

Sources of uncertainty in the extreme value statistics of climate data

Michael Wehner

Received: 1 April 2009 / Revised: 27 January 2010 /
Accepted: 11 February 2010 / Published online: 25 February 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract We investigate three sources of uncertainty in the calculation of extreme value statistics for observed and modeled climate data. Inter-model differences in formulation, unforced internal variability and choice of statistical model all contribute to uncertainty. Using fits to the GEV distribution to obtain 20 year return values, we quantify these uncertainties for the annual maximum daily mean surface air temperatures of pre-industrial control runs from 15 climate models in the CMIP3 dataset.

Keywords Extreme temperature · Return value · Uncertainty · Climate models

AMS 2000 Subject Classification 62G32

1 Introduction

Although not specifically designed to reproduce extremes of temperature and precipitation, general circulation models are increasingly being used to assess the climate change risk associated with such extreme events (Karl et al. 2008). Proper assessment of this risk depends on accurately quantifying both projected changes in extreme events and the uncertainty behind these projections. In this paper, we will examine several sources of uncertainty in the description of climate model produced extreme value statistics. The models in the study are from the CMIP3 database of coupled climate models (Meehl et al. 2007) used in the preparation of the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC). We focus on surface air temperature extremes in this paper as our recent work suggests that the

M. Wehner (✉)
Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., MS-50F, Berkeley, CA 94720, USA
e-mail: mfwehner@lbl.gov

next generation of climate models will much better simulate precipitation extremes than the CMIP3 models due to increases in horizontal resolution (Wehner et al. 2009). We expect that surface air temperature extremes will also be better simulated in the next generation of models, but that the effect of the higher resolution will be less pronounced than it is for precipitation.

Block-maxima formulations of Generalized Extreme Value (GEV) theory have been shown to adequately capture the behavior of both climate model and observed surface air temperature and have been used to make projections of future changes (Zwiers and Kharin 1998; Kharin and Zwiers 2000; Wehner 2005; Kharin et al. 2007). In these formulations, seasonal or annual maxima are extracted from the parent distribution of all daily mean surface air temperatures forming a subsidiary set of random variables. Under very general assumptions regarding the parent distribution, values in its tail must obey well defined distribution functions (Leadbetter et al. 1983; Coles 2001; Castillo et al. 2004). In the block maxima approach described here, that distribution is the three parameter GEV distribution, $F(x)$,

$$F(x) = \begin{cases} e^{[1-k(x-\xi)/\alpha]^{1/k}} & k \neq 0 \\ e^{-e^{-(x-\xi)/\alpha}} & k = 0 \end{cases} \quad (1)$$

where ξ , α and k are called the location, scale and shape factors. The Gumbel distribution is a special case where the shape parameter, k , is zero. Formally, $F(x)$ is the limiting cumulative distribution function of the block maxima of a sample of independently and identically distributed random variables (Leadbetter et al. 1983). The three parameters of the GEV distribution may be estimated from a sample of extreme values by a variety of techniques. In this study, a technique based on L-moments (Hosking and Wallis 1997) is deemed preferable to the more commonly used maximum likelihood estimator (MLE) techniques due to its better efficiency in estimating the GEV parameters for the short samples typical of climate datasets (Hosking et al. 1985; Coles and Dixon 1999; Kharin et al. 2007). A comparison by Kysely (2002) of the L-moments and MLE techniques revealed no significant differences when compared to a limited set of *stationary* 20 year temperature datasets. However, this may not be the case for non-stationary datasets as Kharin and Zwiers (2005) found significant biases when the L-moments method are used in such instances. However, in this study the datasets are forced to be stationary and the GEV parameters can be accurately calculated with confidence using the LMOMENTS package supplied by Hosking (2003). Furthermore, goodness-of-fit tests performed first by Kharin and Zwiers (2000) on an early version of the Canadian model (see Table 1) and later by Kharin et al. (2007) on the transiently forced runs in the CMIP3 database of climate model integrations reveal that the GEV distribution adequately describes both annual temperature and precipitation extremes. They note only for precipitation in certain overly dry biased models is the GEV distribution fit poor and make no statements about poor fit of temperature extremes for any models. Since only the stationary runs in the CMIP3 database are used in this study, one can generally assume that the GEV distribution is a good fit to this somewhat simpler case. This issue is explored further in the [Appendix](#).

