NOTES AND CORRESPONDENCE

On the Verification and Comparison of Extreme Rainfall Indices from Climate Models

Cheng-Ta Chen

Department of Earth Sciences, National Taiwan Normal University, Taipei, Taiwan

Thomas Knutson

Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 15 May 2006, in final form 21 June 2007)

ABSTRACT

The interpretation of model precipitation output (e.g., as a gridpoint estimate versus as an areal mean) has a large impact on the evaluation and comparison of simulated daily extreme rainfall indices from climate models. It is first argued that interpretation as a gridpoint estimate (i.e., corresponding to station data) is incorrect. The impacts of this interpretation versus the areal mean interpretation in the context of rainfall extremes are then illustrated. A high-resolution (0.25° × 0.25° grid) daily observed precipitation dataset for the United States [from Climate Prediction Center (CPC)] is used as idealized perfect model gridded data. Both 30-yr return levels of daily precipitation (P_{30}) and a simple daily intensity index are substantially reduced in these data when estimated at coarser resolution compared to the estimation at finer resolution. The reduction of P_{30} averaged over the conterminous United States is about 9%, 15%, 28%, 33%, and 43% when the data were first interpolated to 0.5° × 0.5°, 1° × 1°, 2° × 2°, 3° × 3°, and 4° × 4° grid boxes, respectively, before the calculation of extremes. The differences resulting from the point estimate versus areal mean interpretation are sensitive to both the data grid size and to the particular extreme rainfall index analyzed. The differences are not as sensitive to the magnitude and regional distribution of the indices. Almost all Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) models underestimate U.S. mean P_{30} if it is compared directly with P_{30} estimated from the high-resolution CPC daily rainfall observation. On the other hand, if CPC daily data are first interpolated to various model resolutions before calculating the P_{30} (a more correct procedure in our view), about half of the models show good agreement with observations while most of the remaining models tend to overestimate the mean intensity of heavy rainfall events. A further implication of interpreting model precipitation output as an areal mean is that use of either simple multimodel ensemble averages of extreme rainfall or of intermodel variability measures of extreme rainfall to assess the common characteristics and range of uncertainties in current climate models is not appropriate if simulated extreme rainfall is analyzed at a model's native resolution. Owing to the large sensitivity to the assumption used, the authors recommend that for analysis of precipitation extremes, investigators interpret model precipitation output as an area average as opposed to a point estimate and then ensure that various analysis steps remain consistent with that interpretation.

1. Introduction

Changes in the frequency or intensity of extreme weather and climate events could have profound impacts on both human society and the natural environment (Easterling et al. 2000b; Meehl et al. 2000). Indicators based on the observed daily precipitation during the second half of the twentieth century suggest that, on average, wet spells produce significantly higher rainfall totals now than a few decades ago (Frich et al. 2002; Alexander et al. 2006). Heavy rainfall events have become more frequent over the past 50 yr even in locations where the mean precipitation has decreased or is unchanged (Easterling et al. 2000a; Folland et al. 2001; Groisman et al. 2005).

It is of great interest to evaluate the ability of the current generation of coupled atmosphere–ocean gen-

Corresponding author address: Prof. Cheng-Ta Chen, National Taiwan Normal University, Department of Earth Sciences, 88, Section 4, Ting-Chou Road, Taipei 116, Taiwan.
E-mail: chen@rain.geos.ntnu.edu.tw

DOI: 10.1175/2007JCLI1494.1

eral circulation models (AOGCMs) to simulate observed extreme rainfall distributions and their trends. But the lack of comparable long-term global gridded daily observations often leads to a deferral of the model evaluation (e.g., Hennessy et al. 1997; Watterson and Dix 2003; Wehner 2004; Tebaldi et al. 2006) or limited evaluation of only the mean precipitation climatology (McGuffie et al. 1999; Semenov and Bengtsson 2002; Voss et al. 2002). Other studies have used station gauge and/or reanalyses data for the assessment (Zwiers and Kharin 1998; Hegerl et al. 2004; May 2004; Kharin et al. 2005) and show some level of agreement between models and observations. The Third Assessment Report (TAR) of the Intergovernmental Panel on Climate Change (IPCC) concluded that comparatively low model resolution is an inhibiting factor for detailed evaluation of extreme rainfall (McAvaney et al. 2001). Studies of daily precipitation characteristics from climate models have generally concluded that simulated precipitation tends to occur more frequently but is less intense than observed heavy rainfall (Osborn and Hulme 1998; Dai 2006; Sun et al. 2006). There is little demonstrated model skill in simulating observed past trends in precipitation extremes (Kiktev et al. 2003).

Despite the problems that current models have in reproducing present-day precipitation frequency distributions and heavy events, projected changes in extreme rainfall are receiving increased attention in view of the serious consequences from possible changes of frequency or intensity of extreme rainfall events (Kharin and Zwiers 2000; Semenov and Bengtsson 2002; Voss et al. 2002; Watterson and Dix 2003; Wehner 2004). Recently, building upon several model intercomparison projects, climate studies of extreme events have used multiple models to further address the issue of possible model dependence and to provide a range of uncertainty from different model formulations (Hegerl et al. 2004; Kharin et al. 2005; Tebaldi et al. 2006). It has been acknowledged that the comparison between model grid output and station data is not straightforward (Kiktev et al. 2003; Wehner 2004; Hegerl et al. 2004) and that calculations of precipitation extreme indices could be sensitive to model resolution (Iorio et al. 2004; Kharin et al. 2005; Emori et al. 2005). For example, Kharin et al. (2005) note that

> Precipitation extremes obtained from individual station records are essentially point estimates and are not directly comparable to the gridded model output that presumably represents precipitation variability on much coarser spatial scales . . . The proper coarse graining of station data is not a trivial task . . . and is beyond the scope of the present paper.

In this study, we will expand on this by quantitatively illustrating the impact on extremes analysis of using the point estimate versus areal mean interpretation of model precipitation output. This will demonstrate that the assumption of model precipitation data as point estimates, or the inconsistent handling of the extreme analysis and comparisons, could easily lead to misinterpretation of model performance and differences. It should be noted that we regard these two interpretations as representing two extreme ends of a range of possible interpretations. We illustrate the impact across this entire range, since a number of previous extreme precipitation analysis studies mentioned above handled the data comparison with procedures that implicitly assumed either a gridpoint or areal average interpretation.

As noted above, a key initial issue to be addressed in model extreme rainfall analysis is the interpretation of the model gridded output. Does the model data represent point (station) estimates, areal mean values, or somewhere in between? With rainfall observations, it is straightforward to treat station data as point estimates and the gridded rainfall analysis as interpolated and weighted means from a set of surrounding stations. The choice of observed data grid size could have a strong impact on the daily rainfall amounts for individual grid boxes if the nearby stations do not have similar precipitation in the same day (i.e., short correlation length scales). In fact this is very likely to occur for many synoptic situations considering the small-scale spatial and temporal variation of precipitation.

All the models use some type of grid system for calculation of atmospheric dynamics and physics. One source of ambiguity in interpreting the meaning of model precipitation output arises from the fact that a model's numerical schemes can be interpreted as a mixture of both gridbox and gridpoint approaches. While some model parameterizations are implicitly areal in implementation, to aggregate the subgrid-scale variations, the finite difference or spectral methods for numerics produce point values at model grid locations. Skelly and Henderson-Sellers (1996) discuss these alternative views in more detail. Mass-flux-based moist convection parameterizations (e.g., Arakawa and Schubert 1974; Moorthi and Suarez 1992) assume the presence of ascending and descending motions within the grid cell (i.e., subgrid-scale variability of convection). Similarly, large-scale (stratiform) cloud parameterizations (e.g., Tiedtke 1993) may explicitly track the fraction of grid cell in which rain is occurring. Clearly the rainfall totals obtained from these parameterizations represents an area average of smaller-scale features (i.e., of the ascending and descending regions or of raining and nonraining parts of the grid cell) and not a point estimate. We thus consider the interpretation of model precipitation gridbox output as representing a point estimate to be inappropriate in the context of extreme precipitation analysis. From the viewpoint of the large-scale water budget, one must generally assume that the model output represents an areal average. While different interpretations of model grid precipitation data may not be particularly important to studies focused on understanding and improving the model simulation of large spatial-scale precipitation features, these interpretations strongly affect the comparison of extreme precipitation indices between observations and models and among different models. Data analysis methods that either implicitly or explicitly assume either the point estimate or areal mean interpretation have been variously applied in previous studies.

If one assumes (incorrectly, we would argue) that model grid data represent a set of point estimates, like station data, one would then compute the daily rainfall extremes of different model data on their native grids and then interpolate to a common grid for intercomparison. The same procedure would also apply to observed station data and gridded estimates from radar, satellite, or combined analyses. Examples of recent intercomparison studies of daily rainfall characteristics and extremes using this approach (albeit with strong caveats) include Kharin et al. (2005, 2007), Sun et al. (2006), and Tebaldi et al. (2006). It is noteworthy that gridded observational analyses, even with relatively high resolution, already involve spatial interpolation and thus should represent an underestimate of extreme daily rainfall as compared to the point measurements.

If the model grid precipitation data are treated as "areal averages" assigned to the center point of the model grid boxes, one should first interpolate the model and observation data to a common grid and then compute the extreme rainfall indices for model evaluation. The same procedure should apply when two models with different resolutions are compared. Otherwise, the disagreement between two datasets could be solely due to the different grid size. The studies by Osborn and Hulme (1997, 1998), Booij (2002), and Iorio et al. (2004) adopted this assumption for model evaluation. This second approach also leads to the general notion that daily station rainfall data, by their nature as point measurements, are not directly comparable to the gridded model output (Hegerl et al. 2004; Kharin et al. 2005). Subgrid-scale variability is not explicitly represented in the model gridded output under this assumption.

