Clim Dyn (2012) 38:1375–1387
DOI 10.1007/s00382-011-1091-4

Statistical significance of trends in monthly heavy precipitation over the US

Salil Mahajan · Gerald R. North · R. Saravanan · Marc G. Genton

Received: 25 October 2010 / Accepted: 28 April 2011 / Published online: 11 May 2011

Abstract Trends in monthly heavy precipitation, defined by a return period of one year, are assessed for statistical significance in observations and Global Climate Model (GCM) simulations over the contiguous United States using Monte Carlo non-parametric and parametric bootstrapping techniques. The results from the two Monte Carlo approaches are found to be similar to each other, and also to the traditional non-parametric Kendall's τ test, implying the robustness of the approach. Two different observational data-sets are employed to test for trends in monthly heavy precipitation and are found to exhibit consistent results. Both data-sets demonstrate upward trends, one of which is found to be statistically significant at the 95% confidence level. Upward trends similar to observations are observed in some climate model simulations of the twentieth century, but their statistical significance is marginal. For projections of the twenty-first century, a statistically significant upwards trend is observed in most of the climate models analyzed. The change in the simulated precipitation variance appears to be more important in the twenty-first century projections than changes in the mean precipitation. Stochastic fluctuations of the climate-system are found to be dominate monthly heavy precipitation as some GCM simulations show a downwards trend even in the twenty-first century projections when the greenhouse gas forcings are strong.

S. Mahajan (✉)
Computational Earth Sciences, Oak Ridge National Laboratory, One Bethel Valley Road, PO Box 2008, Oak Ridge, TN 37830-6301, USA
e-mail: mahajans@ornl.gov

G. R. North · R. Saravanan
Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843-3150, USA

M. G. Genton
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA

1 Introduction

Extreme events of precipitation have a potential for impacting our social and economic activities. It is therefore essential to determine if there has been a systematic change in the extremes over the past years and what awaits us in the future owing to global warming. Various studies indicate an increasing trend in the extremes of precipitation over the past few decades over the US (Kunkel et al. 1999; Groisman et al. 2001; Karl and Knight 1998; Easterling et al. 2000; Groisman et al. 2004). While some argue that global warming induced by anthropogenic forcings may lead to an increase in the extremes of precipitation (Trenberth et al. 2003; Groisman et al. 1999), natural variability of extreme precipitation also seems to play an important role (Kunkel et al. 2003). The climate system is inherently stochastic in nature and thus it is imperative to assess the statistical significance of the results regarding trends in extreme events and isolate them from stochastic fluctuations.

Precipitation has a very incoherent structure. The probability density function of precipitation is skewed and is clearly non-gaussian. Efforts have been made to empirically approximate precipitation with gamma (e.g. Groisman et al. 1999; Wilby and Wigley 2002), exponential (e.g. Zhang et al. 2004), log-normal (e.g. Kedem et al. 1990) and log-skew-elliptical distributions (Marchenko and Genton 2010) but without physical justification. Extremes by nature form the tail of the distribution. The analysis of the behavior of the extremes of precipitation then becomes a more difficult problem, as it entails the analysis of the tail of a distribution whose inherent structure is unknown. It has been pointed out that the lack of physical justification for the use of various distributions to represent precipitation adds significant uncertainty to the study of extremes (Trenberth 1999; Trenberth et al. 2003; Allen and Ingram 2002). Further issues arise when trend analysis is performed on non-gaussian data, and extreme precipitation events clearly do not support normality. Traditional methods for assessing the statistical significance of linear trends in data with known or unknown non-normal distributions call for the use of rank-based non-parametric methods, as used by Kunkel et al. (1999). In addition to the non-normality, the assumption of a linear trend in the data is also a matter in question. While linear models of trend do not necessarily provide an accurate picture of the true trend in the extremes (Lopez-Diaz 2003), they do provide a reasonable first order measure. A few methods for the analysis of trends in the extremes have recently been investigated which deviate from the linear trend assumption. For example, Frei and Schär (2001) apply the logistic linear regression model to estimate the detection probability of trends in the extremes of precipitation, and Benestad (2003, 2006) examines the number of record breaking events in a time-series for analyzing the trends in the extremes.

In this study, we investigate bootstrapping as an alternative approach for the analysis of linear trends in heavy precipitation. Extremes when defined as a maximum or minimum in a block of a sample of an independent and identically distributed random variable, regardless of the underlying distribution of the population, have been shown to asymptotically converge to one of the three Generalized Extreme Value (GEV) distributions when appropriately normalized (Galambos 1978). And, extremes when defined as threshold exceedences, as used in this study, belong to the generalized Pareto family of distributions and share the shape parameter with the corresponding GEV distribution (Coles 2001). Kharin and Zwiers (2000) exploited this property of extremes to compare different GCM simulations for extremes. The rate of convergence of the extremes to the asymptotic GEV distribution however has been noted to be slow with increase in sample size for distributions other than the exponential distribution (Davis 1982; Leadbetter et al. 1983), thus adding considerable uncertainty in the estimation of the correct GEV parameters for finite sample sizes. Bootstrapping does not require estimation of any parameters of the extreme value distributions and hence is more robust. However, bootstrapping for analysis of extreme values too, is limited by sample size. A bootstrap distribution of extremes is weakly consistent with the GEV of the original sample extremes (Athreya and Fukuchi 1997), and converges to a random probability distribution (Angus 1993). Angus (1993) proved it for non-parametric bootstrapping, where the bootstrap sample is obtained from the original sample data by reshuffling with replacement. Though we cannot be completely sure about the accuracy of bootstrap results asymptotically, a difficulty shared with other methods, the bootstrap procedure does provide an alternative measure of the uncertainty of trend in the extremes.

Here, we apply a non-parametric bootstrap approach to assess the statistical significance of the linear trend in monthly heavy precipitation using the Monte-Carlo scheme. A parametric bootstrap approach, where a theoretical distribution is empirically fit to the data and bootstrap samples are generated from the estimated distribution is also implemented. We use a multi-variate two-parameter log-normal model as an estimate of the distribution of precipitation over the US. Since we lack the physical justification for using a log-normal model, as we would have had for applying any theoretical distribution model, the results from the approach could only be considered as suggestive. The parametric approach is thus used here only to probe if it supports the results from the non-parametric approach. As will be shown in the Sect. 5, we do not find a difference in the results from the two approaches.

This study is focused on heavy precipitation over larger spatial and temporal scales. The motivation for the usage of large spatial and temporal scales of precipitation comes from Groisman et al. (2005). He summarized the results of various studies on extremes of precipitation and concluded that for statistical analysis of extremes of precipitation, area-averages over spatially homogeneous regions should be used to avoid the noise associated with daily weather systems. So, unlike others, who performed an extensive study using daily station data from thousands of sites to assess field significance, we use easily accessible gridded monthly precipitation data. We analyze observational gridded data from two different sources and look for consistency. In this study, we focus on the trends in monthly heavy precipitation with a return period of one year over the US. We do not segregate heavy precipitation by seasons to prevent shrinkage of already scant sample sizes. We refer the interested reader to Kunkel and Andsager (1999) for seasonal analyses of extremes of precipitation for different return periods.