Table 1 The climate models used in this study

CMIP3 name	Institution
bccr_bcm2_0	Bjerknes Centre for Climate Research, Bergen, Norway
ccma_cgcm3_1	Canadian Centre for Climate Modelling and Analysis, Victoria, Canada
ccma_cgcm3_1_t63	Canadian Centre for Climate Modelling and Analysis, Victoria, Canada
cnrm_cm3	Centre National de Recherches Meteorologiques, Toulouse, France
csiro_mk3_0	Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia
csiro_mk3_5	Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia
gfdl2.0	Geophysical Fluid Dynamics Laboratory, Princeton, USA
gfdl2.1	Geophysical Fluid Dynamics Laboratory, Princeton, USA
giss_aom	Goddard Institute for Space Studies, New York, USA
giss_model_e_h	Goddard Institute for Space Studies, New York, USA
giss_model_e_r	Goddard Institute for Space Studies, New York, USA
ingv_echam4	National Institute of Geophysics and Volcanology, Bologna, Italy
ipsl	Institut Pierre Simon Laplace, Paris, France
miroc_hires	Center for Climate System Research, Tokyo, Japan/National Institute for Environmental Studies, Ibaraki, Japan/Frontier Research Center for Global Change, Kanagawa, Japan
miroc_medres	Center for Climate System Research, Tokyo, Japan/National Institute for Environmental Studies, Ibaraki, Japan/Frontier Research Center for Global Change, Kanagawa, Japan
miub_echo	University of Bonn, Bonn, Germany
mri_cgcm2_3_2a	Meteorological Research Institute, Tsukuba, Ibaraki, Japan
ncar_ccsm3_0	National Center for Atmospheric Research, Boulder, USA
ncar_pcm1_0	National Center for Atmospheric Research, Boulder, USA

Data is available from www.pcmdi.llnl.gov

Once the GEV parameters have been determined, the return value for a given period, T , can be readily calculated. In the current context, the return value can be thought of as that value of daily mean surface air temperature that can be expected to be realized on average every T years over a very long period of time in a stationary climate. Alternatively, the return value is expected to occur in any single year with a probability of $1/T$. Formally, this is straightforwardly defined as

$$F(X_T) = 1 - 1/T \quad (2)$$

Solving for X_T using the definition of the GEV distribution yields,

$$X_T = \begin{cases} \xi + \alpha \left[1 - \{-\ln(1 - 1/T)\}^k \right] / k & k \neq 0 \\ \xi - \alpha \ln(-\ln(1 - 1/T)) & k = 0 \end{cases} \quad (3)$$

Hence, return values of annual maximum daily mean surface air temperatures are readily obtained by this inversion of the GEV distribution after its three parameters have been estimated.

The interested reader is referred to the textbooks by Coles (2001) and Castillo et al. (2004) for further details of the GEV formulation as well as justification for using the technique on block maxima modeled and observed climate data.

The length of climate records is almost always an issue and particularly complicates a statistical description of rare events. The observed and simulated climate of the past and future centuries are not stationary due to anthropogenic forcing. For the early part of the twentieth century, anthropogenic climate change is slow enough that a decade or two may be considered quasi-stationary (Easterling and Wehner 2009). However, this is a relatively short period in the context of quantification of long term return values. For model-based analyses, short periods from the different realizations of an ensemble of independent and identically distributed calculations can be combined to generate a longer dataset to improve the statistical confidence in the analysis. However, there is only one realization of the real world which limits confidence in any fit of observed extreme value distributions. For later in the twentieth century and any credible scenario of the twenty-first century, decadal temperature trends are significantly biased to be positive (Easterling and Wehner 2009). Either of two approaches can address this issue: detrending within the analysis period ensuring stationarity or estimating non-stationary extreme value distributions allowing for time dependent parameters. This latter approach has the advantage that trend and distribution parameters are estimated at the same time, reducing estimation errors (Coles 2001). On the other hand, usage of the L-moments technique to estimate the GEV parameters assumes stationarity of the annual maxima. Kharin and Zwiers (2005) demonstrated that violation of this assumption may introduce biases in the return value comparable in magnitude to sampling variance. Most of the CMIP3 models archived only 40 years of daily averaged quantities for most categories of simulations. The uncertainties investigated in this paper, mostly using 40 years of pre-industrial control run data, is comparable to concatenating detrended single decades of four independent model realizations.

