Temperature Extremes in the Community Atmosphere Model with Stochastic Parameterizations*

FELIPE TAGLE

Cornell University, Ithaca, New York

JUDITH BERNER

National Center for Atmospheric Research, Boulder, Colorado

MIRCEA D. GRIGORIU, NATALIE M. MAHOWALD, AND GENNADY SAMORODNITSKY

Cornell University, Ithaca, New York

(Manuscript received 29 April 2015, in final form 29 August 2015)

ABSTRACT

This paper evaluates the performance of the NCAR Community Atmosphere Model, version 4 (CAM4), in simulating observed annual extremes of near-surface temperature and provides the first assessment of the impact of stochastic parameterizations of subgrid-scale processes on such performance. Two stochastic parameterizations are examined: the stochastic kinetic energy backscatter scheme and the stochastically perturbed parameterization tendency scheme. Temperature extremes are described in terms of 20-yr return levels and compared to those estimated from ERA-Interim and the Hadley Centre Global Climate Extremes Index 2 (HadEX2) observational dataset. CAM4 overestimates warm and cold extremes over land regions, particularly over the Northern Hemisphere, when compared against reanalysis. Similar spatial patterns, though less spatially coherent, emerge relative to HadEX2. The addition of a stochastic parameterization generally produces a warming of both warm and cold extremes relative to the unperturbed configuration; however, neither of the proposed parameterizations meaningfully reduces the biases in the simulated temperature extremes of CAM4. Adjusting warm and cold extremes by mean conditions in the respective annual extremes leads to good agreement between the models and reanalysis; however, adjusting for the bias in mean temperature does not help to reduce the observed discrepancies. Based on the behavior of the annual extremes, this study concludes that the distribution of temperature in CAM4 exhibits too much variability relative to that of reanalysis, while the stochastic parameterizations introduce a systematic bias in its mean rather than alter its variability.

1. Introduction

In recent decades, there has been increasing interest in the analysis of extreme climate events given their significant impact on human and natural systems (Kunkel et al. 1999; Easterling et al. 2000). Such events typically account for thousands of deaths and billions of dollars in damages globally each year, as population and infrastructure continue to expand in areas that are vulnerable to extremes such as flooding, storm damage, and extreme heat or cold (Easterling et al. 2000). The Intergovernmental Panel for Climate Change (IPCC) in their Fifth Assessment Report (AR5; Hartmann et al. 2013) concluded that most global land areas have experienced significant warming of both maximum and minimum temperature extremes since about 1950. Simulations from global coupled climate models are the primary tools for forecasting potential future changes in extreme climate statistics (Kharin et al. 2007). Because an important aspect of the evaluation of the reliability of these forecasts is an assessment of the models' ability to simulate observed climate extremes, the release of a new

* Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JCLI-D-15-0314.s1.

Corresponding author address: Felipe Tagle, Rhodes Hall 291, Cornell University, Ithaca, NY 14850.
E-mail: fit4@cornell.edu

DOI: 10.1175/JCLI-D-15-0314.1

generation of climate models is usually accompanied by studies focusing on this topic (e.g., Kharin et al. 2007, 2013; Sillmann et al. 2013). Evaluations of the models participating in phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5) have shown that temperature extremes are reasonably represented, as compared to reanalysis and observations (Flato et al. 2013), with greater uncertainties in the simulation of cold extremes (Kharin et al. 2007, 2013). However, performance in representing temperature extremes is strongly dependent on the choice of verification dataset, particularly reanalyses, as discrepancies between these can be as large as the intermodel spread between the CMIP ensemble (Sillmann et al. 2013).

Owing to the multitude of extreme events in the climate system and how the extreme nature of a climate phenomenon is usually dependent on the affected region (Stephenson et al. 2008), most studies of climate extremes rely on the use of extreme indices (e.g., Tebaldi et al. 2006; Alexander and Arblaster 2009; for further references, see Seneviratne et al. 2012; Hartmann et al. 2013). Other studies focus on more extreme climate statistics, typically relying on results from extreme value theory to approximate the distribution of annual extremes (e.g., Kharin et al. 2007, 2013; Brown et al. 2008; Wehner 2004; Wehner et al. 2010). In this study we follow the latter approach and evaluate the performance of the NCAR Community Atmosphere Model, version 4 (CAM4) (Neale et al. 2010), in simulating observed temperature extremes, as measured by 20-yr return levels, against reanalysis and observational datasets and, furthermore, investigate whether the observed discrepancies are climatological in nature. Indeed, it is not unusual for climate models to exhibit systematic errors in mean temperatures; for instance, a longstanding error that is present in the NCAR Community Earth System Model (CESM) is the so-called warm bias over land, which refers to temperatures over land being too warm in summer (Neale et al. 2010). Here, we will examine to what extent differences in mean temperature between the model and the verification datasets explain the observed discrepancies in temperature extremes.

Despite the continuing increase of computing power, which allows climate models to be run with ever-higher resolution, many important physical processes (e.g., tropical convection, gravity wave drag, microphysical processes) are still not resolved (e.g., Shutts 2005; Franzke et al. 2015). Some subgrid-scale processes are altogether unrepresented or represented very crudely; for example, flow over mountains is a source of turbulent kinetic energy at small scales, but in the model it is represented by a drag coefficient (e.g., Palmer 2001; Shutts 2005). Since finescale climate processes have been shown to regulate not only the mean but also the tails of the daily temperature and precipitation distributions (Diffenbaugh et al. 2005), numerous studies have highlighted the importance of horizontal grid resolution on the simulation of climate extremes (e.g., Wehner et al. 2010; Rauscher et al. 2010; Jung et al. 2012; Kopparla et al. 2013). Stochastic-dynamics prediction is an alternative way to represent the effect of fluctuating subgrid-scale processes (Palmer 2001). The omission of variability of unresolved subgrid-scale processes has been proposed as one reason for persistent biases across different models (Jung et al. 2010; Palmer and Weisheimer 2011; Berner et al. 2012). For instance, Berner et al. (2012) showed that including a stochastic representation leads to improvements in the Northern Hemispheric circulation comparable to increasing horizontal resolution.

One such stochastic parameterization is the stochastic kinetic energy backscatter scheme (SKEBS), whose origin lies in large-eddy simulation modeling (Mason and Thomson 1992). It is based on the rationale that a small fraction of the model dissipated energy interacts with the resolved-scale flow and acts as a systematic forcing. Recently, it was adapted by Shutts (2005) to numerical weather prediction. Its impact on weather and seasonal forecasts is reported, for example, in Berner et al. (2008, 2009, 2011, 2015), Palmer et al. (2009), and Doblas-Reyes et al. (2009). On seasonal scales, in integrations with the European Centre for Medium-Range Weather Forecasts (ECMWF) model, SKEBS has been shown to reduce the bias in the circulation over the North American continent and improve the occurrence of Northern Hemispheric blocking (Jung et al. 2005; Berner et al. 2008). In the tropics, SKEBS positively influences the representation of convectively coupled waves (Berner et al. 2012). Another stochastic parameterization is the stochastically perturbed parameterization tendency scheme (SPPT) (Buizza et al. 1999; Palmer et al. 2009), which samples the physical tendencies from an assumed subgrid-scale probability density function. The present study will provide the first assessment of the impact of stochastic parameterizations on climate extremes.

The paper is organized as follows. The datasets and a brief review of the stochastic parameterizations are presented in section 2. Section 3 describes the methodology, while section 4 compares the model-simulated extremes with the observational evidence. Section 5 provides a discussion and conclusions.

2. Data and experimental design

CAM is the atmospheric component of the Community Climate System Model (CCSM) and the new CESM, developed at the National Center for Atmospheric Research (NCAR), under the support of the National Science

Foundation and the U.S. Department of Energy. In this study we consider a simulation of CAM with prescribed SSTs and sea ice according to the Atmospheric Model Intercomparison Project (AMIP) protocol (Gates 1992). Specifically, we use a so-called 1° IPCC-AMIP simulation, spanning the years 1979–2010, at a horizontal resolution of 0.9° × 1.25°. Simulated temperature extremes are evaluated against ERA-Interim (hereafter ERA) (Dee et al. 2011), the latest global atmospheric reanalysis produced by ECMWF, covering the years from 1979 to present on a 1.5° × 1.5° regular grid. Maximum and minimum 2-m temperature since previous postprocessing at a 12-h time step from 1979 to 2010 were downloaded from the ECMWF archive (http://apps.ecmwf.int/datasets/data/interim-full-daily/), from which daily temperature extremes were constructed as the maximum and minimum of the two respective 12-h daily values. For each dataset, annual extremes of daily maximum and minimum 2-m temperature are computed at each grid point over land.

Although reanalyses offer the advantages of gridded output with global spatial coverage, they are nonetheless observationally constrained model output. Their output may be classified into four categories depending on the relative influence of the observational data and the numerical model (Kalnay et al. 1996). Near-surface temperature belongs to the type B category, indicating that observational data exists that directly affects its value; however, the model component still exercises considerable influence. Significant discrepancies in extreme temperature statistics between reanalyses have been documented in Kharin et al. (2007, 2013) and Sillmann et al. (2013), particularly for cold extremes. However, ERA has been shown to adequately capture recent temperature extreme trends over Europe (Cornes and Jones 2013). Moreover, in a comparison of several reanalysis products, Donat et al. (2014) reported that temperature extremes in ERA exhibit the highest temporal and spatial correlations with those of gridded observations over the past 30 years.

As an additional verification dataset, we consider the gridded land-based Hadley Centre Global Climate Extremes Index 2 (HadEX2) observational dataset. HadEX2 consists of the comprehensive set of indices of temperature and precipitation extremes defined by the Expert Team on Climate Change Detection and Indices (ETCCDI), which are calculated directly from station data and interpolated onto a regular grid using a modified version of Shepard's angular distance weighting interpolation algorithm (for details, see Donat et al. 2013). In many countries, these provide the only publicly available information about temperature and precipitation extremes. Because the HadEX2 indices are derived solely from station data, they are free from biases originating from model specification error, as is potentially the case for reanalysis products; however, direct comparison with model output becomes encumbered by the fact that the latter corresponds to area averages, not point values. Systematic biases introduced by this spatial-scale mismatch, however, should be minor given the smoothness of temperature fields. The dataset is available on a 2.5° × 3.75° grid, with a temporal coverage of 1901–2010. We use the TXx and TNn indices, corresponding to the annual extremes of daily maximum and minimum 2-m temperature, respectively, but restrict the domain of analysis to those grid points with complete temporal coverage of the 1979–2010 period.

To investigate the impact of stochastic parameterizations on temperature extremes, two experiments were performed.

a. SKEBS

SKEBS aims to represent model uncertainty arising from unresolved subgrid-scale processes by introducing random perturbations to streamfunction and potential temperature tendencies. SKEBS is based on the rationale that a small fraction of the model dissipated energy interacts with the resolved-scale flow and acts as a systematic forcing.