The climate system is a superposition of numerous stochastic processes and the observational data provides us just one realization of these random processes. Thus, it is difficult to distinguish between natural variability (noise) and forced variability (signal) through analysis of observations alone. Therefore, we turn to sate-of-the-art Global Climate Models (GCMs), which capture many aspects of climate variability. Wilby and Wigley (2002) studied two coupled GCMs, HadCM2 and NCAR-CSM, and found that they capture some of the important physical processes that control the statistics of precipitation, indicating that GCM simulations can be used as a proxy for stochastic realizations of the climate system. A comparison of observational data over the last century and climate models indicates that the variation of precipitation in the current climate system cannot be attributed to anthropogenic forcings alone (Lambert et al. 2004). However, climate model projection studies suggest that intense precipitation would be on the rise as global temperatures increase due to increased green house gas forcings in the future (Meehl et al. 2000a, b; Cubasch et al. 2001; Zwiers and Kharin 1998; Kharin and Zwiers 2000). In this study, we analyze various coupled GCM simulations of the twentieth century and the twenty-first century for trends in monthly heavy precipitation and seek their statistical significance using our bootstrap Monte Carlo techniques.

The rest of the paper is divided into the following sections. Section 2 briefly describes non-parametric and parametric methods for the Monte Carlo tests and the multi-variate nature of the problem. In Sect. 3, we define the monthly heavy precipitation as related to our study. Section 4 outlines the data-sets used in the analysis of monthly heavy precipitation. In the first part of Sect. 5, we present results from the observational data. And, in the second part, we report on results from the analysis of GCM simulations of the twentieth century and projections of the twenty-first century. Finally, we summarize our results in Sect. 6.

2 Bootstrapping methods

To perform a Monte-Carlo significance test on the trends in monthly heavy precipitation over the US, we generate numerous realizations (typically 500 in this study) of precipitation over the contiguous US. To realistically generate realizations it is essential to capture the probability distribution of precipitation, and its spatial and temporal structure. In accordance with Groisman et al. (2005), we are interested in the heavy precipitation over large spatial and temporal scales. The contiguous US is broadly divided into 9 climate regions based on the consistency of climate within them (Karl and Koscielny 1982) as shown in Fig. 1. We use this division of the US for our analysis, i.e we spatially average precipitation within each of these 9 climate regions. In addition to this spatial averaging, we also perform temporal averaging by considering monthly precipitation. It is found that for all the climate regions, the total monthly precipitation is approximately temporally independent once the annual cycle is removed. So, we can assume temporal independence while generating realizations. We still however, have to account for the spatial structure and other statistical properties of precipitation.

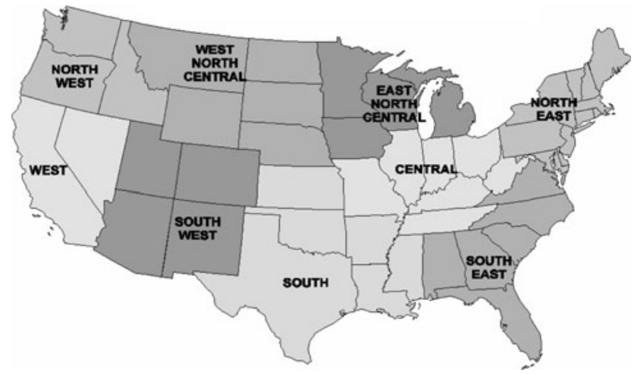

Fig. 1 The 9 Climate regions of contiguous US. (Courtesy: http://www.cpc.noaa.gov/products/analysis_monitoring/regional_monitoring)

To check if there is significant spatial correlation between the monthly precipitation of the nine climate regions, we perform a log-likelihood ratio test (Mardia et al. 1979, 137–138) under the null hypothesis that there is no correlation between the nine climate regions. The test statistic is given by:

$$-n' \log \mid R \mid \tag{1}$$

where, $n' = n(2p + 11)/6$, $\mid R \mid$ is the determinant of the sample correlation matrix, n is the length of data and p is the number of variables (climate regions). For p (=9) variables, this test statistic has an asymptotic chi-square distribution with $p(p - 1)/2$ (=36) degrees of freedom. We find that the correlation between the nine climate regions is statistically significant at the 95% level for all the different observational data-sets and GCM integrations considered in this study.

2.1 Non-parametric approach

Non-parametric bootstrapping essentially involves randomly reshuffling with replacement of the available time-series data to generate new realizations. Since we are just rearranging the time-series we keep the statistical nature of precipitation intact. To maintain the original spatial correlation, we reshuffle all the 9 time-series in the same order. This process of reshuffling is repeated 500 times to generate different realizations. Reshuffling does not preserve temporal correlation. But, as stated before, monthly precipitation averaged over climate regions is found to be essentially temporally independent.

2.2 Parametric approach

Parametric bootstrapping requires the assumption of the probability distribution structure of precipitation. Precipitation has a mixed probability distribution which is a combination of a discrete distribution and a continuous one. The discrete part describes the chance of the occurrence of precipitation, and the continuous part describes the probability of the amount of precipitation given that precipitation does occur. Spatial and temporal averaging though, yields a continuous distribution of precipitation eliminating the discrete part because large-scale averaging ameliorates the problem of no precipitation. Here, we assume that a multivariate log-normal distribution function is an approximate density function of monthly precipitation over the US. We chose the log-normal distribution because it appears to describe the monthly averaged data fairly well as discussed later in Sect. 5.1. The multivariate aspect of the distribution incorporates the spatial correlation of precipitation. Again, temporal independence is assumed. Parameters of the multivariate log-normal distribution are estimated from the data. Realizations of precipitation over the US in the 9 climate regions are then generated by obtaining random samples from the estimated distribution.

3 Monthly heavy precipitation

The definition of extremes of precipitation varies considerably in the literature. In this study, we use a threshold exceedence definition of monthly heavy precipitation similar to the one used by Kunkel et al. (1999). Area-averaged monthly time-series of each of the 9 climate regions are considered. Heavy events of monthly precipitation in a time-series are defined in terms of the return period. A threshold value is associated with a return period (r) in years (e.g., 1, 5, 10 years, etc.) for each of the 9 time-series. The threshold value for each of the time-series is determined in the following manner. First, the number of years of data to be analyzed is established (M) (typically 100 years here). Secondly, the monthly precipitation in the time-series is ranked based on the amount of precipitation in this period. The amount of precipitation during the Nth ranked event is considered as the threshold value for the specified r, where $N = M/r$, for that time-series.