2 Uncertainty from model formulation

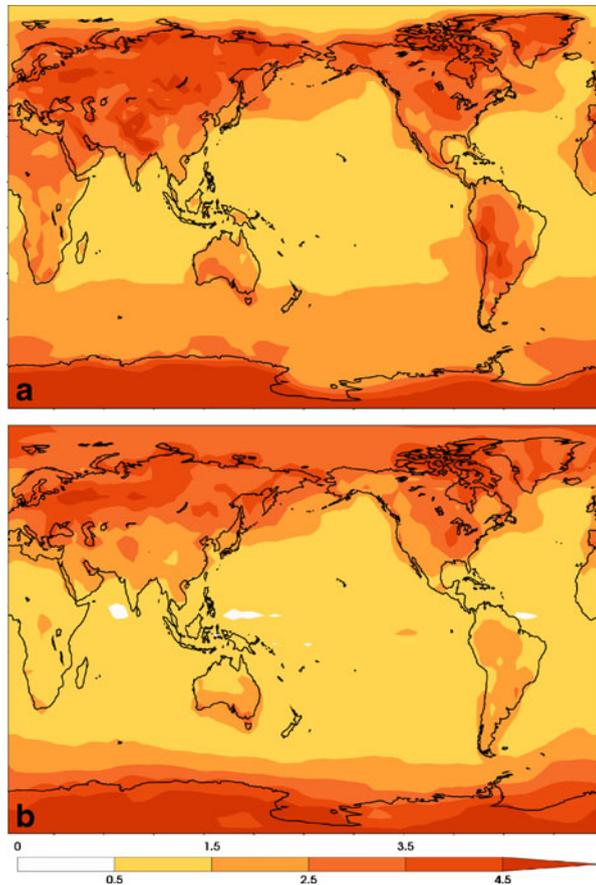
Often the largest source of uncertainty in model projections of future climate comes from differences in the models themselves. Climate models reproduce the actual climate with varying degrees of success. Both known and unknown biases can be large and structured in different ways for different models. Ascertaining model bias can be difficult on the global scale due to observational deficiencies. Robust measurements of extreme values are hampered both by an incomplete spatial coverage as well as the limited duration of the observational record. Furthermore, to the extent that climate models can be tuned to reproduce the recent past, model developers focus on the mean values of climatic observations, not the tails of their distributions. To quantify this model uncertainty, we utilized the pre-industrial (picntrl) control run integrations in the CMIP3 database of climate model simulations to calculate the 20-year return

value of the annual maxima of daily surface air temperature. In control run integrations, all atmospheric trace constituents, including greenhouse gases, are held at constant concentrations. In the absence of climate model drift, these simulations are stationary, lending themselves to straightforward application of the GEV theoretical constructs. We considered 15 different models, listed in Table 1. Most of these modeling groups archived daily averaged quantities for only 40 years, as specified by the CMIP3 protocols. For those models where longer control run datasets are available, we used only the first 40 years of data in this particular part of the present study. Although this period is relatively short, Kharin and Zwiers (2000) demonstrated that in one of these models, the GEV distribution is a good fit to the distribution of annual maxima of daily surface air temperature. Also, 40 years is short enough so that climate model drift is not a factor, eliminating the need to detrend the model control run output. This will prove not to be the case later when we consider longer control run datasets.

Figure 1a shows the intermodel standard deviation in kelvins of the 20-year return value of the annual maxima of daily surface air temperature. In this calculation, we calculated a fit to the GEV distribution followed by a calculation of the return values for each model separately. To ensure uniformity, we began the calculation by regridding each model's parent data set of all daily mean surface temperatures to a $4^\circ \times 5^\circ$ mesh, the resolution of the coarsest model in Table 1. The regridding tool used was the Climate Data and Analysis Tool (CDAT) which uses a standard area-weighted average method that preserves the global mean of the field (available at www2-pcmdi.llnl.gov/cdat/). Annual maxima, the GEV parameters and the return values all were then calculated on this coarse mesh. Examination of Fig. 1a reveals that much of the uncertainty coming from different model formulations is over land and can be as high as 5 K. Kharin and Zwiers (2005) noted that a modification to the GEV distribution may need to be made when the extrema of surface air temperature cluster near the freezing point. Although usually only an issue when considering temperature minima, the maxima over the Arctic sea ice exhibit this clustering near freezing. Maximum temperatures in the Greenland and Antarctic interiors remain well below freezing. This constraint near the freezing point explains in part the relatively low intermodel variability over the Arctic sea ice in Fig. 1a.

