Surface Temperature Probability Distributions in the NARCCAP Hindcast Experiment: Evaluation Methodology, Metrics, and Results

PAUL C. LOIKITH,* DUANE E. WALISER,+ HUIKYO LEE,* JINWON KIM,# J. DAVID NEELIN,@ BENJAMIN R. LINTNER,& SETH MCGINNIS,** CHRIS A. MATTMANN,+ AND LINDA O. MEARNS**

*Jet Propulsion Laboratory/California Institute of Technology, Pasadena, California
+Jet Propulsion Laboratory/California Institute of Technology, Pasadena, and Joint Institute for Regional Earth System Science and Engineering, University of Los Angeles, Los Angeles, California
#Joint Institute for Regional Earth System Science and Engineering, University of Los Angeles, and Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California
@Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California
&Department of Environmental Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey
**Institute for Mathematical Applications to the Geosciences, National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 30 July 2013, in final form 12 August 2014)

ABSTRACT

Methodology is developed and applied to evaluate the characteristics of daily surface temperature distributions in a six-member regional climate model (RCM) hindcast experiment conducted as part of the North American Regional Climate Change Assessment Program (NARCCAP). A surface temperature dataset combining gridded station observations and reanalysis is employed as the primary reference. Temperature biases are documented across the distribution, focusing on the median and tails. Temperature variance is generally higher in the RCMs than reference, while skewness is reasonably simulated in winter over the entire domain and over the western United States and Canada in summer. Substantial differences in skewness exist over the southern and eastern portions of the domain in summer. Four examples with observed long-tailed probability distribution functions (PDFs) are selected for model comparison. Long cold tails in the winter are simulated with high fidelity for Seattle, Washington, and Chicago, Illinois. In summer, the RCMs are unable to capture the distribution width and long warm tails for the coastal location of Los Angeles, California, while long cold tails are poorly realized for Houston, Texas. The evaluation results are repeated using two additional reanalysis products adjusted by station observations and two standard reanalysis products to assess the impact of observational uncertainty. Results are robust when compared with those obtained using the adjusted reanalysis products as reference, while larger uncertainties are introduced when standard reanalysis is employed as reference. Model biases identified in this work will allow for further investigation into associated mechanisms and implications for future simulations of temperature extremes.

1. Introduction

As a result of anthropogenic climate warming, mean temperatures are expected to rise; however, changes in temperature extremes are expected to be associated with more substantial climate impacts (Field et al. 2012). In particular, extreme warm events are expected to become more common and severe while the opposite is true for cold extremes (Solomon et al. 2007; Meehl and Tebaldi 2004; Tebaldi et al. 2006; Meehl et al. 2007). Such changes will likely expose populations to extreme heat events that are unprecedented in the current climate (Meehl et al. 2009).

One particularly noteworthy example, the European heatwave of 2003, caused widespread heat-related illness and claimed tens of thousands of lives (Luber and McGeehin 2008). Events of this magnitude, while virtually unprecedented in the current climate, are projected to become more frequent in the future because of climate warming (e.g., Beniston 2004; Schär et al. 2004; Stott et al. 2004). More recently, the 2011 Russian heatwave was also associated with drastically elevated mortality and morbidity resulting from heat stress and

Corresponding author address: Paul C. Loikith, Jet Propulsion Laboratory/California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91101.
E-mail: paul.c.loikith@jpl.nasa.gov

DOI: 10.1175/JCLI-D-13-00457.1

poor air quality associated with wildfires: some studies have speculated that the extreme nature of this event was related to a combination of natural variability and anthropogenic climate forcing (Dole et al. 2011; Rahmstorf and Coumou 2011; Otto et al. 2012). Recent anomalous heat, including the hottest month on record in the United States, coupled with severe drought has had severe impacts on the U.S. agriculture sector (Karl et al. 2012).

Because the relationship between changes in mean temperature and its extremes is often nonlinear, relatively small changes in the mean may be associated with disproportionately large changes in extremes (Hegerl et al. 2004; Griffiths et al. 2005). Therefore, proper simulation of the probability distribution of temperature anomalies is essential for a realistic representation of extremes. Ruff and Neelin (2012, hereinafter RN2012) analyzed surface temperature (T_s) probability distributions from station data and documented several examples of non-Gaussian, often asymmetric long-tailed distributions. They further note the importance of daily T_s distribution shape, especially the distribution tails, in relation to future global warming. In estimating future T_s threshold exceedances, they demonstrate that places exhibiting near Gaussian tails are more sensitive to incremental warming than places with long tails.

Observational evidence points to a recent increase in temperature variance in the tropics as well as a tendency toward more positive skewness globally (Donat and Alexander 2012). On the other hand, Rhines and Huybers (2013) suggest that observed changes in summertime extremes are primarily attributable to changes in the mean rather than the variance. Lau and Nath (2012) demonstrate shifts in the probability distribution functions (PDFs) of daily maximum temperature in two high-resolution global climate models (GCMs) by the middle of the twenty-first century with only small changes in PDF shape exhibited in some places.

To quantify uncertainty in simulations of future climate, it is important to bring as much observational scrutiny as possible to historical climate model runs. Model evaluation is critical for identifying the range of error (magnitude, geographic distribution, sign) across models for the same region. Comprehensive evaluation of GCMs archived as part of phase 3 of the Coupled Model Intercomparison Project (CMIP3) was performed by Gleckler et al. (2008); however, the demand for more geographically specific climate projections has increased the prominence of limited domain regional climate models (RCMs). While the body of systematic RCM evaluation work is less mature than that for GCM evaluation, some studies have evaluated important variables in RCMs. Kjellström et al. (2011) analyze a suite of RCM hindcast and future projections driven by reanalysis and multiple GCMs over Europe. Kim et al. (2013) evaluate mean surface temperature, precipitation, and insolation using monthly mean data over the conterminous United States using models from the North American Regional Climate Change Assessment Program (NARCCAP).

Specifically focusing on PDFs, Perkins et al. (2007) introduced a PDF skill score to evaluate global models and applied this method over Australia, using climatologically homogenous subregions to compute PDFs of temperature and precipitation. Kjellström et al. (2010) used this method to evaluate temperature and precipitation PDF structure over Europe while also evaluating daily temperature at multiple percentiles of the distribution. Their results show that while some models perform better than others, no model is systematically better or worse in every region or season, suggesting substantial variability in the way RCM bias is manifested.

Comprehensive evaluation of PDF morphology is expected to provide information regarding model representation of extremes and to enhance mechanistic understanding of processes responsible for genesis of extremes. To this end, the present study evaluates RCM-simulated PDF characteristics over North America. The remainder of this paper is organized as follows. Section 2 describes the data and methodology used. Section 3 presents daily temperature bias at different percentiles of the distribution, and section 4 evaluates model variance and skewness across the domain and for select example locations exhibiting non-Gaussian long-tailed PDFs. The sensitivity of the evaluation results to the choice of reference dataset and interpolation procedure is presented in the discussion in section 5, followed by concluding remarks in section 6.

2. Data and methodology

a. NARCCAP data

NARCCAP was designed to serve the high-resolution climate modeling needs of the United States, Canada, and Mexico. All six RCMs used in this paper are dynamically downscaled hindcast experiments performed for NARCCAP (Mearns et al. 2009, 2012, 2013; http://www.narccap.ucar.edu). Information about the individual RCMs is presented in Table 1. In this work, all hindcast model simulations were driven by large-scale forcing from the National Centers for Environmental Prediction (NCEP)–U.S. Department of Energy (DOE) Reanalysis 2 (Kanamitsu et al. 2002).

While the official NARCCAP hindcast time period spans 1979–2004, the period 1980–2003 is used to span

TABLE 1. The RCMs and corresponding references evaluated in this study.

Model	Model name	References
CRCM	Canadian Regional Climate Model	Caya and LaPrise (1999)
ECP2	Scripps Experimental Climate Prediction Center Regional Spectral Model	Juang et al. (1997)
HRM3	Third-generation Hadley Centre Regional Climate Model	Jones et al. (2004)
MM5I	Fifth-generation Pennsylvania State University (PSU)–NCAR Mesoscale Model, Iowa State University version	Grell et al. (1993)
RCM3	International Centre for Theoretical Physics Regional Climate Model version 3	Giorgi et al. (1993a,b)
WRFG	Weather Research and Forecasting Model, Pacific Northwest National Laboratory version	Skamarock et al. (2005)

the longest possible time period for which all models have available T_s. The simulation domain covers most of North America and some of the adjacent Pacific and Atlantic Oceans. Each model is originally provided on a 50-km native curvilinear grid at 3-hourly temporal resolution.

b. *Reference data*

The Wang and Zeng (2014) 2-m temperature dataset based on the National Aeronautics and Space Administration (NASA) Modern Era-Retrospective Analysis for Research and Applications (MERRA) reanalysis is used as the primary reference. Described in Wang and Zeng (2013), this is one of four datasets of hourly gridded T_s based on four reanalysis products and the Climate Research Unit Time Series version 3.10 (CRU TS3.10; Mitchell and Jones 2005). To produce these products, data from MERRA, the NCEP–National Center for Atmospheric Research (NCAR) Reanalysis 1, the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40), and ECMWF Intermim Re-Analysis (ERA-Interim) products are first interpolated to the 0.5° CRU TS3.10 grid. With the exception of MERRA, which is originally available at hourly resolution, the reanalysis is temporally interpolated to hourly resolution. Finally, the data are bias corrected using the CRU TS3.10 monthly mean daily maximum and minimum temperature. The bias correction reduces the uncertainty inherent in the reanalysis.

No additional uncertainty is introduced by the temporal interpolation for the MERRA-based product (MERRA–CRU) and therefore this dataset is chosen as the primary reference. Results were also computed using products based on the ERA-Interim and NCEP–NCAR (ERA-40 did not have complete overlap with the NARCCAP period) to assess uncertainty across the suite of datasets. The sensitivity of the evaluation results to the choice of dataset is discussed further in section 5a.