Thus, an event is considered as a heavy event when the amount of precipitation during the event is more than the threshold in that time-series. For example, when analyzing 100 years of data for heavy events with return period of 1 year, 100 events would be identified as heavy. Now, an event of heavy precipitation with a return period of 1 year should ideally occur once a year in those 100 years. But, occasionally it would occur more than once and sometimes not at all. The number of heavy precipitation events occurring in a year in a region, which is a count variable is termed as the Extreme Precipitation Index (EPI) of that region. The EPI time-series is evaluated for each climate region separately with the precipitation time-series of each climate region having a different threshold value. The EPIs of the 9 climate regions are averaged, with area-weighting, to obtain the United States-EPI (US EPI). In this paper, we focus on monthly heavy precipitation with a return period of 1 year, and analyze the trends in the EPIs and the US EPI. The results for other return periods are found to be similar to that of the 1 year return period events. Results from larger sample sizes are expected to be more robust. The choice of heavy precipitation events with return periods of 1 year presented here is based on the associated larger sample sizes, as compared to those of events with a greater return period, while still laying in the realm of extremes. We also do not segregate precipitation by seasons for the same reason.

Assuming that the trend in the extremes is linear, we perform linear regression on the US EPI and individual EPIs to examine trends. A trend in the US EPI would thus be representative of the trend in the monthly heavy precipitation over the US. And, a trend in individual EPIs would be representative of the trend in the monthly heavy precipitation of the respective region. The statistical significance of these trends is assessed using the Monte Carlo test. A caveat needs to be mentioned here. The actual trend in the monthly heavy precipitation may not be linear and may have a more complex temporal structure. However, we are only interested in the gross features of the trend, which are suitably captured by a linear model. Also, theoretically a logistic linear regression model is more apt for count variables like EPIs (e.g. Frei and Schär (2001)). However, the US EPI series deviates from being a count variable as we average EPIs to obtain US EPI. Since we fit a linear regression for trend analysis to the US EPI, we also fit linear models to the EPIs to maintain consistency.

4 Data

We use gridded monthly precipitation data from two sources. The first one is obtained from the University of East Anglia (Hulme 1992) (henceforth, Hulme data-set), and the second from NASA-GISS (Dai et al. 1997) (henceforth, Dai data-set). Gridded precipitation data suffers from errors caused by instrumentation, inhomogeneous station coverage and method of interpolation (Dai et al. 1997). The two data-sets have different methods of correcting these biases, and thus consistent results from analysis of data from the two data-sets would increase our confidence in the results as being robust. The Hulme data-set, which provides estimates of area-averaged precipitation over $5° \times 5°$ boxes, was examined for the time period from 1900 to 1998. And, the Dai data-set, which provides estimates of point precipitation at the center of $2.5° \times 2.5°$ grid-boxes, was analyzed for the time period from 1900 to 1995.

Global Climate Model integrations simulating the climate of the twentieth century from 5 different models were analyzed. Table 1 shows the list of models, resolution and the number of realizations available for each of these models. The model integrations were obtained from the World Climate Research Program's (WCRP's) Coupled Model Inter-comparison Project phase 3 (CMIP3) multi-model dataset. GCM integrations were also obtained for simulations of the Special Reports on Emission Scenarios (SRES) A2 scenario for the twenty-first century from the same 5 models. The SRES A2 scenario for the twenty-first century (henceforth, twenty-first century) represents a 'heterogeneous world based on self-reliance and preservation of local identities with continuously increasing population, fragmented technological change and regionally oriented economic development' (Cubasch et al. 2001). For the Hulme data-set and the GCM integrations, area averaged times-series of monthly precipitation for the 9 climate regions are computed from the gridded data for the Monte Carlo test. For the Dai dataset, the point precipitation at the grid box centers are assumed to represent area averaged precipitation over the grid box, and are used to compute the area averaged precipitation over the 9 climate regions.

5 Results

We analyze the Hulme data-set, Dai data-set, GCM integrations simulating the twentieth century, and the twenty-first century for linear trends in monthly heavy precipitation over the contiguous US and individually for the 9 climate regions within the US. The trends estimated in each of these data-sets are then tested for significance using the Monte Carlo approach described earlier.

5.1 Hulme data-set

5.1.1 Non-parametric bootstrap test

For the analysis of trends in monthly heavy precipitation in the Hulme data-set, the EPI time-series are evaluated for each climate region as described in Sect. 3. The EPI time-series are then area-averaged over all the 9 regions to obtain the time-series of the EPI representative of the entire US, and is referred to as the US EPI time-series hereafter. The EPI time-series for each of the 9 regions and the US EPI time-series are also evaluated for each of the 500 realizations generated by reshuffling. Linear regression is performed on the US EPI time-series for the observational data and each of the Monte Carlo realizations. Figure 2 shows the US EPI time-series for the Hulme data-set (dotted line) and its regression fit (black, dash-dot line). Also shown are the 500 regression fits of the US EPI (gray, thin lines) of each of the 500 realizations. The simulation envelope (black, solid line) that establishes the 95% confidence level bounds for the Monte-Carlo test is also shown. The 95% confidence level bound means that 95% of the synthetic realizations have a trend smaller than that

Table 1 List of GCM models analyzed and a summary of results

GCM	Resolution	20th century		21st century	
		No. of realizations	No. of realizations with trends	No. of realizations	No. of realizations with trends
CCSM	1.405° × 1.405°	8	1 significantly up 5 up 2 down	5	5 significantly up
GFDL	2.0° × 2.5°	3	0 significantly up 2 up 1 down	2	0 significantly up 2 down
PCM1	2.81° × 2.81°	4	0 significantly up 1 up 3 down	3	3 significantly up
HadCM3	2.5° × 3.75°	2	0 significantly up 0 up 2 down	1	0 significantly up 1 up
MPI-ECHAM	1.875° × 1.875°	3	0 significantly up 2 up 1 down	3	3 significantly up
Total		20	1 significantly up	14	11 significantly up

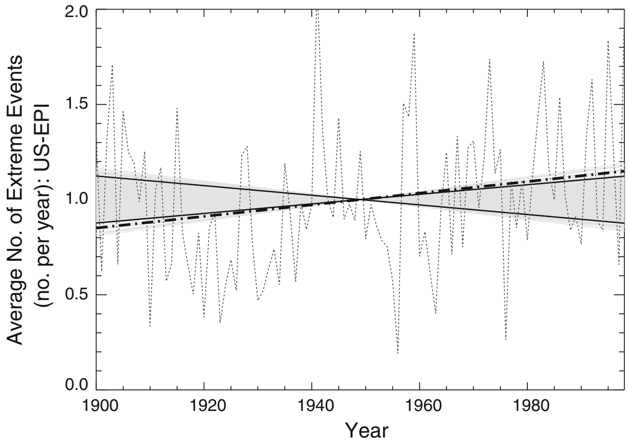

Fig. 2 Trends in monthly heavy precipitation over the US (averaged over the 9 climate regions) in the Hulme data-set. Shown above are the US EPI (Extreme Precipitation Index) time-series of the data (*dotted*, *thin*), trend in the US EPI time-series of the data (*dash-dot*, *thick*), trends in the US EPI of 500 realizations of the data obtained by reshuffling (envelope of *gray thin lines*) and the 95% confidence limit simulation envelope (*black*, *solid*). Note that the *dash-dot line* lies outside the 95% simulation envelope indicating that the trend is statistically significant. The EPI is defined here for monthly heavy precipitation events with return period of one year

of the bound. Thus, if a trend is found to be greater than that of the bound, it implies that the trend is significant.

