger (85%, cf. 39% for extreme events). For daily precipitation, unusual events become 1.2 times as likely in the global- and ensemble-mean, compared to the doubling found for extreme events. In the case of July maximum daily screen temperature, we show that the variation in RFEE with event threshold can be well reproduced, to first order, by approximating the simulated response to

doubled CO_2 as a uniform shift in an assumed Gaussian distribution of daily values. Attempts to approximate the response of simulated distributions of daily precipitation by applying a uniform scaling of wet day precipitation amounts are less successful. This illustrates the importance of changes in the shape of the distributions under $2 \times CO_2$, such as the decreases in wet day frequency and increases in skewness reported above (see also Wehner 2004; Kharin and Zwiers 2005). We also note that the extent to which changes in shape influence changes in extremes depends on how events are defined. In a companion paper, Clark et al. (2006) describe changes in summer heat extremes from the same ensemble, but by considering changes in the temperature associated with given percentiles, rather than changes in the exceedance of a fixed threshold. They find correlations typically in the range 0.5–0.8 between the simulated change in temperature extreme and the change predicted on the basis of a uniform shift with no change in shape, whereas we find correlations in simulated and predicted RFEE exceeding 0.8 at most locations.

Extremely wet seasons are predicted to become more common over wide areas, though the uncertainty in the response is generally larger than the ensemble-mean change. Increases are particularly marked in some equatorial regions, in middle and high latitudes of the winter hemisphere and in the Asian monsoon region in boreal summer, extremely wet monsoon seasons occurring typically ten times more often under $2 \times CO_2$. Extremely wet seasons are predicted to become less common over parts of Amazonia and northern Africa, and over southern Europe in summer, with a corresponding increase in the frequency of extremely dry seasons. Extremely warm seasons are robustly predicted to become much more common in response to doubled CO_2, with a global mean relative frequency of around 60.

If climate changes in daily and seasonal extreme events are calculated with respect to a consistent threshold, the relative frequency of extreme seasons is found to be larger than for extreme daily events, for both screen temperature and precipitation. This occurs because the $1 \times CO_2$ frequency distributions for seasonal data are narrower than for daily data, as the seasonal averaging removes the short-term component of natural climatic variability associated with synoptic weather systems. Therefore, a given change in time-averaged climate is more significant compared to the width of the $1 \times CO_2$ distribution in the case of seasonal data, leading to a larger increase in RFEE.

6 Discussion and further work

Our analysis of uncertainties should be expanded to consider indicators of extreme events such as storminess (e.g. Jones et al. 1999), heatwaves (e.g. Meehl and Tebaldi 2004) or droughts (e.g. Gregory et al. 1997; Frich et al. 2002; Kiktev et al. 2003). It is also important to identify the physical mechanisms driving the range of simulated changes in extremes, both to increase confidence in the predictions and also to inform strategies for forming optimal predictions from an ensemble of model versions of varying quality (see below). In our ensemble the range of simulated changes is driven both by uncertainties in the physical parameterisations of the model, and by the effects of natural variability. Parameterisation uncertainties can, to some extent, be reduced by developing improved models, whereas uncertainty arising from natural variability is unavoidable. It is therefore important to quantify their relative contributions to the range of simulated changes in extremes, for instance by comparing the spread of changes in a perturbed physics ensemble to that from different periods of a long simulation with one particular model version (Murphy et al. 2004).

Our ensemble consists of versions of the HadSM3 atmosphere/mixed layer ocean model in which only one parameter is varied at a time. The ensemble is thus heavily biased towards the 'standard' version of the model, which does not necessarily occupy any special position in parameter space. Ensembles containing multiple parameter perturbations are needed to explore more fully the range of possible outcomes (Stainforth et al. 2005). We have now produced a larger perturbed physics ensemble of equilibrium doubled-CO_2 simulations in which multiple parameters are perturbed simultaneously (Webb et al. 2006). This ensemble is designed to span the space of possible parameter values and climate sensitivities as comprehensively as possible, subject to the constraint that the model versions must provide high quality simulations of present day climate. It will give an improved basis for quantifying uncertainties in the response of both time-averaged climate and extreme events to doubled CO_2. However, it is also necessary to consider 'structural' model uncertainties, such as alterations to the fundamental physical assumptions on which the parameterisations are based, or changes in resolution. To some extent, structural uncertainties can be sampled from multi-model ensembles consisting of GCMs developed at different modelling centres (e.g. Palmer and Räisänen 2002; Tebaldi et al. 2006; Weisheimer and Palmer 2005), although an approach systematically combining combinations of structural options and parameter perturbations may be necessary to provide a thorough sampling of modelling uncertainties. Ultimately the large ensembles required should be produced using comprehensive Earth system models, which represent the effects of climate change feedbacks arising from the ocean circulation (e.g. Manabe et al. 1991) and biogeochemical cycles (e.g. Cox et al. 2000), in addition to the surface and atmospheric physical processes considered in our ensemble. As a step towards this, we have produced an ensemble of simulations of the transient climate response to increasing CO_2, consisting of multiple versions of the HadCM3 coupled ocean-atmosphere GCM (Collins et al. 2006) created by coupling different versions of the atmospheric