The scheme introduces at each time step and grid point additive perturbations to the streamfunction tendency:

$$\dot{\psi}_p(\phi,\lambda,t) = \dot{\psi}_{\mathrm{dyn}}(\phi,\lambda,t) + f(\phi,\lambda,t),$$

where $\dot{\psi}_{\mathrm{dyn}}$ and $\dot{\psi}_p(\phi, \lambda, t)$ are the streamfunction tendency before and after perturbation and $f(\phi, \lambda, t)$ the perturbation tendency. Here, λ and ϕ denote longitude and latitude in physical space and time t. Furthermore, we let the perturbation tendency forcing be expressed in a triangularly truncated spherical harmonics expansion:

$$f(\phi,\lambda,t) = \sum_{m=-N}^{N} \sum_{n=|m|}^{N} f_n^m(t) P_n^{|m|}(\cos\phi) e^{im\lambda}.$$

Here, m and n denote the zonal and total wavenumbers, N is the truncation wavenumber of the numerical model, and P_n is the associated Legendre function of degree n and order m. The spherical harmonics $Y_n^m = P_n^m e^{im\lambda}$ form an orthogonal set of basis functions on the sphere. If they are nonvanishing for at least one $n \leq N$ and do not follow a white-noise spectrum, the pattern perturbations will be spatially correlated in physical space.

Since the physical processes mimicked by this forcing have finite correlation times, temporal correlations are introduced by evolving each spectral coefficient as a first-order autoregressive (AR1) process:

$$f_n^m(t+\Delta t) = \alpha f_n^m(t) + \sqrt{(1-\alpha)}g_n\varepsilon(t),$$

where α is the linear autoregressive parameter determining the temporal decorrelation time, g_n the wavenumber-dependent noise amplitude, and ε a Gaussian white-noise process with mean zero and variance η. The noise amplitude g_n is chosen to have power-law behavior, given by $g_n = bn^p$, and determine the variance spectrum of the forcing.

The pattern $f(\phi, \lambda, t)$ is interpreted as a streamfunction tendency forcing. In the case of perturbing potential temperature, a second perturbation pattern is created analogously but with a different power-law behavior and potentially a different temporal correlation. The behavior of this scheme is determined by the following parameters: the exponent of the power law p, the wavenumber perturbation range $n_1 - n_2$, and the amplitude of forcing energy, which determines the normalization constant b.

In the original implementation, the streamfunction pattern is subsequently weighted with the normalized total instantaneous dissipation rate from numerical dissipation, deep convection, and gravity and mountain wave drags (Shutts 2005; Berner et al. 2009) so that the perturbations are largest in regions with large dissipation and have little effect in regions where and when the dissipation is small. A simplified version of SKEBS assumes the dissipation rate to be spatially and temporally constant, resulting in a state-independent (additive) stochastic forcing. This simplification relies on underlying model dynamics to determine which perturbations will grow and which ones will be damped (Berner et al. 2011). Here, we use the simplified version with constant dissipation rate.

b. SPPT

SPPT is a revision of the original stochastic diabatic tendency scheme of (Buizza et al. 1999) and perturbs the parameterized tendency of physical processes with multiplicative noise. It is based on the notion that, especially with increasing numerical resolution, the equilibrium assumption no longer holds and the subgrid-scale state should be sampled rather than represented by the equilibrium mean. Consequently, SPPT multiplies the accumulated physical tendencies $\dot{x}$ of temperature, zonal and meridional winds, and humidity (T, u, v, q) at each grid point and time step with a multiplicative random coefficient $r(\phi, \lambda, t)$:

$$X_p = (1+r)\dot{x}, \quad \text{with} \quad x = u, v, T, q.$$

Here, X_p is the perturbed parameterized tendency for the variables $x = u$, v, T, q and $r(\phi, \lambda, t)$ a random pattern with spatial and temporal correlations. By design, the perturbations are large where the physical tendencies, and presumably their uncertainty, is large and has very little effect where and when the tendencies are small. SPPT uses the same pattern generator as SKEBS (see above) but a different normalization.

The stochastic pattern evolves in spectral space as

$$r_n^m(t+\Delta t) = \alpha r_n^m(t) + \sqrt{(1-\alpha)}g_n\varepsilon(t),$$

where all variables are as defined above. The temporal correlations are given by the decorrelation time τ defining $\alpha = \exp(-\Delta t/\tau)$.The noise amplitudes are given as follows:

$$g_n = F_0 \exp - L\frac{n(n+1)}{2}, \quad \text{with}$$

$$F_0 = \left\{\frac{\sigma^2(1-\alpha)^2}{2\sum_{n=1}^{N}(2n+1)\exp[-Ln(n+1)]}\right\}^{1/2},$$

where L is a horizontal length scale defining the spatial correlations and σ^2 the perturbation variance at each grid point. The normalization constant F_0 is chosen so that the variance at any grid point σ^2 is given by the total variance in spectral space (Weaver and Courtier 2001). The resulting stochastic pattern follows at each grid point a Gaussian with mean zero and variance σ^2.

3. Methodology

In this study we characterize the extreme behavior of the annual extremes of daily maximum $T_{\max}$ and minimum $T_{\min}$ temperature in terms of 20-yr return values. A T-yr return value can be informally interpreted as the value that is exceeded by an annual extreme on average once every T years. More precisely, it is defined as the quantity that is exceeded in any given year with probability $p = 1/T$, which corresponds to the $(1-p)$ quantile of the distribution of annual extremes. We assume that the distribution of annual extremes can be approximated by a generalized extreme value (GEV) distribution, given by

$$F(x;\mu,\sigma,\xi) = \begin{cases} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right] & \text{for} \quad \xi \neq 0, \quad 1+\xi(x-\mu)/\sigma > 0 \\ \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right] & \text{for} \quad \xi = 0 \end{cases},$$

with μ the location parameter, $\sigma > 0$ the scale parameter, and ξ the shape parameter. The shape parameter controls the tail behavior of the distribution. If $\xi < 0$, the probability density function (PDF) is bounded above, with the upper endpoint given by $\mu - \sigma/\xi$. If $\xi > 0$, the GEV distribution is heavy tailed and its PDF decays "slowly" (polynomially) as $x \rightarrow \infty$; when $\xi = 0$, the tails of the PDF decay relatively faster. These cases give rise to three distinct families of extreme value distributions: the Weibull, Fréchet, and Gumbel families, respectively. We fit a GEV distribution at each grid point over land to the sample of annual extremes and derive the 20-yr return value from the quantile function:

$$z_T = \begin{cases} \mu - \sigma/\xi\{1 - [-\log(1 - 1/T)]^{-\xi}\} & \text{for} \quad \xi \neq 0, \\ \mu - \sigma\log[-\log(1 - 1/T)] & \text{for} \quad \xi = 0, \end{cases} \tag{1}$$

evaluated at $T = 20$. For cold extremes, we fit the GEV distribution to the negative of the sample of annual $T_{\min}$ extremes and reverse the sign of the estimated return value. To compare model parameters and return level estimates with estimates from the verification datasets, the former estimates were regridded to match the coarser grids of the latter datasets.

The justification of the GEV distribution as an appropriate representation of the behavior of annual extremes follows from an important result in extreme value theory, which states that the limiting distribution of the maximum of a sufficiently large random sample belongs to only one of the three extreme value distributions (Leadbetter et al. 1983). However, the daily observations of maximum and minimum temperature, from which the samples of annual extremes are computed, exhibit features typical of environmental datasets, such as an annual cycle and serial correlation that violate the assumptions defining a random sample. Indeed, the annual cycle negates the condition that the observations are identically distributed, while the presence of serial correlation implies dependence among consecutive observations. Despite this, empirical studies have shown the GEV distribution to be a good candidate to describe environmental extremes (e.g., Kharin and Zwiers 2005; Kharin et al. 2007, 2013; Wehner et al. 2010).

Since the model simulations are forced with observed boundary conditions that exhibit trends over the simulations period, we use the nonparametric Mann–Kendall test (Chandler and Scott 2011) to investigate whether such trends are captured in the annual extremes of daily temperature. Unsurprisingly, the test confirms the existence of trends at many grid points, in agreement with the large number of observational studies that have identified trends in such temperature statistics at both global (e.g., Donat et al. 2013) and regional scales (e.g., Zhai and Pan 2003; Klein Tank and Können 2003; Bukovsky 2012).

A common approach to address this form of nonstationarity is to assume that the parameters of the GEV distribution are time dependent (Coles 2001). Kharin and Zwiers (2005) examined the time dependence of GEV parameters in a transient climate setting and found that a model that allowed for time variation in the location and scale parameters (in the form of a linear and log-linear trend, respectively) best represented the transient behavior of temperature extremes. In a similar manner, we fit several models with and without time-varying parameters, and as in Kharin and Zwiers (2005) likelihood ratio tests rejected models that assumed time variation in the shape parameter; however, unlike their work, the datasets did not provide evidence in support of a time-varying scale parameter. Consequently, we performed the analyses described below assuming a GEV distribution with and without a linear trend in the location parameter and found that the differences were generally minor and did not alter qualitatively the conclusions. Therefore, in what follows we present the results pertaining only to the stationary GEV model.

Hosking et al. (1985) compared the short-sample performance of maximum likelihood and the method of L moments in the estimation of the upper quantiles of the GEV distribution and demonstrated that for a broad range of shape parameter values, the estimates by the latter method showed lower root-mean-square error relative to those of the former. Since then, the method of L moments has been widely used in observational and simulation studies requiring the estimation of return values from short samples (e.g., Kharin et al. 2007; Wehner et al. 2010). However, this method does not permit the estimation of time-varying GEV parameters since the computation of L moments requires that the random sample be identically distributed. The above nonstationary GEV distribution was therefore fit by maximum likelihood, and for consistency, it was also used in the estimation of the stationary GEV parameters. One of the benefits of maximum likelihood is that approximate standard errors of the estimated parameters can be obtained by using the inverse of the observed information matrix, but these approximations tend to be unreliable for small sample sizes (Kharin and Zwiers 2005).

An alternative approach to quantify the uncertainty in parameter estimates is through resampling (Efron and Tibshirani 1994). In this study, we use a bootstrapping technique to generate 500 resampled replicates from the original sample of annual extremes and to each replicate fit a GEV distribution and derive return levels, from which standard errors may be computed. Assuming that the original dataset is represented by an $m \times n \times N$

array, with m, n, and N denoting longitude, latitude, and time, respectively, each new sample corresponds to one of N $m \times n$ matrices, in order to preserve the spatial dependence structure.