The NCEP North American Regional Reanalysis (NARR; Mesinger et al. 2006) and standard MERRA (Rienecker et al. 2011) products are also employed to compare how the results change when using original forms of reanalysis versus the CRU TS3.10 adjusted products. Implications are discussed further in section 5a. NARR is produced on a Lambert conformal grid with 32-km resolution. Developed by NASA's Global Modeling and Assimilation Office and disseminated by the Goddard Earth Sciences Data and Information Services Center (GES DISC), MERRA is originally on a global 0.5° × 0.67° latitude–longitude grid.

c. *Data processing methodology*

The analyzed model and reference data comprise daily means of the NARCCAP 3-hourly and the MERRA–CRU hourly output respectively. All NARCCAP data are interpolated to a common 0.5° latitude–longitude grid mesh, the same as the MERRA–CRU grid. Data were interpolated using a kriging algorithm implemented with a thin plate spline (TPS) routine (Fields Development Team 2006) using surface elevation as a covariate and performed only over land grid points. The sensitivity of the evaluation results to the choice of regridding scheme is also evaluated using linear and cubic methods based on Delaunay triangulation (Lee and Schachter 1980) and discussed in section 5b.

All data are subset to a common land-only domain, which covers the maximum possible spatial overlap of all datasets. Temperature anomalies, obtained by subtracting the daily climatological average from each daily value, are used in the computation of several metrics in this paper. Bukovsky (2012) found reasonable temporal trend agreement between the NARCCAP models and observations over this period so it was decided to not remove any trends. Evaluation was performed for the seasons of summer [June–August (JJA)] and winter [December–February (DJF)]. The multimodel ensemble mean is calculated by concatenating the daily data from each of the six RCMs into one time series consisting of six data points (one for each RCM) at each grid point at each time step.

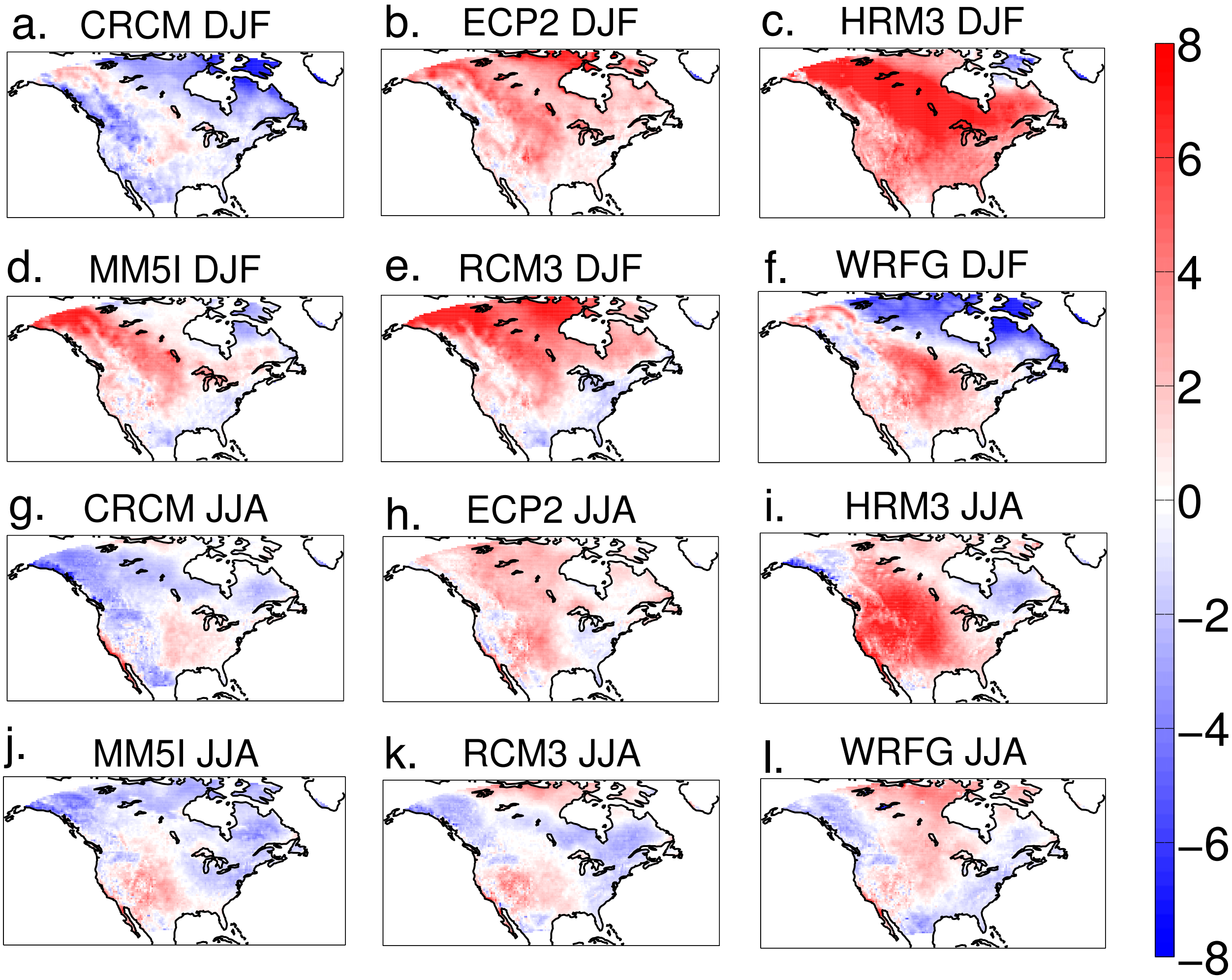

FIG. 1. Bias (°C) of the 50th percentile of the daily surface temperature distribution for (a)–(f) DJF and (g)–(l) JJA for each RCM with respect to MERRA–CRU as discussed in section 3.

3. Percentile-based T_s evaluation

Temperature biases at three percentile thresholds of the daily temperature distribution (5th, 50th, and 95th) were calculated for each model with respect to MERRA–CRU. The 5th (95th) percentiles are chosen to be representative of cold (warm) extremes and approximate temperature in the tails of the distribution. Figures 1a–f show the bias in median DJF temperature for each of the six NARCCAP RCM hindcasts. While the errors differ in sign and magnitude, all models exhibit a warm bias over the central and northern Great Plains. HRM3 exhibits the largest warm biases with magnitudes exceeding 8°C over much of the domain while CRCM has an overall cold bias.

All models have an area of positive median temperature bias in JJA (Figs. 1g–l) over a portion of the Great Plains with HRM3 again being the warmest (~6°–8°C). All RCMs also exhibit a warm bias along the Pacific coast of California and Baja California in JJA. This systematic warm bias suggests that the RCMs are not properly capturing the moderating influence of the relatively cool Pacific Ocean along the coast. Overall, the biases shown in both DJF and JJA are in qualitative agreement with other studies that calculated bias in the mean (Kim et al. 2013; Sobolowski and Pavelsky 2012; Mearns et al. 2012) and daily maximum and minimum temperature (Rangwala et al. 2012), suggesting that these biases are robust features of the overall temperature distribution and throughout the diurnal cycle. Much of the warm bias over the western and southwestern United States is coincident with a dry bias in mean precipitation during the summer (Mearns et al. 2012). Bukovsky et al. (2012) found that the NARCCAP RCMs were unable to properly develop a realistic monsoon structure, in particular over Arizona, contributing to a dry bias here. It is

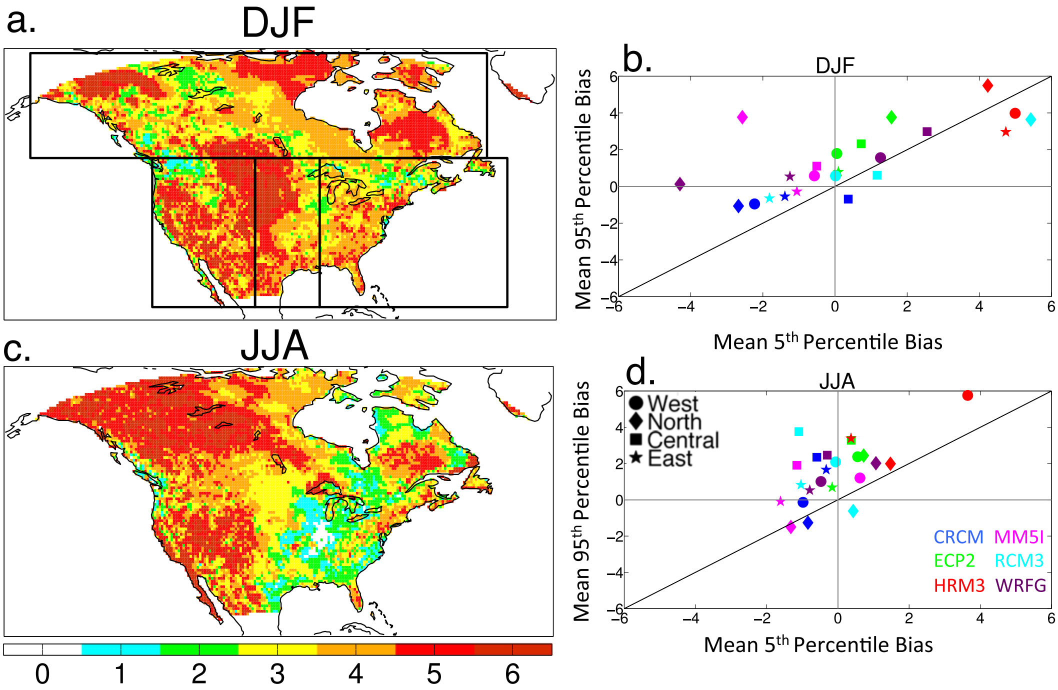

FIG. 2. (left) Number of ensemble members out of six that have the same sign bias at the 5th, 50th, and 95th percentiles of the temperature distribution for (a) DJF and (c) JJA. A value of 6 (0) indicates that all (no) RCMs have positive or negative bias for all three percentiles. The black boxes in (a) outline the four subregions used in subsequent analysis and are described in section 3. (right) The mean bias (°C) of the 5th vs the 95th percentiles of the temperature distribution for each ensemble member for each subregion as outlined in (a), for (b) DJF and (d) JJA. The color and symbol legends are in (d). The black diagonal line indicates where the symbol would lie if the bias were the same for both percentiles. Results are discussion in section 3.

likely that the systematic warm bias is related to a dry soil moisture bias resulting, at least in part, from an unrealistically low production of convective precipitation associated with the North American monsoon and central plains mesoscale convective systems. Such processes are likely not properly resolved at the 50-km resolution of the NARCCAP simulations.