It is noteworthy that the spatial interpolation schemes used can also have important impacts on the analysis results. Since the interpolation needed for comparison in our analysis is generally from a fine grid to a coarse grid, the application of conservative remapping would be more consistent with the "areal mean" assumption for model output. The area-averaged rainfall is conserved using this type of interpolation (Jones 1999). The commonly used bilinear or bicubic interpolation schemes would be more appropriate under a "point estimate" assumption (not recommended). These interpolation schemes are not conservative. The destination grid is mainly determined by the nearby quadrilateral points of the input grid.

Depending on the assumption used, the order of two operations (i.e., data interpolation and extremes analysis) applied to daily precipitation will differ, and one could obtain very different conclusions in an assessment of a climate model's ability to simulate extreme rainfall. Intuitively, when the second (areal average) assumption is used, one would expect a lower-resolution model (all other things being equal) to have higher wet-day frequency, reduced daily intensity, and weaker extreme events. On the other hand, if the point estimate assumption were used, a low-resolution model would not necessarily be expected to produce weaker daily rainfall extremes in the analysis. We will try to illustrate more quantitatively the differences in extreme rainfall indices that result from the assumptions about model output. Note that these different assumptions have minimal impact on the long-term seasonal climatology of precipitation.

Treating model rainfall as an areal average, it is straightforward to reinterpolate the daily gridded model data or observational rainfall analysis before calculating the extremes (Iorio et al. 2004). It is not as simple to transform data assumed to be point estimates to areal mean values. Neither objective analysis of gauge data and other rainfall measurements (Higgins et al. 2000; McCollum and Krajewski 1998; Hewiston and Crane 2005) nor the gridding of daily rainfall characteristics and extreme precipitation indices derived from station data (Osborn and Hulme 1997; Sivapalan and Blöschl 1998; Booij 2002; Kiktev et al. 2003; Alexander et al. 2006) is trivial. Almost all previous work in converting different extreme rainfall indices and their past trends from point (station) observations to areal mean (grid box) form is based primarily on the empirical spatial statistical structure of time series of station extremes. A gridding methodology is developed from the spatial correlation structure of the time series of neighboring station data (Kiktev et al. 2003; Alexander et al. 2006). This statistical structure depends not only on the geographical location but also on the particular indices selected. Additional assumptions are necessary for the search radius and the calculation of weighting functions for each station involved in the gridding process. The characteristic decorrelation length scales are determined from sampled data and curve fitting. Since the time series of extreme indices at stations are the basis for gridding, the gridded result from this method could be rather similar to the station data if the distances of the stations from the gridbox center are short and the decorrelation length scale is small.

Instead of attempting to transform rainfall station data to areal averages here, the starting point for our analysis is a relatively high-resolution gridded daily precipitation analysis for the continental United States (Higgins et al. 2000). An objective analysis scheme was used by Higgins et al. (2000) to interpolate the station data to grids. The daily precipitation regional distribution and movement are generally well preserved in the analysis. Local rainfall intensity in this dataset should be somewhat less than that for gauge data due to the interpolation. In our analysis, we further conservatively interpolate the data to different gridbox sizes and derive the extreme indices based on these interpolated daily rainfall observations. Although both techniques for gridding (based on objective analysis of station daily rainfall or based on the spatial structure of station data time series) attempt to derive grid data from station data, they are different in methodology. However, the objective analysis (spatial interpolation) and, subsequently, conservative remapping of observed data before the calculation of extreme indices is presumably more comparable to the model output under the assumption that model grid data represent areal means.

Presently the IPCC data archive includes extreme precipitation indices calculated on the native grid of the climate models that participated in the modeling activities of the IPCC Fourth Assessment Report (AR4). If studies of extremes and their future changes for multimodel ensembles are based on these archived indices (e.g., Meehl et al. 2005; Tebaldi et al. 2006), then these studies implicitly assume that model output is more like a point estimate. Alternatively, if one assumes that model data represent areal averages, then the archived model indices with different resolutions cannot be compared or averaged directly due to the different data grid sizes.

Other recent assessments of modeled extreme precipitation and daily rainfall characteristics (e.g., Kharin et al. 2005, 2007; Sun et al. 2006; Dai 2006) also use multiple climate model outputs from the Atmospheric Model Intercomparison Project (AMIP)/Coupled Model Intercomparison Project (CMIP) or IPCC AR4 future climate scenario runs. Although Kharin et al. (2005) argued that the model results should not be directly comparable to station data, extreme analysis and other daily rainfall characteristics (i.e., mean intensity or frequency) are commonly done on the original model grid before interpolation to a common grid for multimodel ensemble mean and model intercomparison (e.g., Kharin et al. 2005; Sun et al. 2006; Dai 2006), which is consistent with the assumption that model data are more like point measurements. Hegerl et al. (2004) and Sun et al. (2006) also recognize the possible impact of the different data scale but argue that it should not substantially affect results in their study. On the other hand, the different model formulations and diversity of precipitation-related parameterizations could obscure this apparent scaling effect when multiple model simulations are involved in the comparison.

The main purpose of this note is to further explore the scaling–aggregation issue of extreme rainfall indices derived from climate models and to quantitatively assess the impact of a range of different assumptions. We limit the focus to precipitation indices due to their common usage and the stochastic and noisy nature of their spatial distribution. We discuss the data and extreme rainfall indices used in this study in section 2. In section 3, we explore how different assumptions about the data (gridpoint value versus areal mean) affect the extreme rainfall indices, using idealized data from observational rainfall analysis. In section 4, we discuss the extreme rainfall indices derived from IPCC AR4 participating models and their evaluation under different model grid output assumptions. In section 5, we briefly discuss the possible impact of the spatial interpolation scheme on the estimation of extreme rainfall indices. Our summary and conclusions are presented in section 6.

2. Data and extreme indices

Station gauge measurements are the main data source for previous studies on observed changes in extreme daily precipitation and other extreme climate indices related to precipitation (Frich et al. 2002; Alexander et al. 2006). To study various extreme rainfall characteristics, high-resolution (e.g., hourly to daily) data are required. Although many efforts have been devoted to providing better coverage of historical daily station data over the globe, data quality control and homogeneity checks are also extremely important issues (Groisman et al. 2005; Alexander et al. 2006). In addition, deriving gridded data from time series of individual stations involves detailed analysis of the spatial statistical structure of rainfall (Kiktev et al. 2003; Alexander et al. 2006). Both are beyond the scope of this note, where we ignore possible data quality issues and directly use the Climate Prediction Center (CPC)

Daily U.S. Unified Precipitation gridded dataset as an example of high-resolution precipitation data. The dataset provides daily rainfall estimates on a 0.25° × 0.25° grid for the period 1961–90. Simulated daily rainfall data from different climate models were downloaded from the IPCC AR4 data archive at the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory. Two extreme rainfall indices, the 30-yr return level of daily rainfall and a simple daily intensity index (SDII), are chosen for our study. These have been frequently used for previous studies of extreme precipitation in observations and models (e.g., Zwiers and Kharin 1998; Semenov and Bengtsson 2002; Voss et al. 2002; Wehner 2004; Kharin et al. 2005, 2007; Tebaldi et al. 2006).

a. *CPC Daily U.S. Unified Precipitation*

The continental United States has a relatively dense array of in situ (hourly and daily) rain gauge data. Thus studies of this region have good potential to provide a relatively useful high-resolution precipitation analysis. Over the past few years, the CPC has developed the U.S. Precipitation Quality Control (QC) system and analysis to fulfill this need. The gridded observed daily rainfall data we used in this study are based on this work—the CPC Daily U.S. Unified Precipitation. The detailed description of the data sources, compilation, analysis, and verification are documented in Higgins et al. (2000). The daily analyses are gridded at a horizontal resolution of 0.25° × 0.25° over the domain 20°–60°N, 140°–60°W. A Cressman scheme with modifications was used for the objective analyses. The historical precipitation reanalysis covers the period from 1948 to 1998 with no missing values. Several quality control procedures, including standard techniques (duplicate, buddy, and deviation checks) and intercomparison with radar and satellite estimates, are applied to the gauge data. There is a small possibility that real station extreme values could be erroneously discarded when the "buddy check" standard QC step is applied. We assume that this procedure would not significantly affect our results, although one should be aware of such a possibility. We interpolated the data to grids of increasing grid size (from 0.25° to 4° at an interval 0.25°). This range of grid sizes covers the range used by most current climate model grid sizes in the IPCC AR4 model data archive. For these interpolated observed precipitation analyses (using various grid sizes), only grids consisting exclusively of land are analyzed. This is done to avoid ambiguity of mixing results with grids having partial land cover. For our sensitivity test, we also interpolated the observed daily precipitation to all the model grids [i.e., from 4° × 5° to ~1.1° × 1.1° (T106)] found in the IPCC AR4 model data archive before computing the extreme indices. The spatial interpolation scheme used was based on a conservative remapping technique (Jones 1999). Thus, the interpolated daily precipitation analysis at a coarser grid is consistent with an "area-average" assumption.

b. *Daily precipitation from IPCC AR4 models*

The model-simulated daily precipitation data for present-day climate conditions was extracted from various simulations developed for the IPCC AR4. The data archive at PCMDI (available online at http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php) consists of output from a large sample of current generation coupled AOGCMs. At the time we began this study, 17 models were available with at least one ensemble member and having daily precipitation data for the 1961–2000 period from their twentieth-century climate simulations (20c3m) for the IPCC AR4. Table 1 briefly summarizes the characteristics of these models, including the horizontal grid spacing of the atmospheric components. The global model output was available on longitude–latitude grids ranging in grid spacing from 5° × 4° to T106 with a median spacing of ~2.8° × 2.8° (T42). Only one ensemble member was obtained for each model for validation. Since our model evaluation is focused on the conterminous United States, we masked out model grid boxes that were not entirely land according to the land–sea mask of the CPC dataset. Two modeling groups provide experiments at more than one model resolution. The Center for Climate System Research/National Institute for Environmental Studies/ Frontier Research Center for Global Change (CCSR/NIES/FRCGC) Model for Interdisciplinary Research on Climate 3.2 (MIROC3.2) runs are available for both T106 (hires) and T42 (medres) resolutions. The Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (CGCM3.1) simulations are done at both T63 and T47 resolutions (data available on 128 × 64 and 96 × 48 grids). In these cases we also conservatively interpolated the daily precipitation data from the higher-resolution model run to the model grid boxes of the lower-resolution model. The paired high–low-resolution runs from the same basic model might be useful for examining effects of the different analysis procedures on the model extreme statistics output. However, scale-dependent parameters used in the model and model tuning could be the sources of further differences even for the case of a "single" model framework (Iorio et al. 2004).