The disagreement between models in simulating the mean surface air temperature is not small, with an inter-model standard deviation of the 40-year average surface air temperature exceeding 1.5 K nearly everywhere over land surfaces. To ascertain the effect of uncertainty in mean temperature on the extreme temperature uncertainty shown in Fig. 1a, we subtracted each model's 40-year average temperature from its 20 year return value and recalculated an inter-model standard deviation. This is equivalent to removing the mean from the parent datasets causing all parent data sets to have the same average, defined to be zero. This modified measure of model uncertainty, shown in Fig. 1b, reflects the differences in the widths of the surface air temperature distributions at the tails. In most places, this centering of the distribution reduces the uncertainty due to the high positive correlation between mean and the tails of the temperature distribution. This reduction is particularly noticeable in the southern hemisphere and over Asia. The global mean of the inter-model standard deviation is reduced from 1.97 to 1.59 K. Consistent with our explanation

Fig. 1 **a** The inter-model standard deviation of the 20-year return value of the annual maxima of daily surface air temperature. Calculated from 40 years of daily data from 15 different CMIP3 models. **b** Same as **a** except that the 40-year mean surface air temperature is subtracted from each model's return value prior to the calculation of standard deviation.



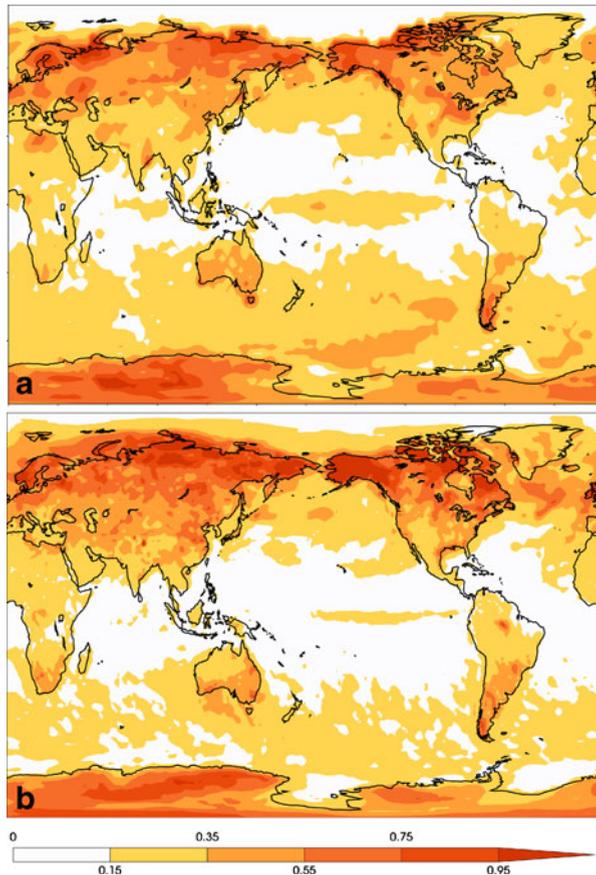
of extreme surface air temperatures being constrained near the freezing point in the Arctic, removal of the mean actually increases inter-model uncertainty of the return value in this region. Rather large inter-model differences in mean Arctic temperatures occur because the Arctic mean temperature is very much below freezing and not clustered like the annual daily maximum. Hence over the Arctic sea ice, the differences shown in Fig. 1b are mostly due to the intermodel mean temperature differences. Antarctic daily temperatures stay very much below the freezing point and do not show this clustering effect. This study did not use the GEV modification of Kharin and Zwiers (2005), but it is not expected that the details of Fig. 1 are impacted greatly in the regions of largest uncertainty.

3 Uncertainty from unforced internal variability

A second source of uncertainty in quantifying extreme climate statistics comes from the unforced internal variability of the climate system itself. This source of uncertainty can be estimated from long control integrations. Two models integrated in

the pre-industrial control configuration for long periods have enough archived daily data that we can estimate and compare the models' internal variability on the multi-decadal time scale by generating sizable samples of 40 year periods. This source of uncertainty should be considered intrinsic to the climate system itself and depends mostly on the length of the period analyzed. Of course, the model control integrations are only a proxy for the real climate system and this uncertainty will vary from model to model. The Parallel Climate Model (PCM) has 640 years of daily data in a single integration and the Community Climate System Model (CCSM3.0) has 560 years of daily data in two separate integrations. Data is freely available from earthsystemgrid.org. Unlike the 40 year segments of control run data used in Fig. 1, climate drift is an issue for these two long datasets. For both of the models used in this part of the study we first calculated the annual maxima of the daily surface air temperature and then applied a linear trend removal. However, for CCSM3.0, we found that there is a 0.1 K bias in the global mean between the two different simulations probably related to the compiler differences of the two different machines used. To remove this bias, we removed the trend from each integration separately, corrected the bias