Figures 2a and 2c show maps of the total number of RCMs that have bias of the same sign at the 5th, 50th, and 95th percentiles (i.e., systematically cold or warm biased across the distribution as estimated at these thresholds). In general, much of the central and western portions of the domain show a plurality of models with a systematic bias in DJF while the same tendency occurs for the northern and western regions in JJA. All RCMs show a systematic warm bias in DJF over the Central Valley of California except for CRCM, which has a systematic cold bias, as evident in Fig. 1. RCM counts are low in the U.S. Midwest or over much of Ontario and northern Quebec in Canada for JJA. These regions tend to have warm biases at the 50th and 95th percentiles, but cold biases at the 5th percentile (not shown).

For additional perspective, the domain is further decomposed into four subregions (black boxed in Fig. 2a), chosen to broadly represent climate regimes and defined as follows: West, including the U.S. Pacific Coast and Rocky Mountains; North, including the Canadian Rockies east to Newfoundland; Central, including most of the U.S. Great Plains; and East, covering the Great Lakes region, the U.S. Southeast, and the Atlantic coast. Figures 2b and 2d show scatterplots of the mean bias for each RCM and each subregion at the 5th percentile versus the 95th percentile. The diagonal black line indicates where the RCMs would lie if they had the same bias at both percentiles, indicating a completely symmetrical shift of the distribution tails, as estimated at these percentiles. RCMs to the left of the diagonal line have a wider PDF than MERRA–CRU while RCMs to the right have a narrower PDF. Similarly, RCMs falling in the lower left and upper right quadrants are colder and warmer at both tails while the biases are of opposite sign for each tail in the upper left and lower right quadrants. In both DJF and JJA, most models have a net widening with fewer models showing a net narrowing. In DJF, many of the RCMs fall near the one-to-one line, consistent with the generally high values in Fig. 2a. In the North subregion, MM5I and WRFG are outliers with large net widening. The Central (squares) and East

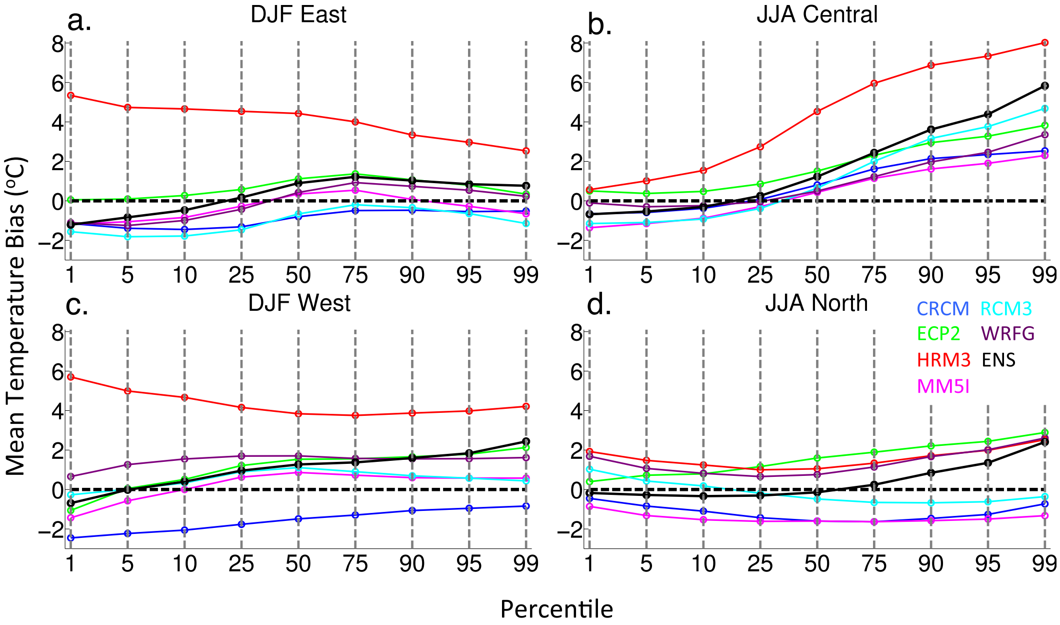

FIG. 3. Mean bias averaged over the (a) East and (c) West subregions for DJF and (b) Central and (d) North subregions for JJA for the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles. The line colors correspond to the legend in (d) and the black line is the multimodel ensemble. The subregions are defined in Fig. 2a.

(stars) subregions during JJA are the farthest away from the one-to-one line, consistent with the low values in Fig. 2c for the same regions.

Figure 3 shows the mean bias for nine percentiles across the PDF for each RCM and the multimodel ensemble in a similar format to that used in Kjellström et al. (2010). Mean biases are computed for the East and West subregions in DJF and the Central and North subregions in JJA. The outstanding HRM3 warm bias persists across the distribution in both DJF and JJA Central subregion while CRCM is systematically among the coldest RCMs. The intraensemble spread increases from low to high percentiles for JJA Central subregion with very large positive biases in the warm tail. While all RCMs produce extreme heat events that are unrealistically severe over the Central subregion, HRM3 stands out with a mean warm bias of nearly 8°C at the 99th percentile.

In some cases, the spread in bias within a subregion can dampen the mean values plotted in Fig. 3. For example, the mean bias at the 50th percentile for HRM3 JJA North subregion in Fig. 3d is small; however, in Fig. 1i it is evident that some of this low bias is a result of averaging warm and cold biases and not an indication of superior model performance. Figure 4 uses a box-and-whisker format to show the extent of the spread in biases at the 5th and 95th percentiles. In many cases, the mean biases in Fig. 3 are representative of the region. For example, HRM3 shows a predominantly warm bias for all four panels in Fig. 3. In many cases the RCMs showing low mean bias in Fig. 3 also have a narrow range of bias values, indicating high model fidelity. CRCM, ECP2, and MM5I for DJF East subregion and JJA Central subregion exemplify this. In other cases, the box-and-whisker plot identifies cases where the low mean bias can be deceptive. The WRFG 5th percentile mean biases are low for DJF East and West subregions; however, there are a considerable number of grid points with relatively large negative and positive biases. This is also the case for HRM3 JJA North subregion 5th and 95th percentiles, similar to the 50th percentile bias in Fig. 1i.

4. Evaluation of variance and skewness

Whereas the analysis in section 3 focused on temperature bias at multiple percentiles to estimate differences in model-simulated PDFs, this section uses higher-moment statistics to evaluate the shape of the distributions. Because of the important relationship between temperature variability, the length of the distribution tails, and extremes, the standard deviation (SD) and skewness of model simulated T_s are compared against MERRA–CRU. In what follows, all analyses use temperature anomalies to allow for easy comparison of PDF shape between datasets as all have a mean of 0.

a. *Standard deviation*

The ratios between the SD of daily T_s for each model and MERRA–CRU in DJF are displayed in Fig. 5.

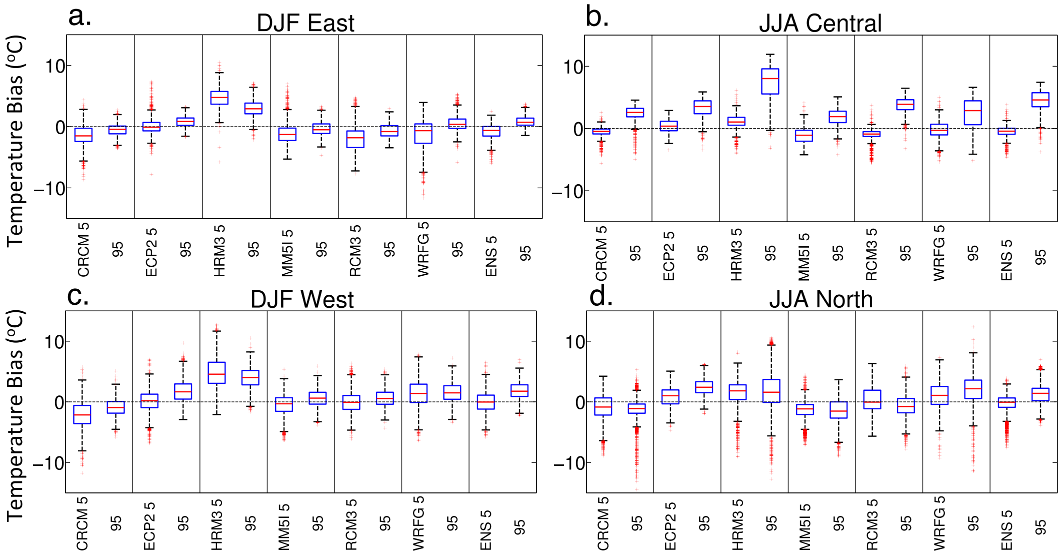

FIG. 4. Box-and-whisker plots showing the temperature bias (°C) for every grid cell within the indicated subregion for the 5th and 95th percentiles of the temperature distribution (indicated with a 5 and 95 after the RCM name respectively). Plots are for seasons and subregions (a) DJF East, (b) JJA Central, (c) DJF West, and (d) JJA North as in Fig. 3. The red line indicates the median temperature bias and edges of the blue box are at the 25th and 75th percentiles. The black dashed lines extend to 1.5 times the interquartile range with outliers plotted as red plus signs.

Values greater (less) than one indicate where the model has a higher (lower) SD than MERRA–CRU. To test for significance, a bootstrapping procedure is applied as follows. For a given RCM/MERRA–CRU pair for a given year, the data pair is randomly determined to remain the same or be swapped. In other words, if a coin flip resulted in heads, the data pair would remain the same. If a coin flip resulted in tails, the entire MERRA–CRU and RCM 90-day (for DJF) season would be swapped so that MERRA–CRU would have the RCM data and the RCM would have the MERRA–CRU data. This is repeated for each year, keeping the entire seasons intact. Once a new randomly generated pair of datasets is constructed, the ratio of the SD of the new RCM to the new MERRA–CRU is computed. This is repeated 1000 times and the ratio is determined to be significant if the two-tailed p value is less than 0.01. Only significant grid cells are shaded in Fig. 5. The RCMs generally show values greater than one in the northern portion of the domain, with the exception of RCM3, and lower than or near one over the southern Great Plains and southeastern United States.

In many examples (MM5I, CRCM, HRM3, and WRFG) the most striking areas of positive bias are present to the north of the region of maximum SD in MERRA–CRU (Fig. 5a). The band of high variance (stretching from the northwest corner of the domain southeast along the eastern edge of the Canadian Rockies and into the northern Great Plains) is in an area highly influenced by large temperature fluctuations due to synoptic-scale weather events associated primarily with warm advection from lower latitudes and cold advection from higher latitudes (Loikith et al. 2013). Areas north of this region are among the coldest in the hemisphere, limiting daily temperature variability on the cold side of the PDF and leading to lower variance. For this reason, much of the variability in daily temperature occurs only on the warm side of the PDF here. This is in contrast to the band of higher SD to the south, which is characterized by a PDF that is more symmetrical about the mean. The tendency for the models to have positive SD bias north of this region of climatologically large variance indicates that models expand this high variance region substantially northward compared with MERRA–CRU. One possible mechanism for this feature is a storm track that is displaced or extended too far north. The notable negative bias in HRM3 in the southern half of the domain and positive bias in the northern third indicates that the band of high variance apparent in MERRA–CRU is diminished in the south and enhanced or extended in the north. Coupled with the outstanding warm bias (Fig. 1c), this may suggest a storm track that is displaced substantially northward.