TABLE 1. The list of IPCC AR4 model simulations, with daily precipitation of twentieth century available, analyzed in this study. Model resolution is the size of a horizontal grid on which model output was available. Spectral truncations are shown with "T numbers" referring to the triangular truncation wavenumber for various spectral models. Model documentation, references, and links can be found online at http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php.

Model	Resolution	Modeling center
CCCma CGCM3.1 T47	96 × 48 (T47)	Canadian Centre for Climate Modeling and Analysis
CCCma CGCM3.1 T63	128 × 64 (T63)	Canadian Centre for Climate Modeling and Analysis
CNRM-CM3	128 × 64 T42	Centre National de Recherches Météorologiques, Météo-France
CSIRO Mk3.0	192 × 96 T63	CSIRO Atmospheric Research, Australia
GFDL CM2.0	144 × 90	Geophysical Fluid Dynamics Laboratory
GFDL CM2.1	144 × 90	Geophysical Fluid Dynamics Laboratory
GISS-ER	72 × 46	Goddard Institute for Space Studies
IAP FGOALS-g1.0	128 × 60 T42	LASG/Institute of Atmospheric Physics, China
INM-CM3.0	72 × 45	Institute for Numerical Mathematics, Russia
IPSL CM4	96 × 72	L'Institut Pierre-Simon Laplace, France
MIROC3.2(hires)	320 × 160 T106	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan
MIROC3.2(medres)	128 × 64 T42	
MIUB ECHO-G	96 × 48 T30	Meteorological Institute of the University of Bonn, Germany; Meteorological Research Institute of KMA, Korea
MPI ECHAM5/MPI-OM	192 × 96 T63	Max-Planck-Institut für Meteorologie, Germany
MRI CGCM2.3.2	128 × 64 T42	Meteorological Research Institute, Japan
NCAR CCSM3	256 × 128 T85	National Center for Atmospheric Research
NCAR PCM1	128 × 64 T42	National Center for Atmospheric Research

c. *Extreme daily rainfall indices*

1) 30-YR RETURN LEVEL OF DAILY PRECIPITATION

A frequently used extreme precipitation index is the return level associated with a given time scale as derived from the estimated parameters of the generalized extreme value (GEV) distribution. Several recent studies have focused on model projections of future changes in the return level of daily precipitation (Kharin and Zwiers 2000; May 2004; Wehner 2004; Kharin et al. 2005, 2007). The theory and application of statistical modeling to analysis of extreme values using the GEV distribution are well established (e.g., Coles 2001). Similar to Kharin et al. (2005), we estimate the parameters for the GEV distribution using L moments (Hosking 1990) with 30-yearly maximum data from the 1961–90 period. Having fit the GEV distribution to a sample of annual extreme precipitation, the T-year return values X_T can be estimated from the quantile function (inverse of the cumulative distribution function) as

$$X_T = \begin{cases} \xi - \dfrac{\alpha}{\kappa}\left[1 - \left\{-\ln\left(1 - \dfrac{1}{T}\right)\right\}^{-\kappa}\right], & \text{if} \quad \kappa \neq 0; \\ \xi - \alpha \ln\left[-\ln\left(1 - \dfrac{1}{T}\right)\right], & \text{if} \quad \kappa = 0. \end{cases} \tag{1}$$

Here ξ is the location parameter, α is the scale parameter, and κ is the shape parameter of GEV distribution. We have selected the 30-yr return level (i.e., $T = 30$) as the main extreme precipitation indicator in this study. The same procedure was used for analyzing both the observed and model data. Other methods (e.g., maximum likelihood) can also be used to estimate these parameters. It is argued that the uncertainty and bias due to the use of different methods, as estimated by a Monte Carlo technique, is relatively small for this type of precipitation extreme statistic (Kharin and Zwiers 2005; Wehner 2004). Goodness-of-fit tests by a bootstrap procedure and the Kolmogorov–Smirnov test have been used to demonstrate that a GEV distribution fits the annual precipitation extremes satisfactorily (Karin and Zwiers 2000). Using the same procedure, we obtained similar outcome in the goodness of statistical model fit.

2) SIMPLE DAILY INTENSITY INDEX

Another extreme precipitation index often used in previous studies is the SDII (Semenov and Bengtsson 2002; Meehl et al. 2005; Tebaldi et al. 2006; Alexander et al. 2006). This is one of 27 indices recommended by the World Climate Research Programme/Climate Variability and Predictability (WCRP/CLIVAR) Expert Team on Climate Change Detection, Monitoring, and Indices (ETCCDMI). The SDII is defined as the mean intensity of daily rainfall of all wet days with daily precipitation exceeding 1 mm day^{-1}. This index is not as extreme as the 30-yr return level of daily precipitation derived from yearly maximum rainfall. Nevertheless, this "not as extreme" characteristic could be advantageous for model comparison studies (Tebaldi et al. 2006; Sun et al. 2006). The SDII, using the mean value from all wet days, might be less affected by strong spatial variability of individual rainfall events. However, the rain-day frequency from different spatial-scale data would still be affected and require some attention (Osborn and Hulme 1997; Sun et al. 2006). It is still of interest to examine how our assumption (point value versus areal mean) about the input gridded data affects the SDII when the grid size changes.

3. Data spatial scale and extreme precipitation indices

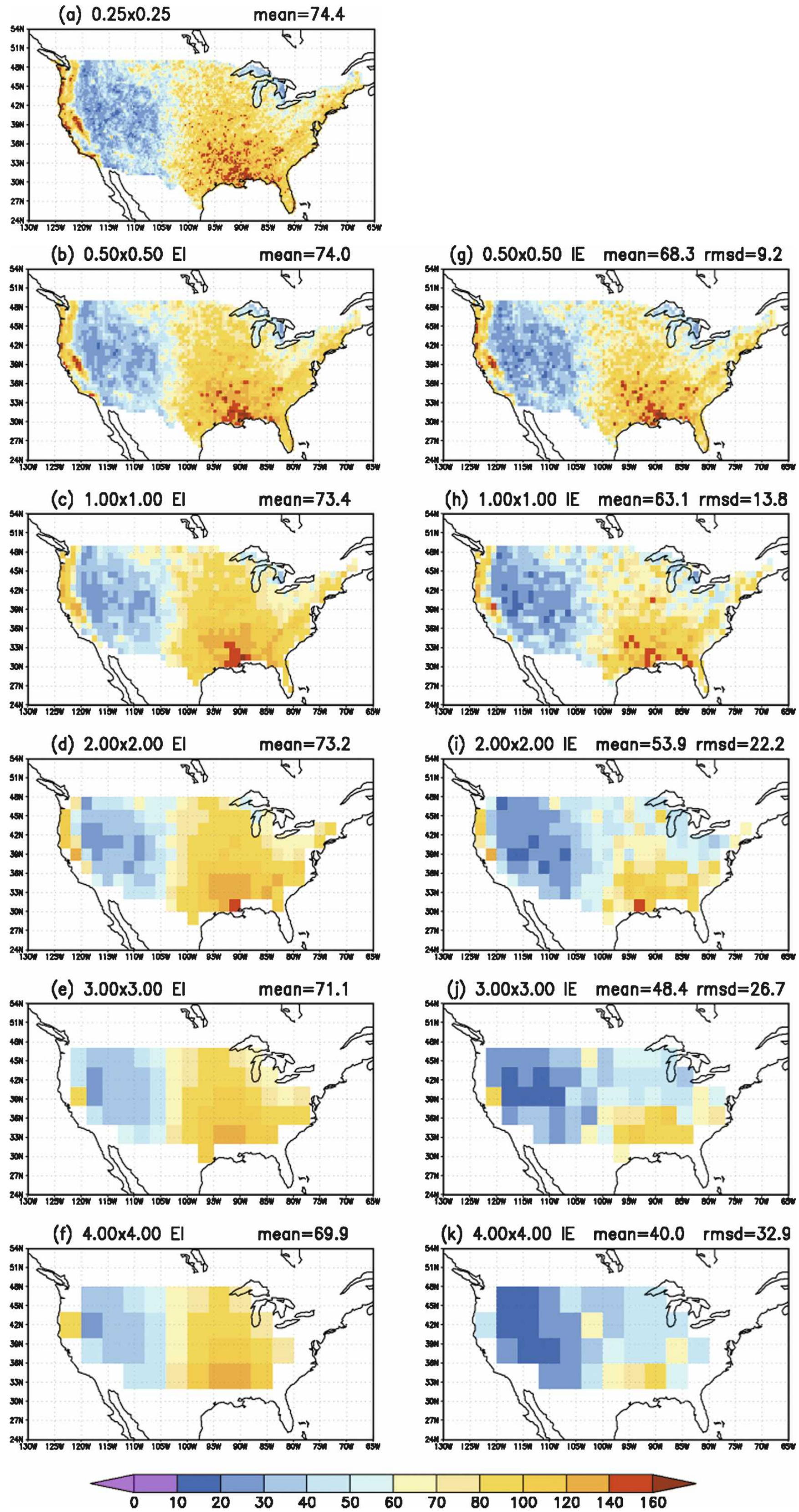
The differences resulting from the two assumptions (point value or areal mean) on gridded precipitation data can be explored by reversing the order of two operations (i.e., data interpolation and extreme indices calculation). We can simply interpolate the high-resolution observed analysis data (treating it as "perfect model" grid output) to illustrate the impact of these assumptions and how sensitive the impacts are to the spatial scale.