Several of the analyses rely on hypothesis tests performed at each grid point, as is the case, for example, of likelihood ratio tests where the validity of one model is measured relative to another or in the evaluation of the statistical significance of observed trends as in the Mann–Kendall tests above. However, the interpretation of these results in a spatial context, where there is spatial dependence, is often misleading (Livezey and Chen 1983). A field significance test is a popular statistical technique designed for the simultaneous evaluation of multiple hypothesis tests, usually specified over geographic areas. Such a test may be interpreted as a type of metatest, as the data being tested are the results of individual or local tests, and the null hypothesis is that all of the individual null hypotheses are true (Wilks 2011). Because of the complex dependence structures found in environmental datasets, the sampling distribution of test statistics, such as the number of tests that are significant, is difficult to derive analytically. Fortunately, good approximations may be obtained by means of resampling. We first compute the desired test statistic based on the results of the original collection of annual extremes and subsequently generate 500 replicates of annual extremes in the manner described above; then, on each replicate we perform the individual hypothesis tests and compute the test statistic. The procedure yields 500 values of the test statistic, from which the significance under the null hypothesis of the original value may be determined.

4. Results

a. Demonstration on a simple example

We begin the analysis with a simple example that helps illustrate how differences in the mean and variance of an idealized distribution of temperature affect the distribution of the associated extremes (Fig. 1). Similar schematics have been developed since the IPCC Third Assessment Report to characterize the effect of a changing climate on the daily temperature distribution (e.g., Fig. 2.32 in Folland et al. 2001). We assume that the distribution of temperature is well described by a Gaussian distribution and consider a scenario with two distributions having the same variance but with the mean of one slightly larger than the other, as would be the case, for instance, for a warm bias of CAM relative to ERA or where the means are the same but the distribution of CAM exhibits a higher variance. Analytically or through Monte Carlo simulation, we can derive the GEV distribution of the maximum and minimum of a large sample from each of these distributions, representing the distributions of the annual maximum and minimum of $T_{\max}$ and $T_{\min}$, respectively. The GEV PDFs of warm and cold extremes for the first scenario are shown in Figs. 1c and 1e, respectively. In each plot, the distributions of CAM and ERA are identical except that of CAM is shifted to the right by a quantity equal to the difference of the means of T. Thus a shift in the distribution of T induces an identical shift in the distributions of both warm and cold extremes. In the second scenario, despite the means of T being the same, the PDF of CAM for warm extremes exhibits both a shift toward warmer temperatures and an increase in variability relative to that of ERA, while that of cold extremes is shifted toward colder temperatures with an identical increase in variability (Figs. 1d,f). An aspect that we will examine repeatedly in this study is the role of differences in the $T_{\max}$ and $T_{\min}$ climatologies on the respective PDFs of extremes. Because we have implicitly neglected any form of nonstationarity, such as the presence of an annual cycle, these climatologies correspond here to the means of the respective GEV distributions. Adjusting for these mean differences in the first scenario (Figs. 1g,i) removes the shift observed above and renders the PDFs identical, while in the second, the mean differences are canceled but the discrepancies in variability persist (Figs. 1h,j).

We apply these ideas to the GEV PDFs of $T_{\max}$ annual extremes of CAM4 and ERA over land regions, with parameters set to area averages of gridpoint estimates (Fig. 2). Parameter values are denoted in the panel. At this spatial scale, shape parameter estimates are remarkably similar across the two datasets—namely, -0.27 and -0.28 for CAM4 and ERA, respectively—suggesting that the tail behavior is well captured by the model. The negative values indicate that the distributions of $T_{\max}$ annual extremes can be approximated by those from the Weibull family, which is characterized by a bounded upper tail. We note that the quantile function of the GEV distribution, as depicted in Eq. (1), for the case of $\xi \neq 0$, is a function of all three GEV parameters, but by holding ξ constant, quantile differences may be conveniently decomposed in terms of differences in location and scale parameter estimates. Furthermore, given the functional dependence of the variance of the GEV distribution (not shown) on only the shape and scale parameters, a constant shape parameter implies that differences in scale parameter estimates may be interpreted as differences in interannual variability. Therefore, differences in 20-yr return levels may be analyzed in terms of differences in the location parameter, which is a measure of central tendency, and interannual variability of temperature extremes. Here we observe that the return level estimate of CAM4 exceeds that of ERA by 2.25°C (Fig. 2a), which is largely explained by the difference of 1.88° in the location

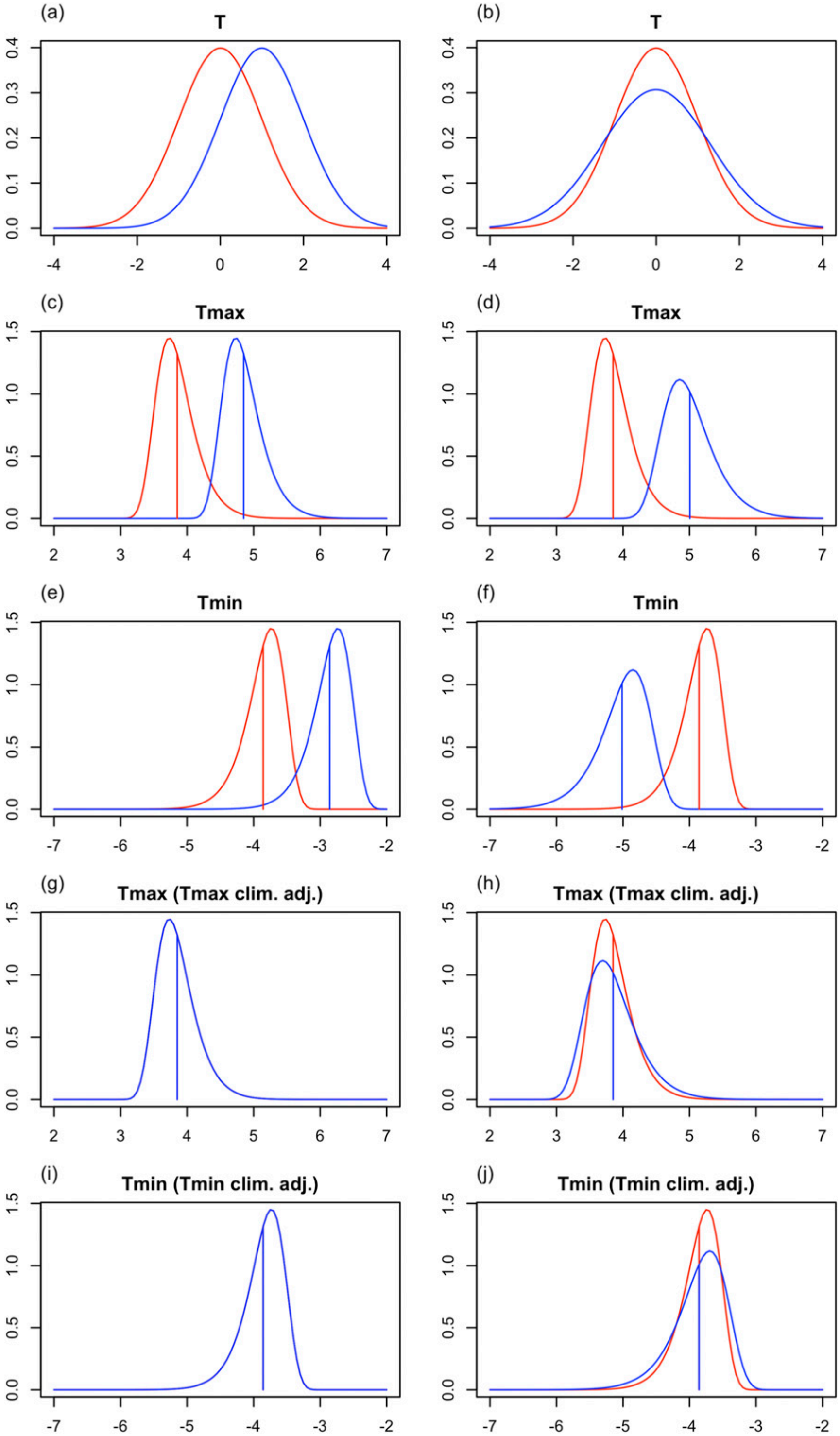

FIG. 1. Idealized Gaussian distributions of (a),(b) temperature T for CAM (blue) and ERA (red) and (c)−(f) the corresponding GEV probability density functions obtained by drawing 5000 samples of length 10 000 from the distributions above and fitting GEV distributions to the resulting 5000 extreme observations. A standard Gaussian distribution is assumed for ERA T. (left) T distributions with the same standard deviation but the climatological mean of CAM is shifted by 1. (right) T distributions with the same mean but the climatological standard deviation of CAM is 1.3. Vertical lines denote the mean of the GEV distribution. GEV distributions adjusted by their (g),(h) $T_{\max}$ and (i),(j) $T_{\min}$ climatologies.

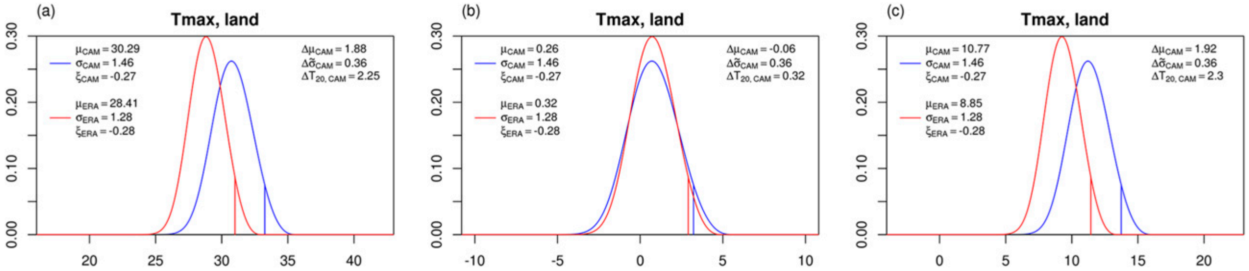

FIG. 2. (a) GEV PDFs of CAM4 (blue) and ERA (red) with parameter values (left legend) set to the area-averaged estimates over land regions, with gridpoint estimates corresponding to averages of parameter values obtained from 500 resampled replicates of the annual extremes of daily maximum temperature. Location and 20-yr return level differences are indicated in the right legend, along with differences in a modified scale parameter, defined as $\Delta\tilde{\sigma}_s = -(\sigma_s - \sigma_{\mathrm{ERA}})f(\xi_{\mathrm{ERA}})$, with f obtained from Eq. (1) and s representing the CAM dataset. Probability density functions adjusted by the area-averaged maximum of the annual cycle (b) $\max\overline{T}^{\mathrm{ac}}_{\max}$ and (c) $\max\overline{T}^{\mathrm{ac}}_{\mathrm{mean}}$. Vertical lines denote 20-yr return levels.

parameter estimates, while the contribution from the scale parameter difference is only 0.36°C. Note how this value does not correspond to the actual difference in scale parameters; rather, it is the difference scaled by a factor that depends on the value of the shape parameter, which we assume to be the same for both datasets. However, owing to the slight difference of 0.01 in the shape parameter estimates, the decomposition is not exact, as seen by the 0.01°C discrepancy between the actual 20-yr return level difference and the sum of the two contributions. Comparing the two distributions, the GEV distribution of CAM4 appears shifted to the right and slightly wider than that of ERA, reflecting the larger magnitude of the location and scale parameter estimates in the CAM4 simulation.