Figure 6 shows the SD ratios for JJA. While the daily temperature variability is lower in the summer compared with winter, resulting in overall lower SD values, the ratio is generally higher than for DJF. Overall, SD is higher in all six models over most of the domain with values below one in the north in CRCM, MM5I, and RCM3. All RCMs experience a notable positive variance

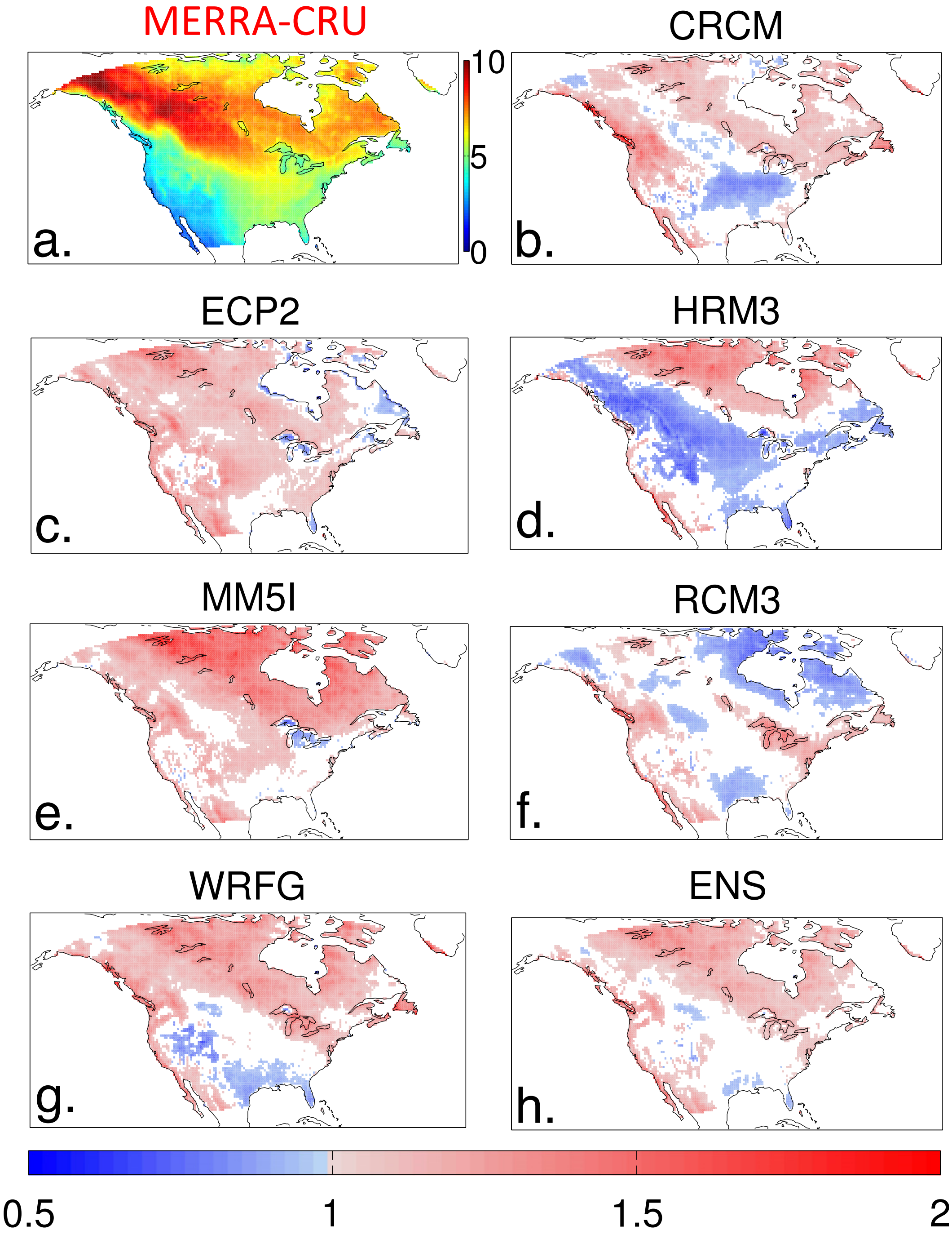

FIG. 5. (a) DJF standard deviation (°C) for MERRA–CRU. (b)–(h) Standard deviation ratios for DJF computed as the ratio of the standard deviation of daily temperature anomalies for the RCM to the standard deviation of the daily temperature anomalies for MERRA–CRU. Only grid cells determined to be statistically significant (two-tailed p value <0.01) according to a bootstrapping procedure are shaded (see section 4a for discussion).

bias along and to the north of the Gulf of Mexico coast where MERRA–CRU shows relatively small standard deviations (~1°–2°C).

Notably, all datasets have higher SD along the Pacific Coast. Climate variability here is influenced largely by occasional offshore wind events producing anomalously warm T_s values (e.g., Santa Ana events in southern California; Hughes and Hall 2010). It is possible that the positive SD bias is indicative of a tendency for more frequent and/or intense offshore wind events. Ratios are

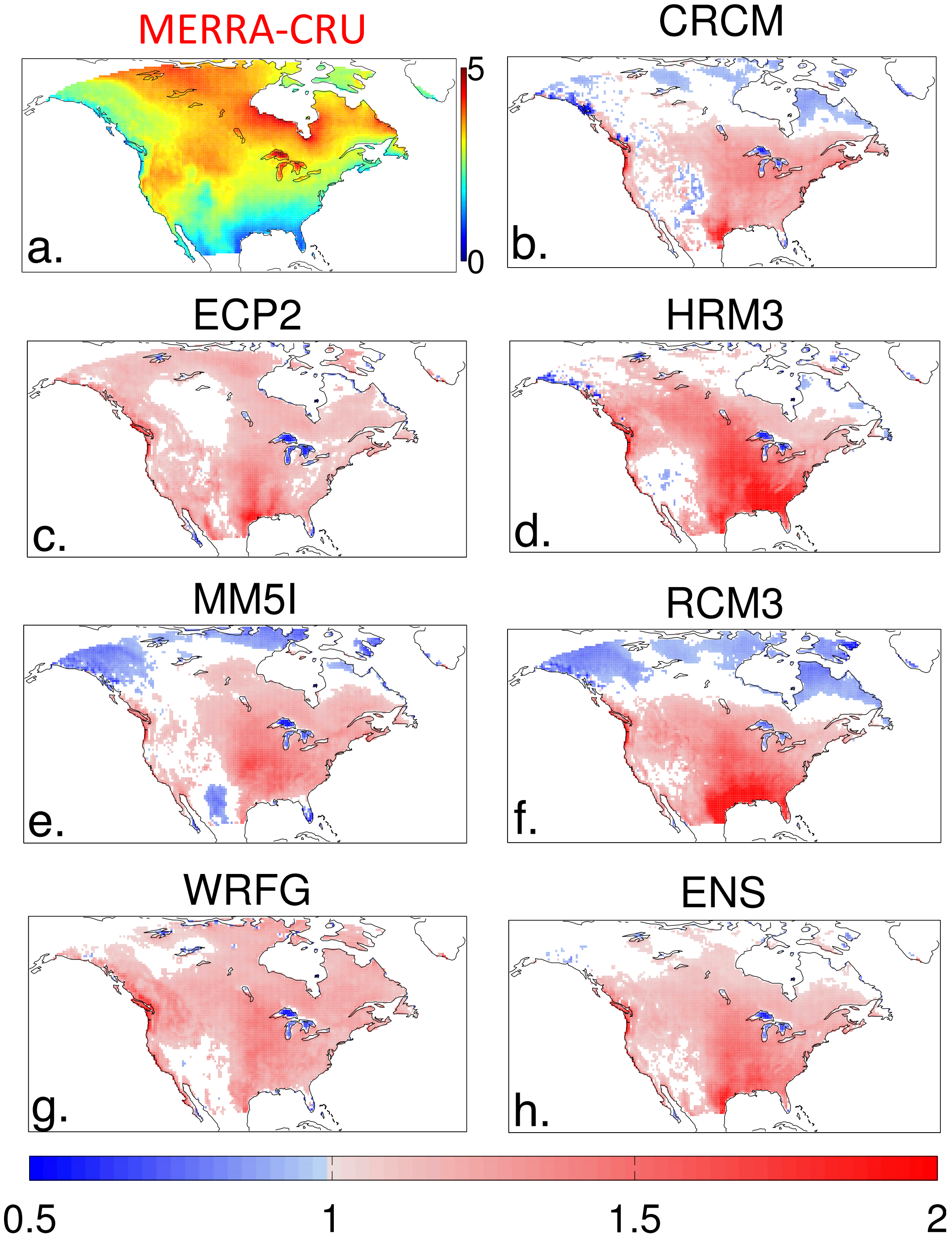

FIG. 6. As in Fig. 5, but for JJA.

generally close to one and largely nonsignificant over the U.S. Southwest and the Rocky Mountains.

Results of the evaluation of SD are summarized for the four subregions in Table 2. The values represent the spatial mean of the ratios as calculated and plotted in Figs. 5 and 6, except all grid cells contribute to the mean, not just significant ones. When averaged over the subregions, variance ratios are generally very close to one indicating that the variance over- or underestimates are not too severe in most cases. The relatively large ratios in JJA over the south-central United States are reflected in the elevated mean values for the JJA Central and East subregions. Of all regions and seasons, DJF for the Central subregion shows the ratios closest to one

TABLE 2. Mean standard deviation ratios by subregion for DJF and JJA. All values are used including those from grid cells that are not determined to be statistically significant.