Fig. 2 The standard deviation of the 20-year return value of the annual maxima of daily surface air temperature calculated from non-overlapping 40 year periods of climate model long control runs. This map is a measure of the uncertainty due to internal variability of the climate system. **a** Result from the Parallel Climate Model (PCM). **b** Result from the Community Climate System Model (CCSM3.0).



and concatenated the two integrations to form a uniform 560 year dataset. To quantify the effect of model internal variability in a manner comparable to the previous section, we divided these corrected datasets into 40 year segments and calculated GEV parameters and return values for each of the segments. We also calculated the GEV parameters and return values for the entire multi-century datasets to determine if there is any systematic bias arising from a shorter dataset. The difference between that calculation and the average over all the 40 year segments was miniscule with global mean differences in the return values of 0.0058 K for PCM and 0.0024 K for CCSM3.0.

Figure 2a shows the standard deviation of this calculation of the 20 year return value of the annual maximum daily mean surface air temperature for the PCM and Fig. 2b shows the same for the CCSM3.0. The patterns of uncertainty appear to be similar for these two related models with large differences between high and low latitudes as well as between land and ocean areas. The centered (mean removed) spatial correlation between Fig. 2a and b is 0.81. Comparison with Fig. 1 reveals that the uncertainty in extreme temperatures due to internal variability is much smaller than the uncertainty due to differences in model formulation. Note that the color scale in Fig. 2 is five times smaller than Fig. 1. However, the spatial distributions are somewhat similar. The centered spatial correlation between the PCM internal uncertainty (Fig. 2a) and the inter-model uncertainty with the means removed (Fig. 1b) is 0.74. The same quantity for CCSM3.0 is 0.71.

4 Uncertainty from statistical models

A third source of uncertainty in quantifying extreme climate statistics comes from uncertainty in the fit of the GEV parameters. Following the lead of Kharin and Zwiers (2000, 2005), we investigate this source of uncertainty using a bootstrapping method detailed by Hosking and Wallis (1997). In this method, GEV parameters are first estimated from the actual available sample data. Then, a set of random samples distributed according to this GEV distribution is generated. GEV parameters and associated return values are then calculated for each of these random samples (of the same size as the actual sample). Using the long CCSM3.0 control, we generated 50 random samples at each grid point on the native grid of varying sample sizes. Figure 3a shows the standard deviation of the return values calculated over the random samples for a sample size of only 20 years. Figure 3b shows the same uncertainty calculated from a sample size of 40 years and Fig. 3c shows it calculated from a sample size of 100 years. Not surprisingly as the sample size increases, the uncertainty resulting from the fit of the statistical model decreases. The 40-year sample size result is directly comparable to Figs. 1 and 2b. We note that this uncertainty is much smaller than the inter-model uncertainty shown in Fig. 1 but is comparable to the uncertainty resulting from this model's unforced internal variability shown in Fig. 2b. Furthermore, we note significantly larger uncertainty over land than over ocean. Figure 3d shows this statistical fit uncertainty averaged over land for CCSM3.0 as a function of sample size (the solid line). Additionally, we show as dots the same quantity for all the other models using a 40-year sample size. All results were calculated on the