Earlier studies have discussed the precipitation scale issue. Sivapalan and Blöschl (1998) derived empirical areal reduction factors based on the statistical properties of the transformation from point rainfall to areal mean rainfall. For daily precipitation, they found that the factor mainly depends on the scaled catchment area and return period from parameters of the Gumble distribution. Booij (2002) further modified their expression for the more general case of the GEV distribution. Both studies assumed that gridded data represent areal mean precipitation. From their results, for a given basin area, the ratio between the area-averaged rainfall intensity over a duration D with return period T and the point rainfall intensity for the same D and T is smaller when the area size and return period are increased. This is expected since very extreme events are unlikely to occur simultaneously at different locations within a large grid domain. Many assumptions and parameters are involved in their empirically derived relationships. Therefore, we did not try to directly apply their relationships here. Further, we are not directly using station measurements but rather the high-resolution gridded.

Figure 1a shows the 30-yr return levels of daily precipitation (P_{30}) estimated from the original CPC Daily U.S. Precipitation dataset on a 0.25° × 0.25° grid. Figures 1b–f are the interpolation of Fig. 1a to 0.5° × 0.5°, 1° × 1°, 2° × 2°, 3° × 3°, and 4° × 4° grid boxes, respectively. The 30-yr return levels were interpolated to the larger grids *after* performing the extremes analysis. In contrast, Figs. 1g–k are P_{30} of the same resolution corresponding to Figs. 1b–f, but the interpolation of rainfall data to 0.5° × 0.5°, 1° × 1°, 2° × 2°, 3° × 3°, and 4° × 4° grid boxes is done *before* the calculation of GEV parameters and $P_{30.}$ For the case of different models with the same resolution from 0.5° to 4° grid, we would use Figs. 1b–f for the evaluation if we assume that the raw model outputs represent point estimates. However, since we argue that model precipitation output represents areal means, this implies we should use Figs. 1g–k for the evaluation. Some effects of spatial smoothing from interpolation are evident from Figs. 1a–f. Nevertheless, the domain-averaged return level remains approximately the same. When the same interpolation is applied to the daily precipitation first and then the return levels estimated based on 30-yearly maxima of the coarser grid precipitation data, the domain-averaged P_{30} is reduced from 74.4 to 68.3, 63.1, 53.9, 48.4, and 40.0 mm day^{-1} as the grid size is increased from 0.25° to 0.5°, 1°, 2°, 3°, and 4°, respectively. Also the domain-averaged root-mean-square (rms) difference between results using the two approaches increases to 9.2, 13.8, 22.2, 26.7, and 32.9 mm day^{-1}, respectively.

Figure 2 is a box plot of P_{30} obtained using the preferred approach (interpolation before extreme value analysis) at all U.S. grid points analyzed at various resolutions from 0.5° to 4°. They are all positively skewed due to the nature of the probability distribution of the daily precipitation. There is an obvious trend of a decrease in the distribution values (25th, 50th, 75th percentiles, etc.) when the data gridbox size is increased. Note that the sample size for the coarse-resolution data is much smaller, which might lead to some irregularity in the general trend.

The areal reduction factors (ARFs; defined as the ratio between area-averaged rainfall indices when extreme analysis is done before versus after the interpolation) for various resolutions are shown in Figs. 3a–e. These depict the ratio of Figs. 1g–k to 1b–f for each corresponding grid size. The ARFs are strongly dependent on the grid size. The different intensities of P_{30} in the eastern United States versus the western mountain region do not lead to a similar difference in the ARFs. The domain-averaged values are 0.91, 0.85, 0.72, 0.67, and 0.57 for grid size increases to 0.5°, 1°, 2°, 3°, and 4°, respectively. Similarly, the ARFs for SDII at different resolutions are shown in Figs. 3f–j. This is the ratio of SDIIs between two approaches, with the wet-day frequency (areal daily rainfall larger than 1 mm day^{-1}) analysis and SDII calculation applied either after or before the spatial interpolation of data. Even for a much "less extreme" index like SDII, the ARFs are reduced drastically as the grid size increases, with domain-averaged values of 0.93, 0.87, 0.79, 0.74, and 0.69


(a) 0.25x0.25 mean=74.4
(b) 0.50x0.50 EI mean=74.0
(c) 1.00x1.00 EI mean=73.4
(d) 2.00x2.00 EI mean=73.2
(e) 3.00x3.00 EI mean=71.1
(f) 4.00x4.00 EI mean=69.9
(g) 0.50x0.50 IE mean=68.3 rmsd=9.2
(h) 1.00x1.00 IE mean=63.1 rmsd=13.8
(i) 2.00x2.00 IE mean=53.9 rmsd=22.2
(j) 3.00x3.00 IE mean=48.4 rmsd=26.7
(k) 4.00x4.00 IE mean=40.0 rmsd=32.9
0 10 20 30 40 50 60 70 80 100 120 140 160

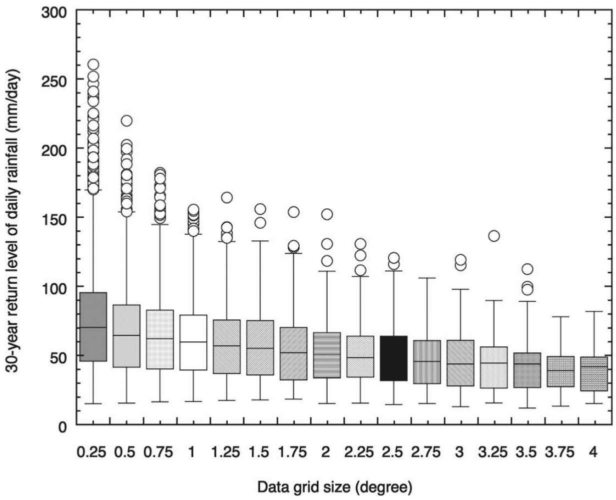


FIG. 2. Box plots of 30-yr return levels of P_{30} over all grid points in the conterminous United States at various grid box sizes from 0.25° to 4°. P_{30}s are calculated by interpolating the data to different resolutions before estimating P_{30} from parameters of the GEV distribution. Each box encloses 50% of the data with the median value of the variable displayed as a line within each box. The lines extending from the top and bottom of each box denote the minimum and maximum values within the dataset that fall within 1.5 times the interquartile distance (between upper and lower quartiles) from the median. Any value outside of this range, called an outlier, is displayed as an individual point.

for grid sizes of 0.5°, 1°, 2°, 3°, and 4°, respectively. The ARFs for SDII are larger on average than ARFs for P_{30}. This is as expected since SDII is not as extreme as P_{30}. The ARFs for SDII, analogous to P_{30} ARFs, are not directly linked to the magnitude of SDII and are primarily affected by the gridbox size.

4. Impact of different interpretations of model grid rainfall on model evaluation

For the evaluation of IPCC AR4 models over the U.S. domain, again we illustrate both approaches for deriving P_{30} from the CPC Daily U.S. Precipitation analyses for various model resolutions. It should be noted that there are model biases in rainfall simulation that are not related to the interpretation of model grid output. We do not try to address the sources of those model biases here. Rather, we emphasize here that model versus observation differences that arise from the spatial scale of the model grid (resolution) can be better accounted for in model evaluation studies if the (more appropriate) areal mean assumption is used.

Figure 4 shows the U.S. domain-averaged P_{30} calculated from CPC gridded observations and from years 1961 to 1990 of the 17 different IPCC AR4 climate model runs. The closed circles are estimated by first interpolating the observed rainfall data to various resolutions before conducting the extreme value analysis. These are consistent with the assumption that the model precipitation output represents areal means. The original observed rainfall analysis data at 0.25° × 0.25° resolution results in a domain averaged P_{30} of 74.4 mm day^{-1}. It should also be expected that the domain-averaged P_{30} would be higher than 74.4 mm day^{-1} if the extreme value analysis had been based on point measurements (station data), in view of the general trend of scale impacts and theoretical considerations (Sivapalan and Blöschl 1998).

Figure 4 also shows U.S. domain-averaged P_{30} for precipitation output from various models. The extreme indices were calculated based on each model's native grid data, consistent with the (not recommended) point estimate assumption if one made direct comparison among models. Comparing these model results to the observations under this assumption, most of the model would appear to underestimate the magnitude of $P_{30.}$ Of the 17 models, only the MIROC3.2(hires) (T106) run, the Max Planck Institute (MPI) ECHAM5/MPI Ocean Model (MPI-OM), and the L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4) produce present-day P_{30} values of similar magnitude to the high-resolution observed rainfall analysis (i.e., >70 mm day^{-1}). Note that one cannot exclude the possibility that the underestimation is mainly due to the common errors in model formulation or parameterization. But from the analyses in section 3, a certain part of underestimation can be attributed to the spatial-scale difference between models and observation.

If the model output are assumed to represent areal means (as we recommend), then we compare with the observed result obtained after regridding to the model's resolution, and in this case about half of models overestimate the domain-averaged P_{30}. Under this assumption, seven models simulate a reasonable range of P_{30}.