	CRCM		ECP2		HRM3		MM5I		RCM3		WRFG		ENS	
DJF 5th	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA
West	1.1	1.1	1.2	1.2	1.0	1.2	1.1	1.1	1.1	1.2	1.1	1.2	1.1	1.2
North	1.1	1.0	1.1	1.1	1.1	1.1	1.3	1.0	1.0	1.0	1.2	1.2	1.2	1.1
Central	1.0	1.3	1.1	1.3	0.9	1.7	1.1	1.4	1.0	1.7	1.0	1.3	1.0	1.5
East	1.1	1.3	1.1	1.1	0.9	1.4	1.1	1.2	1.1	1.3	1.1	1.2	1.1	1.3

across the ensemble where synoptic-scale meteorology, generally unimpeded by complex topography or coastal zones, dominates daily temperature variability. This suggests that the RCMs simulate storm strength and frequency with reasonable fidelity here.

b. T_s skewness

As opposed to variance, which primarily describes the width of the PDF, skewness is more directly related to extreme values as it describes the shape of the tails and the degree of PDF symmetry. The models capture the MERRA–CRU (Fig. 7a) large-scale skewness pattern, with positive skewness in the northeast, a large coherent region of relatively strong negative skewness extending from the northwest to the Great Lakes, and modest skewness over the southeastern United States. The most notable differences lie in the magnitude of skewness. HRM3 is an outlier and has negative skewness extending much farther north than MERRA–CRU, which may have a physical relation to the fact that it is also the warmest RCM at all percentiles (e.g., Fig. 1c). It is interesting to note that the transition zone from primarily negative (south) to positive (north) skewness corresponds to the band where relatively few models have bias of the same sign at all percentiles (Fig. 2). Models may have difficulty accurately capturing this spatial transition in T_s regime, leading to errors in the simulated PDF shape. Notably, no RCMs have the negative to positive skewness transition biased to the south.

Model fidelity in simulating skewness in winter is likely indicative of differences in the simulation of large-scale climate mechanisms, including mechanisms associated with extremes (Loikith et al. 2013). Details of these mechanisms and their relationship to extremes are examined in an observational study by Loikith and Broccoli (2012). For example, in winter the PDFs in the northern region have long warm tails resulting from advection of relatively warm air from lower latitudes. Advection of cold anomalies of comparable magnitude from the north rarely occurs because the air in this region is climatologically among the coldest in the hemisphere, reducing variability on the cold side of the tail as discussed in section 4a. Models that show more restricted regions of positive skewness (e.g., HRM3 and MM5I) would generate extreme warm events less frequently than in MERRA–CRU; models that show stronger positive skewness in this region (e.g., RCM3 and WRFG) may simulate the occurrence of warm advection events too frequently in the region. In addition to having skewness that is more positive than MERRA–CRU, WRFG also has a colder background climate in this region (Figs. 1, 3), with a warm bias to the south. Under conditions of northward advection into the cold-biased region, extreme warm anomalies may occur that contribute to the positive skewness bias. RCM3 has similar skewness error as WRFG, but with warm biases over this region and cold biases to the south, making it more difficult to propose a mechanism here.

Another illustrative example in DJF is the region of negative skewness encompassing Oregon, Washington, and British Columbia. Climate in this region is generally dominated by cool maritime air that suppresses the occurrence of extreme warm events, especially close to the coast. Extreme cold events occur when air originating from high continental latitudes is advected into the region. Such events are rare, however, because inland mountain ranges prevent cold, dense, and often shallow Arctic air masses from advecting westward. Many RCMs exhibit skewness that is more negative than the reference here. This suggests that these datasets may generate more frequent, severe, and extensive extreme cold air outbreaks than in reality. HRM3 and CRCM capture this feature with the highest fidelity. Here, HRM3 is the only model that generates substantially warm biases at the 5th percentile (Fig. 3c), supporting the hypothesis that the more negative skewness simulated by most models results from unrealistically frequent cold outbreaks. In all cases except for HRM3, the biases at the 95th percentile are warmer than at the 5th percentile, further contributing to an asymmetry in model error that disproportionately affects days in the cold tail.

Figure 8 shows skewness for JJA. MERRA–CRU shows predominantly negative or weak skewness throughout the domain with positive skewness along the Pacific coast. The differences between the models and the reference data are more substantial than in DJF; however, many

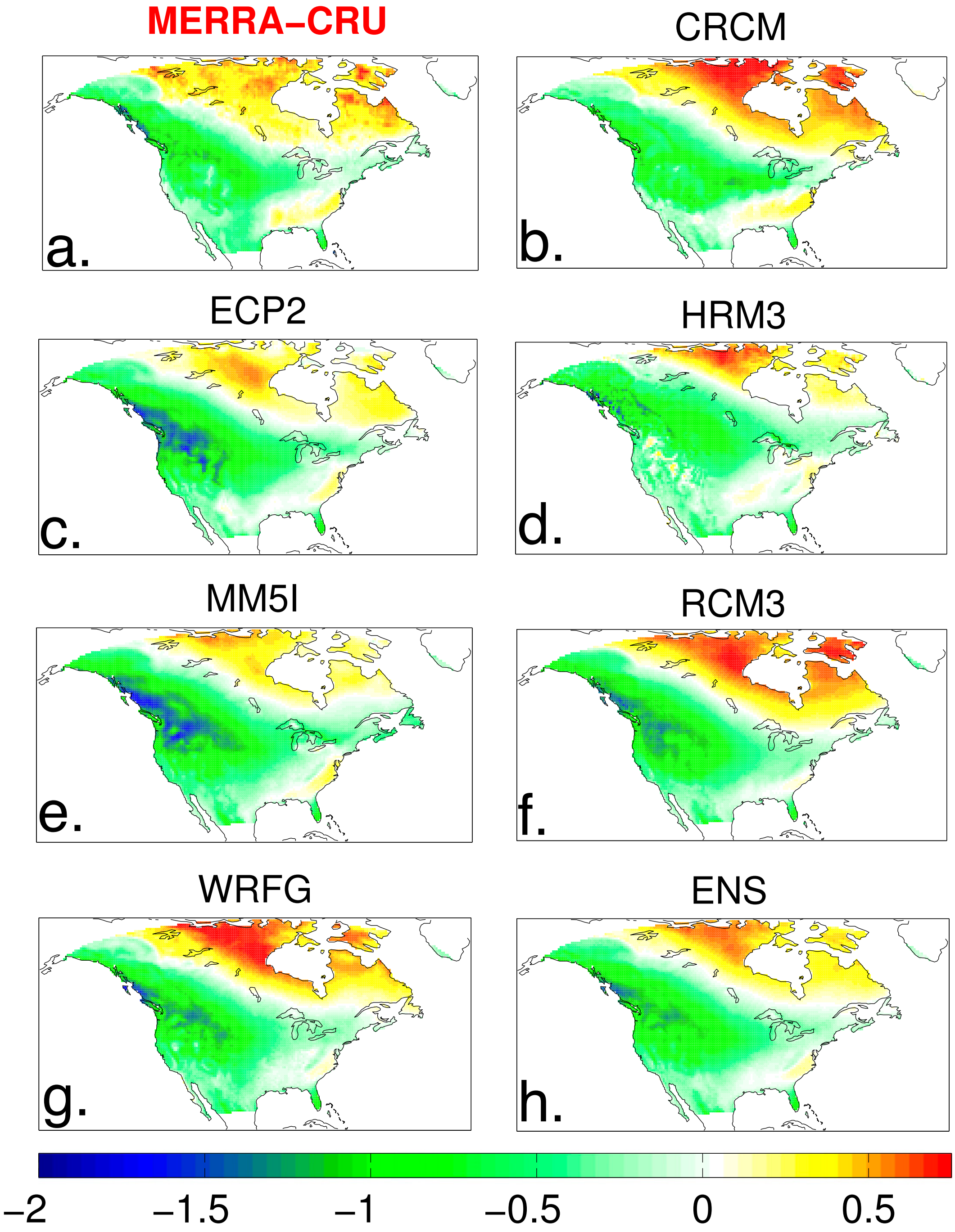

FIG. 7. Skewness of DJF daily temperature anomalies for (a) MERRA–CRU (as reference), (b)–(g) the RCMs, and (h) the multi-RCM ensemble. See section 4b for discussion.

features are realistically reproduced. For example, the negative skewness over the Rocky Mountains is captured by all RCMs. This is coincident with low standard deviation ratios in Fig. 6, suggesting that the RCMs are reproducing the PDFs with skill here. All ensemble members also reasonably capture the band of positive skewness along the Pacific coast. Here, the moderating effects of the Pacific Ocean inhibit both cold and warm extremes most of the time. Occasional offshore wind events block the moderating effects of the ocean and lead to large excursions on the warm side of the distribution. The agreement here suggests that the RCMs are able to realistically capture these rare events.