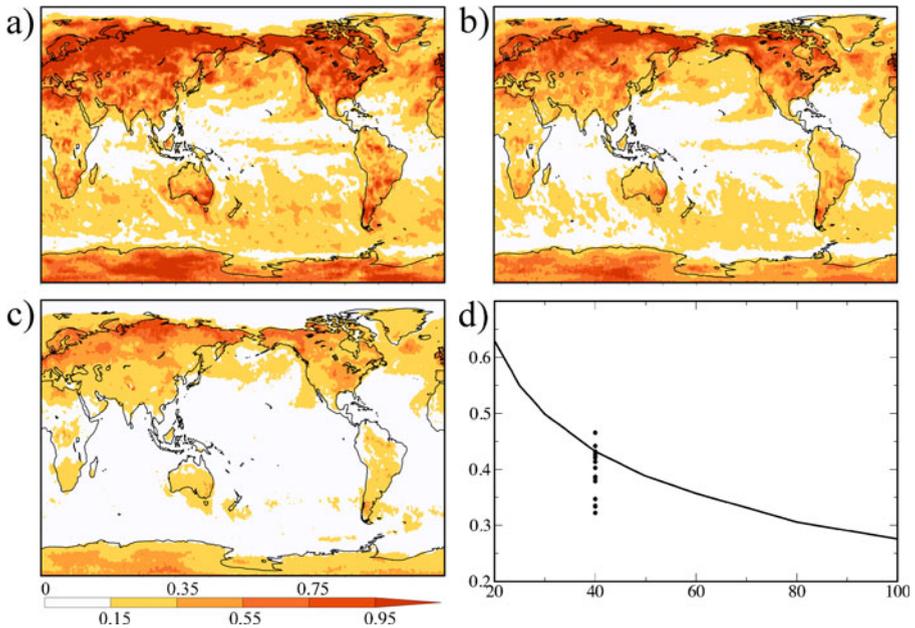


Fig. 3 The standard deviation (kelvins) of the 20-year return value of the annual maxima of daily surface air temperature calculated due to the uncertainty in the fit of the GEV parameters calculated by the method of Hosking and Wallis (1997). **a** Results from CCSM3.0 using 20 year sample size. **b** Same using a 40 year sample size. **c** Same using a 100 year sample size. **d** The average over land as a function of sample size for CCSM3.0 (solid line) and the 40 year sample size results for the rest of the climate models considered in this study (dots). Same color scale as Fig. 2.

model's native grids. We note that this uncertainty (at 0.43 K) is larger for CCSM3.0 than for most of the other models which range from 0.32 to 0.46 K.

5 Discussion

We have identified and compared three sources of uncertainty in the calculation of extreme value statistics for observed and modeled climate data. Using fits to the GEV distribution to obtain 20 year return values, we analyzed the annual maximum daily mean surface air temperature of pre-industrial control runs from the CMIP3 dataset of climate models. We find that uncertainty resulting from differing model formulations is much larger than that due to both unforced internal climate variability and the fit of the statistical model. A consistent spatial pattern of higher uncertainty over land than over ocean is true for all three types of uncertainties. The uncertainties due to unforced internal climate variability and the fit of the statistical model are similar in both magnitude and spatial pattern with both exhibiting a strong poleward gradient over land, especially in the northern hemisphere. Subtraction of the mean value from the analysis generally reduces the inter-model uncertainty and reveals

a similar poleward gradient. However, the technique is problematic over the ice-covered Arctic Ocean due to a clustering of daily maximum temperatures near the freezing point and biases in the models simulation of mean temperatures.

Each of these three uncertainties would be better quantified by the usage of more data. Although they could be reduced by considering spatial autocorrelation effects, the inter-model uncertainty would remain the largest. From a practical view, given the relative magnitude of the inter-model uncertainty compared to the two types of uncertainties resulting from a small sample size for any single model, the need for very large ensembles of transiently forced climate model runs is diminished for the individual modeling groups. This is an important consideration in the choice of the optimal ensemble size of transiently forced simulations given computer resource limitations. Furthermore, while the single model return value uncertainties considered here are much smaller than anticipated future changes (Wehner 2005; Kharin et al. 2007), the inter-model uncertainty is not. Hence, robust methods of discriminating climate model quality (Santer et al. 2009; Tebaldi et al. 2005) are critically important.

Acknowledgments This work was performed under the auspices of the US Department of Energy (DOE) by the Lawrence Berkeley National Laboratory (LBNL) under contract No. DE-AC03-76SF00098 (LBNL) and with support from the DOE Regional and Global Climate Modeling Program. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. The author thanks Ben Santer (LLNL) for useful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Kharin et al. (2007) subjected the CMIP3 models to a Kolmogorov–Smirnov (KS) goodness of fit test and concluded that GEV distributions of block extrema temperatures are generally acceptable if temperatures are not clustered near freezing. However, as first noted in Kharin and Zwiers (2000), extreme temperatures, especially minima are often clustered around the freezing point of water for hydrological reasons. For the maximum temperatures considered in the current study, this kind of clustering occurs only in the extreme cold of the polar regions. In fact, the interior of the Antarctic is too cold and rarely reaches temperatures as high as the freezing point. Rather, clustering is more likely over the Arctic sea ice and possibly the Antarctic sea ice due to somewhat warmer temperatures resulting from the heat flux through the ice from the ocean. Additionally, the likelihood of cross-polar flows is much higher in the Arctic than the Antarctic. These events can warm the polar temperatures considerably.

The KS goodness of fit test is often used to determine whether a specified distribution function adequately describes a sample data set and its test statistic is conveniently independent of the distribution function. However, it is more sensitive to discrepancies in the center of the distribution rather than to those in the tails (NIST/SEMATECH 2006). Alternatively, the Anderson–Darling (AD) test provides a similar test but is more influenced by data points near the tails of the distribution (NIST/SEMATECH 2006). Unlike the KS test, specific critical values of the AD test statistic must be calculated for each class of distribution function to be tested. Fortunately, values for the Weibull distribution (the GEV distribution with a negative shape parameter) have been tabulated (Stephens 1977). As the ultimate purpose of using a GEV distribution in this study is to calculate return values from relatively short samples, the AD test is somewhat more relevant than the KS test.