←

FIG. 1. Illustration of the impact of reversing the order of operations (data interpolation vs extreme indices calculation) for the estimate of 30-yr return levels of P_{30} for different grid sizes starting from the 0.25° × 0.25° CPC Daily U.S. Unified Precipitation dataset. (a) P_{30} estimated at original data resolution and interpolated to (b) 0.5°, (c) 1°, (d) 2°, (e) 3°, and (f) 4° grid boxes. P_{30} estimated with the data interpolation performed prior to calculation of P_{30} based on daily data interpolated to (g) 0.5°, (h) 1°, (i) 2°, (j) 3°, and (k) 4° grid boxes. The domain-averaged P_{30} and rms difference (rmsd) between the two approaches are shown in the upper-right corner of the panels. Unit is mm day^{-1}. Only grids with all land points in the original data are shown.

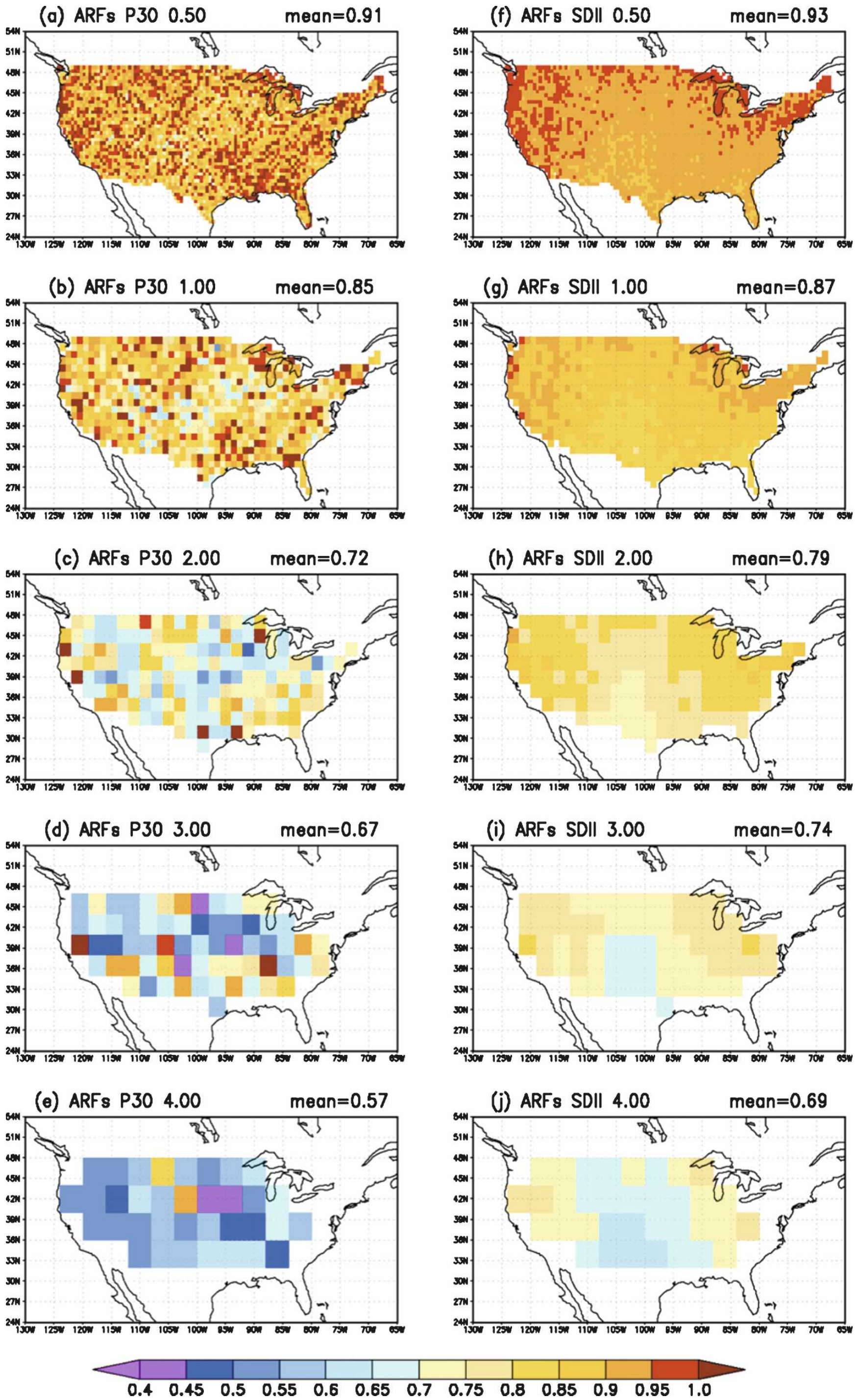


FIG. 3. The distribution of ARFs for 30-yr return levels of P_{30} at data grid sizes of (a) 0.5°, (b) 1°, (c) 2°, (d) 3°, and (e) 4°. The ARF is defined as the ratio of P_{30} estimated after vs before the data are interpolated from the original 0.25° resolution. ARF distributions for the SDII at data grid sizes of (f) 0.5°, (g) 1°, (h) 2°, (i) 3°, and (j) 4°. The domain-averaged ARFs are shown in the upper-right corner of each panel.

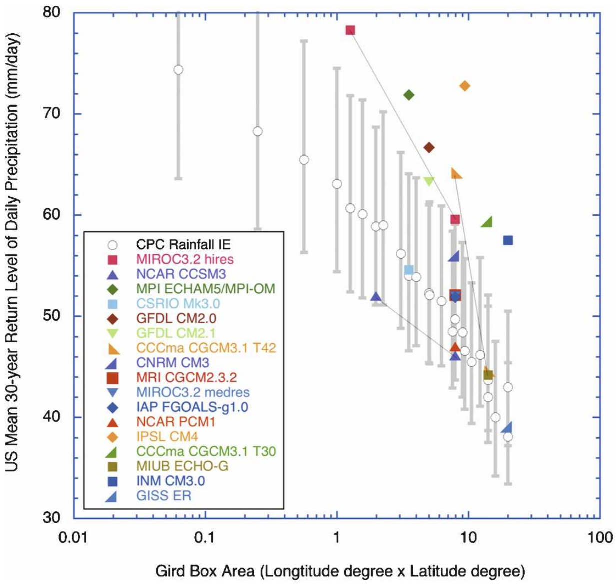


Fig. 4. Area-averaged 30-yr return levels of P_{30} over the conterminous United States for different grid box sizes as derived from CPC Daily U.S. Unified Precipitation and available IPCC AR4 models. For the CPC observations, the P_{30}s are calculated using data interpolation to different resolutions before estimating the P_{30} values. P_{30}s from all climate models are estimated at their native resolution. Additional P_{30} estimates are presented for the MIROC3.2(hires) (T106), NCAR CCSM3 (T85), and CCCma CGCM3.1 (T63) models using model data interpolated to T42, T42, and T30 grids, respectively, before the calculation of P_{30}. Solid lines link the pairs of P_{30} estimates obtained from these model runs. A log scale is used for the x axis (gridbox area). The CCCma CGCM3.1 T63 (T47) run used a 128 × 64 (96 × 48) global transform grid, the same transform grid as used in other T42 (T30) models. The error bars associated with CPC values are the U.S. area mean of the 5th and 95th percentiles across all the grid points. The percentiles are estimated at each grid point using a Monte Carlo resampling method (n = 1000) in which for each trial, 30-yearly maximum daily precipitation values are randomly selected from the 1956–95 period without replacement before calculating the 30-yr return level.

Only the National Center for Atmospheric Research (NCAR) Community Climate System Model version 3 (CCSM3) integration with T85 resolution produces a domain-averaged P_{30} much smaller than the observed value. It is clear that the two assumptions about the nature of the model precipitation output lead to totally different conclusions regarding the ability of the models to simulate extreme precipitation events.

To assess the significance of the differences between the models and observation, we use a Monte Carlo method to construct 1000 samples at each grid by randomly selecting 30-yearly maximum daily rainfall amounts out of the 40-yr period from 1956 to 1995 for the P_{30} calculation. The random resampling is done without replacement to ensure that the same yearly maximum daily rainfall is not sampled more than once. The U.S. domain average of the 5th and 95th percentiles of P_{30} value at each grid is used as a rough estimate of the possible range of U.S. mean P_{30} under this statistically unusual condition (all the grids are at the 5th or 95th percentiles of 1000 samples). The 5th and 95th percentiles are shown as the lower and upper error bars associated with P_{30} from the CPC observed analysis in Fig. 4. Such a measure of uncertainty does not allow for cancellation of overestimates and underestimates as occurred in almost all the models shown here. The 5th and 95th percentiles of the 1000 resampled U.S. domain-averaged P_{30} (allowing cancellation of differences among grid points due to data sampling) deviate only slightly from the median value (less than 0.5 and 1.5 mm day^{-1} for T106 and T30 resolutions, respectively).

Taking all the model results into consideration, there is some tendency for climate models to produce smaller P_{30} values at lower resolution. Kharin et al. (2005) also found a similar trend in the AMIP II models. This tendency provides more support for the view of treating model output as an areal mean.

For two of the climate models (CCCma CGCM3.1 and MIROC3.2), a pair of numerical simulations were run at both high and low resolutions. Although scale-dependent parameters can affect the model integrations and model tuning is often done to minimize the model biases, these simulation pairs still used basically identical dynamical and physical structures for each specific model. Comparing the rainfall extreme events simulated at these different resolutions could reveal more about the fundamental characteristics of model grid output. Therefore, in this section we will examine in more detail the spatial distribution of P_{30} over the United States from these simulations. Further, we can use the second approach again to interpolate the higher-resolution results to lower resolution before calculating the extreme indices. This attempts to illustrate the impact on the model simulated extreme events resulting solely from interpolation.

