Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations

Seung-Ki Min · Xuebin Zhang · Francis W. Zwiers · Petra Friederichs · Andreas Hense

Abstract This study assesses the detectability of external influences in changes of precipitation extremes in the twentieth century, which is explored through a perfect model analysis with an ensemble of coupled global climate model (GCM) simulations. Three indices of precipitation extremes are defined from the generalized extreme value (GEV) distribution: the 20-year return value (P_{20}), the median (P_m), and the cumulative probability density as a probability-based index (*PI*). Time variations of area-averages of these three extreme indices are analyzed over different spatial domains from the globe to continental regions. Treating all forcing simulations (ALL; natural plus anthropogenic) of the twentieth century as observations and using a preindustrial control run (CTL) to estimate the internal variability, the amplitudes of response patterns to anthropogenic (ANT), natural (NAT), greenhouse-gases (GHG), and sulfate aerosols (SUL) forcings are estimated using a Bayesian decision method. Results show that there are *decisively* detectable ANT signals in global, hemispheric, and zonal band areas. When only land is considered, the global and hemispheric detection results are unchanged, but detectable ANT signals in the zonal bands are limited to low latitudes. The ANT signals are also detectable in the P_m and *PI* but not in P_{20} at continental scales over Asia, South America, Africa, and Australia. This indicates that indices located near the center of the GEV distribution (P_m and *PI*) may give better signal-to-noise ratio than indices representing the tail of the distribution (P_{20}). GHG and NAT signals are also detectable, but less robustly for more limited extreme indices and regions. These results are largely insensitive when model data are masked to mimic the availability of the observed data. An imperfect model analysis in which fingerprints are obtained from simulations with a different GCM suggests that ANT is robustly detectable only at global and hemispheric scales, with high uncertainty in the zonal and continental results.

1 Introduction

Detection of anthropogenic influences in the observed climate extremes is very important. This is because extreme events have potentially devastating effects on human society and the economy, and because the detection of human influence in observations will enhance our confidence in projected changes in extremes. One of the major obstacles to the detection of external influence in the extremes, especially precipitation extremes, is the limited availability of daily observations.

Anthropogenic influence has recently been detected in some temperature extreme indices. Using a gridded data set (Caesar et al. 2006) that includes a large part of the Northern Hemisphere land area and Australia, Christidis et al. (2005) and Shiogama et al. (2006) have detected an anthropogenic influence on extreme indices for warm nights, cold nights, and cold days during the second half of the twentieth century. These results from single model analyses are consistent with earlier detectability analysis results for temperature extremes (Hegerl et al. 2004). Meehl et al. (2007b) compared the observed and multi-model simulated changes in the temperature extremes

S.-K. Min (✉) · X. Zhang · F. W. Zwiers
Climate Research Division, Environment Canada,
4905 Dufferin Street, Downsview, ON M3H 5T4, Canada
e-mail: seung-ki.min@ec.gc.ca

P. Friederichs · A. Hense
Meteorological Institute, University of Bonn, Bonn, Germany

averaged over the continental United States. They showed that the observed trends (decreases in frost days; increases in growing season length, warm nights, and heat wave intensity) over 1975–1999 are accounted for by anthropogenic forcing, but not by natural forcing.

Using an atmospheric general circulation model forced by observed sea surface temperatures (SST), Kiktev et al. (2003) found that the inclusion of anthropogenic forcing significantly improved the model performance in simulating observed trends in temperature extremes for 1950–1995. Kiktev et al. (2007) provided an updated analysis using five coupled climate models that included anthropogenic forcing. They confirmed moderate skill of the models in simulating trends of temperature extremes during the second half of the twentieth century. Hegerl et al. (2004) carried out a model-to-model detection study in which the fingerprint from one model is compared to *observations* from the same model (perfect) or from another model (imperfect). They found that anthropogenic influence in temperature extremes is robustly detectable with a signal-to-noise ratio comparable to that in mean temperature changes.

Precipitation extremes are expected to increase globally as the climate warms constrained by moisture availability or the Clausius-Clapeyron relationship, and the increases are likely to be larger than those in mean precipitation (Allen and Ingram 2002; Trenberth et al. 2003; Held and Soden 2006; Pall et al. 2007; Kharin et al. 2007). Most coupled climate model simulations project an increase of extreme precipitation over large parts of the globe under the greenhouse warming (Kharin and Zwiers 2000; Meehl et al. 2000; Cubasch et al. 2001 and references therein; Semenov and Bengtsson 2002; Allen and Ingram 2002; Watterson and Dix 2003; Hegerl et al. 2004; Wehner 2004; Kharin and Zwiers 2005; Emori and Brown 2005; Tebaldi et al. 2006; Kharin et al. 2007). Observed changes in precipitation extremes are qualitatively consistent with model projections, (e.g., Groisman et al. 2005; Alexander et al. 2006, Hegerl et al. 2007), but detecting anthropogenic influence in precipitation extremes has not yet been achieved for a number of reasons.

Daily precipitation observations are very limited both spatially and temporally. Observed extreme precipitation at sparsely located stations represents point estimates. This hinders a direct comparison with model simulated precipitation considered to be area estimates (Osborn and Hulme 1997). Furthermore, disagreements in simulated extreme precipitation between GCMs are large, especially in the tropics where uncertainty in the parameterization of convection affects the simulated precipitation (Hegerl et al. 2004; Kharin et al. 2005, 2007).

Consequently, there has been little success in detecting anthropogenic signals in precipitation extremes. Kiktev et al. (2003, 2007) found that anthropogenic forcing contributed little to the simulation of trends in precipitation extremes, unlike in temperature extremes. Hegerl et al. (2004) compared the observed and model simulated trends of precipitation extremes represented by annual maximum daily or 5-day precipitation amount over land and found that changes in heavy precipitation might be more detectable than changes in annual mean precipitation. Both studies seem to provide some evidence that detection of an anthropogenic signal in precipitation extremes in the instrumental period may not yet be possible, but a more comprehensive analysis is needed to draw such a conclusion.

In this paper, we present a further study of the detectability of the precipitation extremes response to external forcing. We follow the perfect model approach of Hegerl et al. (2004), but undertake a more comprehensive analysis. We consider decadal-scale changes of extremes, rather than the long-term trends used in previous studies, and we take the availability of observational data into account in our analysis. We also consider various combinations of external forcing including greenhouse-gases, sulfate aerosols, natural, and anthropogenic (greenhouse-gases and sulfate aerosols combined). Our analyses are conducted over various spatial domains ranging from the globe including the oceans to individual continents. In addition, we also consider a probability-based extremes index that gives equal weight at all locations after normalizing precipitation variability.

The remainder of this paper is organized as follows. Section 2 describes the model simulations. Indices for precipitation extremes, a Bayesian method for signal analysis, and calculation details are explained in Sect. 3. Spatial patterns of simulated precipitation extremes under different forcings are qualitatively compared in Sect. 4. The results of our detectability analyses are described in Sect. 5. Robustness of detection results to the availability of observational data and to the use of fingerprints obtained from another GCM is examined in Sect. 6. Conclusions are presented in Sect. 7.

2 Model simulations

We use an ensemble of climate simulations performed with the ECHO-G coupled climate model (Legutke and Voss 1999; Min et al. 2005a, b, 2006). The atmospheric component, ECHAM4, has T30 ($\sim 3.75°$) horizontal resolution with 19 pressure levels in the vertical. Its oceanic component, HOPE-G, has horizontal resolution equivalent to approximately T42 ($\sim 2.8°$) with meridional refinement toward the equator up to 0.5°. HOPE-G has 20 vertical layers. ECHO-G applies an adjustment to the annual mean fluxes of heat and fresh water, but momentum is not

adjusted. The use of heat and moisture flux adjustments may have some implications for the model's responses to external forcing, although flux adjusted and non-adjusted models appear to respond similarly at large scales (Cubasch et al. 2001). CO_2, CH_4, N_2O, and 16 minor industrial gases are treated as greenhouse-gases (Roeckner et al. 1999). The direct and first-indirect effects of sulfate aerosols are considered using an interactive sulfur cycle model (Feichter et al. 1997). The large-scale precipitation is parameterized based on relative humidity following Sundqvist (1978) and Sundqvist et al. (1989). The convective precipitation is parameterized using the mass flux scheme of Tiedtke (1989) with an adjustment closure by Nordeng (1994).

This study uses the annual maximum daily precipitation from several ensembles of forced simulations performed with the ECHO-G model under different external forcing factors (Table 1). Natural (NAT), greenhouse-gas (GHG), sulfate aerosol (SUL), anthropogenic (ANT, GHG and SUL combined), and all forcing (ALL, ANT and NAT combined) simulations were produced for the period 1860–2000. The simulations use three different initial conditions selected at 100-year intervals from a long preindustrial control simulation. GHG concentrations are provided by the ENSEMBLES project (Jean-Francois Royer, personal communication 2006). Sulfate aerosols emissions and tropospheric ozone concentrations are obtained from Roeckner et al (1999). The solar and volcanic forcing is introduced by varying the solar constant following Crowley (2000). Min et al. (2006) provide more details on the external forcing. A 341-year long preindustrial control simulation (CTL) provides data for estimation of the internal variability.

We also use a three-member ensemble simulation performed with the Third Generation of the Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled Global Climate Model (CGCM3, Flato 2005) for the imperfect model analysis. The CGCM3 has an atmospheric horizontal resolution of T47 ($\sim 3.75°$) with 31 vertical levels and oceanic resolution of about $1.85°$ with 29 levels in the vertical. The simulations were forced with ANT-only forcing for 1860–2000. We shall refer those simulations as ANT*. A 500-year preindustrial control simulation (CTL*) was also used to define a reference set necessary to estimate a probability-based index (PI) (Table 1; see below).

3 Methodology

3.1 Indices for extreme precipitations

We assume that annual maximum daily precipitation follows the generalized extreme value (GEV) distribution that incorporates Gumbel, Frechet, and Weibull distributions. The GEV has a cumulative distribution function (CDF) given by

$$
\begin{aligned}
&F(x; \mu_t, \sigma_t, \xi_t) \\
&= \begin{cases}
\exp\left[-\exp\left\{-\frac{x-\mu_t}{\sigma_t}\right\}\right], & \xi_t = 0 \\
\exp\left[-\left\{1 + \xi_t \frac{x-\mu_t}{\sigma_t}\right\}^{-\xi_t^{-1}}\right], & \xi_t \neq 0, \quad 1 + \xi_t \frac{x-\mu_t}{\sigma_t} > 0.
\end{cases}
\end{aligned}
\tag{1}
$$

Here μ, σ, and ξ are the location, scale, and shape parameters, respectively. By inverting the CDF for a given probability p, quantiles of the GEV distribution can be obtained as

$$
X_p(t) = \begin{cases}
\mu_t - \sigma_t \ln[-\ln(p)], & \xi_t = 0, \\
\mu_t - \frac{\sigma_t}{\xi_t} \ln\left[1 - (-\ln(p))^{-\xi_t}\right], & \xi_t \neq 0.
\end{cases}
\tag{2}
$$

We assume that the GEV parameters are time-dependent as denoted by subscript t, so that the GEV distribution and its quantiles can vary with time (see below).