The AD test statistic is defined as

$$AD = \sum_{k=1}^n \frac{2k-1}{n} [\ln(F(Y_k)) - \ln(1-F(Y_{n+1-k}))] - n \quad (4)$$

where F is the distribution to be tested and Y are the sample data points. A correction for small samples alters the test statistic to

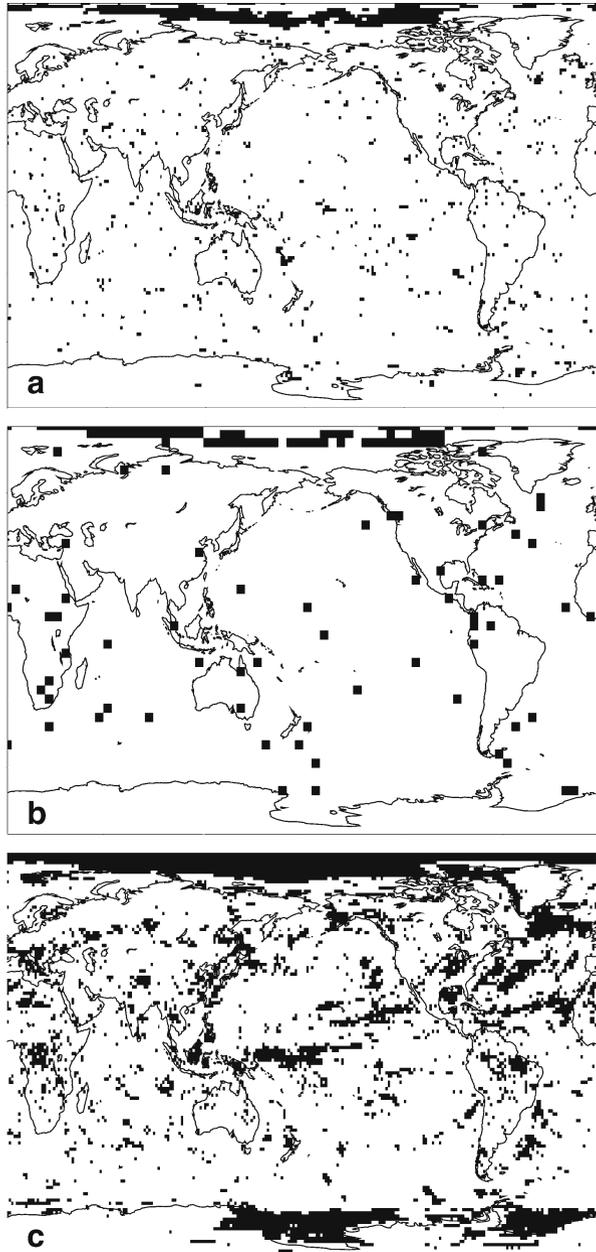
$$AD^* = AD(1 + 0.2/\sqrt{n}) \quad (5)$$

The critical values defining the significance levels for rejection of the null hypothesis that the Weibull distribution describes a sample were first calculated by Stephens (1977). A fit of these observed significance levels (p-values) is given as (Romeu 2003)

$$OSL = 1/\{1 + \exp[-0.1 + 1.24 \ln(AD^*) + 4.48AD^*]\} \quad (6)$$

A value of OSL less than 0.05 rejects the Weibull distribution with a chance of erroneously doing so of less than 5%. Figure 4 shows where this occurs for three samples selected from the CCSM3.0 control run. The upper panel (a) shows the 5% rejection level for a 40 year sample at the original T85 resolution. Of most prominence is the spatially coherent shaded region in the Arctic. Annual maximum temperatures in these cells are likely to exhibit significant clustering near the freezing point. The scattered cells outside the Arctic are not of concern and should be expected. Regridding to the coarser $4^\circ \times 5^\circ$ grid has little effect on either the Arctic region or the scattered cells as shown in the middle panel (b). Of the CMIP3 models considered in this study, many exhibit similar behavior to CCSM3.0. However, some models do not exhibit any clustering in the Arctic. This is likely due to model biases in polar temperatures. As sample size increases, goodness of fit tests become more strict. In the lower panel (c), the 5% rejection of the Weibull distribution is shown for the full 600 year sample of the CCSM3.0 control run at the original T85 resolution. Rejection in the Arctic region is considerably more widespread in this case. Additionally,

Fig. 4 Rejection at the 5% level of the null hypothesis that a GEV distribution fits the annual maximum daily surface air temperature from CCSM3.0 as determined by an Anderson–Darling test. **a** 40 year sample at the original T85 resolution. **b** 40 year sample regridded to $4^\circ \times 5^\circ$. **c** 600 year sample at the original T85 resolution.



rejection over Antarctic sea ice regions is noted. Spatially coherent regions outside the polar regions (mostly over oceans) also occur.