Figure 5 compares the MIROC3.2 model simulations of P_{30} with observations under different assumptions. Figures 5a–d use the point estimate assumption where we analyze the extreme statistics of the data on their native grid, then interpolate P_{30} obtained from observed data to T106 and T42 grids for comparison to the models. Under that assumption, the MIROC3.2 T106 run tends to overestimate extreme rainfall over the Rocky Mountain region and underestimates P_{30} in the southeastern United States (cf. Figs. 5a and 5c). The MIROC3.2 T42 run considerably underestimates P_{30} for the eastern United States while reasonable P_{30} values are simulated in the western United States, where

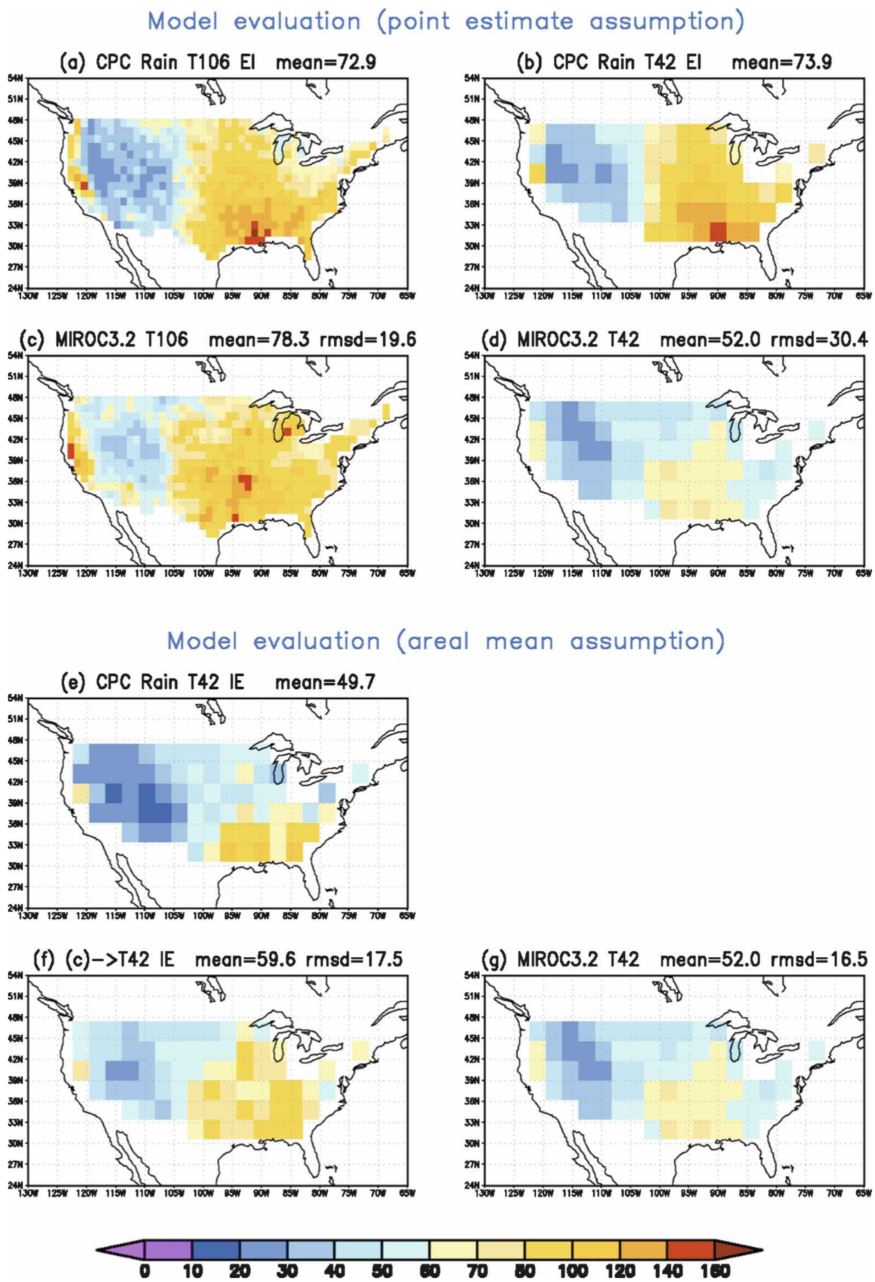


FIG. 5. Comparison of 30-yr return levels of P_{30} estimated from the CPC Daily U.S. Unified Precipitation dataset and the MIROC3.2 model simulations. P_{30}s estimated from observations on a 0.25° × 0.25° grid are interpolated to (a) T106 and (b) T42 model grids for comparison with the MIROC3.2 (c) T106 and (d) T42 runs. Based on data interpolated to a common T42 grid before estimation of P_{30}, (e) observed P_{30} estimates are compared to P_{30} estimated from the (f) T106 and (g) T42 runs. Domain-averaged P_{30} are reported in the upper-right corner of the panels. Average rmsds are also evaluated for the model data with respect to P_{30} estimates from observations. Unit is mm day^{-1}.

relatively weak extreme events are found in the observations (cf. Figs. 5b and 5d). Again, the difference shown here is not only due to the less appropriate assumption but could be caused by model bias. If the (recommended) areal-average interpretation is used and both observed and simulated daily data are interpolated to T42 resolution before calculating the P_{30} (Figs. 5e–g), the MIROC3.2 T106 run still overestimates P_{30} except for a few areas in the Southeast (cf. Figs. 5e and 5f). Similar model biases are found in the MIROC3.2 T42 run (cf. Figs. 5e and 5g), except the general overestimation is smaller and the underestimate of P_{30} in the Southeast is not as large as with the first approach (cf. Figs. 5b and 5d). Assuming the model outputs represent areal means, the rms differences for the T106 and T42 runs are comparable at 17.5 and 16.5 mm day^{-1}, respectively. Note that after interpolating all data to the same grid size, we regard the model biases (compared lower panels of Fig. 5) as mainly arising from model formulation and not from the impact of the data spatial scale.

When the T106 data are interpolated to T42 prior to computing P_{30}, the results are fairly similar to those from the original model run with T42 resolution. Part of the P_{30} differences between the original data output from the high- and medium-resolution model runs (Figs. 5c versus 5d) can therefore be attributed to this spatial-scale effect under the second assumption. The assumption that model output represents a point estimate is inconsistent with the results from MIROC3.2, since the precipitation parameterization, if designed as a gridpoint process, should be expected to generate similar rainfall regardless of model grid size, which it does not. The CCCma CGCM3.1 uses a different approach in the calculations of model's physical tendencies (including precipitation). Both T63 and T47 versions of CCCma CGCM3.1 overestimate P_{30} under the areal mean assumption (Fig. 4). The overestimations are found in most of the United States except the Southeast where larger observed P_{30} values are found. If we interpolate the CCCma T63 (128×64) daily precipitation to 96×48 grids before calculating the P_{30}, the United States domain-averaged P_{30} become very similar to observed P_{30}. However, it is much smaller than the P_{30} from the CCCma CGCM3.1 T30 run. Interestingly, the behavior shown by the CCCma CGCM3.1 at different resolutions seems more consistent with the assumption of model output as a point estimate according to this analysis (i.e., the P_{30} estimates do not seem to depend very strongly on the resolutions), although we do not regard this as a strong argument for using the point estimate assumption. We also reinterpolated the T85 NCAR CCSM3 run to T42 resolution before calculating the extremes. A reduction of P_{30} upon interpolation from higher- to lower-resolution data again occurs, and the area-averaged P_{30} is similar to that of the NCAR Parallel Climate Model version 1 (PCM1) model run at T42 resolution. Although the CAM3 and CCM3 atmospheric models used in these two coupled models are distinct, part of the difference between NCAR CCSM3 and NCAR PCM1 could be simply due to the spatial-scaling effects of the model data.

Attempts to use different model runs to form a model consensus or an estimate of the range of model uncertainty (Hegerl et al. 2004; Kharin et al. 2005, 2007; Tebaldi et al. 2006) are complicated by these alternative interpretations as discussed in the introduction. Based on our survey of past studies, there are sometimes inconsistencies in handling the observed and simulated data and in performing extremes analysis when multiple models are involved. Based on our analysis, these inconsistencies could lead to the inclusion of artifacts from spatial-scaling effects during the data aggregation. These impacts should be carefully considered in the model evaluation and in comparison studies aimed at examining the impacts of different model formulations and diversity of precipitation related parameterizations.

Using the IPCC data archive of extreme indices calculated from various AR4 model output, Meehl et al. (2005) and Tebaldi et al. (2006) use the interannual standard deviation over the simulation period (after detrending) to standardize the time series of extreme indices and their changes before model aggregation. This approach can adjust for different absolute magnitudes of the simulated indices among the different models. It is not affected by the spatial-scaling effects when the assumption that model output represents a point estimate is used. However, if model data are interpreted as areal means, the impact of the data gridbox size on the intensity of indices is not necessarily equal to the impact on their interannual standard deviation. For example, we found that the reduction of standard deviation of annual SDII in the 1948–98 period after detrending, when the data are first interpolated to 4° before computing the SDII, is about 51% averaged over the U.S. domain (Fig. 6). The area-averaged ARF for SDII over the United States is 31%. These results suggest that the analyses by Meehl et al. (2005) and Tebaldi et al. (2006) would have obtained a smaller standardized change in model ensemble mean SDII than the change estimated if they had interpreted the model data as areal means (first interpolating all the models to the lowest possible model resolution before calculating the SDII). Again this could occur as an artifact of the effect of data grid scale alone, since only observed data are used to produce the effect in Fig. 6.

It is not practical to evaluate different resolution models with a different version of "regridded" observation. Therefore, our recommendation is to interpolate all the model gridded data to lowest possible resolution before extreme precipitation analysis and model verification (or comparison).