Adding a constant to a sample of annual extremes will result in an identical change in the value of the location parameter, as in the mean, while leaving the shape and scale parameters unchanged. Therefore, given the form of the quantile function [Eq. (1)], return levels will reflect any systematic differences in the extreme temperatures between model and reanalysis. We consider two potential sources of systematic biases: $T_{\max}$ and T_{mean} climatological differences. The $T_{\max}$ (T_{mean}) annual cycle is defined as the 1979–2010 average for each calendar month of monthly $T_{\max}$ (T_{mean}). Because annual warm extremes tend to coincide with the maximum of the warm extreme climatology and mean temperature climatology, adjusting for these climatologies can be done by subtracting from these $\max\overline{T}^{\mathrm{ac}}_{\max}$ and $\max\overline{T}^{\mathrm{ac}}_{\mathrm{mean}}$, respectively. After adjusting for the $T_{\max}$ climatology (Fig. 2b) the distributions are quite similar, as the difference in location parameter estimates of Fig. 2a becomes negligible and the difference of 2.25°C in the return values is reduced to 0.32°C, which coincides almost exactly with the discrepancy in scale parameters. Adjusting for the T_{mean} climatology, however, does little to improve the correspondence in return values between the two datasets; on the contrary, the difference is increased to 2.3°C, as the discrepancy in location parameters increases from 1.88° to 1.92°C. The similarity in the distributional pattern depicted in Fig. 2a with that of Fig. 1d, together with the relative invariance of the location parameter difference to the adjustment in mean biases, provides the first indications that the distributions of T from CAM4 and ERA are related qualitatively as in Fig. 1b.

b. Temperature extremes in CAM4, ERA, and observations

Warm and cold extremes of CAM4 and differences with the corresponding temperature extremes of ERA and HadEX2 are displayed in Fig. 3. Significant positive differences are seen over land regions in comparison to ERA, in particular over the midlatitudes in the Northern Hemisphere and in the subtropics of South America, with the exception of central Africa and the Arabian Peninsula where slight negative differences emerge (Fig. 3c). Over the midwestern United States return levels exceed 40°C in CAM4, while the values in ERA are roughly 10–15°C smaller. The limited spatial coverage of the HadEX2 TXx index restricts the scope of the comparison primarily to the Northern Hemisphere. Positive differences of similar magnitude can be seen over central North America and western Eurasia, where return value differences between model and observations also differ by up to 10°C (Fig. 3e). However, coherent regions of colder warm extremes emerge over the Himalayas and North America extending into the Greenland ice sheet.

Cold extremes are well simulated over Australia and most of South America but are significantly colder over most of the Northern Hemisphere (Fig. 3d). Area-averaged discrepancies over North America and Asia exceed 5°C. Similar differences appear in the comparison with HadEX2, with the exception of Greenland (Fig. 3f). Indeed, simulated cold extremes are generally

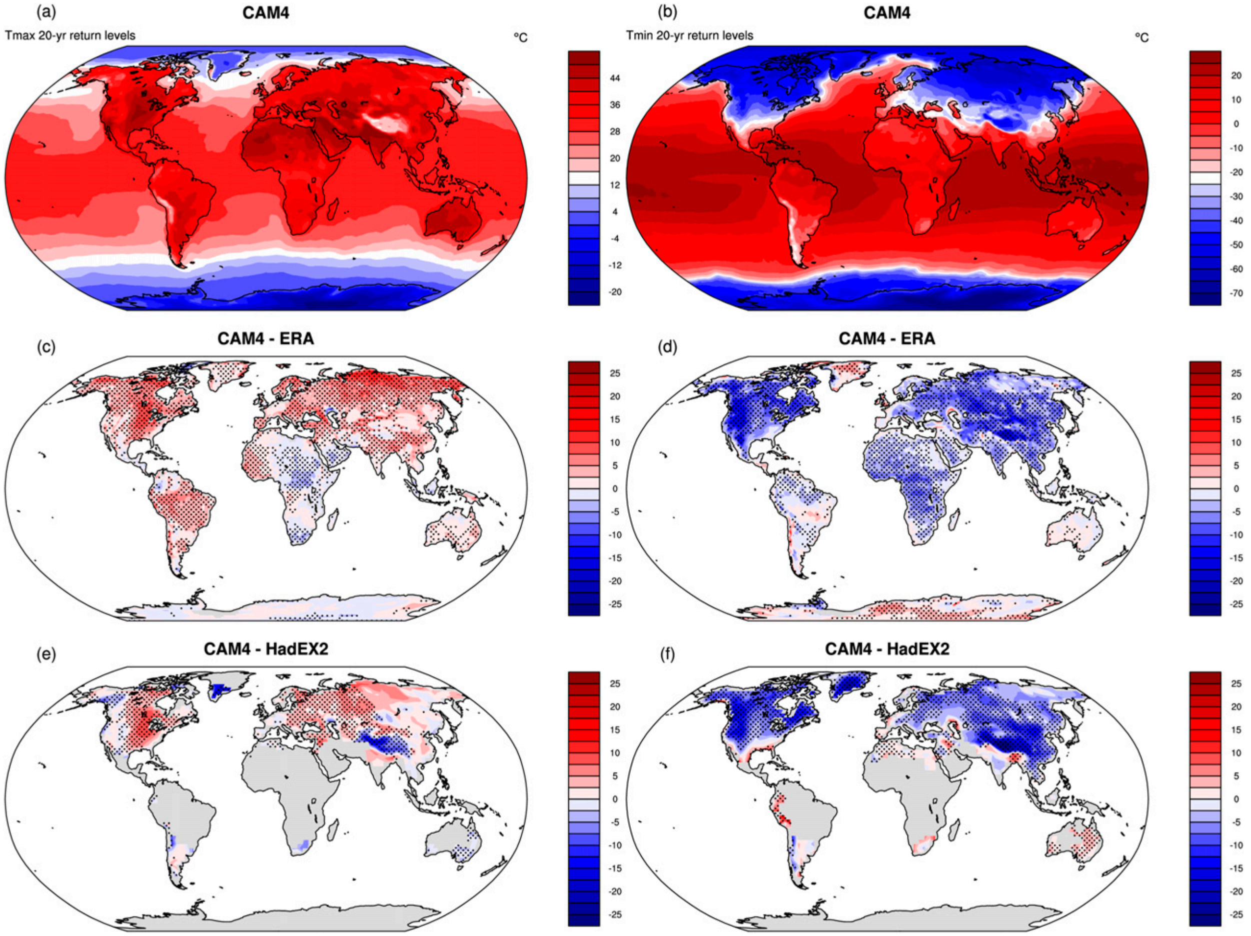

FIG. 3. (a),(b) CAM4 20-yr return level estimates of 1979–2010 annual extremes of daily (left) maximum and (right) minimum temperature and the corresponding differences with estimates from (c),(d) ERA-Interim and the (e),(f) HadEX2 observational dataset. Stippling indicates return level differences are significant at a 5% level.

in good agreement over ice-covered land regions relative to reanalysis, in contrast with HadEX2 where severe negative differences extend into the Greenland ice sheet. We note that the spatial coverage is better for the HadEX2 TNn index.

We investigate the degree to which the differences in warm and cold extremes between CAM4 and ERA can be explained by $T_{\rm max}/T_{\rm min}$ climatological differences. Figure 4 displays CAM4 warm and cold extremes (adjusted by $\max\overline{T}^{\rm ac}_{\rm max}$ and $\min\overline{T}^{\rm ac}_{\rm min}$, respectively) and the differences with the corresponding adjusted ERA extremes. Warm extremes are in good agreement after the adjustment, although over the midwestern United States, Australia, and China the climate model indicates return values up to 6°C higher than ERA; however, these values are not always statistically significant. In contrast, large coherent regions where CAM4 overestimates cold extremes remain after the adjustment, particularly over western Eurasia and North America. However, area-averaged differences of 0.8°, 1.2°, and 2.1°C, over North America, Europe, and Asia, respectively, highlight the notable reductions in cold extreme return level discrepancies between the two datasets.

In summary, we note that over large areas, particularly in the Northern Hemisphere, excluding the Greenland ice sheet and the Himalayas, the climate model generally overestimates both warm and cold extremes. This is consistent with our hypothesis that CAM4 exhibits greater variability in T relative to ERA, as this overestimation is qualitatively depicted in Figs. 1d and 1f. Adjusting by the respective extreme climatologies leads to substantial reductions in the noted discrepancies, although slight differences remain, especially for cold extremes.

c. *Impact of stochastic parameterizations on extreme events*

Next, we analyze the impact of adding a stochastic parameterization on the model issues discussed above.

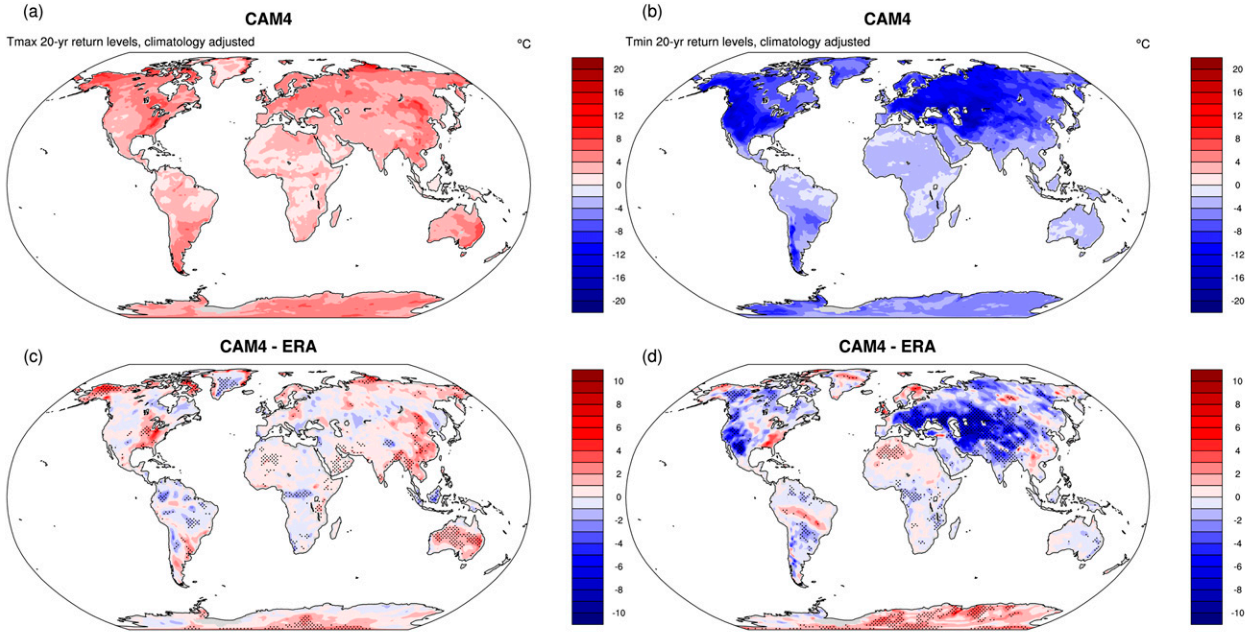
(a)
CAM4
Tmax 20-yr return levels, climatology adjusted
°C
20
16
12
8
4
0
-4
-8
-12
-16
-20
(b)
CAM4
Tmin 20-yr return levels, climatology adjusted
°C
(c)
CAM4 - ERA
10
8
6
4
2
0
-2
-4
-6
-8
-10
(d)
CAM4 - ERA