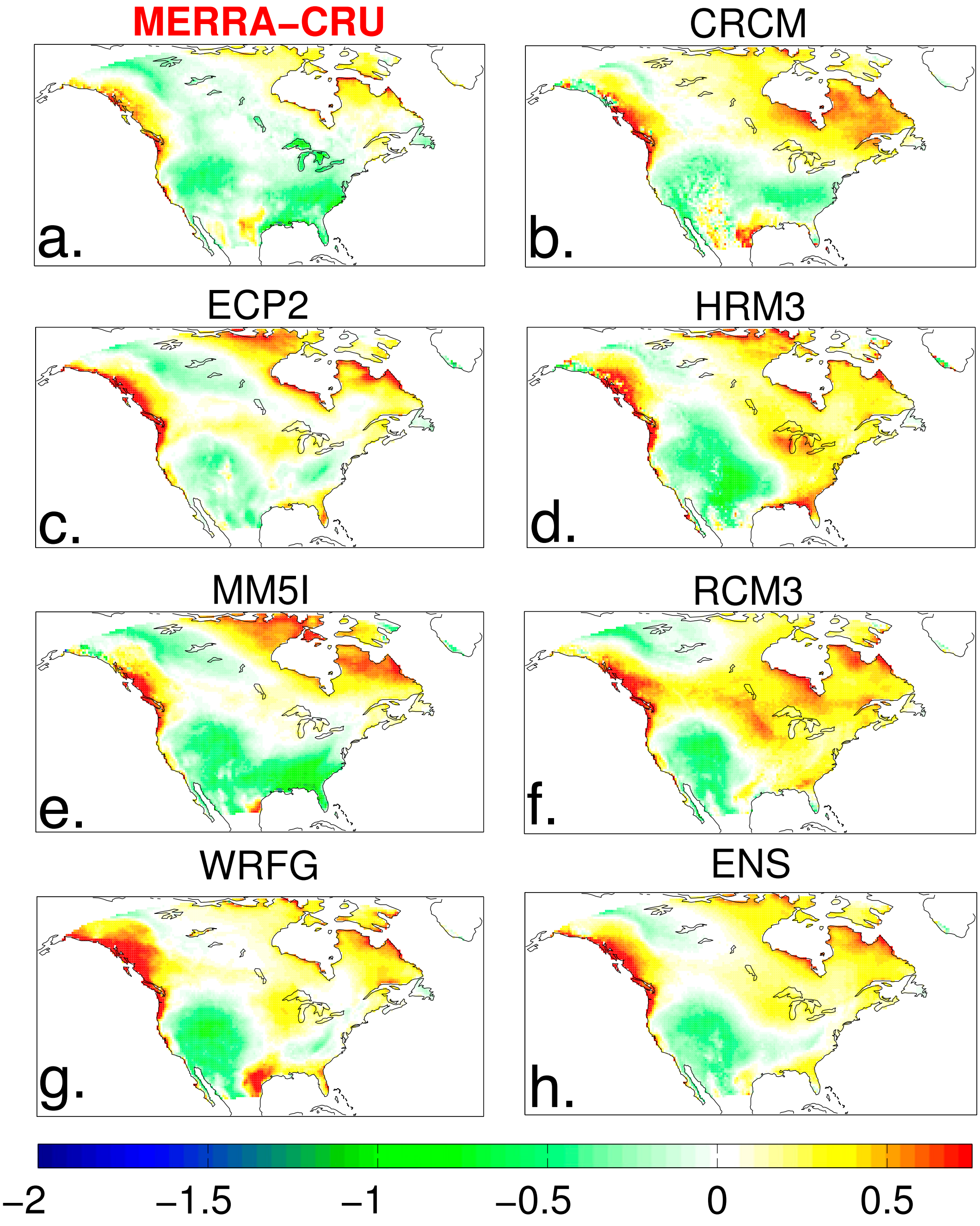

FIG. 8. As in Fig. 7, but for JJA.

Skewness is not as well reproduced along the Gulf of Mexico coast and over the eastern United States. MERRA–CRU shows predominantly negative skewness along the U.S. Gulf Coast, while all RCMs but MM5I show predominantly positive skewness. This area also coincides with large positive variance biases in Fig. 6, with the most far reaching and strongest biases for HRM3 and RCM3. These RCMs also show the largest disagreement for skewness, indicating substantial error in the overall RCM-simulated PDFs here. Along the U.S. Gulf Coast during summer, daily temperature variability is low and the occurrence of synoptic-scale weather events that are often associated with advection of anomalous T_s are rare. As a result, the tails of the distribution are likely influenced largely by variations in insolation, precipitation, and land surface conditions. For example, soil moisture has been associated with the occurrence and implicated as a source of amplification and persistence for heat waves (Hong and Kalnay 2000; D'Odorico and Porporato 2004; Fischer et al. 2007; Loikith and Broccoli 2014). On the other hand, decreased insolation because of clouds and evaporative cooling from rain can result in anomalously cool temperatures and climatologically humid air originating from the Gulf of Mexico may enhance latent heat flux sufficiently to limit extreme warm events here. This thermodynamic limitation on extreme warmth combined with more opportunity for unusually cool days likely results in the negatively skewed PDF here. The notable RCM disagreement may result from difficulties in producing realistic convective clouds and precipitation that result in relatively large excursions on the cold side of the PDF.

Figures 7 and 8 are quantitatively summarized using a Taylor diagram in Fig. 9. Consistent with the qualitative discussion above, all RCMs perform similarly well in DJF with the exception of HRM3. While all ensemble members exhibit larger spatial variance of skewness than MERRA–CRU, as indicated by variance ratios greater than one, the spatial variances of skewness for CRCM and the ensemble are the closest to the reference. Figure 9 reflects the large differences between MERRA–CRU and the RCMs for JJA skewness, with HRM3 showing the largest error for JJA as well. All RCMs in JJA also have variance ratios greater than one except for ECP2 while the multi-RCM ensemble (ENS) has nearly the same variance as MERRA–CRU.

c. Individual cases

The PDFs for four individual grid points are plotted in Fig. 10. Each case corresponds to an example identified in RN2012 as having non-Gaussian behavior in at least one tail. All locations are chosen as the closest grid point to the actual observation station located at the major airport for each city used in RN2012. The 0.5°C resolution of the data makes it difficult to make a fair quantitative grid point to station comparison, especially in areas of complex terrain. Therefore these examples are intended as a qualitative evaluation of the ability of the RCMs to reproduce key features of the probability distributions documented in RN2012. All distributions are defined as frequencies of occurrence computed from temperature anomalies binned every 0.5°C. For reference, Gaussian PDFs are plotted with the same SD as MERRA–CRU.

All datasets exhibit a short warm tail and a long cold tail in DJF for Seattle (Fig. 10a) and Chicago (Fig. 10b), supported by the negative skewness values. In both of these locations, RN2012 show long cold tails, with the asymmetry more pronounced in Seattle. In the scenario of a uniform shift in the PDF toward warmer conditions, both locations would experience a greater increase in warm extremes compared with locations with a long warm tail and a smaller decrease in cold extremes compared with a location with a Gaussian or short cold tail. In general, the multimodel ensemble variance and skewness are very similar in both cases. Figure 8 indicates that in all datasets, Seattle is positioned near the strongest (coastal) part of a long, large-scale feature of negative skewness that stretches from the U.S. West Coast to near Chicago. This suggests a substantial role of large scales in the air mass advection creating these long cold tails. While this may make it less surprising that the models do qualitatively well at capturing the long tail in this region, it also helps to boost confidence in using these models to predict changes in this feature.

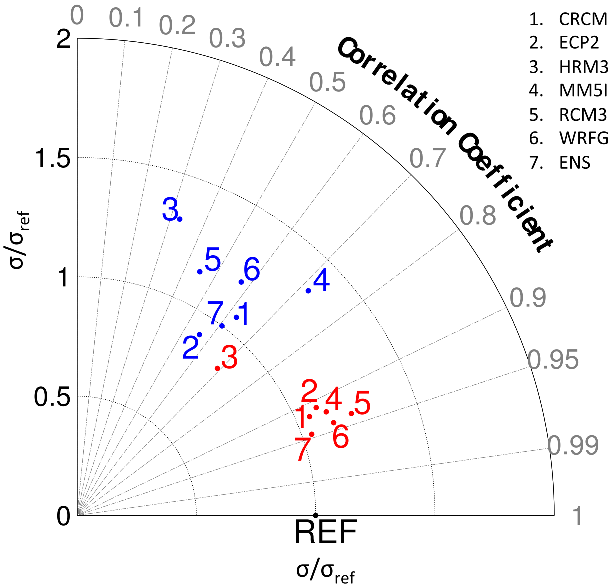

FIG. 9. Taylor diagram comparing the spatial patterns of skewness for DJF (red) and JJA (blue). The values along the radial axes are the ratio of the RCM spatial standard deviation of skewness to the reference spatial standard deviation of skewness. The azimuthal axis is the pattern correlation and the distance from the REF point (MERRA–CRU) is equivalent to the centered root-mean-square error, normalized by the reference spatial standard deviation. Skewness values are weighted by the square root of the cosine of latitude before computing the Taylor diagram metrics. Each RCM is labeled by a number as defined in the legend in the upper-right corner of the figure.

Figure 10c is for Houston, Texas, where RN2012 show a wide cold tail, similar to MERRA–CRU. For this case, all RCMs show a wider distribution at both tails (with the exception of WRFG on the cold side) with a fairly symmetrical multi-RCM ensemble distribution. The larger variance and wider tails suggests that the RCMs may oversimulate conditions such as anomalously low soil moisture associated with extreme warmth while also oversimulating days with heavy rainfall and low insolation associated with cool conditions.

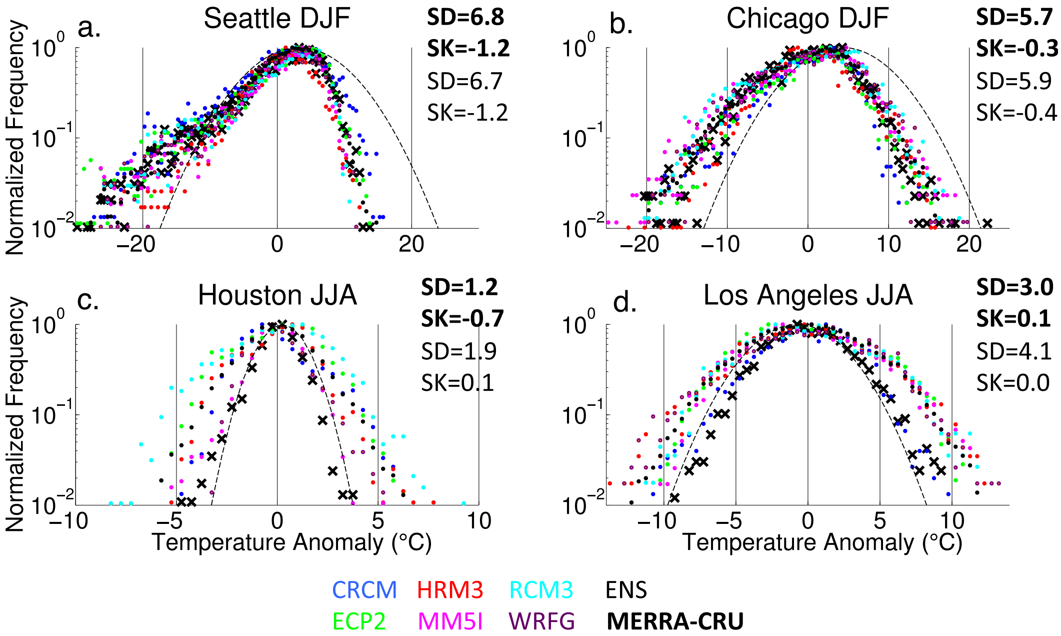

FIG. 10. Sample probability distributions of temperature anomalies for four locations corresponding approximately to examples presented in RN2012, as discussed in section 4c. Temperature anomalies are binned every 0.5°C and the bin counts are normalized by the maximum bin count for each dataset and plotted on a log scale. The black X's are for MERRA–CRU and the colored dots are for the individual RCMs as defined in the legend at the bottom of the figure. The ENS is represented by black dots. The dashed curve is a Gaussian fit to the MERRA–CRU distribution using the standard deviation of the entire distribution. The skewness (SK) and the standard deviation (SD) of MERRA–CRU (in boldface) and the multi-RCM ensemble are indicated at the top right of each panel.