It is thus observed from Fig. 2 that the trend in monthly heavy precipitation averaged over the US lies just outside the 95% confidence limit bounds and so is statistically significant, though only marginally. We cross-check our results with that obtained from the traditional rank-based Kendall's τ test. It was observed that the trend is also statistically significant at the 95% confidence level when Kendall's τ test is applied implying the robustness of our test.

The significance of the trends in each of the 9 climate regions is also tested individually using the Monte-Carlo approach. Simulation envelopes for 95% confidence level for each region of the Hulme data-set are evaluated from the EPI regression trends of the 500 realizations in the respective regions. Figure 3 shows the number of heavy precipitation events (EPI) occurring in each of the regions and their regression fits along with the regression fits of the EPIs of the 500 realizations of synthetic data of that region. Also shown are the 95% confidence limits based on these realizations. Upward trends are observed in most of the regions. As a result, area averaging of the EPIs over the 9 regions leads to an upward trend. Spatial averaging also leads to a narrowing of the confidence limits, allowing the overall trend to become significant. Note that although upward trends are observed in most of the regions, a statistically significant upwards trend is observed in only one of the 9 regions, the North-East. The statistically significant upwards trend in the North-East climate region is consistent with Groisman et al. (2004), who find a statistically significant upwards trend there in the winter in very heavy daily precipitation, defined as the top 0.3 % of precipitation. The upward trends in the East-North-Central and the Central climate regions are also partly consistent with Groisman et al. (2004) where they found statistically significant upward trends, in those regions. Kunkel et al. (1999) also find a significant upwards trend in the East-North-Central region in extreme precipitation defined as 7-day duration events with return period of 1 year analyzed for the period 1931–1996. Groisman et al. (2004) also find a significant upwards trend in the South climate region, where we fail to discern a trend. The differences in our results from others is likely due to the different definitions of heavy precipitation events, and also because of the coarser spatial and temporal resolution of our data-set, where we are smoothing out the finer structure of extremes.

5.1.2 Parametric bootstrap test

The Monte-Carlo test of trends in monthly heavy precipitation in the Hulme data-set is also performed using realizations obtained from the multivariate log-normal distribution model of precipitation estimated from the data. In order to examine if our empirical model is realistically modeling precipitation we look at the density function and quantile-quantile plots of observed data and the simulations from the model. Figure 4a shows the density function of monthly precipitation of the 9 climate regions from the Hulme data-set. Density functions of each of the 500 synthetic realizations, generated using the model, are also plotted for each of the regions. The density functions of the data and the realizations are normalized so that the area under each of the curve equals unity. It is observed that multivariate log-normal distribution is quite a good fit for monthly precipitation over the US. Notice that the skewness in precipitation data is well replicated in the synthetic data. However, examining the right end of the density functions and the quantile-quantile plots (not shown) reveals thicker right tails of simulated data as compared to the observations over all the 9 regions. This discrepancy of the log-normal model implies greater variance in the tail of the simulated density function. Ordinarily, this would imply that our results based on the log-normal model are conservative. But, we define heavy precipitation events as threshold exceedences where each time-series, observed or simulated, has its own unique threshold for extremes. Consequently, simulated data would thus only have a larger threshold value as compared to observational data. The biases introduced on the linear trends in heavy precipitation due to a larger extreme threshold cannot be stated conclusively and needs further investigation. Our results based on the parametric approach thus could only be suggestive

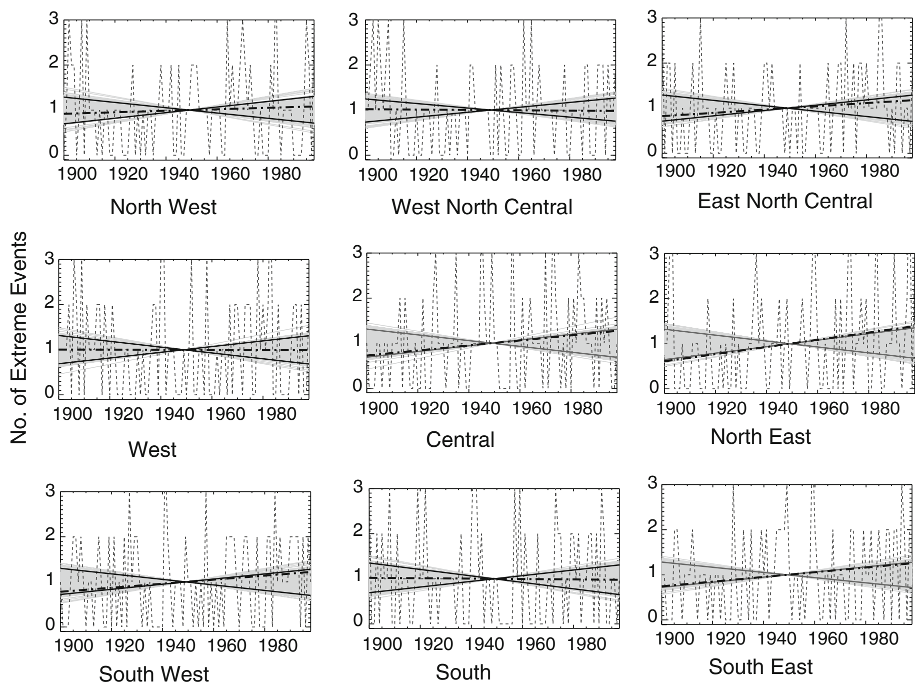

Fig. 3 Trends in monthly heavy precipitation in the Hulme data-set in the 9 climate regions. The line styles are the same as Fig. 2. Only one (North-East) of the 9 regions displays a statistically significant upwards trend, that too, marginally

of the statistical significance of the trends in the extremes of the observations and would not be accurate.

Figure 5 shows the Monte-Carlo test for the parametric bootstrap method. The confidence intervals evaluated from the parametric bootstrap strongly match those evaluated from the non-parametrically generated realizations and hence support the results from the non-parametric bootstrap approach both over the contiguous US and at each of the climate regions (not shown), further confirming the robustness of our approach. For further reference it is stated here that the results from the non-parametric bootstrap and the parametric bootstrap, although not shown here, also match well for the Dai data-set, and all of the GCM integrations analyzed. Further discussion and figures, unless otherwise mentioned, are all for the non-parametric bootstrap Monte-Carlo technique.

5.2 Dai data-set

The analysis in Sect. 5.1 is repeated using the Dai data-set instead of the Hulme data-set. Figure 6 shows the result of the analysis. The trend in the US EPI of the Dai data-set, and the associated 95% confidence limits evaluated using the non-parametric bootstrap Monte-Carlo technique are shown. The 500 synthetic realizations are generated by reshuffling precipitation time-series derived from the Dai data-set. This trend is not statistically significant, unlike the Hulme data-set, although it fails the significance test only marginally. Kendall's τ test was also performed on the Dai data-set and a strong agreement between the two tests was observed.