Table 1 Coupled climate model simulations used in this study

Model	Abbreviation	Forcing	Period (ensemble member)	Role in perfect model analysis
ECHO-G	CTL	Preindustrial control run with no external forcing	1860–2200 (1)	Scenario 1 (internal variability)
	ALL	Natural plus anthropogenic forcing	1860–2000 (3)	*observations*
	NAT	Natural forcing (solar and volcanic)	1860–2000 (3)	Scenario 2
	ANT	Anthropogenic forcing (greenhouse-gases and sulfate aerosols)	1860–2000 (3)	Scenario 3
	GHG	Greenhouse-gas forcing	1860–2000 (3)	Scenario 4
	SUL	Sulfate aerosol forcing	1860–2000 (3)	Scenario 5
CGCM3.1 (T47)	CTL*	Preindustrial control run with no external forcing	1850–2350 (1)	Used to define a reference set for PI (for imperfect model analysis)
	ANT*	Anthropogenic forcing (greenhouse-gases and sulfate aerosols)	1860–2000 (3)	Scenario 3 (for imperfect model analysis)

Using the GEV distribution, we select two distribution-based indices for analyses. One is the median P_m located near the center of the GEV distribution and the other is the 20-year return value P_{20} positioned in the tail. P_m is the quantile that corresponds to $p = 0.5$ while P_{20} is the quantile for $p = 0.95$. From Eq. (2), P_m and P_{20} are defined as

$$P_m(t) = X_{0.5}(t),$$
$$P_{20}(t) = X_{0.95}(t). \qquad (3)$$

Zhang et al. (2004) and Kharin and Zwiers (2005) considered trends in location and scale parameters for long-term changes because they found that treating shape parameter as a constant was useful. Here, dealing with 140 years (1861–2000), we allow for the decadal fluctuations in all three parameters rather than linear trends. To reduce the dimensionality of the time series to be analyzed, the GEV parameters are estimated for non-overlapping 20-year periods of the model simulations separately, assuming a fixed distribution within each 20-year period. This is appropriate since the non-stationary component should be sufficiently small when compared to internal climate variability within such short time periods, given that the twentieth century forcing is also small relative to that used in the twenty-first century scenarios (Kharin and Zwiers 2005, Kharin et al. 2007). We combine the three ensemble members for each forced experiment to produce samples of size 60 for each 20-year period, thereby reducing the uncertainty in parameter estimates. Additionally, 100 samples of 60 annual maxima are constructed from the CTL simulation by repeatedly choosing three 20-year periods at random from the 17 available non-overlapping 20-year chunks (340 years). The GEV distribution is fitted to each sample of 60 annual extremes. Possible underestimation of internal variability resulting from the use of a relatively short control simulation is taken into account by manually inflating covariance matrices (see below).

Variation of regional average extreme precipitation tends to be dominated by subareas of higher extreme values because of the long-tailed nature of extreme precipitation. One way to improve the representativeness of areas with smaller extreme values would be to introduce a normalization of extreme values at the grid-point scale before calculating regional averages. Here we utilize the CDF in Eq. (1) for normalizing extreme precipitation, which ranges from 0 to 1. Then a normalized index PI for each 20-year period is defined as:

$$PI(t) = [F(P_a; \mu_r, \sigma_r, \xi_r)], \qquad (4)$$

where [ ] denotes a 20-year time mean, P_a is the annual extreme of precipitation in year a, and the subscript r represents a reference data set. Because PI is based on the probability integral transform, it also has the advantage of

having similar amplitudes across different GCMs even if the GCMs have different extreme precipitation climatologies. In contrast to P_{20} and P_m, we do not vary the GEV parameters for PI with time, but rather estimate the parameters from a reference data set. Otherwise, long-term changes would be difficult to identify in PI due to the normalization between 20-year periods. In the perfect model analysis we utilize 956 samples as a reference data set, where samples consist of annual maxima collected from all forced runs for the period 1860–1920 (ALL, NAT, ANT, GHG and SUL, 615 annual extremes in total) as well as CTL (341 years). For the imperfect model analyses, 683 reference samples are collected from the three member ANT* ensemble for 1860–1920 (183 years) and a 500-year CTL* simulation (Table 1). The main results reported below are insensitive to the use of the control run only or forced runs only for the estimation of parameters. The larger sample that is obtained by combining both types of runs allows us to avoid possible biases and discontinuities at the boundaries of the reference period (Zhang et al. 2005).

The method of maximum likelihood (ML) is employed for fitting the GEV distribution to the samples from the model simulations. Following Kharin and Zwiers (2005), a simplex function minimization procedure is applied after taking L-moment estimates as the initial values for the maximization. We did not encounter difficulties in fitting the GEV distribution to annual maximum daily precipitation at any grid point.

3.2 Bayesian decision method

We use a Bayesian decision method (Min et al. 2004) to detect external influence. Given the observational data vector **d** (here area-averaged extreme indices P_{20}, P_m, and PI) and the possible forcing scenarios m_i ($i = 1, …, 5$) CTL, NAT, ANT, GHG, and SUL, the Bayesian process classifies the observed changes into the most likely scenario defined as the one with the maximum posterior $P(m_i|\mathbf{d})$ likelihood. If all scenarios are considered to be equally likely a priori, which we assume here for simplicity, the Bayesian decision depends only on the Bayes factors defined as the likelihood ratios:

$$B_{i1} = \frac{l(\mathbf{d}|m_i)}{l(\mathbf{d}|m_1)}, \ i = 2, ..., 5, \qquad (5)$$

where m_1 is a reference scenario which we take to be CTL. The Bayes factor B_{i1} represents the observational evidence in favor of the scenario m_i against m_1. The evidence is said to be *substantial*, *strong*, or *decisive* when the logarithm of the Bayes factor is larger than 1, 2.5, or 5 respectively, that is to say, when the assessed scenario m_i is 3, 12, or 150 times more

probable than m_1 (Kass and Raftery 1995). Several recent Bayesian detection analyses have used this approach for the assessment of evidence of anthropogenic influence on climate (e.g., Min et al. 2004; Schnur and Hasselmann 2005; Lee et al. 2005; Min and Hense 2006, 2007).

Assuming multivariate Gaussian distributions (see below for discussion of the validity of this assumption), the likelihood function has a simple form:

$$l(\mathbf{d}|m_i) = \frac{1}{\sqrt{(2\pi)^q}} \sqrt{\frac{\det \mathbf{A}_i^{-1}}{\det \mathbf{\Sigma}_i \det \mathbf{\Sigma}_0}} \exp\left(-\frac{1}{2} \Lambda_i\right) \quad (6)$$

where q is the dimension of the data vector, $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_i$ are the covariance matrices of the observation and scenario respectively, $\mathbf{A}_i = \mathbf{\Sigma}_i^{-1} + \mathbf{\Sigma}_0^{-1}$, and $\Lambda_i = (\mathbf{d} - \boldsymbol{\mu}_i)^{\mathrm{T}} (\mathbf{\Sigma}_i + \mathbf{\Sigma}^0)^{-1} (\mathbf{d} - \boldsymbol{\mu}_i)$ where $\boldsymbol{\mu}_i$ is mean of the scenario m_i (see Min et al. 2004 for more details). The covariance matrices can be spatial, temporal, or spatio-temporal depending on the analyzed variable. In this study, they are temporal covariance matrices obtained from area averaged time series. Note that the likelihood is an exponential function of a generalized distance measure Λ_i. This means that the Bayesian decision is equivalent to measuring a distance between observational and scenario mean vectors ($\mathbf{d}$ and $\boldsymbol{\mu}_i$) taking the relevant covariance structures into account and then searching for the scenario that is closest to the observations. The scenario mean is obtained from forced simulations while the covariance matrices are estimated using CTL data (see below for detailed methods).

Even though we are dealing with extreme precipitation indices, the Gaussian assumption can be applied with little concern. This is because the variables analyzed are the spatial averages of the extreme indices over large regions, i.e., the mean of a large number of samples. As discussed by Hegerl et al. (2004, 2006), and supported by the central limit theorem, the distribution of those mean values should be very close to Gaussian. We also conducted the Shapiro-Wilk normality test on CTL samples, and found that the null hypothesis of normality can not be rejected at the 5% significance level for the large scale area-averaged extreme precipitation indices used in this paper.

3.3 Detailed method for a perfect and an imperfect model analysis

For a perfect model analysis, we take time series of extreme indices from the ALL experiment as the observational vector $\mathbf{d}$ and evaluate Bayes factors for the other forced scenarios B_{i1} ($i = 2, 3, 4, 5$) using Eqs. (5) and (6). This corresponds to calculating signal amplitudes for NAT, ANT, GHG, and SUL with respect to CTL (Table 1). The detection variables are anomaly time series of P_{20}, P_{m}, or PI for 1861–2000

(seven 20-year intervals) relative to the 1861–1920 mean (the first three intervals) of the forced experiments. Taking different reference periods does not affect the main results given below because we assess detectability by measuring a generalized distance between two anomaly time series vectors of observation and scenarios (see above) and this distance is not much affected by the selected reference period. In other words, the main signals from the external forcing factors are generally associated with long-term components; selecting different reference periods only affects the time mean but does not alter the temporal fluctuations that are of interest. For CTL, a sample of 140-year (seven interval) time series of extreme indices are obtained as follows. This sample is used to estimate the CTL covariance matrix $\mathbf{\Sigma}_1$ in Eq. (6). First, we manually construct one time series consisting of 100 GEV parameters which have been estimated above using 100 samples of 60 (three 20 years) annual maxima. This corresponds to a 2,000-year time series. In order to obtain 140-year time series samples, we apply moving windows with a shift of 40 years. This produces 47 CTL samples for which anomalies are subsequently constructed as for other data vectors.