FIG. 4. (a),(b) CAM4 20-yr return level estimates of 1979–2010 annual extremes of daily (left) maximum and (right) minimum temperature and the corresponding differences with estimates of (c),(d) ERA-Interim. Return level estimates are adjusted by the corresponding extremes of the annual cycle, $\max\overline{T}^{\mathrm{ac}}_{\max}$ and $\min\overline{T}^{\mathrm{ac}}_{\min}$. Stippling in (c) and (d) indicates statistical significance at a 5% level.

Overall, warm and cold extremes are quite similar to those of unperturbed CAM4 (Fig. S1; see supplemental material). The signature of SPPT and SKEBS on extreme events is remarkably similar, especially when considering that SKEBS is most active in the extratropics while SPPT has the largest impact on the near-surface fields in the tropics (Berner et al. 2009, 2015). This suggests that the stochastic parameterizations may excite modes of variability already present in CAM4 rather than impose their specific characteristics on the distribution of extreme events, which is further substantiated in a forthcoming study. Discrepancies are apparent only by direct comparison with CAM4 (Fig. 5). Warm extremes simulated by SKEBS are considerably warmer over parts of the Northern Hemisphere, particularly over Asia (Fig. 5a), in some parts in excess of 8°C. Replacing SKEBS with SPPT results in a nearly identical spatial pattern, except that the positive differences are less pronounced (Fig. 5c). Cold extremes for both SKEBS and SPPT relative to CAM4 present a less coherent spatial pattern, but statistically significant positive differences are found over most continents, which help reduce the overestimation of cold extremes mentioned above and thus lead to a better agreement with reanalysis.

To assess if these differences between CAM4 and the stochastic parameterizations are collectively statistically significant or could be the result of sampling variability perhaps as a result of the limited sample sizes, we test the null hypotheses that the CAM4 and SKEBS as well as the CAM4 and SPPT simulations are realizations from a single data-generating process with the same extreme value distribution. This analysis is carried out at each grid point by performing a likelihood-ratio test involving the likelihood of a GEV distribution fit separately to each time series of annual extremes and the likelihood of a GEV distribution fit to the concatenated time series. A field significance test, as described in section 3, is then performed to assess whether the results of the individual tests are significant at a regional level. We note that under the null hypothesis the concatenated time series represents a sample drawn from the same GEV distribution; therefore, in the resampling stage, replicates are generated by selecting at random $2N$ $m \times n$ matrices from the concatenated $m \times n \times 2N$ array. We use seven continental regions: North America, South America, Europe, Africa, Asia, Australasia, and Antarctica; displayed in Fig. 6, each of these is a combination of subcontinental-scale regions defined in Seneviratne et al. (2012). The null hypotheses that warm extremes simulated by the two pairs of simulations belong to the same GEV distribution are strongly rejected in all regions except Antarctica. The analogous hypotheses for cold extremes are strongly rejected in all regions.

We further assess, as was done earlier for the return level discrepancies between CAM4 and ERA, to what degree the observed warm and cold extreme differences between reanalysis and the stochastic parameterizations can be attributed to the respective climatologies (Fig. S2

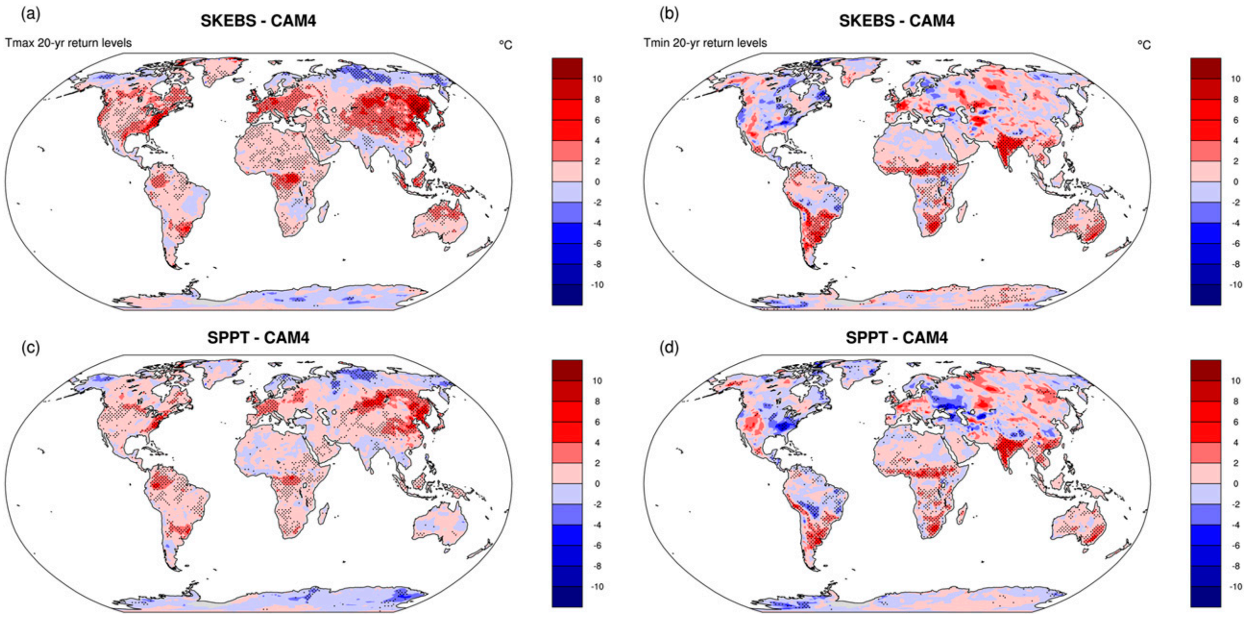

FIG. 5. Differences of 20-yr return level estimates of 1979–2010 annual extremes of daily (left) maximum and (right) minimum temperature between CAM4 and the stochastic parameterization schemes (a),(b) SKEBS and (c),(d) SPPT.

in the supplemental material). The spatial patterns after adjusting for the respective extreme climatologies closely resemble those of CAM4 relative to reanalysis. A comparison with CAM4 is displayed in Fig. 7. Warm extremes of both parameterizations show very close agreement with those of CAM4, as do cold extremes in the Southern Hemisphere. However, a heterogeneous spatial pattern of areas of warm and cold differences, though mostly not statistically significant, emerges for cold extremes over the Northern Hemisphere, with the exception of a large area of warmer extremes over northwestern Asia, which is most clearly depicted in the SPPT and CAM4 comparison.

Overall, the effect of introducing a stochastic parameterization, with its tendency to enhance the overestimation of warm extremes in CAM4, while mildly reducing that of cold extremes is conceptually consistent with Figs. 1c and 1e, as the first effect may be represented by a distribution of warm extremes shifted toward warmer temperatures while the second by a similar though more moderate shift of the distribution of cold extremes. This would suggest that SKEBS introduces a warm bias in the distribution of T of CAM4 rather than augmenting its variability.

d. *Regional analysis of GEV PDFs*

In this section we aim to characterize the discrepancies of temperature extremes observed above in terms of distributional differences in global and regional-scale GEV PDFs, as was done in section 4a. Figures 8a and 8b display the PDFs of CAM4, SKEBS, and ERA for $T_{\max}$ and $T_{\min}$ annual extremes, respectively, over land regions. The GEV parameters at each grid point correspond to averages over the 500 resampled estimates. We present results only for SKEBS as those of SPPT are comparable. Differences in parameter estimates are highly significant as indicated by two-sample t tests (Wilks 2011) except for most of the shape parameter estimates. Henceforth all parameter differences should be interpreted as statistically significant at a 5% level except where otherwise noted. As was noted in section 4a, at these scales the shape parameter estimates for warm extremes exhibit little variation across datasets; here we confirm that this similarity extends to cold extremes as well, indicating that the tail behavior is well captured by the models in a more general sense. This close agreement between shape parameter estimates supports the use of the decomposition of return level

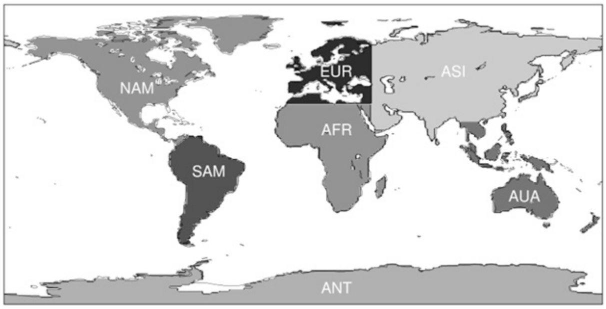

FIG. 6. Continental regions adapted from the subcontinental definitions of Seneviratne et al. (2012) (as in Fig. 10.7 of Bindoff et al. 2013).