What are the consequences in these regions where the Weibull (and hence GEV) distribution are rejected on the estimation of return values? In the 40 year samples,

where clustering near freezing occurs, the return value is essentially 0°C for a wide range of periods, T . So errors in the fitted distribution, while statistically significant are not reflected in the return values with any physical significance. In the 600 year sample, 20 year return values can be accurately estimated directly from the sample without usage of a distribution function.

References

- Castillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M.: *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, New York (2004)
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer, Berlin (2001)
- Coles, S.G., Dixon, M.J.: Likelihood-based inference for extreme value models. *Extremes* **2**, 5–23 (1999)
- Easterling, D.R., Wehner, M.F.: Is the climate warming or cooling? *Geophys. Res. Lett.* **36**, L08706 (2009)
- Hosking, J.R.M.: The L-moments page. <http://www.research.ibm.com/people/h/hosking/lmoments.html> (2003)
- Hosking, J.R.M., Wallis, J.R.: *Regional Frequency Analysis, An Approach Based on L-Moments*. Cambridge University Press, Cambridge (1997)
- Hosking, J.R.M., Wallis, J.R., Wood, E.F.: Estimation of the generalized extreme-value distribution by the method of probability weighted moments. *Technometrics* **27**, 251–261 (1985)
- Karl, T.R., Meehl, G.A., Miller, C.D., Hassol, S.J., Waple, A.M., Murray, W.L. (eds.): *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research*, Washington, DC (2008)
- Kharin, V.V., Zwiers, F.W.: Changes in the extremes in an ensemble of transient climate simulation with a coupled atmosphere–ocean GCM. *J. Clim.* **13**, 3760–3788 (2000)
- Kharin, V.V., Zwiers, F.W.: Estimating extremes in transient climate change simulations. *J. Clim.* **18**, 1156–1173 (2005)
- Kharin, V.V., Zwiers, F.W., Zhang, X., Hegerl, G.C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419–1444 (2007)
- Kysely, J.: Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series. *Clim. Res.* **20**, 211–222 (2002)
- Leadbetter, M.R., Lindgren, G., Rootzen, H.: *Extremes and related properties of random sequences and processes*, 336 p. Springer, New York (1983)
- Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J., Taylor, K.E.: The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**, 1383–1394 (2007)
- NIST/SEMATECH: e-Handbook of Statistical Methods (2006). <http://www.itl.nist.gov/div898/handbook>
- Romeu, J.L.: Anderson–Darling: a goodness-of-fit test for small samples assumptions. Selected Topics in Assurance Related Technologies, vol. 10, no. 5. System Reliability Center, Rome, New York, USA (2003). http://src.alionscience.com/pdf/A_DTest.pdf
- Santer, B.D., Taylor, K.E., Gleckler, P.J., Bonfils, C., Barnett, T.P., Pierce, D.W., Wigley, T.M.L., Mears, C., Wentz, F.J., Brueggemann, W., Gillett, N.P., Klein, S.A., Solomon, S., Stott, P.A., Wehner, M.F.: Incorporating model quality information in climate change detection and attribution studies. *Proc. Nat. Acad. Sci.* **106**, 14778–14783 (2009). doi:10.1073/pnas.0901736106
- Stephens, M.A.: Goodness of fit for the extreme value distribution. *Biometrika* **64**, 583–588 (1977)
- Tebaldi, C., Smith, R.L., Nychka, D., Mearns, L.O.: Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. *J. Clim.* **18**, 1524–1540 (2005)
- Wehner, M.: Changes in daily precipitation and surface air temperature extremes in the IPCC AR4 models. *US CLIVAR Variations* **3**, 5–9 (2005)
- Wehner, M.F., Smith, R., Duffy, P., Bala, G.: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Clim. Dyn.* **34**, 241–247 (2009)
- Zwiers, F.W., Kharin, V.V.: Changes in the extremes of the climate simulated by CCC GCM2 under CO₂ doubling. *J. Clim.* **11**, 2200–2222 (1998)