Noting the difference in the linear trend in the US-EPI of the Hulme data-set and the Dai data-set, it is logical to test if the trends displayed in the two data-sets are statistically similar. For this purpose, we fit a linear regression model to the difference of the Hulme US EPI and Dai US EPI, and test the null hypothesis that the regression coefficient is zero. To establish the confidence interval of the regression coefficient we again use a Monte-Carlo approach (Chernick 1999) as the statistical characteristics of the regression coefficient of the difference are not known a-priori. The bootstrap samples of the difference of Hulme US EPI and Dai US EPI are obtained by the non-parametric approach as discussed in Sect. 5.1. Hulme data-set and the Dai data-set are both reshuffled with replacement with the same order of reshuffling. The same order of reshuffling maintains the correlation between the two data-sets in the bootstrap sample. The regression coefficient of the difference in the reshuffled Hulme US EPI and the Dai US EPI is estimated. This process is repeated 500 times. From these 500 estimates of the regression coefficient of the difference of the Hulme US EPI and Dai US EPI, we evaluate the 95% confidence interval of the regression coefficient. It is found that the null hypothesis cannot be rejected and thus the linear trends in the Hulme US EPI and Dai US-EPI are statistically similar.

The results from the analysis of Hulme data-set and the Dai-dataset hence support each other. The marginal difference in the results from the two data-sets is due to the different methods of generation of the data-sets. While the two data-sets share the same station data as described in Eischeid et al. (1991), differing interpolation techniques

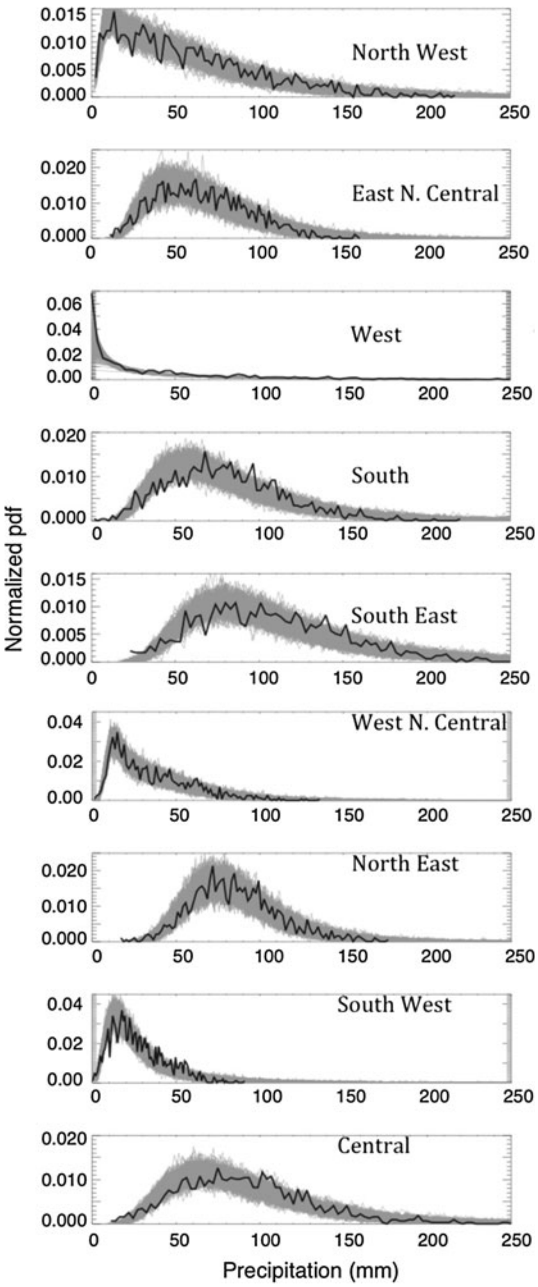

Fig. 4 Probability Density Functions (PDFs) of monthly precipitation over the 9 climate regions in the US from the Hulme observational data (*black*) for the 20th century for the Monte Carlo tests. The *gray* lines are the PDFs of various synthetic data realizations from the multivariate log-normal distribution model of precipitaion over the US based on the Hulme data-set

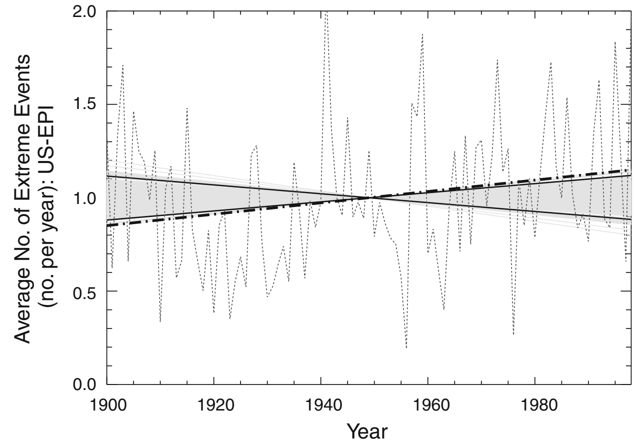

Fig. 5 Same as Fig. 1, but the 500 realizations are generated from the multivariate log-normal model of precipitation over the US and the 95% confidence limit simulation envelope is evaluated using these realizations

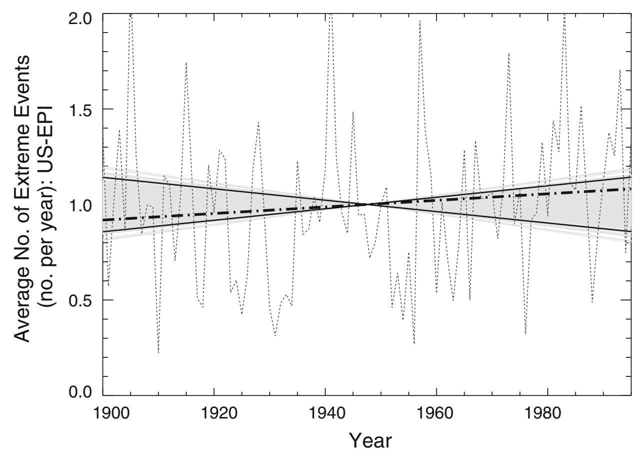

Fig. 6 Same as Fig. 2, but for the Dai data-set

cause differing estimates of the mean and the variability of monthly precipitation (New et al. 2001). The Hulme gridded data-set is created by weighted averaging of station data within a grid box. The weights are assigned based on the area the station represents. The Dai data-set on the other hand estimates precipitation at grid point centers from station data using an inverse distance weighting scheme, where the weights to station data are assigned based on its distance from a grid box center. Our results above, where we see a weaker trend in heavy precipitation in the Dai dataset, suggest that the interpolation scheme used for the creation of the Dai dataset probably leads to a lower variability in heavy precipitation as compared to the Hulme data-set. Identification of the exact causes of the difference in the heavy precipitation events of the two data-sets requires an in-depth study, not pursued here.