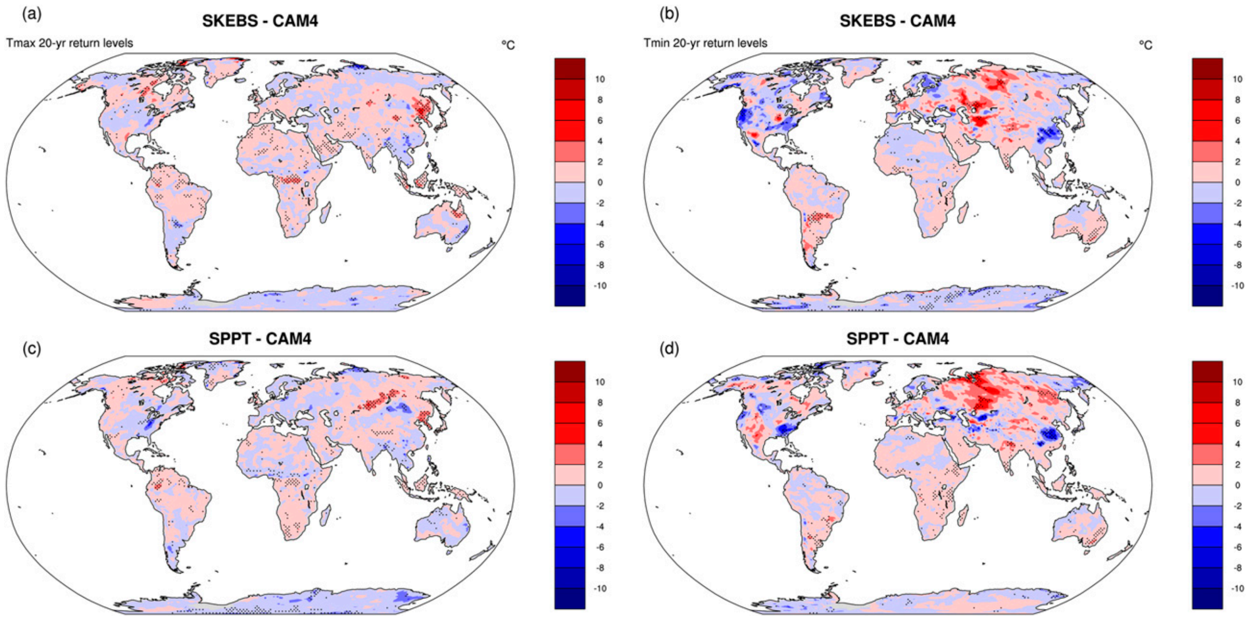

FIG. 7. As in Fig. 5, but simulations have been adjusted by the respective extremes of the $T_{\max}$ and $T_{\min}$ annual cycles.

differences in terms of location and scale discrepancies, as discussed above.

Figure 8a adds to Fig. 2a the SKEBS GEV PDF curve, which appears displaced to the right of that of CAM4, with only a minor increase in the scale parameter. Thus the overestimation of warm extremes relative to CAM4 reported in Fig. 5a can be interpreted as the result of a shift in the distribution of CAM4 toward higher temperatures rather than an increase in interannual variability. The displacement gives rise to a return level difference of 3.3°C against ERA compared to the 2.3°C difference of CAM4. Disagreements of similar magnitude apply to cold extremes, with differences in return levels of −3.4° and −3.2°C for CAM4 and SKEBS, respectively, relative

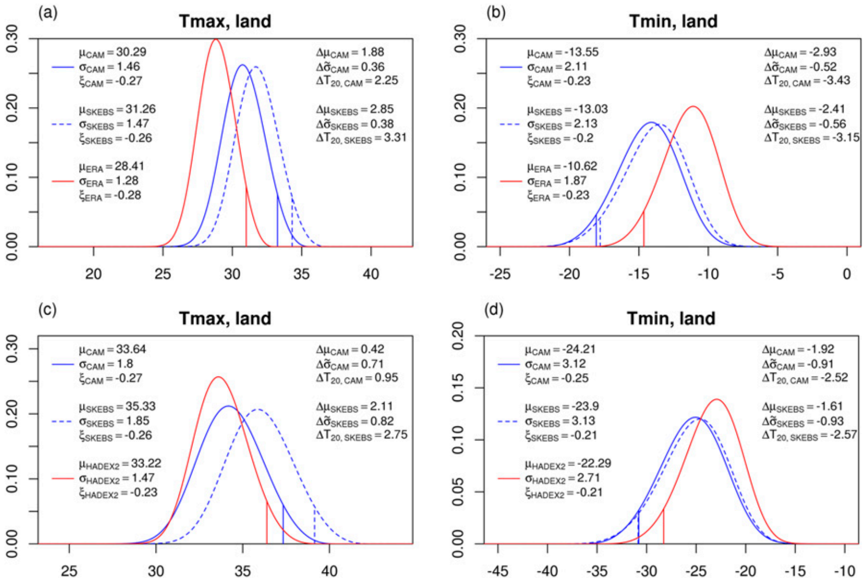

FIG. 8. As in Fig. 2a, but for GEV PDFs of CAM (solid blue), and SKEBS (dashed blue) for (left) warm and (right) cold extremes over land, with respect to (a),(b) ERA (red) and (c),(d) HadEX2 (red). Parameter values in the HadEX2 comparison correspond to area averages within 15°–72.5°N, excluding Greenland.

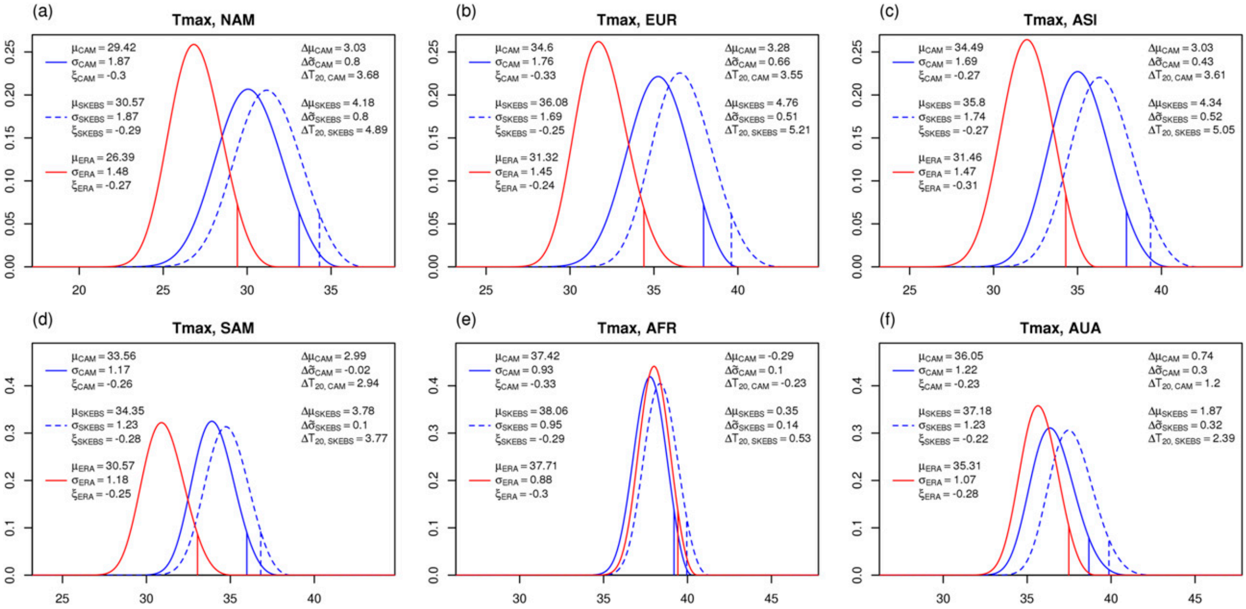

FIG. 9. As in Fig. 8a, but for GEV PDFs of $T_{\max}$ annual extremes with parameter values area averaged over continental regions defined in Fig. 6.

to ERA, which are related predominantly to discrepancies in location parameter values. Despite the considerable increase in interannual variability in the models and reanalysis, as compared with warm extremes, the distributional differences between CAM4 and ERA are consistent with Fig. 1f, while the slight displacement to the right of the SKEBS distribution is consistent with Fig. 1e.

In Figs. 8c and 8d we repeat the analysis above using HadEX2 as the reference dataset. To minimize potential distortions arising from observational uncertainties, the averages of the GEV parameters were taken over land regions within 15°–72.5°N, excluding Greenland. Despite the difference in land definition, the comparison with HadEX2 yields qualitatively similar results. However, because of its Northern Hemispheric focus, extremes show significantly higher interannual variability across the three datasets. Discrepancies in scale parameter estimates also increase in magnitude, particularly for cold extremes, which leads to these differences playing a more important role in explaining differences in return level estimates.

A more detailed examination of regional differences in warm extremes between the models and ERA is presented in Fig. 9. The regions correspond to those displayed in Fig. 6, but we exclude Antarctica. With the exception of Africa, where there is a notable similarity between the various PDFs, the distributional pattern observed for warm extremes (Fig. 8a) is largely reproduced across the regions. Discrepancies in return levels are more severe over the Northern Hemisphere but display greater uniformity compared to those of the Southern Hemisphere. For instance, differences between CAM and ERA range between 3.55° and 3.68°C in the former compared to a range between −0.23° and 2.94°C in the latter. The hemispheric differences in the magnitude of return level discrepancies are tied to larger differences in both location and scale parameter estimates. Note that mild discrepancies in shape parameter estimates arise in some regions, such as in Europe and Australasia, which diminish the accuracy of the decomposition. However, it is of little concern because in such cases the variation in return levels is still primarily driven by location parameter differences.

The distributional patterns for cold extremes over the continental regions in Fig. 10 are qualitatively similar to that observed at the global scale (Fig. 8b), with minor differences over South America and Australasia where the discrepancy in return levels between CAM4 and ERA is so small that the right shift induced by SKEBS either cancels it or enhances it after it switches sign. The signature of SKEBS on the distribution of annual cold extremes is almost indistinguishable from that of CAM4 over the Northern Hemisphere, particularly over North America where the close agreement in the location and scale parameter estimates is such that the difference in return levels results is largely due to the slight discrepancy in the shape parameter estimates.