In order to consider the possibility of underestimation of the internal variability due to the use of a short control run or structural error, we test the sensitivity of the Bayes factors by inflating the covariance matrix by a factor α for different α's. We further assume that the covariance matrix of the *observations* $\mathbf{\Sigma}_0$ (ALL here) and those of the other forced scenarios $\mathbf{\Sigma}_i$ ($i = 2, 3, 4, 5$ for NAT, ANT, GHG, and SUL) are identical to that of CTL ($\alpha\mathbf{\Sigma}_1$). That is, we assume that external forcing has not substantially affected the internal variability of precipitation extremes over the twentieth century. Applying this assumption under stronger external forcings can lead to overestimated detectability due to underestimation of noise (Min et al. 2004). In this special case, one can easily see from Eqs. (5) and (6) that increasing the internal variability results in decreased Bayes factors $(B_{i1})^{1/\alpha}$ when there is an evidence for detection (i.e., when the Bayes factor is greater than one). This is equivalent to reducing the logarithm of the Bayes factor (i.e. signal amplitude) by a factor of α or to enlarging the decision criterion by the same amount. Consequently, doubling the internal variability effectively increases the thresholds for declaring Bayes factors as indicating *strong* or *decisive* evidence to 5 and 10 respectively (see above).

The Bayes factors are calculated using anomaly time series for the whole twentieth century changes (five 20-year intervals). Including 1861–1900 does not change the main results. This Bayesian analysis is repeated over different spatial scales ranging from the global mean to hemispheric, zonal, and continental regional means.

An imperfect model analysis is carried out by replacing the fingerprint in the perfect model analysis with that from

the CGCM3 ensembles (ANT*, see Table 1). Here we restrict our analysis to *PI* because it is a standardized index that enables a more reasonable intercomparison between models and regions.

4 Simulated patterns of extreme precipitation

4.1 Control experiment

Figure 1 shows the spatial distribution of averaged values of the P_m, P_{20}, the scale and the location parameters of GEV distribution, computed from the CTL run. The spatial patterns of P_m and location parameter (Fig. 1a, b) strongly resemble the spatial distribution of annual mean precipitation (not shown) characterized by stronger precipitation over the tropical western Pacific, Indian, and equatorial Atlantic Oceans, and a well-organized intertropical convergence zone (ITCZ) and South Pacific convergence zone (SPCZ). P_m less than 10 mm is found over the eastern subtropical South Pacific and South Atlantic, Sahara Desert, Arctic, and Antarctica.

P_{20} exhibits a spatial pattern similar to that of P_m, but stronger values (>60 mm) are broadly evident over the tropical Pacific and Indian Oceans (Fig. 1d)—note that P_{20} is always larger than P_m by definition. This can be explained in part by larger scale parameters over these regions (Fig. 1c) because P_{20} is more affected by the scale

parameter than P_m. For example, in the case of a Gumbel distribution, the contribution of the scale parameter to P_{20} is about eight times larger than that to P_m (second term of the right hand side of Eq. (2)). A maximum of the scale parameter is visible over the central equatorial Pacific which seems to be related to the strong and frequent El Niño and Southern Oscillation (ENSO) simulated by the model (Min et al. 2005b).

4.2 Forced experiments

Simulated changes in extreme precipitation are compared among the different external forcing factors. Figure 2 displays differences in P_{20}, P_m, and *PI* between the recent 20-year (1981–2000) period and the reference period (1861–1920). The ALL pattern is characterized by an overall increase of daily precipitation extremes over the western Pacific and Indian Ocean and a weak decrease over the eastern subtropical South Pacific and South Atlantic. When compared with model climate patterns (Fig. 1), areas of increasing extreme precipitation coincide well with those of larger mean amounts, and vice versa, although there are some exceptions e.g., over high latitudes and the Sahara Desert.

The spatial pattern of the difference in the location parameter resembles that of P_m, but changes in the scale parameter are not well-structured (not shown). This means,

Fig. 1 Climate patterns of P_m, location and scale parameters, and P_{20} obtained from ECHO-G CTL experiments
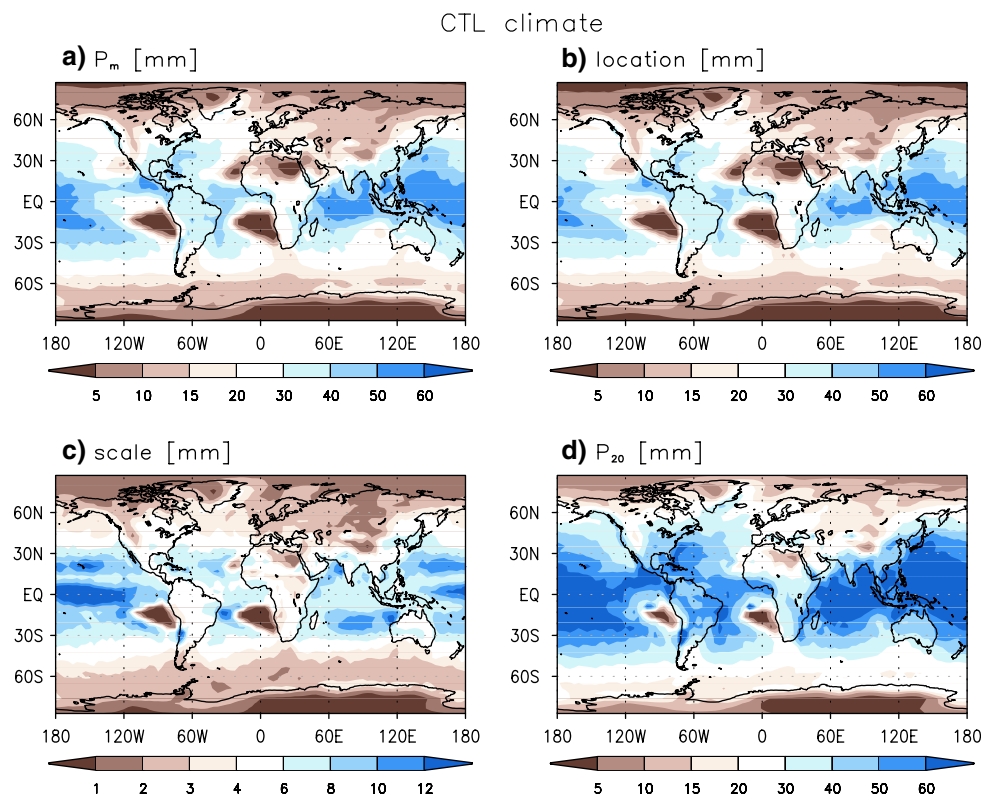
in general, that most of the response to external forcing is explained by a shift of the GEV distribution. However, scale parameter change cannot be neglected regionally. For example, in the ALL simulations P_{20} has a positive pattern over the central equatorial Pacific in contrast with the negative anomalies in P_m and PI over this region. This seems to be related to an increase in the scale parameter arising from an intensified ENSO-like GHG warming pattern as well as enhanced interannual variation through two major volcanic eruptions during 1981–2000 (see below; cf. Gillett et al. 2004). Another difference in the pattern of precipitation extremes is that PI is strengthened over higher latitudes and land areas. This is because PI represents relative change at each grid point based on the probability by which one can measure the changes in the risk of extremes. The effect of this standardization on the detectability will be explored below.

GHG response patterns are characterized by more dominant areas of increasing extremes with larger amplitudes than for ALL forcing. Other characteristics of change are similar. The broad increase in precipitation extremes is in accord with many previous studies of coupled model simulations cited above which indicate a spatially consistent increase of precipitation extremes across different regions under CO_2 warming. This is in contrast with changes in the annual mean precipitation that are characterized by a mixed pattern of increase and decrease over different regions (e.g., Cubasch et al. 2001; Douville et al. 2006; Meehl et al. 2007a; Zhang et al. 2007). The SUL response pattern resembles that of GHG except with opposite sign, indicating that sulfate aerosol in this model offsets the effect of GHG forcing in extreme precipitation (cf. Shiogama et al. 2006).

Considering that the GHG response pattern for P_{20} and P_m is most pronounced over the tropical Pacific and the Indian Ocean, one might think of a possible impact of the ENSO-like mean state change in the ECHO-G model (cf. Meehl et al. 2007a). Figure 3 shows patterns of change in total precipitation and SST over the tropics in the different forced simulations. Note that each pattern is expressed as a deviation from the corresponding zonal mean change in order to identify the ENSO-like pattern more clearly. Note also that the corresponding zonal mean change pattern is plotted at the right side of each panel. An El Niño-like pattern is seen in the SST changes that occur under sulfate aerosol forcing while GHG forcing produces a more La Niña-like response pattern in this model. This is different from multi-model patterns that show a more El Niño-like response to GHG forcing (Yamaguchi and Noda 2006). Corresponding precipitation changes in the GHG simulations are characterized by increases over the western Pacific and Indian Ocean and a reduction over the central Pacific. The SUL experiment exhibits a pattern of precipitation

change that is opposite to the GHG result. The similarity of patterns between Figs. 2 and 3 indicates the important role of this model's ENSO-like SST change in determining changes of mean and extreme precipitation in the low latitudes. However, it should be noted that the ENSO-like base state changes are still uncertain due to large inter-model differences (Collins and the CMIP modelling group 2005; van Oldenborgh et al. 2005; Paeth et al. 2008).

The ANT pattern of change in extreme precipitation in Fig. 2 has almost the same structure as the corresponding GHG response pattern, but the amplitude is smaller as might be expected from the offsetting effect of sulfate aerosol forcing. The ANT response is similar to the ALL response in pattern and amplitude. In the NAT experiment, extreme precipitation decreases over central equatorial Pacific except for P_{20}, which is different from the decreasing pattern for SUL. Interestingly, the NAT SST and precipitation response patterns are closer to the GHG response pattern than that for SUL over the equatorial Pacific (Fig. 3). This appears to be a specific feature of the ECHO-G response to volcanic forcing which was implemented by varying solar constant rather than volcanic aerosols. During 1981–2000 there were two pronounced volcanic events (El Chichón in 1982 and Pinatubo in 1991) resulting in a reduction of solar constant that might give rise to a cooling. This cooling is enhanced over the tropics, particularly the cloud-free eastern equatorial Pacific (cf. Cubasch et al. 1997). The different NAT response pattern in P_{20} over this region is explained by an increase in the scale parameter (not shown) arising from the two pronounced volcanic eruptions. An in-depth analysis would be required to isolate the localized volcanic effect, but this is beyond the scope of this paper.

5 Signal detectability at different spatial scales

5.1 Global and hemispheric scales

Time series of area averaged extreme indices over seven hemispheric domains are shown in Fig. 4. Grey bands represent internal variability as obtained from CTL. The indices for the ALL simulation (black lines) are characterized by an early increase from 1910 to 1950 and a recent increase since 1970. The GHG simulations have a monotonic increasing trend in the indices while the SUL simulations have a decreasing trend. In the NAT simulations, there is a maximum near 1950, but the variations in the other periods are within the range of internal variability. ANT results capture the ALL response pattern especially in the latter half of the twentieth century.