5.3 GCM integrations: 20th century

Each of the 20 different GCM integrations of the twentieth century are analyzed in exactly the same manner as described in Sect. 5.1, except that instead of using the Hulme data-set, data from each of the GCM integrations are used as independent realizations of the Earth's climate, with the bootstrap confidence intervals evaluated for each of the integrations individually. All models exhibit biases in simulating precipitation distribution, which also extends into the extremes. However, GCMs seem to capture the skewness of the density function of precipitation fairly well, as is seen in Fig. 7 for one of the CCSM twentieth century simulations with the multi-variate log-normal distribution estimations adequately representing the skewness, much like the Hulme data-set (Fig. 4). GCMs also capture the spatial structure of precipitation. A significant spatial correlation between the climate regions is observed in each of the GCM integrations similar to observations based on the hypothesis test stemming from Eq. 1, though the strength of the correlation among the 9 regions varies in each of the models.

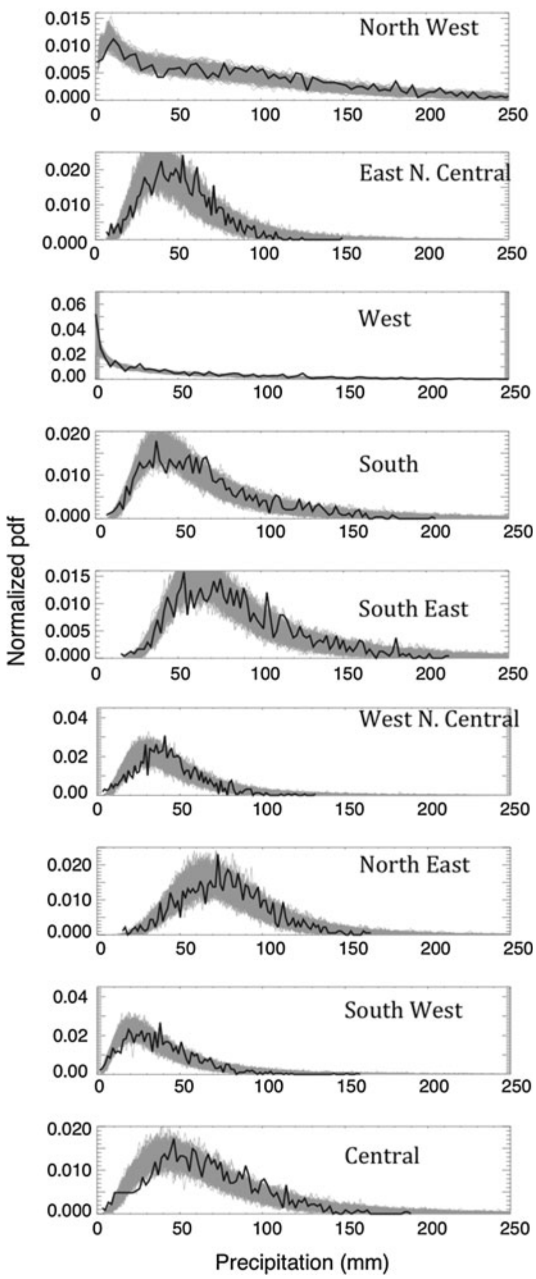

Fig. 7 Same as Fig. 4, but for a realization of the CCSM simulation of the twentieth century. The *gray lines* are the PDFs of various synthetic data realizations from the multivariate log-normal distribution model based on the CCSM simulation

Only 1 (CCSM) of the 20 different GCM integrations analyzed demonstrates a statistically significant (based on the non-parametric test, Sect. 5.1) upwards trend, that too marginally, in monthly heavy precipitation over the US, when analyzed in the same manner as the observed precipitation. Table 1 summarizes the results for the significance of trends in monthly heavy precipitation over the US obtained from all the GCM integrations analyzed. Figure 8 shows a typical result from one CCSM integration. The US EPI time-series and the confidence limits for each of the GCM integrations are evaluated in the same manner as for the real world data. Both upward and downward trends in monthly heavy precipitation are observed in various GCM runs. Lambert et al. (2004) using optimal finger-printing were unable to isolate a signal from the natural variability of the climate associated with the changes in precipitation from the climate model. Our results indicate a similar predicament for monthly heavy precipitation, as we do not observe a significant trend in all but one of the GCM integrations. We also investigate if the trends in the monthly heavy precipitation over the US as simulated by GCMs are statistically different from that of the Hulme data-set, using a test as described in Sect. 5.2 It is found that 9 of the 20 integrations (45%) are statistically dissimilar to the real world data when trends in monthly heavy precipitation are analyzed indicating either a poor simulation of heavy precipitation event in GCMs or strong stochastic fluctuations extending into the extremes of the climate-system replicated in the GCMs. These inconsistencies are also reflected at the regional scale. A comparison between the trends in the EPIs of individual climate regions in the GCM simulations and the Hulme data-set illustrates little correspondence. For example, the North-East region in the Hulme data-set displays a significant upwards trend, but in GCM integrations of the twentieth century, some runs show an upward trend, whereas others display a downward trend.

5.4 GCM integrations: 21st century

The analysis described in Sect. 5.1 is also applied to each of the 14 GCM projections of the twenty-first century, replacing the Hulme data-set with 100 years of data from the projected twenty-first century. Significant spatial correlation between the climate regions is also observed in each of these GCM projections of the twenty-first century. Strongly significant upward trends in monthly heavy precipitation over the US in the twenty-first century are

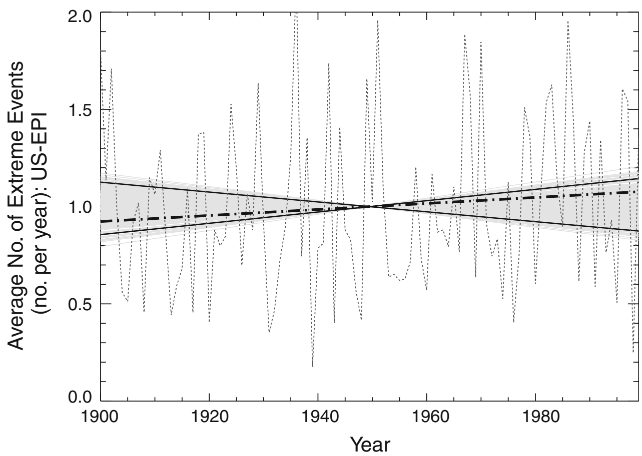

Fig. 8 Same as Fig. 2, but for a typical GCM integration of the twentieth century (CCSM)

observed. Integrations were analyzed for the twenty-first century for the same five GCMs as in the twentieth century. 11 of the 14 integrations demonstrate significant upward trends in monthly heavy precipitation in the twenty-first century. Figure 9 shows the trends in heavy precipitation of one such realization from CCSM. Table 1 summarizes the results obtained from all the GCM integrations of the twenty-first century. A strong signal is thus projected in the future associated with anthropogenic forcings, which is consistent with the IPCC assessment that increases in heavy precipitation are very likely in the twenty-first century. Surprisingly, 2 of the 14 integrations, both GFDL runs, display downward trends even in the twenty-first century when the greenhouse gas forcing signal is strong.