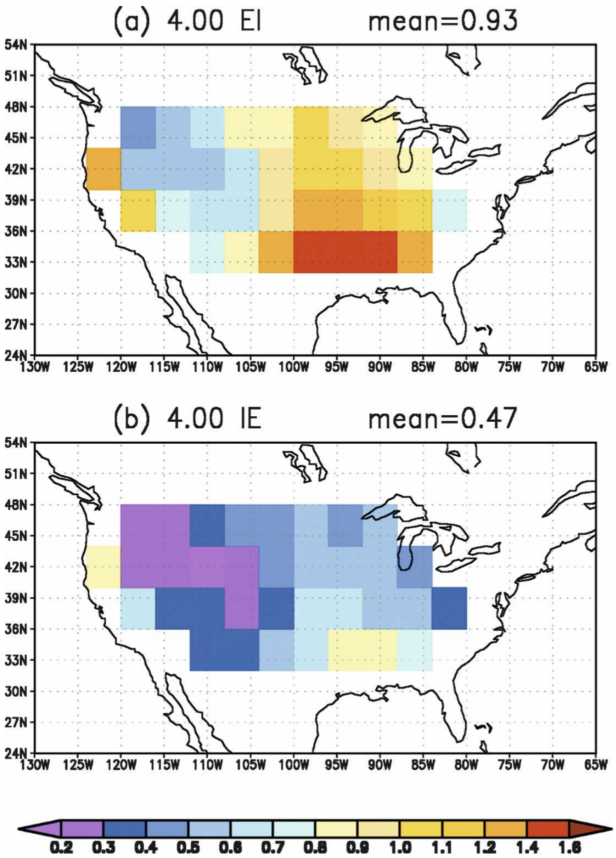


FIG. 6. The interannual std dev of the SDII from the period 1948–98. (a) The SDII is derived from the original 0.25° × 0.25° resolution data and then interpolated to 4° × 4° grid boxes. (b) The data are interpolated to 4° × 4° grid boxes before the calculation of the SDII for individual years. Unit is mm day^{-1}.

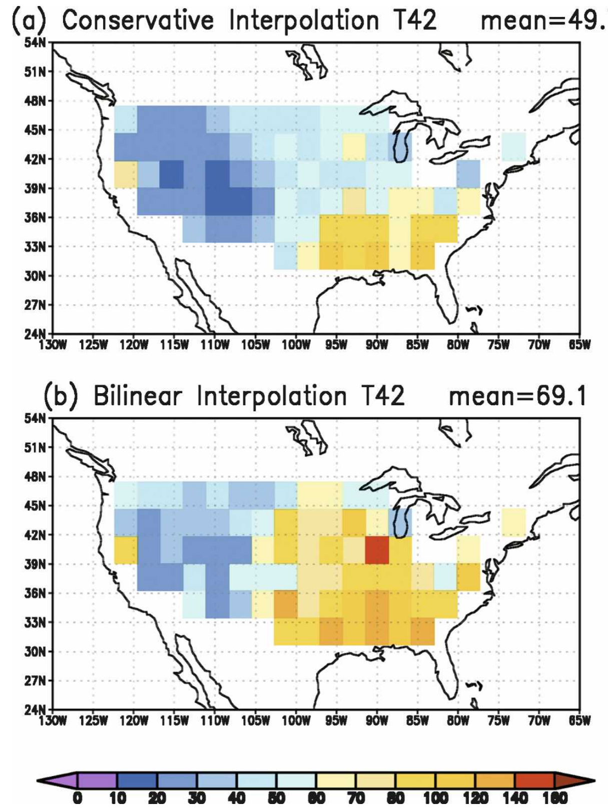


FIG. 7. The 30-yr return levels of P_{30} calculated from the CPC Daily U.S. Unified Precipitation data interpolated to a T42 grid with (a) conservative remapping or (b) bilinear remapping schemes before the estimation of P_{30}. Domain-averaged P_{30} are reported in the upper-right corner of the panels. Unit is mm day^{-1}.

5. Impact of spatial interpolation scheme

Another relevant issue for extreme rainfall analysis at different spatial resolutions is the remapping scheme used for interpolation. A conservative remapping scheme was used in our study (Jones 1999). We did not use nonconservative remapping schemes because they are not consistent with an "areal average" assumption for model output. Nevertheless, one should also be aware of the possible impact of different interpolation methods on extreme rainfall analyses. When a nonconservative remapping scheme (e.g., bilinear, bicubic, and distance weighted) is used for interpolation from a fine grid to a coarse grid, only the input grids nearest to the output grids are involved in the interpolation. This allows for the possibility of more extreme local values being obtained on the output grid in comparison to the additional smoothing involved with the conservative scheme. Figure 7 compares P_{30} estimated by applying different interpolation schemes (conservative versus bilinear), before the estimation of P_{30}. The results show that higher P_{30} values are obtained when this nonconservative interpolation scheme is used. Other nonconservative remapping schemes, such as bicubic and distance-weighted average, lead to similar results to bilinear interpolation. In conclusion, if one uses a nonconservative interpolation scheme, such a procedure can also have an important impact on extreme value statistics, analogous to the order of operations. However, even though nonconservative schemes can produce higher extreme values, their use would appear to be more justifiable in the context of the point estimate paradigm, which is not recommended here.

6. Discussion and conclusions

The assessment of the precipitation extreme indices, in syntheses of multiple climate models, may be strongly affected by the assumption of whether the model grid data represent point estimate or areal averages. We argue that the areal average assumption is more appropriate than the point estimate assumption, which is not recommended. Using the high-resolution CPC Daily U.S. Unified Precipitation and adopting different assumptions about the gridded data (i.e., on the order of the data interpolations and extremes analysis), we highlight the possible impact of these assumptions using high-resolution observations as perfect model data. The intensity of extreme precipitation indices at reduced resolution is much weaker when gridded data are assumed to be areal means. The reduction increased as the grid size increases. The value of ARFs due to areal mean assumptions is sensitive to the extreme indices selected (e.g., P_{30} versus SDII). ARFs are fairly insensitive to the geographical location and the magnitude/distribution of indices. We also recommend that the areal average assumption be used for the model data in model intercomparison studies. Under the areal mean assumption, extreme rainfall events simulated from different models with distinct resolutions should be assessed using observations conservatively interpolated to the same resolution prior to computing extreme statistics. Even though the current generation of climate models has typical spatial resolution that many regard as less than ideal for simulation of extreme events, as compared to higher-resolution regional climate model simulation (Jones and Reid 2001; Räisänen and Joelsson 2001; Fowler et al. 2005; Frei et al. 2006), one can still undertake an assessment of the coarse-resolution models, although our results indicate that appropriate consideration of the spatial scale of the validation data be incorporated.

Using the areal mean assumption, the calculation of multimodel ensemble mean and comparisons of different model simulation of extreme rainfall should be done at the lowest model resolution so that all the data can be interpolated to a common grid before computation of extreme indices. Also a conservative interpolation scheme should be used to maintain the same areal-averaged rainfall before and after the remapping. Our results also imply that attempts to normalize the variability of simulated extreme precipitation indices from different models may also be affected by these issues associated with assumptions about interpretation of model outputs. The direct comparison and ensemble means of derived extreme indices obtained from different models (at their native resolution) in the current IPCC data archive are therefore not possible unless one assumes that the model output represents point estimates, which is not recommended here. One might be able to determine an appropriate spatial-scaling factor for downscaling the lower-resolution model data to higher-resolution model data (Booij 2002). However, one typically needs long-term station or gridded daily precipitation data with near-global coverage to derive the empirical relationship. Regional climate models forced by AOGCMs can dynamically downscale extreme precipitation events to a spatial scale that is more comparable with observed rainfall analysis or station data. The scaling issues identified in the present study should be reduced in this case. There are, however, other issues associated with such regional modeling approaches, including, for example, the nesting method, the lack of two-way ocean coupling, and the consistency of the climate sensitivity and physical parameterizations used in the global versus regional models.

Many observed analyses of past trends in extreme weather and climate events use station data as opposed to gridded data (Frich et al. 2002; Kunkel et al. 2003; Zhai et al. 2005). Unfortunately, these can be readily compared with model data only under the point estimate assumption. Other studies have produced gridded data of observed rainfall extreme indices (or trends in indices) that have been used, or could be used for comparison with models (e.g., Osborn and Hulme 1998; Kiktev et al. 2003; Alexander et al. 2006). In these cases the gridding methods were different from the present study. Typically the extreme indices (time series) were derived using observed station data as a basis. The statistical structure of spatial correlations among neighboring station time series was used to construct the weighting functions for combining the station data into grid boxes (Kiktev et al. 2003; Alexander et al. 2006). As discussed in the introduction, the gridding and interpolation are actually applied after the extreme analysis. Therefore, although the results by Alexander et al. (2006) are gridded extreme rainfall indices derived from observations, it is not clear that their gridded data are suitable for model evaluation without further consideration of the spatial-scale issues identified here. Figure 8 compares the averaged yearly maximum daily precipitation during the 1961–90 period over the United States from the CPC daily precipitation at $0.25° \times 0.25°$ resolution and the gridded HadEX data from Alexander et al. (2006) at $3.75° \times 2.5°$ resolution (available online at http://hadobs.metoffice.com/hadex/). It is apparent that the HadEX data, despite their relatively coarse resolution, show larger yearly maximum daily rainfall values than the CPC daily rainfall. This clearly indicates that the intensities of gridded yearly maximum daily precipitation from HadEX are more comparable to that of station data. Nevertheless, one should note that the gridding methodology in Alexander et al. (2006) is not aimed at creating area averages but rather at reducing errors at the interpolated points by proper weighting of surrounding station data.