As a simple test of the extent to which the Clausius-Clapeyron relationship holds in the twentieth century under

Fig. 2 Change patterns of P_{20}, P_m, and PI for 1981–2000 relative to 1861–1920 mean from ECHO-G ALL, NAT, GHG, ANT, and SUL experiments

Change patterns 1981−2000 minus 1861−1920
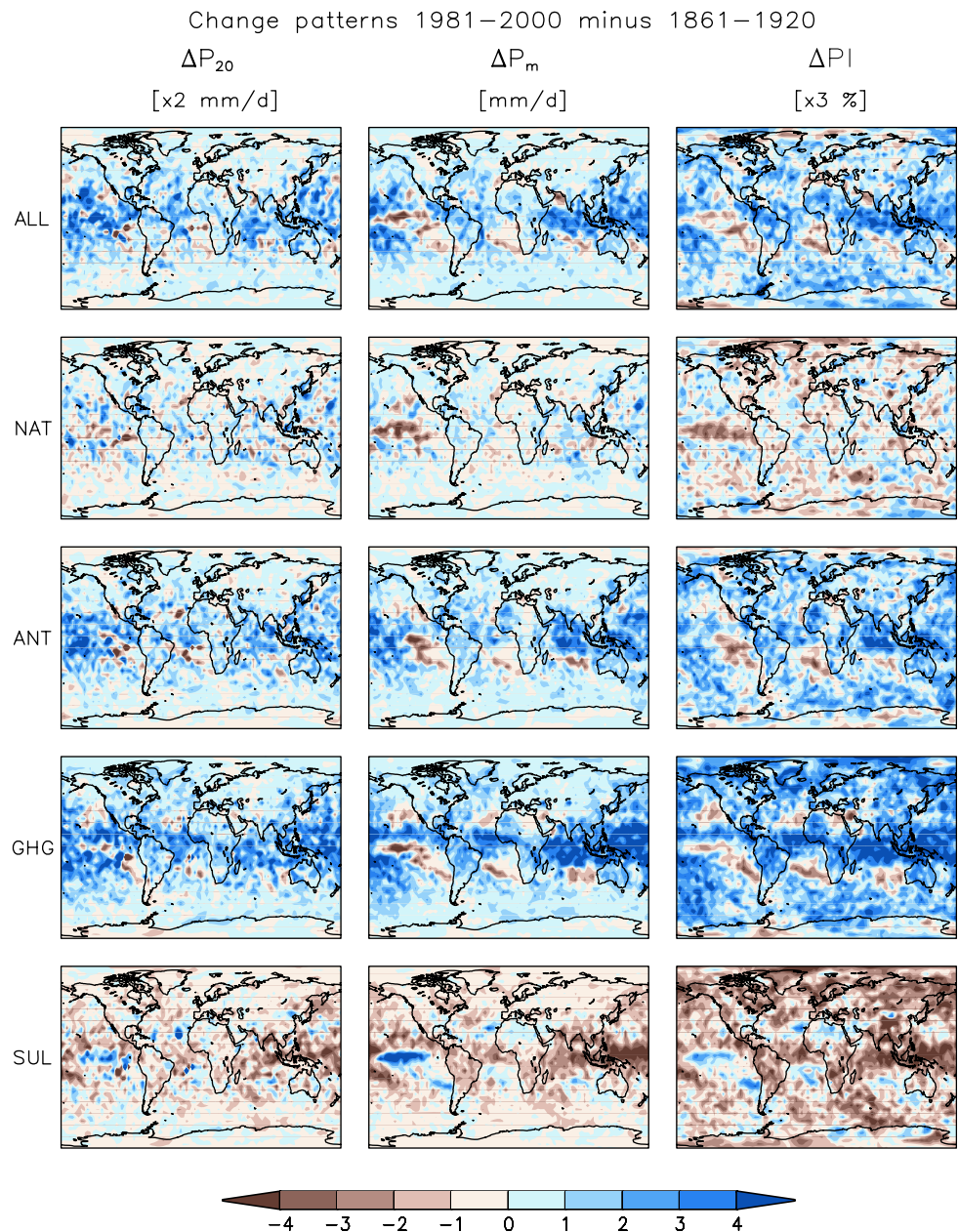
the different external forcings, we present in Table 2 the ratios of global mean changes in the indices (ΔP_{20}, ΔP_m, and ΔPI, %) to global mean surface air temperature changes ($\Delta \bar{T}$, K) in the late twentieth century (1981–2000). Changes are ensemble averages relative to the 1861–1920 mean. The NAT runs are excluded here since the temperature change in 1981–2000 due to NAT forcing is too small (−0.08 K) to provide a reasonable estimate of the sensitivity. Overall the sensitivities of precipitation extremes to global warming are stable within 5.8–8.3% K^{-1} across the extreme indices and the different external forcings. This is close to the sensitivity predicted by the Clausius-Clapeyron relationship (about 7%). This is also in concert with estimates of 6.3–

7.5% K^{-1} obtained from twenty-first century simulations performed with the same model (Kharin et al. 2007), indicating the robustness of the moisture availability constraint to the magnitude of GHG forcing. In contrast, global mean precipitation changes ($\Delta \bar{P}$) with respect to global warming range from 0.1 to 2.4% K^{-1}, which is much smaller than the Clausius-Clapeyron constraint, again in agreement with previous studies (e.g., Allen and Ingram 2002; Pall et al. 2007; Kharin et al. 2007). The SUL results seem somewhat different from the GHG, ANT, and ALL results. This might be associated with higher sensitivity of global mean precipitation response to shortwave forcing rather than to GHG longwave forcing (Hegerl et al. 2007).

Fig. 3 Change patterns of total precipitation (PCP) and sea surface temperature (SST) over low latitudes for 1981–2000 relative to 1861–1920 from ECHO-G ALL, NAT, ANT, GHG, and SUL experiments. Each panel is the mean of three ensemble members. Note that values are expressed as a deviation from zonal mean which is plotted on the right side of each panel. *Contour lines* represent climate patterns obtained from the reference period of 1861–1920
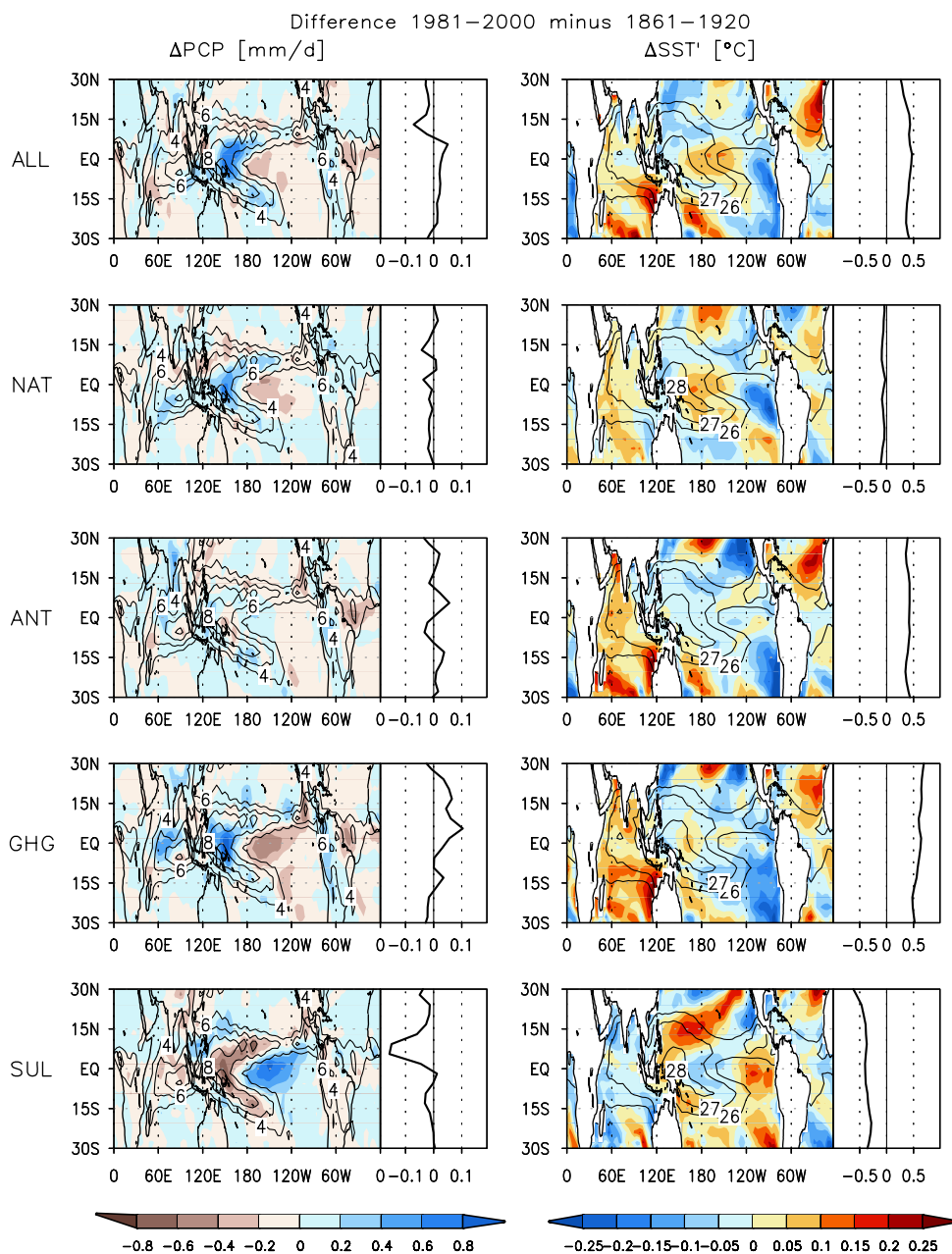
Some differences are recognizable between the extreme indices shown in Fig. 4. The range of the internal variability relative to signals is smaller in PI compared to P_{20} and P_m. Use of the same interval and correspondingly different spatial weighting in PI appears to be responsible for the reduced variability. Another difference is found during 1981–2000 in the NAT simulations where PI has smaller (even negative) values than P_{20} and P_m. This appears to be related to the use of a fixed GEV distribution in the definition of PI, unlike temporally varying parameters as in P_{20} and P_m.