The upward trend in the US EPI in the twenty-first century simulations in most of the models is also reflected

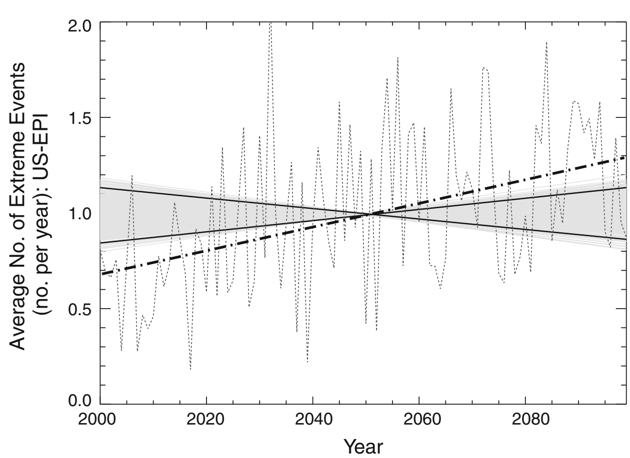

Fig. 9 Same as Fig. 2, but for a typical GCM integration of the twenty-first century (CCSM)

in the individual climate regions. Most of the model simulations of the twenty-first century show a significant upward trend in monthly heavy precipitation in each of the 9 climate regions. Even the regions that showed a downward trend in the twentieth century simulations, show an upward trend in the twenty-first century runs. The one HadCM3 model run of the twenty-first century that does not show a significant upward trend in the US EPI, demonstrates upward trends, though not prominent, in the individual regions. Note that this model showed a declining trend in the US EPI when simulations of the twentieth century were analyzed, but simulation of the twenty-first century does show a substantial, though not significant, upwards trend, signaling the strengthening anthropogenic forcing response.

The changing nature of heavy precipitation is associated with changes in mean state and the variability. To explore the possible reasons for the surprising downward trends in the GFDL simulations of the twenty-first century, we compute the mean and variance of monthly precipitation averaged over the US for each of the models, averaged over all ensemble members, in both the twentieth century and the twenty-first century. Figure 10 shows the mean and variance of monthly precipitation of the 5 GCMs. A slight increase in mean monthly precipitation in the twenty-first century is seen in all the models, except for the GFDL and Hadley model, which show a small decrease. However, there is a substantial increase in the variance of monthly precipitation over the US in all the GCMs, including the GFDL and Hadley models, in the twenty-first century as compared to the twentieth century. A change in the variance has a larger effect on the frequency of heavy events than a change in the mean of the same magnitude (Katz and Brown 1992; Meehl et al. 2000b). An increase in the variance implies that there is a greater spread of the probability distribution function, resulting in an increase in the frequency of extremes. Quantile-quantile plots of twenty-first century model runs against their respective twentieth century model runs also indicate a thickening right-tail of precipitation in the twenty-first century in most of the regions for all models. The presence of downward trends in monthly heavy precipitation in the twenty-first century GFDL runs, in-spite of an increase in the variance of precipitation, thus highlights the stochastic fluctuations and sampling issues associated with the extremes of precipitation, which possibly overshadows the anthropogenic green house gas forcing signal. This also underscores the necessity of using an ensemble of climate projections. It is quite possible that an upward trend will be seen if more realizations from the GFDL model are considered, as the variance of precipitation increases in the twenty-first century integrations.

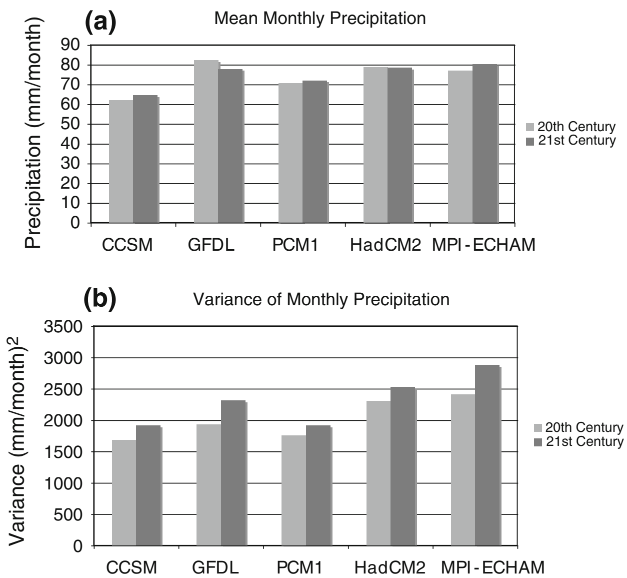

Fig. 10 **a** Mean and **b** Variance of monthly precipitation over the US in GCMs for the twentieth century and twenty-first century

6 Summary and conclusions

Observational precipitation data of the past century from two sources, GCM simulations of the twentieth century and GCM projections of the twenty-first century are assessed for linear trends in monthly heavy precipitation, defined by a return period of one year, over contiguous US. The statistical significance of these trends are assessed from Monte-Carlo tests using non-parametric and parametric log-normal bootstrap samples separately. Individual tests on observational data sources indicate a significant upwards trend in the Hulme data set, and a non-significant upwards trend in the Dai data set irrespective of the method of generation of the bootstrap samples. A significance test reveals that the individual trend magnitudes of the two data sources are not statistically different from each other. The two data sources are constructed from the same station data set. The difference in the trends in monthly heavy precipitation in the two data sets then probably arises from the different ways of accounting for inhomogeneities in the station data set or different methods of generating the gridded data. A deeper analysis of the reasons for different trends in heavy precipitation in the two data sets is beyond the scope of this study.

The trends estimated from the two data sources straddle the margin of statistical significance, and hence a definitive answer to the question of increasing trend of heavy precipitation over the US cannot be arrived at by looking at observational data. GCM simulations forced with twentieth century boundary conditions have been demonstrated to simulate the climate reasonably well. Considering GCM simulations as proxies of the climate system, and perceiving them as independent realizations of the twentieth century climate, a somewhat consistent answer is found. Individual tests on GCM simulations indicate that 19 among 20 integrations have no significant upwards trend implying that the trend in heavy precipitation over the US as seen in the observations might not be real. There is reason to believe the GCM simulations for large scale analysis. Quantile-Quantile plots (not shown) of GCMs against the Hulme dataset while not perfect indicate positive as well as negative biases over the 9 climate regions. Assuming, albeit naively, that these biases cancel out when averaged over the US, trends in a GCM simulation can be considered to be representative of the real-world. However, statistical tests comparing trends in heavy precipitation of the Hulme dataset and the GCM simulations indicate that 9 among the 20 GCM integrations demonstrate trends that are significantly different from the observations, suggesting that GCMs are not yet fully capable of simulating extremes of precipitation at a regional level.

Nevertheless, one GCM simulation of the twentieth century does demonstrate a statistically significant upwards trend in monthly heavy precipitation, implying that anthropogenic forcings of the climatic conditions in the twentieth century were conducive to a rising trend in heavy precipitation over the US. But, in the presence of natural stochastic fluctuations the forced trend might have been overshadowed pointing towards the type II error in statistical analysis. Type II errors are associated with detectability of trends whence the probability of detection of a weak trend is lower, and hence an acceptance of the null hypothesis that no trend is present, does not conclusively imply the absence of a trend but rather that the trend could not be detected. Frei and Schär (2001) quantify the detection probability for trends in extreme precipitation events for station data. Here, we do not delve into detection probability of trends in the extremes but treat it as a caveat and hence we cannot conclusively deny the possibility of the presence of an increasing trend in monthly heavy precipitation over the US during the past century.