component (distinguished by multiple parameter perturbations) to the ocean component. This work will be extended in future to include ensembles driven by realistic scenarios of historical and future anthropogenic emissions, allowing assessment of the implications of uncertainties in model parameters for the range of possible climate changes during the twenty-first century.

Methods of obtaining probabilistic projections of climate change from GCM ensembles are now being suggested, based on the fit of the simulations to observations of either historical climate changes in global surface temperature patterns (Allen et al. 2000; Allen and Ingram 2002; Stott and Kettleborough 2002), or of time averages of global fields or regional spatial averages of selected present-day climate variables (Murphy et al. 2004; Tebaldi et al. 2005). We have not attempted such an approach here, since it is not yet clear whether these criteria represent a relevant constraint on future changes in extreme events. However, the possibility of making probabilistic projections of extreme events should be considered, by weighting the simulations of ensemble members according to the fit to relevant observations. In addition to the variables mentioned above, the list of possible observational constraints should be extended to include others likely to be relevant to regional extreme events, such as regional relationships between changes in different climate variables (which may be related to the strength of climate change feedbacks, see e.g. Williams et al. 2003), or the distributional characteristics of present-day simulations of variables such as those considered in this paper.

Acknowledgements Thanks go to Gareth S. Jones for his assistance with the production of the HadSM3 ensemble, and to two anonymous reviewers for constructive comments which helped us improve the manuscript. This work was funded by the UK Department of the Environment, Food and Rural Affairs, under the Climate Prediction Programme (Contract PECD 7/12/37). This paper is British Crown Copyright.

References

Allen M, Ingram W (2002) Constraints on future changes in climate and the hydrological cycle. Nature 419:224–232

Allen M, Stott P, Mitchell J, Schnur R, Delworth T (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature 407:617–620

Bhaskaran B, Mitchell J (1998) Simulated changes in southeast Asian monsoon precipitation resulting from anthropogenic emissions. Int J Climatol 18:1455–1462

Boer GJ, Yu B (2003) Dynamical aspects of climate sensitivity. Geophys Res Lett 30:1135

Bonsal B, Zhang X, Vincent L, Hogg W (2001) Characteristics of daily and extreme temperatures over Canada. J Climate 14:1959–1976

Brabson B, Lister D, Jones P, Palutikof J (2005) Soil moisture and predicted spells of extreme temperatures in Britain. J Geophys Res 110:D05104

Carnell R, Senior C (1998) An assessment of measures of storminess: simulated changes in Northern Hemisphere winter due to increasing CO_2. Clim Dyn 14:369–383

Christensen J, Christensen O (2003) Severe summertime flooding in Europe. Nature 421:805

Clark R, Brown S, Murphy J (2006) Modelling northern hemisphere summer heat extreme changes and their uncertainties using a physics ensemble of climate sensitivity experiments. J Climate (in press)

Collins M, Booth B, Harris G, Murphy J, Sexton D, Webb M (2006) Towards quantifying uncertainty in transient climate change. Clim Dyn (in press)

Cox P, Betts R, Jones C, Spall S, Totterdell I (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408:184–187

Cubasch U, Santer B, Hellbach A, Hegerl G, Höck H, Maier-Reimer E, Mikolajewicz U, Stössel A, Voss R (1994) Monte Carlo climate change forecasts with a global coupled ocean–atmosphere model. Clim Dyn 10:1–19

Cubasch U, Meehl G, Boer G, Stouffer R, Dix M, Noda A, Senior C, Raper S, Yap K (2001) Projections of future climate change. In: Houghton J, Ding Y, Griggs D, Noguer M, van der Linden P, Dai X, Maskell K, Johnson C (eds) Climate change 2001: the scientific basis. Cambridge University Press, London, pp 525–582

Durman C, Gregory J, Hassell D, Jones R, Murphy J (2001) A comparison of extreme European daily precipitation simulated by a global and regional climate model for present and future climates. Q J R Meteorol Soc 127:1005–1015