Figure 5 shows the results of signal detectability (logarithm of the Bayes factors) for ANT, NAT, GHG and SUL over different hemispheric domains. These results are from a five-dimensional analysis using a time vector of the twentieth century, i.e. $q = 5$ in Eq. (6). It is clearly shown that the ANT signal is decisively detectable over all hemispheric domains and all extreme indices. PI has a stronger signal than P_{20} and P_m which originates from the reduced internal variability as discussed above. ANT detectability is stronger in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH). One possible explanation for the hemispheric asymmetry is that the change in extremes has less spatial uniformity in the SH (drying subtropical regions are larger) which would weaken the signal in the hemispheric mean (Fig. 2). The

Fig. 4 Time series of area averaged extreme values (P_{20}, P_m, PI) over the globe, land, ocean, NH, SH, NH land (NHL), and SH land (SHL) from ECHO-G CTL, ALL, ANT, GHG, and SUL experiments. ANT* represents CGCM3 results for PI. Note that CTL ranges are whole spread from 47 samples
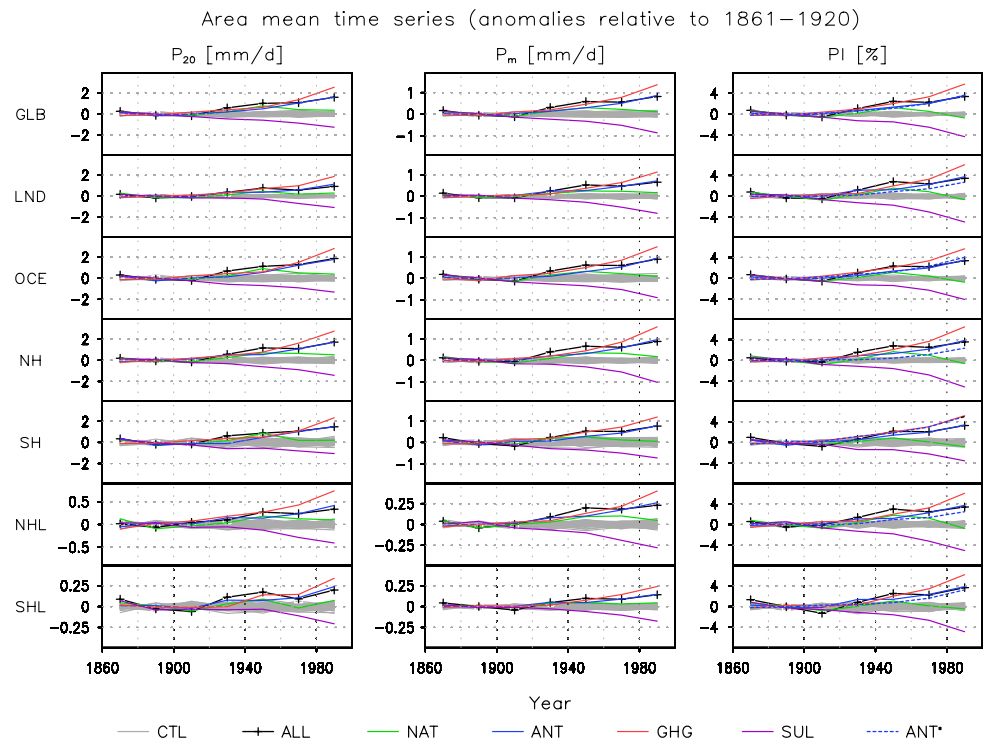

Table 2 Simulated changes of global mean surface air temperatures ($\Delta\bar{T}$, K) and the ratios of globally averaged extreme precipitation (ΔP_{20}, ΔP_m, and ΔPI, %) to $\Delta\bar{T}$ from the ECHO-G twentieth century forced experiments described in Table 1

Forcing	$\Delta\bar{T}$	$\Delta P_{20}/\Delta\bar{T}$	$\Delta P_m/\Delta\bar{T}$	$\Delta PI/\Delta\bar{T}$	$\Delta\bar{P}/\Delta\bar{T}$
ALL	0.48	8.19	6.05	7.12	0.12
ANT	0.50	7.84	6.16	7.21	0.30
GHG	0.77	8.03	6.23	7.50	1.13
SUL	−0.52	5.81	5.91	8.31	2.38

All changes are ensemble means for 1981–2000 with respect to 1861–1920 means. Global mean total precipitation changes ($\Delta\bar{P}$, %) are also given for comparison

detectability becomes weaker when considering land only, due to relatively greater internal variability. Overall, these ANT detectability results are found to be robust even when internal variability is doubled.

GHG signals are detectable over most regions, but with reduced amplitude compared to ANT. This means that the generalized distance (Λ in Eq. (6)) between ALL (the *observation*) and GHG remains large compared to the distance between ALL and CTL, but is less than the distance between ALL and ANT. The NAT signal is also detectable in P_{20} and P_m over most hemispheric domains and in PI over land, the NH, and NH land (NHL) only. This discrepancy in PI seems to be associated with a relative decrease of extreme precipitation for 1981–2000, more dominantly over ocean areas, due to applying the fixed GEV parameters (see above). The GHG and NAT signals
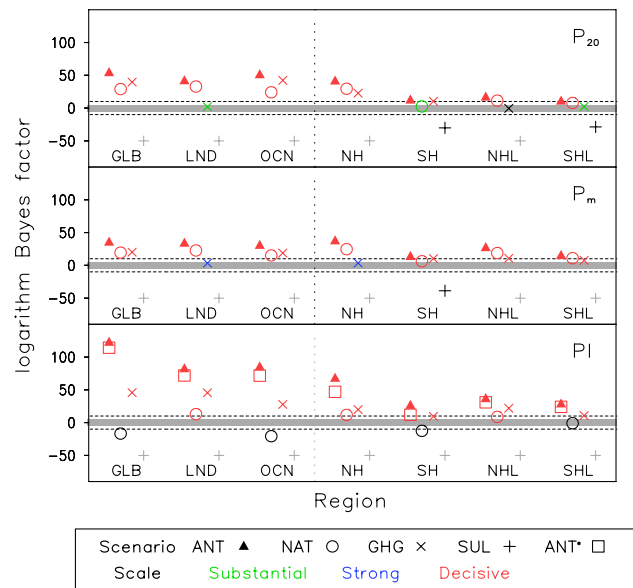

Fig. 5 Signal detectability as assessed by means of Bayes factors for area-averaged extreme precipitation indices P_{20}, P_m, and PI over global and hemispheric areas for the twentieth century (see Fig. 4) which are obtained from a perfect model analysis with ECHO-G regarding ALL simulations as observations and ANT, NAT, GHG, and SUL simulations as fingerprints. ANT* represents an imperfect model analysis with using CGCM3 data as ANT fingerprint. Bayes factors within the *grey shaded bands* indicate less than *decisive* (red) evidence for the forced scenario if larger than 5 and for CTL if smaller than -5. *Dashed lines* indicate the same threshold when internal variability is doubled. Assessments of *strong* (blue) and *substantial* (green) evidence similarly require log Bayes factors greater than 2.5 and 1 respectively. *Grey mark* represents log Bayes factors less than -50

remain detectable even if internal variability is doubled. Overall, the SUL results are characterized by strong negative values of the log Bayes factors, representing very low detectability. For simplicity we omit the SUL results in the detectability plots given below.

5.2 Zonal bands

We divided the globe into six 30° latitudinal bands and examined detectability in time variations of the extreme indices averaged over those zonal bands. ALL forcing runs show an overall increase in the extreme indices in all zonal bands whereas GHG and SUL runs exhibit clear increases and decreases respectively (not shown) as in the hemispheric result. The amplitudes of these changes are larger over low latitudes and are reduced over high latitudes in P_{20} and P_m as would be expected given the latitudinal variation in precipitation variability. Maxima in the NAT runs that occur around 1950 are more pronounced over low latitudes than mid to high latitudes. These features are commonly found in all of the extreme indices examined. Internal variability is weaker in PI as in the hemispheric result.

Figure 6 shows the time series of zonal averages when only land data is included. As a whole, the effects of internal variability are more apparent in these smaller areas, particularly over the southern mid- and high-latitude lands (SMIL and SHIL) where the land area is relatively small. Consequently, extreme precipitation changes fall within the range of internal variability in SMIL and SHIL. On the other hand, compared to the southern tropics (STR), the signal-to-noise ratio in the southern tropical land area (STRL) is larger as the forced responses are stronger. This seems to be caused by removing the areas of decreasing precipitation extremes over the southern tropical ocean (Fig. 2).

The Bayesian decision method was also applied for the zonal bands with and without ocean areas. Results in Fig. 7 show that, when including ocean areas, the ANT signal can be detected over all zonal bands for all three indices, except for P_{20} over the southern mid- and high-latitudes (SMI and SHI). Note that PI and P_m have larger detectability, which is related to their location near the center of the GEV distribution. This suggests there is potential merit in using PI and P_m for detection with real observations. In contrast, P_{20} has larger uncertainty because it represents the tail of the distribution and therefore produces a smaller signal-to-noise ratio. GHG signals are detectable over many regions, but their amplitudes are smaller than for the ANT signal. NAT signals are detectable only over lower latitudes, consistent with the stronger solar influence on the tropical climate (Cubasch et al. 1997; Meehl et al. 2003; Min and Hense 2007). When the internal variability is doubled, *decisive* evidence for ANT remains over the northern tropics and mid-latitude (NTR and NMI) for all three indices.

The land-only result is characterized by reduced signal detectability that is caused mainly by the relatively larger internal variability in smaller area averages. The stronger detection power of PI and P_m over P_{20} still holds. ANT