To project into the future, climate models, despite their demonstrated deficiencies are our best bet. An analysis of GCM integrations simulating the twenty-first century under the SRES A2 scenario reveals that there is unquestionably a significant trend in monthly heavy precipitation in most of the integrations, implying a role for anthropogenic forcing in the systematic increase of heavy precipitation. This upward trend in heavy precipitation in the twenty-first century appears to be associated primarily with an increase in the variance of the precipitation, rather than an increase in the mean precipitation. This result is also consistent with other modeling studies that show that extremes of precipitation are bound to increase as the world warms. Two of the 14 integrations analyzed show a downward trend of monthly heavy precipitation in the twenty-first century, though statistically insignificant, indicating that the impact of natural variability can be dominating even when the anthropogenic greenhouse gas forcing is strong. Overall, while we cannot definitively answer if the heavy precipitation in the US has been increasing in the past century, GCM integrations suggest that it is very likely to increase in the future.

It should be noted however, that heavy precipitation over large spatial and temporal scales as studied here has different ramifications from that of point-extremes. For e.g. long-term planning for flood water levels and river-water runoffs would require analysis of gross extreme features as in other studies and in ours, while planning for point-extreme events, for e.g. flash floods, would require extreme value analysis at a particular location, may be in conjunction with large scale analysis. Planning on a small-scale however, should never solely rely on results of large-scale analyses, as they do not characterize the small-scale features of a specific region completely.

Acknowledgments S.M. and G.R.N. wish to gratefully acknowledge partial support from a NOAA grant whose P.I. is Professor S. S. P. Shen. In addition, S.M. wishes to thank Mr. H. J. Haynes whose endowment provided support. This research also received some support from an NSF/CMG grant ATM-0620624. M.G.G.'s research was partially supported by NSF grant DMS-1007504 and by Award No. KUSC1-016-04 made by King Abdullah University of Science and Technology (KAUST). We acknowledge the modeling groups for making their model output available as part of the WCRP's CMIP3 multi-model dataset, the Program for Climate Model Diagnosis and Inter-comparison (PCMDI) for collecting and archiving this data, and the WCRP's Working Group on Coupled Modeling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multi-model dataset is supported by the Office of Science, US Department of Energy.

References

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

Angus J (1993) Asymptotic theory for bootstrapping the extremes. Commun Stat Theor Methods 22(1):15–30

Athreya KB, Fukuchi J (1997) Confidence intervals for endpoints of a c.d.f. via bootstrap. J Stat Plan Inference 58:299–320

Benestad RE (2003) How often can we expect a record-event? Clim Res 23:3–13

Benestad RE (2006) Can we expect more extreme precipitation on the monthly time scale? J Clim 19:630–637

Chernick MR (1999) Bootstrap methods: a practitioner's guide. Wiley-Interscience, New York

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Berlin

Cubasch U et al (2001) Projections of future climate change. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) Climate change 2001: the scientific basis. Contributions of working group 1 to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp 881

Dai A, Fung I, Del Genio A (1997) Surface observed global land precipitation variations during 1900–1988. J Clim 11:2943–2962

Davis RA (1982) The rate of convergence in distribution of the maxima. Stat Neerland 36:31–35

Easterling DR, Evans JL, Groisman PYa, Karl TR, Kunkel KE, Ambenje P (2000) Observed variability and trends in extreme climate events: a brief review. Bull Am Meteor Soc 81:417–425

Eischeid JK, Bradley RS, Jones PD (1991) A comprehensive precipitation data set for global land areas. DOE Tech. Rep. TR051, DOE ER-69017T-H1, 81 pp

Frei C, Schär C (2001) Detection probability of trends in rare events: theory and application to heavy precipitation in the Alpine region. J Clim 14:1568–1584

Galambos J (1978) The asymptotic theory of extreme order statistics. Wiley, New York

Groisman PY et al (1999) Changes in the probability of heavy precipitation: important indicators of climatic change. Clim Change 42:243–283

Groisman PYa, Knight RW, Karl TR (2001) Heavy precipitation and high streamflow in the contiguous United States: trends in the twentieth century. Bull Am Meteor Soc 82:219–246

Groisman PYa, Knight RW, Karl TR, Easterling DR, Sun B, Lawrimore JM (2004) Contemporary changes of the hydrological cycle over the contiguous United States: trends derived from in situ observations. J Hydrometeor 5:64–85

Groisman PYa, Richard WK, Easterling DR, Karl TR, Hegerl GC, Razuvaev VN (2005) Trends in intense precipitation in the climate record. J Clim 18(9):1326–1350

Hulme M (1992) A 1951-80 global land precipitation climatology for the evaluation of general circulation models. Clim Dyn 7:57–72

Karl TR, Kosciенly AJ (1982) Drought in the united states: 1895–1981. J Climatol 2:313–329

Karl TR, Knight RW (1998) Secular trends of precipitation amount, frequency, and intensity in the USA. Bull Am Meteor Soc 79:231–241

Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. Clim Change 21:289–302

Kedem B, Chiu LS, North GR (1990) Estimation of mean rain rate: application to satellite observations. J Geophys Res 95:1965–1972

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

Kunkel KE, Andsager K, Easterling DR (1999) Long-term trends in extreme precipitation events over the contiguous United States and Canada. J Clim 12:2515–2527

Kunkel KE, Easterling DR, Redmond K, Hubbard K (2003) Temporal variations of extreme precipitation events in the United States 1895–2000. Geophys Res Lett 30:1900. doi:10.1029/2003GL018052

Lambert FH, Stott PA, Allen MR, Palmer MA (2004) Detection and attribution of changes in twentieth century terrestrial precipitation. Geophys Res Lett 31:L10203. doi:10.1029/2004GL019545

Leadbetter MR, Lindgren G, Rootzen H (1983) Extremes and related properties of random sequences and processes. Springer, Berlin

Lopez-D'az J (2003) A nonparametric test for trends in the occurrence of rare events. J Clim 16:2602–2614

Marchenko YV, Genton MG (2010) Multivariate log-skew-elliptical distributions with applications to precipitation data. Environmetrics 21:318–340

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press, London

Meehl GA, Karl T, Easterling DR et al (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 GA, 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

New M, Todd M, Hulme M, Jones P (2001) Precipitation measurements and trends in the twentieth century. Int J Climatol 21:1889–1922

Trenberth KE (1999) Conceptual framework for changes of extremes of the hydrological cycle with climate change. Clim Change 42:327–339

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

Wilby RL, Wigley TML (2002) Future changes in the distribution of daily precipitation totals across North America. Geophys Res Lett 29:1135. doi:10.1029/2001GL013048

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

Zwiers FW, Kharin VV (1998) Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. J Clim 11:2200–2222