In general, the impact of SKEBS on the distribution of annual extremes of CAM4 is largely limited to changes in the location of the distribution, with minor effects on the tail behavior and interannual variability. As noted above, this is consistent with SKEBS shifting the

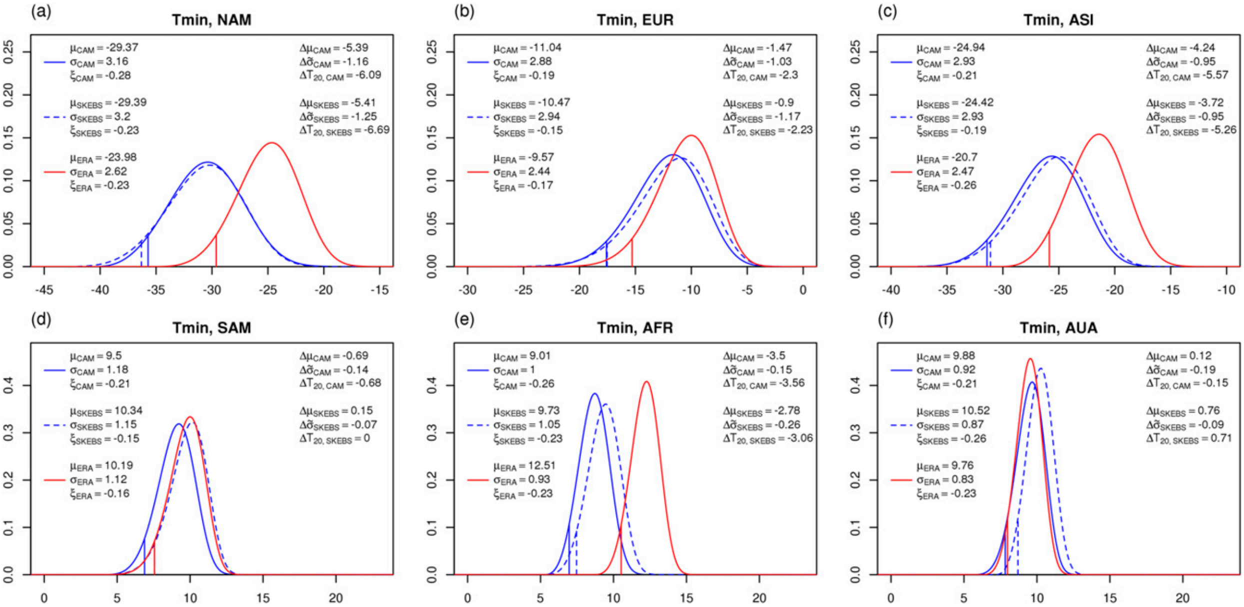

FIG. 10. As in Fig. 9, but with T_{min} annual extremes.

distribution of T of CAM4 toward warmer temperatures (Fig. 1a). We explore this further by adjusting the regional PDFs examined above of both warm and cold extremes by their respective T_{mean} climatologies (i.e., $\max\overline{T}^{\mathrm{ac}}_{\mathrm{mean}}$ and $\min\overline{T}^{\mathrm{ac}}_{\mathrm{mean}}$, respectively). We note that such an adjustment affects only the location of the distributions. In Fig. 11, for space reasons, we focus only on Northern Hemisphere warm extremes and Southern Hemisphere cold extremes. Over the Northern Hemisphere, the reductions in location parameter differences between CAM4 and ERA are at most 1.6°C, but the reductions between SKEBS and ERA are consistently larger, with values in excess of 2°C in Asia and Europe, represented by shifts of the SKEBS distributions toward

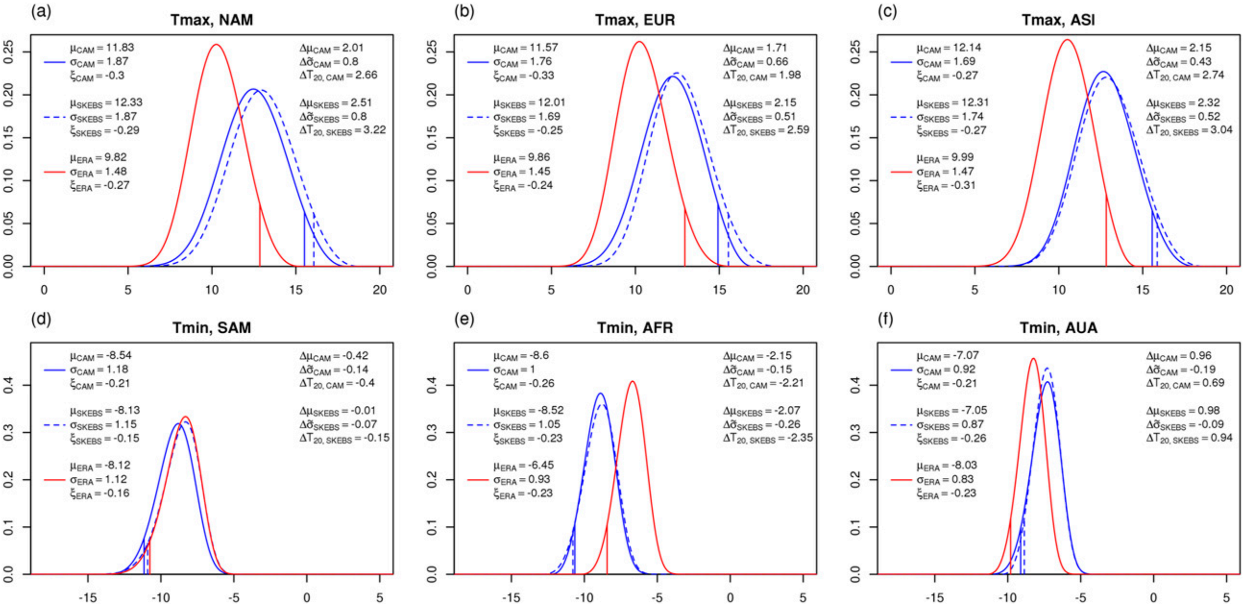

FIG. 11. As in Figs. 9 and 10, but GEV PDFs of (a)–(c) T_{max} annual extremes over Northern Hemisphere land regions and (d)–(f) T_{min} annual extremes over Southern Hemisphere land regions, adjusted by the respective extreme of the monthly mean temperature annual cycle $\overline{T}^{\mathrm{ac}}_{\mathrm{mean}}$.

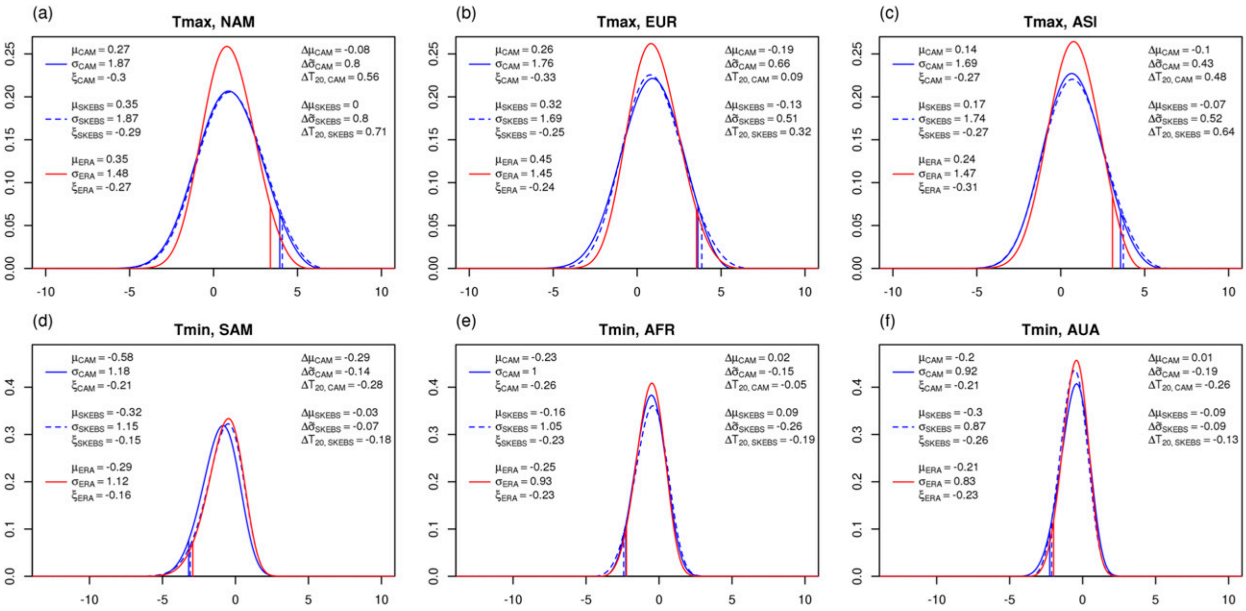

FIG. 12. As in Fig. 11, but for GEV PDFs adjusted by the respective extreme of the $T_{\max}$ and $T_{\min}$ annual cycle.

colder temperatures relative to ERA. This effect considerably reduces the location parameter differences between the SKEBS and CAM4 PDFs to within 0.5°C. Over the Southern Hemisphere, the adjustment also leads consistently to an improved agreement in location parameter estimates between SKEBS and CAM4. These results provide further evidence that SKEBS introduces a warm bias in the distribution of T of CAM4.

In Fig. 12 we extend the analysis of the previous figure by adjusting the warm and cold extreme distributions by their respective $T_{\max}$ and $T_{\min}$ climatologies. Across extremes and regions, the adjustment largely cancels the location discrepancies between the CAM4 and reanalysis distributions, with values less than 0.5°C, and similar reductions apply to the discrepancies between SKEBS and CAM4. The resulting discrepancies in return values relative to reanalysis fall within 0.8°C. Note how the adjustment by the annual extreme climatology has the compound effect on the SKEBS distribution of correcting for both the warm bias in T relative to CAM4 and the enhanced variability of CAM4 relative to ERA.

5. Discussion and conclusions

This work investigates the skill of the general circulation model CAM4 in simulating annual extremes of near-surface temperature. Previous studies have demonstrated that the extreme statistics of temperature over the late twentieth century in many general circulation models and observationally constrained datasets agree reasonably well, although there is substantial sensitivity on the choice of the reference dataset (Kharin et al. 2007, 2013; Sillmann et al. 2013). In general, these studies report significant improvements in the agreement after correcting for systematic differences in the mean conditions of temperature extremes.

Temperature extremes are evaluated in terms of 20-yr return levels and compared against those estimated from ERA-Interim and the gridded land-based HadEX2 observational dataset. Our results indicate that CAM4 overestimates both warm and cold extremes over land regions, particularly over the Northern Hemisphere when compared against reanalysis. These differences appear to be more pronounced than those found by Kharin et al. (2007) in a similar assessment using the ensemble mean of models participating in the IPCC AR4 diagnostic exercise and ERA-40 and more recently by Kharin et al. (2013) using the CMIP5 multimodel ensemble median and ERA-Interim. Similar spatial patterns, though less spatially coherent, emerge relative to the HadEX2 dataset. Colder warm extremes, however, arise in high-terrain regions, such as Greenland and the Tibetan Plateau, but these discrepancies are likely the result of biases in the station data measurements, as similar biases were reported in Sillmann et al. (2013) when comparing CMIP5 and HadEX2 TXx climatologies.

We interpret these return level discrepancies in terms of differences in global and regional-scale GEV PDFs, obtained by spatially averaging gridpoint GEV parameter estimates. At these scales, estimates of the shape parameter, which controls the tail behavior of the GEV distribution, are very similar across all datasets for both cold and warm extremes. This allows the analysis to focus on distributional differences based on the two remaining GEV parameters—namely, the location and scale parameters, where the first represents a measure of central tendency while the second, in the present context, is intimately related to the interannual variability of annual extremes. CAM4 GEV PDFs exhibit more extreme return values relative to those of the verification datasets in agreement with the overestimation found at the gridpoint level. The discrepancies originate primarily from differences in location parameter estimates, as represented by a shift of the CAM4 PDFs toward higher (lower) temperatures for warm (cold) extremes. Interannual variability is significantly higher for cold extremes in all datasets. CAM4 consistently overestimates the magnitude of the scale parameter, indicating that interannual variability in annual temperature extremes in CAM4 is too large when compared to observations and reanalysis.