RN2012 show wide warm tails for JJA in Los Angeles and nearby Long Beach, California, using station data. In this region, the prevailing surface wind trajectory is from the relatively cool Pacific Ocean, preventing large temperature excursions on both sides of the PDF while infrequent offshore wind events can cause large excursions on the warm side. This feature is not well captured by the MERRA–CRU distribution or the majority of the RCMs. This may reflect the complex terrain and sharp climate gradients that lie between the coast (where the station observations are taken in RN2012) and the warmer interior. There is, however, reasonable agreement in the shape of the distributions at this grid cell, with notable symmetry apparent for all datasets. CRCM stands out as very closely matching MERRA–CRU here as well. It is encouraging that the RCMs are able to reproduce many of the observed features of the distribution in this complex region; however, for more societally relevant evaluation and model projections, higher resolution is a necessity here.

5. Discussion

The results presented in this work are based on a single reference dataset and interpolation scheme both deemed to be superior over other possible choices. These choices, while deliberate, introduce a level of subjectivity to the analysis. This section explores the sensitivity of the results to these choices.

a. *Sensitivity to choice of reference data*

This evaluation required a reference dataset providing T_s at relatively high spatial and temporal resolution over North America. Reanalysis meets these criteria; however, the MERRA–CRU dataset was chosen because it has the virtue of being bias corrected with in situ observations. Nonetheless, observational uncertainty can be similar in magnitude to individual model biases, presenting a challenge in model evaluation (Gómez-Navarro et al. 2012). In this section the sensitivity of the evaluation results to the choice of reference is explored using four additional datasets (section 2b). Two of the datasets are computed using the same methods used in creating MERRA–CRU and two are standard reanalysis products. For brevity, this section focuses on the bias of median temperature for DJF and JJA and JJA skewness, although all metrics are impacted to some degree by the choice of reference.

Figure 11 shows the DJF and JJA bias in median T_s for all four datasets in reference to MERRA–CRU. Differences between MERRA–CRU and ERA-Interim–CRU and NCEP–CRU products are generally between

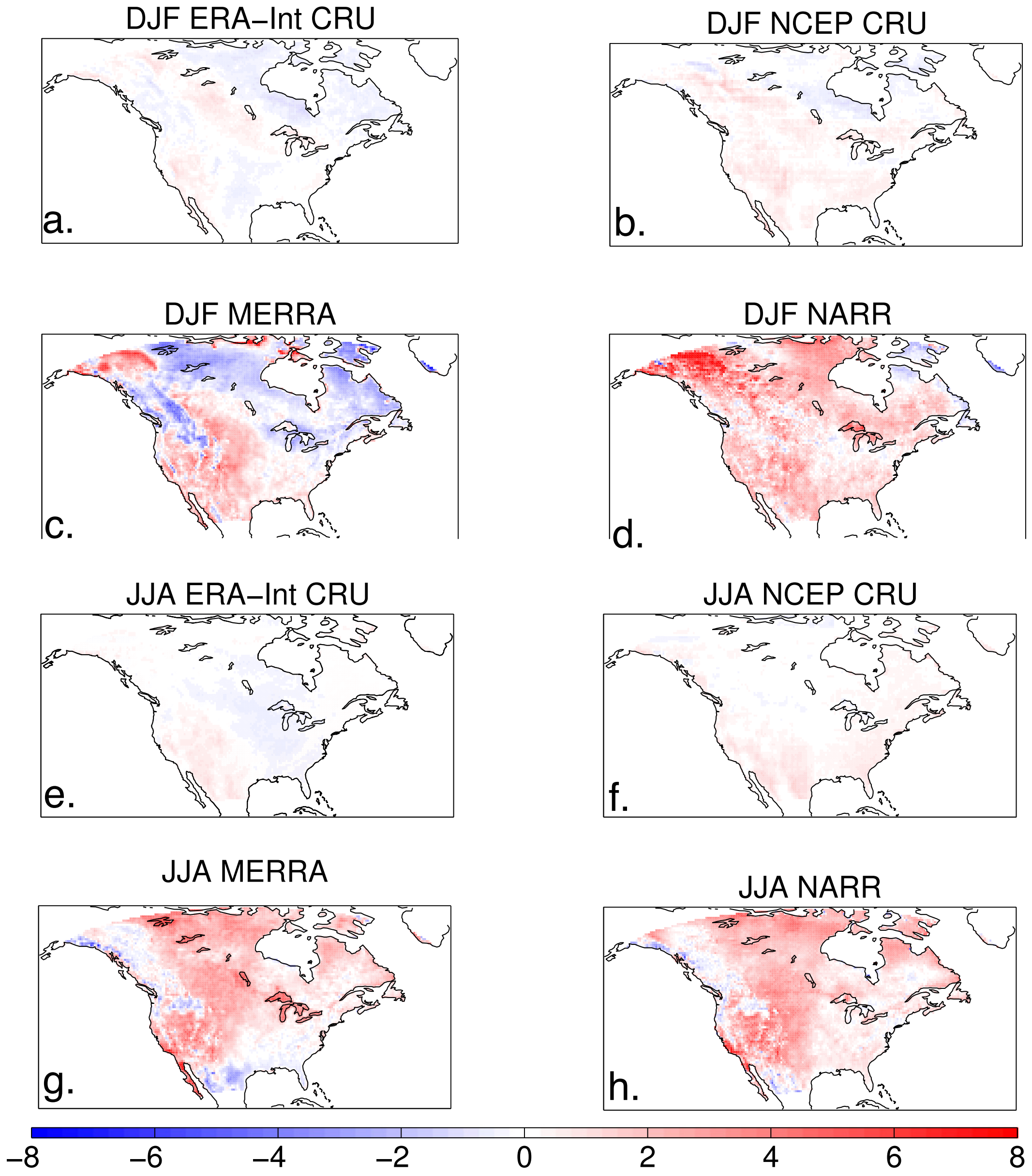

FIG. 11. Bias (°C) of the 50th percentile of the daily surface temperature distribution for (a),(e) ERA-Interim–CRU, (b),(f) NCEP–CRU, (c),(g) MERRA, and (d),(h) NARR in reference to MERRA–CRU. Bias maps of temperature are for DJF in (a)–(d) and JJA in (e)–(h). See sections 2 and 5a for further information on datasets.

−1° and +1°C. While the CRU TS3.10–based bias correction reduces uncertainty relative to the original reanalysis, sources of uncertainty remain. For example, the bias correction is only performed using monthly means from CRU TS3.10, introducing some uncertainty at the daily and subdaily time scales. Additional uncertainty arises from the temporal interpolation used to produce data at hourly time steps for the non-MERRA products. The uncertainty is nevertheless reduced in these datasets compared with the unadjusted reanalyses (Wang and Zeng 2013).

The differences between standard reanalysis and MERRA–CRU are much greater. NARR shows biases as large as 4–6°C in DJF while the unadjusted MERRA T_s is warmer than MERRA–CRU by a similar margin. JJA differences are similar to DJF in magnitude, except both NARR and MERRA show similar primarily warm biases. The biases in NARR and MERRA compared with MERRA–CRU are as large as or larger than some of the RCM biases for median temperature. As such, evaluation results could be quite different if either reanalysis was used as the primary reference. Table 3 shows

TABLE 3. The root-mean-square value of the spatial bias for the entire domain computed relative to each reference dataset (rows) for each RCM (columns). The top rows are for DJF and the bottom for JJA.

	CRCM	ECP2	HRM3	MM5I	RCM3	WRFG
DJF						
MERRA–CRU	2.7	3.2	6.4	2.6	4.0	3.5
ERA-Interim–CRU	2.6	3.3	6.4	2.5	4.1	3.4
NCEP–CRU	2.8	3.2	6.2	2.5	4.0	3.4
MERRA	2.8	3.5	6.6	2.8	4.4	3.1
NARR	3.9	2.2	4.8	2.0	2.7	4.0
JJA						
MERRA–CRU	2.0	1.9	3.8	1.9	1.7	1.9
ERA-Interim–CRU	2.1	1.9	3.8	1.9	1.7	2.0
NCEP–CRU	2.1	1.8	3.6	2.0	1.8	2.0
MERRA	2.9	1.3	2.9	2.8	2.1	1.5
NARR	3.0	1.2	2.7	2.8	2.2	1.7

how the RMS of the spatial bias changes depending on the reference data used. Differences are very small between the CRU TS3.10 adjusted datasets; however, biases computed using NARR or MERRA show greater differences resulting in larger or smaller error depending on the RCM. These results suggest caution should be exercised if using traditional reanalysis as an observational basis for model evaluation of 2-m temperature.

JJA skewness is also associated with relatively large reference data uncertainty in some regions. The left column of Fig. 12 shows JJA skewness for the CRU-based datasets and the right column for traditional reanalysis. All datasets capture the positive skewness along the Pacific coast and the negative skewness over the western mountains of the United States and Canada consistently. These are the same features that the RCMs exhibited high fidelity in reproducing. The largest differences are in the southern United States, especially along the Gulf of Mexico coast. All datasets show negative skewness over this region except for MERRA. The CRU TS3.10 bias-corrected MERRA shows some positive skewness over Texas and northern Mexico, but this feature is greatly diminished over the non-bias-corrected MERRA. If the RCM performance were to be judged using all reference datasets but MERRA, the results would be qualitatively similar, showing consistently low fidelity. However, if MERRA were employed as the reference dataset, the RCMs (WRFG in particular) would show substantially higher fidelity. While not shown here, other work further implicates MERRA as an outlier. Loikith and Broccoli (2012) show negative skewness in this region in July using gridded temperature observations from the Hadley Centre Global Historical Climatology Network–Daily (HadGHCND) dataset (Caesar et al. 2006). Perron and Sura (2013) also show negative skewness over most of the southern United States using NCEP–NCAR Reanalysis 1 and Cavanaugh and Shen (2014) show negative skewness over the same regions using station data. RN2012 show a long cold tail using station data at Houston. It is possible that similar processes contribute to the seemingly spurious positive skewness in MERRA and in the RCMs: improper representation of convective clouds and precipitation and/or incomplete or incorrect coupling with the land surface.

b. *Sensitivity to interpolation procedure*

Kriging is chosen as the primary interpolation scheme for this study because it results in less smoothing of spatial detail than averaging-based interpolation methods and is capable of producing values outside the range of inputs. This is particularly true when using surface elevation as a covariate, as is done here. Kriging also better preserves high-frequency variations in the data and terrain influences on T_s, which makes it an attractive method for this work. However, the data interpolated with kriging yield very similar results to data interpolated using linear- and cubic-based Delaunay triangulation.