Fig. 6 Same as Fig. 4 but for precipitation extremes averaged over six zonal bands with land only: northern high-latitude (NHI, 60–90°N), northern mid-latitude (NMI, 30–60°N), northern tropics (NTR, 0–30°N), southern tropics (STR, 0–30°S), southern mid-latitude (SMI, 30–60°S), and southern high-latitude (SHI, 60–90°S). Land area is named by attaching "L" to the corresponding acronym of zonal bands
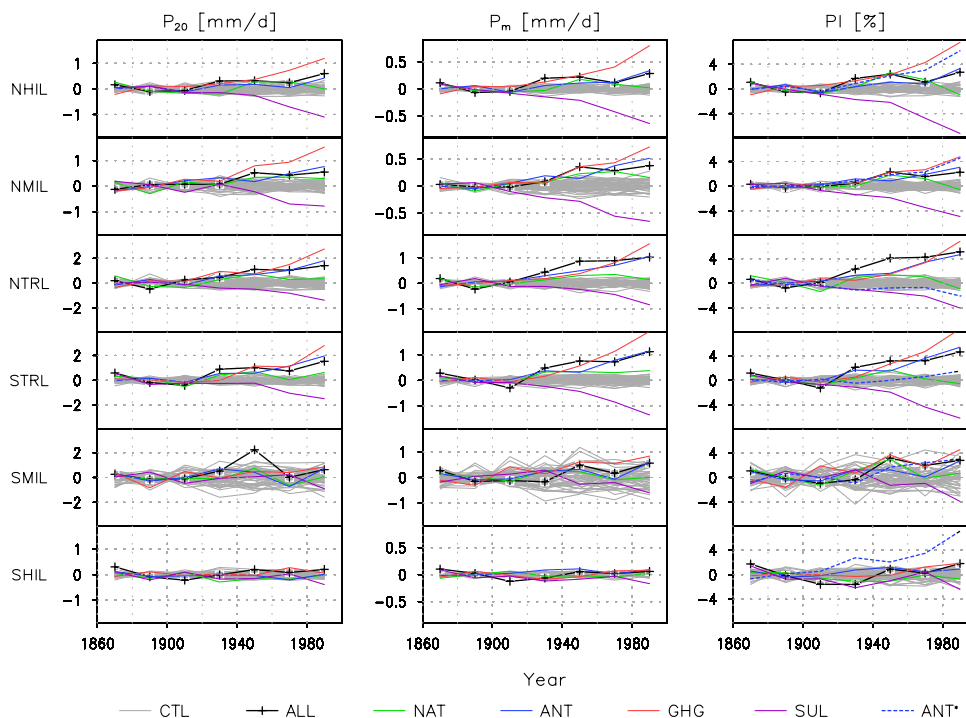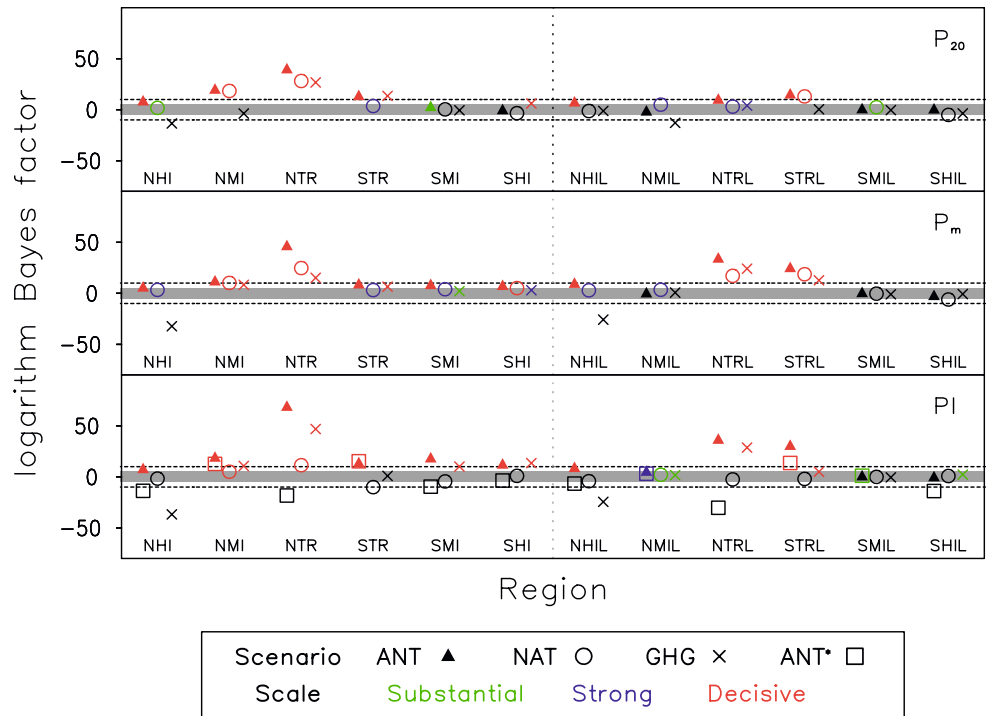
Fig. 7 Same as Fig. 5 but for zonal bands with land plus ocean (*left*) and with land only (*right*). See time series and regional domains in Fig. 6

signals are detectable in some zonal bands, specifically the northern and southern tropical land areas (NTRL and STRL) and the northern high-latitude land (NHIL). Two tropical bands exhibit particularly strong ANT detectability even when the internal variability is doubled, suggesting that tropical land areas might be good candidates for detection if adequate daily observations of precipitation were available (cf. Goswami et al. 2006). The detectability of GHG and NAT signals is not robust in the smaller land areas although they also indicate greater potential

detectability in the tropics. Damping of the NAT signals in the *PI* results (due to the fixed GEV distribution) remains in the latitudinal detection, but not as strongly as in the hemispheric results.

5.3 Continental regions

We also extended our analysis to smaller scales over land. Figure 8 defines several continental scale domains

Fig. 8 Continental domains used in this study and observational availability inferred from Alexander et al. (2006) data set of maximum 5-day consecutive precipitation amounts. The *shaded area* represents grid points where observations are available for longer than 40 years during 1951–1999
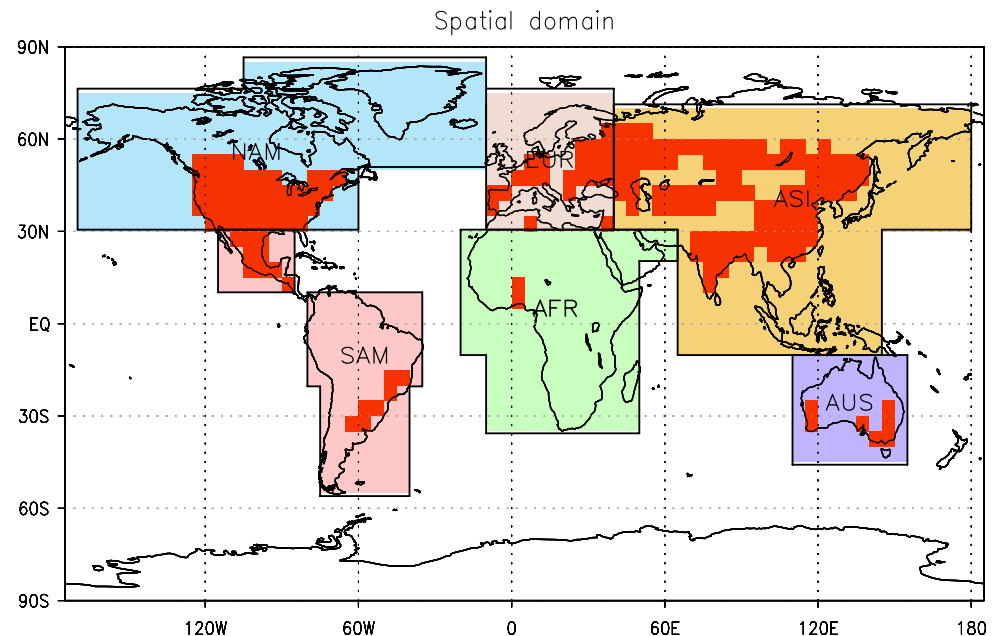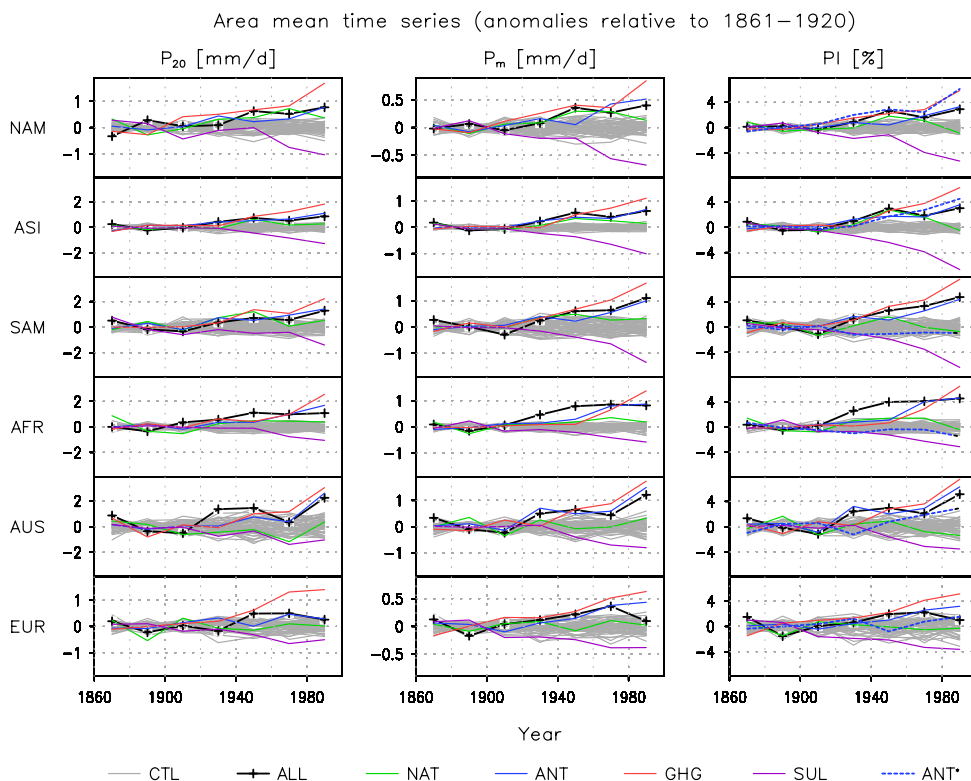
Fig. 9 Same as Fig. 4 but for six continental regions: North America (NAM), Asia (ASI), South America (SAM), Africa (AFR), Australia (AUS), and Europe (EUR)

following Stott (2003) and Min and Hense (2007). Note that the continental analyses here include all grid points, not just the red shaded points in Fig. 8 indicating the availability of observations in the latter half of the twentieth century. Extreme precipitation indices averaged over six continental regions are displayed in Fig. 9. Compared with hemispheric and northern latitudinal areas examined above, there are larger differences in the temporal distributions across these regions. The ALL simulations are characterized by two periods of increasing extreme precipitation, from 1910 to 1950 and after 1970, which are common over all regions. This pattern of change in extreme precipitation indices resembles the behavior of surface temperature changes in the same continental areas (Min and Hense 2007), suggesting higher detectability in extreme precipitation changes than in total precipitation changes. Europe (EUR) is an exception where extreme indices are characterized by a slight decrease in 1981–2000. We speculate that this results from the stronger internal variability related to the North Atlantic Oscillation (NAO) that is reasonably simulated by ECHO-G (Min et al. 2005b). Overall the internal variability (CTL ranges) becomes larger relative to the response to forcing as the size of regions becomes smaller. Clearer increases and decreases appear in the indices from GHG and SUL simulations, respectively. Pronounced NAT forcing responses around 1950 can be found over North America (NAM), Asia (ASI), and South America (SAM) but their structures
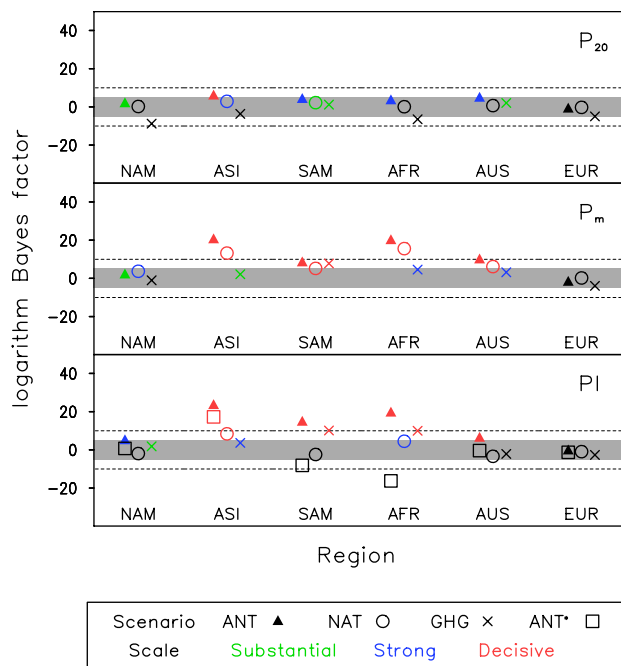

Fig. 10 Same as Fig. 5 but for continental regions using time series shown in Fig. 9

are a bit different among variables and the period of a maximum changes across regions.