The differences in location parameters are negligible if the annual extremes of the model and reanalysis are adjusted by their respective climatologies of monthly extremes. When adjusted in this way, the agreement between the global and regional GEV PDFs is very good, although the overestimation of the scale parameter persists, particularly for cold extremes over Northern Hemispheric regions. We stress that the match with reanalysis data can be achieved only when adjusting with the bias of the $T_{\max}/T_{\min}$ climatologies, which is an extreme statistic. When debiasing with monthly mean temperature the discrepancy between reanalysis and model remains, confirming that it is not the mean warm/cold bias over land which explains the differences in return values. We note that this overestimation of temperature extremes is also reflected in the diurnal temperature range between both datasets. Considering differences of the 1979–2010 average of the monthly mean diurnal cycle between CAM4 and ERA at the month where the $T_{\max}$ climatology is extreme (not shown), a good correspondence exists with warm extreme return level differences over land (spatial correlation of 0.58), though the relationship does not hold as well for cold extremes at the $T_{\min}$ climatology extreme.

To evaluate the impact of missing subgrid-scale variability, the effect of two stochastic parameterization schemes was studied: a stochastic kinetic energy backscatter (SKEBS) scheme and a stochastically perturbed parameterization tendency (SPPT) scheme. Including a stochastic parameterization noticeably increased the magnitude of warm extremes while reducing that of cold extremes. This is contrary to the effect of adding a white or red noise process to a linear system, which would result in an increase in variability and consequently higher return levels for both $T_{\max}$ and $T_{\min}$ annual extremes (e.g., Berner 2005; Gardiner 2009). Since CAM4 already overestimates extremes, the effect of adding a stochastic parameterization is beneficial for cold extremes but adverse for warm extremes. However, neither of the parameterization schemes meaningfully reduces the overestimation of temperature extremes in CAM4. Unexpectedly, the impact of the two schemes was very similar, although SKEBS is typically most active in the midlatitudes, while SPPT tends to have the biggest impact in the tropics, where convection leads to large tendencies, and hence perturbations, in the physical parameterizations. Although different in nature, both schemes seem to excite modes of variability in CAM4 in such a way that the response in the extremes is the same. Jung et al. (2005) and Berner et al. (2008) provide an example of another process—namely, Northern Hemispheric blocking—where adding a stochastic parameterization does not change its structure but rather only its relative frequency.

Our findings can be best summarized using the schematics shown in Fig. 1. Comparing the model and reanalysis, the distributional differences between annual warm and cold extremes closely resemble Figs. 1d and 1f, respectively, suggesting that the distribution of T in the model exhibits too much variability relative to that of reanalysis. In contrast, the addition of a stochastic parameterization to the model does not induce significant changes on either the shape or scale parameter; instead, the distributions of extremes are shifted toward warmer temperatures, as displayed conceptually in Figs. 1c and 1e. This suggests that the stochastic schemes introduce a systematic bias in the mean rather than enhance the variability of T.

Adjusting for biases in the extremes leads to negligible differences in the location parameters of both perturbed and unperturbed simulations and results in extremes comparing much better with reanalysis. However, adjusting for the bias in mean temperature has a muted effect on the discrepancies in extremes between the models and reanalysis. We conclude that CAM4 misses an important aspect of temperature extremes—namely, the mean statistics of temperature extremes. These can be easily adjusted for in historical data but not necessarily for projections in a changing climate; thus, more attention should be given to this aspect when using climate models for predicting extremes.

Acknowledgments. The authors thank Dani Rae Bundy Coleman for help in the development of the datasets and Daniel Wilks and Claudia Tebaldi for helpful comments. The work of the authors was supported in part by the NSF (Grant CCF-1048915). F. Tagle would like to thank the Climate and Global Dynamics Division at NCAR for their support and hospitality during an extended summer visit.

REFERENCES

Alexander, L. V., and J. M. Arblaster, 2009: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol.*, **29**, 417–435, doi:10.1002/joc.1730.

Berner, J., 2005: Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker–Planck equation. *J. Atmos. Sci.*, **62**, 2098–2117, doi:10.1175/JAS3468.1.

——, F. Doblas-Reyes, T. Palmer, G. Shutts, and A. Weisheimer, 2008: Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Philos. Trans. Roy. Soc. London*, **366A**, 2559–2577, doi:10.1098/rsta.2008.0033.

——, G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, doi:10.1175/2008JAS2677.1.

——, S.-Y. Ha, J. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, doi:10.1175/2010MWR3595.1.

——, T. Jung, and T. Palmer, 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25**, 4946–4962, doi:10.1175/JCLI-D-11-00297.1.

——, K. Fossell, S.-Y. Ha, J. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, doi:10.1175/MWR-D-14-00091.1.

Bindoff, N., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952, doi:10.1017/CBO9781107415324.022.

Brown, S., J. Caesar, and C. A. Ferro, 2008: Global changes in extreme daily temperature since 1950. *J. Geophys. Res.*, **113**, D05115, doi:10.1029/2006JD008091.

Buizza, R., M. Milleer, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, doi:10.1002/qj.49712556006.

Bukovsky, M. S., 2012: Temperature trends in the NARCCAP regional climate models. *J. Climate*, **25**, 3985–3991, doi:10.1175/JCLI-D-11-00588.1.

Chandler, R., and M. Scott, 2011: *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. John Wiley and Sons, 388 pp.

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

Cornes, R., and P. Jones, 2013: How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *J. Geophys. Res. Atmos.*, **118**, 10262–10276, doi:10.1002/jgrd.50799.

Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Diffenbaugh, N. S., J. S. Pal, R. J. Trapp, and F. Giorgi, 2005: Fine-scale processes regulate the response of extreme events to global climate change. *Proc. Natl. Acad. Sci. USA*, **102**, 15 774–15 778, doi:10.1073/pnas.0506042102.

Doblas-Reyes, F., and Coauthors, 2009: Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **135**, 1538–1559, doi:10.1002/qj.464.

Donat, M., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, doi:10.1002/jgrd.50150.

——, J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, doi:10.1175/JCLI-D-13-00405.1.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, 2000: Climate extremes: Observations, modeling, and impacts. *Science*, **289**, 2068–2074, doi:10.1126/science.289.5487.2068.

Efron, B., and R. J. Tibshirani, 1994: *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 456 pp.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

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

Franzke, C. L. E., T. J. O'Kane, J. Berner, P. D. Williams, and V. Lucarini, 2015: Stochastic climate theory and modeling. *Wiley Interdiscip. Rev.: Climate Change*, **6**, 63–78, doi:10.1002/wcc.318.

Gardiner, C. W., 2009: *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics, Vol. 13, 4th ed. Springer, 447 pp.

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73**, 1962–1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.

Hartmann, D., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254, doi:10.1017/CBO9781107415324.008.

Hosking, J., J. R. Wallis, and E. F. Wood, 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251–261, doi:10.1080/00401706.1985.10488049.

Jung, T., T. Palmer, and G. Shutts, 2005: Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.*, **32**, L23811, doi:10.1029/2005GL024248.

——, and Coauthors, 2010: The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor. Soc.*, **136**, 1145–1160.

——, and Coauthors, 2012: High-resolution global climate simulations with the ECMWF model in Project Athena: Experimental design, model climate, and seasonal forecast skill. *J. Climate*, **25**, 3155–3172, doi:10.1175/JCLI-D-11-00265.1.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kharin, V. V., and F. W. Zwiers, 2005: Estimating extremes in transient climate change simulations. *J. Climate*, **18**, 1156–1173, doi:10.1175/JCLI3320.1.

——, ——, X. Zhang, 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, doi:10.1175/JCLI4066.1.

——, ——, ——, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, doi:10.1007/s10584-013-0705-8.

Klein Tank, A., and G. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99.

J. Climate, **16**, 3665–3680, doi:10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2.

Kopparla, P., E. M. Fischer, C. Hannay, and R. Knutti, 2013: Improved simulation of extreme precipitation in a high-resolution atmosphere model. *Geophys. Res. Lett.*, **40**, 5803–5808, doi:10.1002/2013GL057866.

Kunkel, K. E., R. A. Pielke Jr., and S. A. Changnon, 1999: Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. *Bull. Amer. Meteor. Soc.*, **80**, 1077–1098, doi:10.1175/1520-0477(1999)080<1077:TFIWAC>2.0.CO;2.

Leadbetter, M. R., G. Lindgren, and H. Rootzén, 1983: *Extremes and Related Properties of Random Sequences and Processes*. Springer, 336 pp.

Livezey, R. E., and W. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59, doi:10.1175/1520-0493(1983)111<0046:SFSAID>2.0.CO;2.

Mason, P., and D. Thomson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78, doi:10.1017/S0022112092002271.

Neale, R. B., and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM 4.0). NCAR Tech. Note NCAR/TN-485+STR, 212 pp. [Available online at http://www.cesm.ucar.edu/models/ccsm4.0/cam/docs/description/cam4_desc.pdf.]

Palmer, T., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, doi:10.1002/qj.49712757202.

——, and A. Weisheimer, 2011: Diagnosing the causes of bias in climate models—Why is it so hard? *Geophys. Astrophys. Fluid Dyn.*, **105**, 351–365, doi:10.1080/03091929.2010.547194.

——, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. European Centre for Medium-Range Weather Forecasts Tech. Memo., 42 pp.

Rauscher, S. A., E. Coppola, C. Piani, and F. Giorgi, 2010: Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Climate Dyn.*, **35**, 685–711, doi:10.1007/s00382-009-0607-7.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, doi:10.1256/qj.04.106.

Sillmann, J., V. Kharin, X. Zhang, F. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, doi:10.1002/jgrd.50203.

Stephenson, D. B., H. Diaz, and R. Murnane, 2008: Definition, diagnosis, and origin of extreme weather and climate events. *Climate Extremes and Society*, H. Diaz and R. Murnane, Eds., Cambridge University Press, 11–23.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes. *Climatic Change*, **79**, 185–211, doi:10.1007/s10584-006-9051-4.

Weaver, A., and P. Courtier, 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1815–1846, doi:10.1002/qj.49712757518.

Wehner, M. F., 2004: Predicted twenty-first-century changes in seasonal extreme precipitation events in the parallel climate model. *J. Climate*, **17**, 4281–4290, doi:10.1175/JCLI3197.1.

——, R. L. Smith, G. Bala, and P. Duffy, 2010: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dyn.*, **34**, 241–247, doi:10.1007/s00382-009-0656-y.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 704 pp.

Zhai, P., and X. Pan, 2003: Change in extreme temperature and precipitation over northern China during the second half of the 20th century. *Acta Geogr. Sin.*, **58**, 1–10.