Easterling D, Meehl G, Parmesan C, Changnon S, Karl T, Mearns L (2000) Climate extremes: observations, modeling and impacts. Science 289:2068–2074

Emori S, Brown S (2005) Dynamic and thermodynamic changes in mean and extreme precipitation under change climate. Geophys Res Lett 32:L17706

Frich P, Alexander L, Della-Marta P, Gleason B, Haylock M, Klein Tank A, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19:193–212

Furrer R, Sain S, Nychka D, Meehl G (2006) Multivariate Bayesian analysis of atmosphere-ocean general circulation models. Environ Ecol Stat (in press)

Giorgi F, Francisco R (2000) Evaluating uncertainties in the prediction of regional climate change. Geophys Res Lett 27:1295–1298

Giorgi F, Mearns L (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. J Climate 15:1141–1158

Giorgi F, Hewitson B, Christensen J, Hulme M, Von Storch H, Whetton P, Jones R, Mearns L, Fu C (2001) Regional climate information - evaluation and projections. In: Houghton J, Ding Y, Griggs D, Noguer M, van der Linden P, Dai X, Maskell K, Johnson C (eds) Climate change 2001: the scientific basis. Cambridge University Press, London, pp 583–638

Gordon C, Cooper C, Senior C, Banks H, Gregory J, Johns T, Mitchell J, Wood R (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 16:147–168

Gregory J, Mitchell J (1995) Simulation of daily variability of surface temperature and precipitation over Europe in the current and $2\times CO_2$ climates using the UKMO climate model. Q J R Meteorol Soc 121:1451–1476

Gregory J, Mitchell J, Brady A (1997) Summer drought in northern mid-latitudes in a time-dependent CO_2 climate experiment. J Climate 10:662–686

Hennessy K, Gregory J, Mitchell J (1997) Changes in daily precipitation under enhanced greenhouse conditions. Clim Dyn 13:667–680

Hewitt C, Mitchell J (1997) Radiative forcing and response of a GCM to ice age boundary conditions: cloud feedback and climate sensitivity. Clim Dyn 13:821–834

Hewitt C, Senior C, Mitchell J (2001) The impact of dynamic sea-ice on the climatology and climate sensitivity of a GCM: a study of past, present and future climates. Clim Dyn 17:655–668

Horton E, Folland C, Parker D (2001) The changing incidence of extremes in worldwide and Central England temperatures to the end of the twentieth century. Clim Change 50:267–295