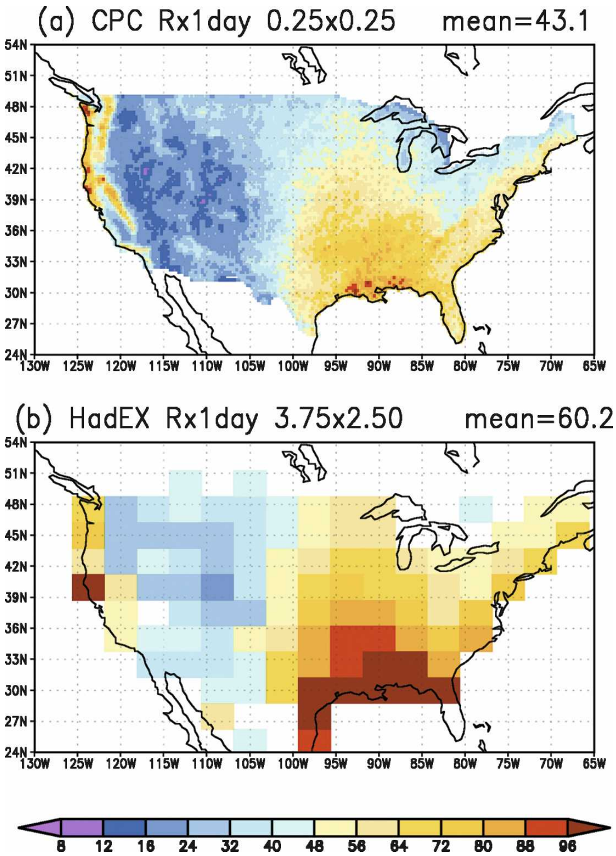


FIG. 8. The averaged yearly maximum daily precipitation for the 1961–90 period from the (a) 0.25° × 0.25° CPC Daily U.S. Unified Precipitation and (b) 3.75° × 2.5° HadEX extreme climate indices dataset. Unit is mm day^{-1}.

We use high-resolution CPC Daily U.S. Unified Precipitation and IPCC AR4 model simulations to highlight the scale dependence of extreme precipitation when different assumptions about the data (i.e., area average versus point estimate) are made for the analysis. Even a less extreme index like the SDII is sensitive to such assumptions. Although our results are based only on U.S. regional data, we speculate that similar effects will occur for other regions of the world. The different daily rainfall characteristics in the tropics could conceivably lead to quantitatively varying results, but lack of long-term daily gridded rainfall data hinders assessment of these issues in the tropical regions at this time. Extreme temperature indices, with typical decorrelation lengths at least several times larger than that of extreme precipitation indices (Alexander et al. 2006), likely exhibit only a minor effect. They are not discussed in this note.

Acknowledgments. We acknowledge the developers of the CPC U.S. Unified Precipitation data, which we obtained from the NOAA/CIRES/ESRL PSD Climate Diagnostics Branch, Boulder, Colorado, from their Web site at http://www.cdc.noaa.gov/. The HadEX gridded extreme climate indices data are available online at http://hadobs.metoffice.com/hadex/, supported by the Met Office Hadley Centre for Climate Change and WMO CCl/CLIVAR ETCCDMI. We also thank the numerous international modeling groups for providing their IPCC AR4 model data for analysis, the PCMDI for collecting and archiving the model data, and the IPCC Data Archive at Lawrence Livermore National Laboratory supported by the Office of Science, U.S. Department of Energy. Also, we wish to thank two anonymous reviewers, John Lanzante, and Sergey Malyshev for helpful and cogent comments on an initial draft. The corresponding author would like to thank the visiting scientist program of the Atmospheric and Oceanic Sciences program at Princeton University and the Geophysical Fluid Dynamics Laboratory (GFDL) for providing facilities needed for this work. The visit of the corresponding author to GFDL was also supported by the National Science Council of Taiwan and National Taiwan Normal University.

REFERENCES

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.,* **111,** D05109, doi:10.1029/2005JD006290.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment. Part I. *J. Atmos. Sci.,* **31,** 674–701.

Booij, M. J., 2002: Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. *Int. J. Climatol.,* **22,** 69–85.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate,* **19,** 4605–4630.

Easterling, D. R., J. L. Evans, P. Ya. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje, 2000a: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.,* **81,** 417–425.

——, G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, 2000b: Climate extremes: Observations, modeling, and impacts. *Science,* **289,** 2068–2074.

Emori, S., A. Hasegawa, T. Suzuki, and K. Dairaku, 2005: Validation, parameterization dependence, and future projection of daily precipitation simulated with a high-resolution atmospheric GCM. *Geophys. Res. Lett.,* **32,** L06708, doi:10.1029/2004GL022306.

Folland, C. K., and Coauthors, 2001: Observed climate variability and change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

Fowler, H. J., M. Ekström, C. G. Kilsby, and P. D. Jones, 2005: New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. *J. Hydrol.,* **300,** 212–233.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale, 2006: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *J. Geophys. Res.,* **111,** D06105, doi:10.1029/2005JD005965.

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

Groisman, P. Ya., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev, 2005: Trends in intense precipitation in the climate record. *J. Climate,* **18,** 1326–1350.

Hegerl, G. C., F. W. Zwiers, P. A. Stott, and V. V. Kharin, 2004: Detectability of anthropogenic changes in annual temperature and precipitation extremes. *J. Climate,* **17,** 3683–3700.

Hennessy, K. J., J. M. Gregory, and J. F. B. Mitchell, 1997: Changes in daily precipitation under enhanced greenhouse conditions. *Climate Dyn.,* **13,** 667–680.

Hewiston, B. C., and R. G. Crane, 2005: Gridded area-averaged daily precipitation via conditional interpolation. *J. Climate,* **18,** 41–57.

Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, 2000: *Improved United States Precipitation Quality Control System and Analysis.* NCEP/Climate Prediction Center Atlas 7, 40 pp. [Available online at http://www.cpc.ncep.noaa.gov/research_papers/ncep_cpc_atlas/7/index.html.]

Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.,* **B52,** 105–124.

Iorio, J., P. B. Duffy, B. Govindasamy, S. L. Thompson, M. Khairoutdinov, and D. Randall, 2004: Effect of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States. *Climate Dyn.,* **23,** 243–258.

Jones, P. D., and P. A. Reid, 2001: Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *Int. J. Climatol.,* **21,** 1337–1356.

Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.,* **127,** 2204–2210.

Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate,* **13,** 3760–3788.

——, and ——, 2005: Estimating extremes in transient climate change simulations. *J. Climate,* **18,** 1156–1173.

——, ——, and X. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate,* **18,** 5201–5223.

——, ——, ——, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate,* **20,** 1419–1444.

Kiktev, D., D. M. H. Sexton, L. Alexander, and C. K. Folland, 2003: Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate,* **16,** 3560–3571.

Kunkel, K. E., D. R. Easterling, K. Redmond, and K. Hubbard, 2003: Temporal variations of extreme precipitation events in the United States: 1895–2000. *Geophys. Res. Lett.,* **30,** 1900, doi:10.1029/2003GL018052.

May, W., 2004: Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. *Climate Dyn.,* **22,** 183–204.

McAvaney, B. J., and Coauthors, 2001: Model evaluation. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 471–523.

McCollum, J. R., and W. F. Krajewski, 1998: Uncertainty of monthly rainfall estimates from rain gauges in the Global Precipitation Climatology Project. *Water Resour. Res.,* **34,** 2647–2654.

McGuffie, K., A. Henderson-Sellers, N. Holbrook, Z. Kothavala, O. Balachova, and J. Hoekstra, 1999: Assessing simulations of daily temperature and precipitation variability with global climate models for present and enhanced greenhouse climates. *Int. J. Climatol.,* **19,** 1–26.

Meehl, G. A., F. W. Zwiers, J. Evans, T. Knutson, L. Mearns, and P. Whetton, 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.

——, J. M. Arblaster, and C. Tebaldi, 2005: Understanding future patterns of increased precipitation intensity in climate model simulation. *Geophys. Res. Lett.,* **32,** L18719, doi:10.1029/2005GL023680.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.,* **120,** 978–1002.

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

——, and ——, 1998: Evaluation of the European daily precipitation characteristics from the atmospheric model intercomparison project. *Int. J. Climatol.,* **18,** 505–522.

Räisänen, J., and R. Joelsson, 2001: Changes in average and extreme precipitation in two regional climate model experiments. *Tellus,* **53A,** 547–566.

Semenov, V. A., and L. Bengtsson, 2002: Secular trends in daily precipitation characteristics: Greenhouse gas simulation with a coupled AOGCM. *Climate Dyn.,* **19,** 123–140.

Sivapalan, M., and G. Blöschl, 1998: Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. *J. Hydrol.,* **204,** 150–167.

Skelly, W. C., and A. Henderson-Sellers, 1996: Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers? *Int. J. Climatol.,* **16,** 1079–1086.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate,* **19,** 916–934.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change,* **79,** 185–211.

Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.,* **121,** 3040–3061.

Voss, R., W. May, and E. Roeckner, 2002: Enhanced resolution modelling study on anthropogenic climate change: Changes in extremes of the hydrological cycle. *Int. J. Climatol.,* **22,** 755–777.

Watterson, I. G., and M. R. Dix, 2003: Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. *J. Geophys. Res.,* **108,** 4379, doi:10.1029/2002JD002928.

Wehner, M. F., 2004: Predicted twenty-first-century changes in seasonal extreme precipitation events in the parallel climate model. *J. Climate,* **17,** 4281–4290.

Zhai, P., X. Zhang, H. Wan, and X. Pan, 2005: Trends in total precipitation and frequency of daily precipitation extremes over China. *J. Climate,* **18,** 1096–1108.

Zwiers, F. W., and V. V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. *J. Climate,* **11,** 2200–2222.