Figure 10 represents Bayesian analysis results for the continental regions. It shows that ANT signals are

decisively detectable over Asia, South America, Africa (AFR), and Australia (AUS) when using P_m and *PI*. ANT signals are also detectable in P_{20} over the same regions, although less convincingly. Stronger detectability appears over Asia and Africa where the internal variability is smaller and simulated response is larger compared to the other regions (Fig. 9). GHG signals are at least *strongly* detectable over South America and Africa only with P_m and *PI*. NAT signals are detectable over many regions with P_m, but only over Asia with P_{20}. The *decisive* ANT signals over Asia and Africa remain detectable when internal variability is doubled.

6 Sensitivity test

6.1 Availability of observational data

In the real world, detection can only be conducted over areas where observational data are available. In order to examine how applicable the results above will be in real world detection studies, we repeated our analyses using only GCM data at grid points where we judge that adequate observational data are available. For this purpose, we used the criterion applied by Alexander et al. (2006) in their analyses of maximum consecutive 5-day precipitation amounts. Grid boxes with at least 40-year of data during 1951–1999 are marked with red squares in Fig. 8. Before
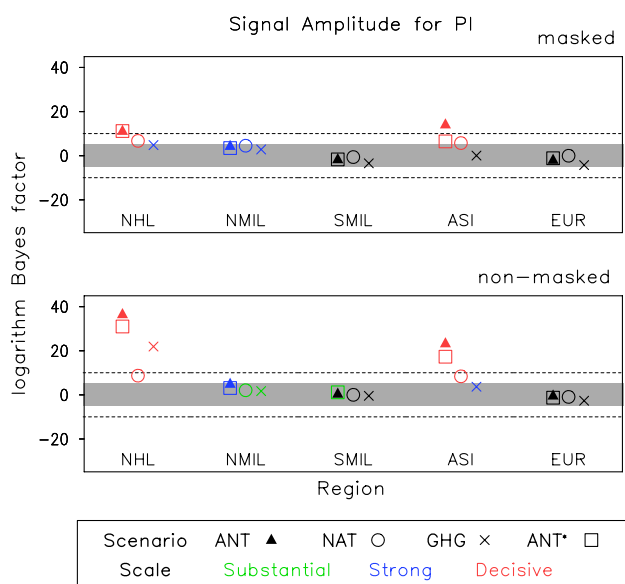

Fig. 11 *Upper* Detectability of ANT, NAT, and GHG signals when observational mask is applied to model data for which analysis is confined to the five regions of NHL, NMIL, SMIL, ASI, and EUR according to the observational availability (at least 30% fraction of land grids) as shown in Fig. 8. (*lower*) Non-masked results for the same regions are repeated for a better comparison. See text for details

masking model fields and calculating spatial averages of extreme indices from them, both models and observations were interpolated to the same $5° \times 5°$ grid. Analysis is confined to five regions where observations are available over reasonably large areas representing at least 30% of fraction of land grid boxes—NH land (NHL), northern mid-latitude land (NMIL), southern mid-latitude land (SMIL), Asia (ASI), and Europe (EUR).

Results are shown in Fig. 11. It indicates that detectability can change substantially if detection is conducted on the regions with available observations. Although signal amplitudes are usually reduced, ANT is *decisively* detectable over the NH land and Asia and *strongly* over the northern mid-latitude land. This suggests that the ANT signal is potentially detectable in the observations in these regions.

6.2 Fingerprint from another model

To examine the sensitivity of detectability to the uncertainty in the structure of fingerprint, we repeated the Bayesian analysis using ANT*, a fingerprint computed from simulations of another GCM, the CGCM3. This analysis is restricted to *PI* which is free of the influence of climatological differences between different models (e.g., Hegerl et al. 2004; Kharin et al. 2007). Results from this imperfect model analysis are also given in Figs. 5, 7, 10, and 11 (denoted as ANT*). For global and hemispheric mean *PI*, ANT* detectability from the imperfect model analysis is very similar to that from the perfect model analysis (Fig. 5). This is very consistent with temporal behavior seen in Fig. 4.

However, results for smaller domains are different. For the zonal bands, imperfect model analysis (Fig. 7) suggests that ANT* is *decisively* detectable only over the northern mid-latitude (NMI) and the southern tropics (STR) if data over both land and ocean is used. If only land data is included, ANT* is still detected over the two zonal bands but detectability for the northern mid-latitude land (NMIL) becomes weaker. In addition, the *PI* time series in Fig. 6 for ANT and ANT* exhibit a pronounced difference over the northern tropical land area (NTRL). Continental-scale results present larger inter-model differences and Asia (ASI) is the only region showing consistent ANT/ANT* detectability (Fig. 10). Detectability results from the imperfect model analysis are not affected by the availability of the observed data (Fig. 11).

Overall, the large inter-model differences suggest that single-model results may not be robust over smaller spatial domains and that one needs to consider large scale patterns so as to detect ANT signals in extreme precipitation changes (Hegerl et al. 2004; Tebaldi et al. 2006; Kharin et al. 2007; Kiktev et al. 2007).

7 Conclusions and discussion

This study examines the extent to which anthropogenic and/or natural influences may be detectable in precipitation extreme indices through a perfect and an imperfect model analysis. Three extreme precipitation indices are defined based on the GEV distribution. They are the 20-year return value (P_{20}), the median (P_m), and the cumulative probability (PI). The P_{20} events are much rarer and more extreme than P_m events. The PI provides relative values based on the probability at each grid point. Regional averages of P_{20} and P_m give higher weight to areas of higher extreme values, while that of PI gives the same weight everywhere. The results from the three indices were compared with each other to explore whether the signal is more readily detectable in any particular index. Fingerprints were obtained from five three-member ensemble experiments with the ECHO-G model under different external forcings: ALL, NAT, ANT, GHG, and SUL. Using the ALL simulations as pseudo observations, we compared signal amplitudes of other experiments individually with the range of internal variability determined from the control simulations. A Bayesian decision method was used to quantify the differences. We analyzed signal detectability for different spatial domains ranging from the globe to individual continents. We also examined the applicability of our results under realistic conditions by conducting our analysis over the regions where there is substantial data coverage from the observations, and by using fingerprints obtained from simulations of another GCM.

As a whole, our analyses suggest that ANT signals should be detectable in extreme precipitation during the twentieth century. The potential becomes weaker as the size of the spatial domain decreases. The ANT signal is consistently detectable in our experiments (with *decisive* evidence) on global to hemispheric scales with all indices, regardless of whether we use the data for the whole domain or for land only. It is also robustly detected in the 30 degree zonal bands, although the detectability is only retained in low latitudes if only land data are used, suggesting that early detection might now be possible over tropical land areas if enough observations were available.

ANT signals are also *decisively* detectable over individual continents in our experiments except for North America and Europe where the larger internal variability associated with the NAO might have weakened the signal-to-noise ratio. Nevertheless, signals remain detectable in P_m and PI. The greater detectability of the ANT signal in P_m and PI is mostly due to the relatively lower internal variability, which is a characteristic of variables located near center of the GEV distribution. In contrast, externally forced signals were less detectable in P_{20}, which is situated in the tail of the distribution. This suggests there is a better

chance to detect ANT signal if P_m or PI is used. GHG and NAT signals are also detectable but less robustly. Note however, that NAT signals were more easily detected in the low latitudes, as in surface air temperatures (e.g., Cubasch et al. 1997; Meehl et al. 2003; Min and Hense 2007).

It is found that the ENSO-like change of mean state under external forcing in the ECHO-G model plays a crucial role in determining extreme precipitation change, especially over the tropical ocean. Since an ENSO-like mean state change in response to ANT forcing is model-dependent (Meehl et al. 2007a), there may be an increased chance for early detection in precipitation extremes if one focuses on the areas and seasons that are less affected by such a response. Furthermore, it is also necessary to consider the effects of atmospheric circulation change on precipitation (Emori and Brown 2005; Meehl et al. 2005; Pall et al. 2007).

Detectability was not much affected when we repeated the analyses on the data grid where there is good observational coverage during the latter half of the twentieth century. However, we found that signal detectability is highly sensitive to inter-model uncertainty. When simulations from another GCM were used to construct the fingerprint, the ANT signal was only detectable on global and hemispheric scales, and results for smaller regions were not very robust, suggesting that the goal of early detection is more realistic at the global and hemispheric scales.

We found that globally averaged extreme precipitation responses in the simulated twentieth century climate under different forcing factors (ALL, ANT, GHG, and SUL) are in overall concert with the Clausius-Clapeyron constraints. This is in agreement with previous studies using future scenario simulations (Allen and Ingram 2002; Trenberth et al. 2003; Pall et al. 2007; Kharin et al. 2007). This robustness of moisture availability constraints on the extreme precipitation changes supports the higher detectability in extreme precipitation than in the mean precipitation (Hegerl et al. 2004).

This study has a few methodological distinctions from previous studies. First, detectability is assessed using transient climate simulations of the entire twentieth century rather than of the latter half of the twentieth century (Kiktev et al. 2003, 2007) or future simulations (Hegerl et al. 2004). Second, we consider temporal variations of precipitation extremes rather than just the long-term trends (Kiktev et al. 2003, 2007; Hegerl et al. 2004). This has the potential to improve the detectability of external signals, especially if decadal variation is substantial. Third, different spatial domains ranging from the globe to individual continents were considered. Fourth, our perfect model analysis is constructed more realistically by considering ALL simulations as observations and the other experiments as possible explanations.

It should be noted that the annual maxima used here may be drawn from particular seasons over the large parts of regions, such as monsoonal areas, where strong seasonal signatures in precipitations exist. Therefore the use of seasonal maxima may not substantially increase detectability. On the other hand, incorporating seasonality could be beneficial on regional scales by separating mechanisms of extreme precipitations into convection versus large-scale process.