Huntingford C, Jones R, Prudhomme C, Lamb R, Gash J, Jones D (2003) Regional climate-model predictions of extreme rainfall for a changing climate. Q J R Meteorol Soc 129:1607–1621
Jones R (2000) Managing uncertainty in climate change projections—issues for impact assessment. Clim Change 45:403–419
Jones P, Reid P (2001) Assessing future changes in extreme precipitation over Britain using regional climate model integrations. Int J Climatol 21:1337–1356
Jones P, Horton E, Folland C, Hulme M, Parker D, Basnett T (1999) The use of indices to identify changes in climate extremes. Clim Change 42:131–149
Karl T, Knight R (1997) The 1995 Chicago heat wave: how likely is a recurrence? Bull Amer Meteor Soc 78:1107–1119
Karl T, Nicholls N, Gregory J (1997) The coming climate. Sci Am 276:54–59
Kharin V, Zwiers F (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. J Climate 13:3760–3788
Kharin V, Zwiers F (2005) Estimating extremes in transient climate change simulations. J Climate 18:1156–1173
Kiktev D, Sexton D, Alexander L, Folland C (2003) Comparison of modelled and observed trends in indices of daily climate extremes. J Climate 16:3560–3571
Lambert S, Boer G (2001) CMIP1 evaluation and intercomparison of coupled climate models. Clim Dyn 17:83–106
Manabe S, Stouffer R, Spelman M, Bryan K (1991) Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2. Part I: annual mean response. J Climate 4:785–818
Marsh TJ (2001) The 2000/01 floods in the UK- a brief overview. Weather 56:343–345
Meehl G, Tebaldi C (2004) More intense, more frequent, and longer lasting heat waves in the 21st century. Science 305:994–997
Meehl G, Karl T, Easterling D, Changnon S, Pielke R, Changnon D, Evans J, Groisman P, Knutson T, Kunkel K, Mearns L, Parmesan C, Pulwarty R, Root T, Sylves R, Whetton P, Zwiers F (2000a) An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts and model projections. Bull Am Meteor Soc 81:413–416
Meehl G, Zwiers F, Evans J, Knutson T, Mearns L, Whetton P (2000b) Trends in extreme weather and climate events: issues related to modeling extremes in projections of future climate change. Bull Am Meteor Soc 81:427–436
Meehl G, Tebaldi C, Nychka D (2004) Changes in frost days in simulations of 21st century climate. cd 23:495–511
Meehl G, Arblaster J, Tebaldi C (2005) Understanding future patterns of increased precipitation intensity in climate model simulations. Geophys Res Lett 32:L18719
Mitchell J, Wilson C, Cunnington W (1987) On CO_2 climate sensitivity and model dependence of results. Q J R Meteorol Soc 113:293–322
Murphy J, Sexton D, Barnett D, Jones G, Webb M, Collins M, Stainforth D (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772
New M, Hulme M, Jones PD (1999) Representing twentieth-century space-time climate variability. part I: development of a 1961-90 mean monthly terrestrial climatology. J Climate 12:829–856
Palmer T (2001) A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. Q J R Meteorol Soc 127:279–303
Palmer T, Räisänen J (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. Nature 415:512–514
Parmesan C, Root T, Willig M (2000) Impacts of extreme weather and climate on terrestrial biota. Bull Am Meteor Soc 81:443–450
Pittock A, Jones R, Mitchell C (2001) Probabilities will help us plan for climate change. Nature 413:249
Pope V, Galliani M, Rowntree P, Stratton R (2000) The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. Clim Dyn 16:123–146
Reilly J, Stone P, Forest C, Webster M, Jacoby H, Prinn R (2001) Uncertainty and climate change assessments. Science 293:430–433
Rowell D, Jones R (2006) The causes and uncertainty in future summer drying over Europe. Clim Dyn (Submitted)
Schär C, Vidale P, Lüthi D, Frei C, Häberli C, Liniger M, Appenzeller C (2004) The role of increasing temperature variability in European summer heatwaves. Nature 427:332–336
Senior C, Jones R, Lowe J, Durman C, Hudson D (2002) Predictions of extreme precipitation and sea-level rise under climate change. Phil Trans R Soc Lond A 360:1301–1311
Simmons A, Burridge D (1981) An energy and angular momentum conserving finite difference scheme with hybrid coordinates. Mon Weather Rev 109:758–766
Stainforth D, Aina T, Christensen C, Collins M, Faull N, Frame D, Kettleborough J, Knight S, Martin A, Murphy J, Piani C, Sexton D, Smith L, Spicer R, Thorpe A, Allen M (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403–406
Stocker T, Clarke G, Le Treut H, Lindsen R, Meleshko V, Mugara R, Palmer T, Pierrehumbert R, Sellers P, Trenberth K, Willebrand J (2001) Physical climate processes and feedbacks. In: Houghton J, Ding Y, Griggs D, Noguer M, van der Linden P, Dai X, Maskell K, Johnson C (eds) Climate change 2001: the scientific basis. Cambridge University Press, London, pp 417–470
Stott P, Kettleborough J (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. Nature 416:723–726
Stott P, Stone D, Allen M (2004) Human contribution to the European heat wave of 2003. Nature 432:610–614
Tebaldi C, Arblaster J, Hayhoe K, Meehl G (2006) Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. Clim Change (in press)
Tebaldi C, Smith R, Nychka D, Mearns L (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Climate 18:1524–1540
Visser H, Folkert R, Hoekstra J, de Wolff J (2000) Identifying key sources of uncertainty in climate change projections. Clim Change 45:421–457
Webb M, Senior C, Sexton D, Ingram W, Williams K, Ringer M, McAvaney B, Colman R, Soden B, Gudgel R, Knutson T, Emori S, Ogura T, Tsushima Y, Andronova N, Li B, Musat I, Bony S, Taylor K (2006) On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Clim Dyn (in press)
Wehner M (2004) Predicted twenty-first-century changes in seasonal extreme precipitation events in the Parallel Climate Model. J Climate 17:4281–4290
Weisheimer A, Palmer T (2005) Changing frequency of occurrence of extreme seasonal temperatures under global warming. Geophys Res Lett 32:L20721
Wetherald R, Manabe S (1999) Detectability of summer dryness caused by greenhouse warming. Clim Change 43:495–511
Williams K, Ringer M, Senior C (2003) Evaluating the cloud response to climate change and current climate variability. Clim Dyn 20:705–721
Xie P, Arkin PA (1998) Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. J Climate 11:137–164
Zwiers F (2002) The 20-year forecast. Nature 416:690–691
Zwiers F, Kharin V (1998) Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. J Climate 11:2200–2222