It should also be noted that this study is more a *resampleable* model experiment as we use three realizations to define the *observations*. That is, the data set used as observations is three times as large as would be realizable in the real world. The major aim of utilizing the three member ensemble is to obtain more reliable estimates of extreme precipitation indices from larger number of samples, but this might reduce noise related to sampling error on the *observations* and affect detectability. In this regard, we conducted a simple test by comparing GEV parameters estimated from a single realization to those from three realizations. The L-moment method (Hosking 1990) was applied for GEV parameter estimation with the single realization of *observations* due to a small number of samples, i.e. 20 annual maxima, because the ML method can occasionally produce unreliable estimates when sample size is too small (Martins and Stedinger 2000, Kharin et al. 2005). We found very similar spatial and temporal patterns of the two GEV parameters obtained from single and three realizations, suggesting that the internal (or intraensemble) variability is relatively weak compared to the mean response in the ALL experiment with the ECHO-G model.

Finally, it should be pointed out that the results presented here probably represent the upper limit of detectability in extreme precipitation. Comparison between model grid data with station-based observations (e.g., Osborn and Hulme 1997) remains a challenge. Reanalyses are also not of sufficient quality in this respect (Kharin et al. 2005). More importantly, multimodel analyses using historical simulations should be carried out to consider the uncertainty arising from different model responses (Tebaldi et al. 2006; Kharin et al. 2007). Also, it would be imperative to include high resolution models that could better resolve regional climate features associated with precipitation extremes and to test the sensitivity of detectability to model resolution: Can the lower detectability at smaller regional scales be improved by increasing model resolutions?

References

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

Alexander LV et al (2006) Global observed changes in daily climatic extremes of temperature and precipitation. J Geophys Res 111:D05109. doi:10.1020/2005JD006290

Caesar J, Alexander L, Vose R (2006) Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded data set. J Geophys Res 111:D05101. doi:10.1029/2005JD006280

Christidis N, Stott PA, Brown S, Hegerl GC, Caesar J (2005) Detection of changes in temperature extremes during the second half of the 20th century. Geophys Res Lett 32:L20716. doi:10.1029/2005GL023885

Collins M, the CMIP Modelling Groups (2005) El Niño- or La Niña-like climate change? Clim Dyn 24:89–104

Crowley TJ (2000) Causes of climate change over the last 1000 years. Science 289:270–277

Cubasch U, Hegerl GC, Voss R, Waszkewitz J, Crowley TJ (1997) Simulation of the influence of solar radiation variations on the global climate with an ocean-atmosphere general circulation model. Clim Dyn 13:757–767

Cubasch U et al (2001) Projections of future climate change. In: Houghton JT et al (eds) Climate change 2001: The scientific basis. Cambridge University Press, New York, pp 525–582

Douville H, Salas-Mélia D, Tyteca S (2006) On the tropical origin of uncertainties in the global land precipitation response to global warming. Clim Dyn 26:367–385

Emori S, Brown SJ (2005) Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. Geophys Res Lett 32:L17706. doi:10.1029/2005GL023272

Feichter J, Lohmann U, Schult I (1997) The atmospheric sulfur cycle in ECHAM-4 and its impact on the shortwave radiation. Clim Dyn 13:235–246

Flato GM (2005) The third generation coupled global climate model (CGCM3) (and included links to the description of the AGCM3 atmospheric model). available at http://www.cccma.bc.ec.gc.ca/models/cgcm3.shtml

Gillett NP, Weaver AJ, Zwiers FW, Wehner MF (2004) Detection of volcanic influence on global precipitation. Geophys Res Lett 31:L12217. doi:10.1029/2004GL020044

Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Prince Xavier K (2006) Increasing trend of extreme rain events over India in a warming environment. Science 314:1442–1445

Groisman PYa, Knight RW, Easterling DR, Karl TR, Hegerl G.C, Razuvaev VN (2005) Trends in intense precipitation in the climate record. J Clim 18:1326–1350

Hegerl GC, Zwiers FW, Stott PA, Kharin VV (2004) Detectability of anthropogenic changes in annual temperature and precipitation extremes. J Clim 17:3683–3700

Hegerl GC, Karl TR, Allen M, Bindoff NL, Gillett N, Karoly D, Zhang X, Zwiers F (2006) Climate change detection and attribution: beyond mean temperature signals. J Clim 19:5058–5077

Hegerl GC et al (2007) Understanding and attributing climate change. In: Solomon S et al (eds) Climate change 2007: the physical science basis. Cambridge University Press, Cambridge, pp 663–745

Held IM, Soden BJ (2006) Robust responses of the hydrological cycle to global warming. J Clim 19:5686–5699

Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J Roy Stat Soc B52:105–124

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795

Kiktev D, Sexton DMH, Alexander L, Folland CK (2003) Comparison of modeled and observed trends in indices of daily climate extremes. J Clim 16:3560–3571

Kiktev D, Caesar J, Alexander LV, Shiogama H, Collier M (2007) Comparison of observed and multimodeled trends in annual extremes of temperature and precipitation. Geophys Res Lett, 34:L10702. doi:10.1029/2007GL029539

Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. J Clim 13:3760–3788

Kharin VV, Zwiers FW (2005) Estimating extremes in transient climate change simulations. J Clim 18:1156–1173

Kharin VV, Zwiers FW, Zhang X (2005) Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. J Clim 18:5201–5223

Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J Clim 20:1419–1444

Lee T, Zwiers F, Hegerl G, Zhang X, Tsao M (2005) A Bayesian approach to climate change detection and attribution assessment. J Clim 18:2429–2440

Legutke S, Voss R (1999) The Hamburg atmosphere-ocean coupled circulation model ECHO-G. Technical report No. 18, German Climate Compute Centre (DKRZ), Hamburg, Germany, p 62

Martins ES, Stedinger JR (2000) Generalized maximum likelihood generalized extreme-value quantile estimators for hydrological data. Water Resour Res 36:737–744

Meehl GA, Zwiers F, Evans J, Knutson T, Mearns L, Whetton P (2000) Trends in extreme weather and climate events: issues related to modeling extremes in projections of future climate change. Bull Amer Meteor Soc 81:427–436

Meehl GA, Washington WM, Wigley TML, Arblaster JM, Dai A (2003) Solar and greenhouse gas forcing and climate response in the twentieth century. J Clim 16:426–444

Meehl GA, Arblaster JM, Tebaldi C (2005) Understanding future patterns of increased precipitation intensity in climate model simulations. Geophys Res Lett 32:L18719. doi:10.1029/2005GL023680

Meehl GA et al (2007a) Global climate projections. In: Solomon S et al (eds) Climate change 2007: The physical science basis. Cambridge University Press, Cambridge, pp 747–845

Meehl GA, Arblaster JM, Tebaldi C (2007b) Contributions of natural and anthropogenic forcing to changes in temperature extremes over the U.S. Geophys Res Lett 34:L19709. doi:10.1029/2007GL030948

Min S-K, Hense A (2006) A Bayesian assessment of climate change using multimodel ensembles. Part I Global mean surface temperature. J Clim 19:3237–3256

Min S-K, Hense A (2007) A Bayesian assessment of climate change using multimodel ensembles. Part II regional and seasonal mean temperatures. J Clim 20:2769–2790

Min S-K, Hense A, Paeth H, Kwon W-T (2004) A Bayesian decision method for climate change signal analysis. Meteorol Z 13:421–436

Min S-K, Legutke S, Hense A, Kwon W-T (2005a) Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. Part I. Near-surface temperature, precipitation and sea level pressure. Tellus 57A:605–621

Min S-K Legutke S, Hense A, Kwon W-T (2005b) Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. Part II. El Niño Southern Oscillation and North Atlantic Oscillation. Tellus 57A:622–640

Min S-K, Legutke S, Hense A, Cubasch U, Kwon W-T, Oh J-H, Schlese U (2006) East Asian climate change in the 21st century as simulated by the coupled climate model ECHO-G under IPCC SRES scenarios. J Meteor Soc Jpn 84:1–26

Nordeng TE (1994) Extended versions of the convective parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. ECMWF Research Department, Technical Memorandum No. 206, p 41

Osborn TJ, Hulme M (1997) Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. J Clim 10:1885–1908

Paeth H, Scholten A, Friederichs P, Hense A (2008) Uncertainties in climate change prediction: El Niño-Southern Oscillation and monsoons. Glob Planet Change 60:265–288

Pall P, Allen MR, Stone DA (2007) Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO_2 warming. Clim Dyn 28:351–363

Roeckner E, Bengtsson L, Feichter J, Lelieveld J, Rodhe H (1999) Transient climate change with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. J Clim 12:3004–3032

Schnur R, Hasselmann K (2005) Optimal filtering for Bayesian detection and attribution of climate change. Clim Dyn 24:45–55

Semenov VA, Bengtsson L (2002) Secular trends in daily precipitation characteristics: greenhouse gas simulation with a coupled AOGCM. Clim Dyn 19:123–140

Shiogama H, Christidis N, Caesar J, Yokohata T, Nozawa T, Emori S (2006) Detection of greenhouse gas and aerosol influences on changes in temperature extremes. Sci Online Lett Atmos 2:152–155

Stott PA (2003) Attribution of regional-scale temperature changes to anthropogenic and natural causes. Geophys Res Lett 30:1728. doi:10.1029/2003GL017324

Sundqvist H (1978) A parameterization scheme for non-convective condensation including prediction of cloud water content. Quart J Roy Meteor Soc 104:677–690

Sundqvist H, Berge E, Kristjansson JE (1989) Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. Mon Wea Rev 117:1641–1657

Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. Clim Change 79:185–211

Tiedtke M (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon Wea Rev 117:1779–1800

Trenberth KE, Dai A, Rasmussen RM, Parsons DB (2003) The changing character of precipitation. Bull Am Meteor Soc 84:1205–1217

van Oldenborgh GJ, Philip SY, Collins M (2005) El Niño in a changing climate: a multi-model study. Ocean Sci 1:81–95

Watterson IG, Dix MR (2003) Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. J Geophys Res 108(D13):4379. doi:10.1029/2002JD002928

Wehner MF (2004) Predicted twenty-first-century changes in seasonal extreme precipitation events in the parallel climate model. J Clim 17:4281–4290

Yamaguchi K, Noda A (2006) Global warming patterns over the North Pacific: ENSO versus AO. J Meteor Soc Jpn 84:221–241

Zhang X, Zwiers FW, Li G (2004) Monte Carlo experiments on the detection of trends in extreme values. J Clim 17:1945–1952

Zhang X, Hegerl G, Zwiers FW, Kenyon J (2005) Avoiding inhomogeneity in percentile-based indices of temperature extremes. J Clim 18:1641–1651

Zhang X, Zwiers FW, Hegerl GC, Lambert FH, Gillett NP, Solomon S, Stott P, Nozawa T (2007) Detection of human influence on twentieth century precipitation trends. Nature 448:461–465