To quantify the similarity, the root-mean-square (RMS) value of the spatial bias for the 5th (95th) percentiles of DJF (JJA) temperature for each subregion is shown in Table 4 for all three interpolation methods. Overall, the results are insensitive to the choice of regridding with most cases showing the same or nearly the same RMS bias for all three methods. There is some indication of kriging resulting in a reduction in overall bias in some regions with complex topography. For example, many RCMs show a cold bias over the Central Valley of California. This bias is reduced in the kriging-based results compared with the Delaunay triangulation-based results, likely resulting from the elevation correction (not shown). In addition to impacting extremes, interpolation can have an effect on variance, especially if averaging is performed. The standard deviation ratios, however, are similarly insensitive to the choice of interpolation based on these three relatively sophisticated techniques.

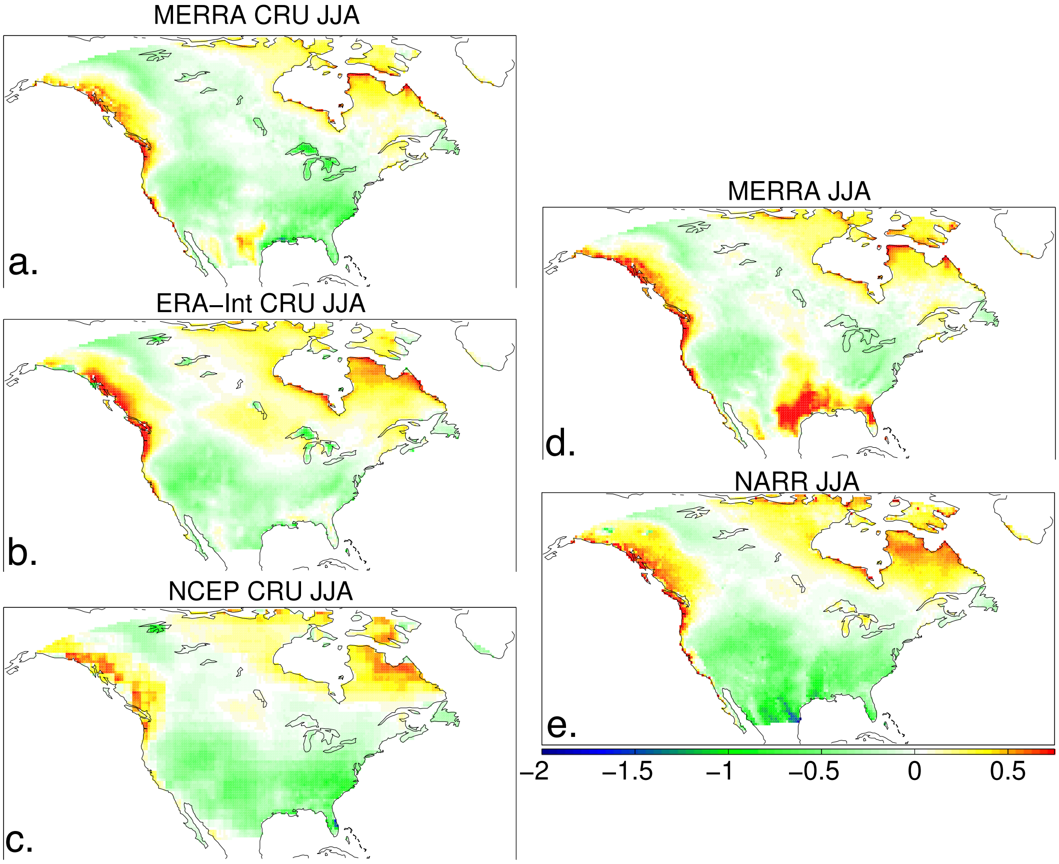

FIG. 12. Skewness of JJA daily temperature anomalies for (a) MERRA–CRU, (b) ERA-Interim–CRU, (c) NCEP–CRU, (d) MERRA, and (e) NARR as discussed in section 5a.

6. Summary and conclusions

Multiple methodologies are employed to evaluate daily surface temperature distributions from a suite of six NARCCAP RCM hindcast experiments against a dataset based on MERRA reanalysis and CRU TS3.10 gridded surface temperature observations, with sensitivity to choice of reference data and interpolation methods also examined. RCM biases are identified and quantified with many RCMs showing systematic, and in some cases large, biases in the temperature distribution at all percentiles (Figs. 1–4). In many cases, additional PDF structure biases are found. While temperature biases, especially those that are systematic across the entire probability distribution, can be accounted for and corrected in model output, error in model-simulated PDF

TABLE 4. The root-mean-square value of the spatial bias for each subregion for data interpolated using (left) kriging-, (center) linear-, and (right) cubic-based interpolation methods. Values for DJF (JJA) are for the 5th (95th) percentiles of the temperature distribution.

	CRCM	ECP2	HRM3	MM5I	RCM3	WRFG
DJF 5th						
West	3.2, 3.2, 3.2	1.8, 2.0, 2.0	5.7, 5.8, 5.8	1.9, 1.9, 1.9	2.0, 2.1, 2.1	2.6, 2.7, 2.7
North	3.9, 3.8, 3.9	3.3, 3.6, 3.6	6.2, 6.2, 6.2	3.8, 3.7, 3.8	6.3, 6.4, 6.4	5.5, 5.4, 5.5
Central	1.6, 1.6, 1.6	1.3, 1.3, 1.3	7.6, 7.6, 7.6	1.4, 1.4, 1.4	2.1, 2.1, 2.1	3.2, 3.2, 3.2
East	2.3, 2.2, 2.3	1.2, 1.3, 1.3	4.9, 5.2, 5.1	2.1, 2.2, 2.3	2.8, 2.8, 2.8	2.6, 2.7, 2.7
JJA 95th						
West	2.3, 2.2, 2.2	3.0, 3.1, 3.1	6.3, 6.2, 6.2	2.1, 2.1, 2.2	2.8, 2.8, 2.9	2.2, 2.2, 2.2
North	2.2, 2.3, 2.3	2.8, 3.0, 3.0	3.8, 3.9, 4.0	2.5, 2.6, 2.6	1.9, 1.9, 1.9	3.0, 3.0, 3.1
Central	2.7, 2.7, 2.7	3.6, 3.6, 3.6	7.9, 7.9, 7.9	2.3, 2.2, 2.3	4.0, 3.9, 4.0	3.4, 3.3, 3.4
East	2.0, 1.9, 1.9	1.2, 1.3, 1.3	4.4, 4.3, 4.3	0.9, 0.9, 1.0	1.9, 1.9, 1.9	1.3, 1.3, 1.3

morphology is more problematic. In particular, error related to the tails of model-simulated PDFs will impact the accuracy with which models simulate extremes.

Variance is generally higher than MERRA–CRU across all RCMs in the northern portion of the domain in winter and throughout the domain in summer while in winter variance is smaller or similar to MERRA–CRU in the south (Figs. 5 and 6). The low variance bias over the U.S. Southwest in JJA coincides with reasonable skewness agreement suggesting PDF shape is well reproduced by the RCMs over this region of complex terrain. Conversely, large positive variance biases over the eastern and southeastern portions of the domain coincide with large RCM-reference disagreement for skewness, indicating difficulty in simulating PDF shape here and suggestive of problems with simulating temperature extremes. Several factors may be related to these discrepancies including differing cloud and precipitation representation and how the air temperature is coupled with land surface characteristics. In the winter, the major patterns in skewness (i.e., positive skewness in the northeastern part of the domain and negative skewness to the south) are realistic in most models (Fig. 7).

Comparison of temperature PDFs for selected locations to those previously analyzed from station data (Fig. 10) can be particularly useful when interpreted in light of these skewness maps. Long cold tails in the distribution of wintertime daily temperature anomalies seen for locations such as Seattle and Chicago are reasonably well simulated in the models. These are part of a coherent region of negative skewness that stretches from the U.S. Northwest to the Great Lakes region that is likewise reproduced in the models with reasonable fidelity. Long warm tails in the summer temperature distribution for the Los Angeles region are not captured by the RCMs or reference data, likely because the grid resolution is too coarse to accurately reflect the coastal climate of Los Angeles. The RCMs are unable to capture the key features of the distribution tails for Houston. For such features that validate reasonably well, the models may be used in future work to further analyze the dynamics yielding the long tails. Predictions of changes in extreme temperature occurrences, such as under global warming, may also be more reliable for these regions where the tail characteristics for present climate are comparable to observations. On the other hand, identifying regions such as along the Gulf of Mexico in the summer where the skewness and tail characteristics do not validate well can help pinpoint regions where confidence would currently be lower in statements about extreme temperature occurrences, and where model development efforts might productively be focused.

The impact that the choice of reference dataset can have makes interpretation of evaluation results difficult if not properly assessed. Based on three reanalysis–CRU TS3.10 combined datasets, results appear robust with little difference depending on which of the three references are used. Results differ considerably in some cases if standard MERRA or NARR datasets are used as reference. These differences include larger or smaller total temperature bias and substantial differences in JJA skewness, with meaningful implications for the interpretation of model performance. The use of unadjusted reanalysis alone would make it difficult to constrain reference data uncertainty as different reanalysis assimilation procedures can result in large biases (e.g., Wang and Zeng 2013). This further suggests that caution in the choice of reference data should be exercised and indicates a need for more high temporal and spatial resolution T_s observation products.

An important future direction in understanding RCM PDF uncertainty, and the inherent relationship this uncertainty has to temperature extremes, is to use this information to investigate mechanisms that are linked to model error. While evidence exists connecting extreme temperature events to larger-scale, low-frequency modes of climate variability such as El Niño–Southern Oscillation and the Arctic Oscillation (Kenyon and Hegerl 2008), which largely occur outside of the domain of these RCMs, Loikith and Broccoli (2014) show that in many places extreme temperatures are also associated with local, amplified, transient weather events that could be examined on an RCM domain. Evaluation of such mechanisms will further identify discrepancies in dynamical processes. Additional analysis of model-simulated soil moisture, cloud cover, and precipitation will also be useful for understanding error in summertime extremes.

Acknowledgments. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Part of this research was funded by NASA National Climate Assessment 11-NCA11-0028 and AIST AIST-QRS-12-0002 projects and the NSF ExArch 1125798 (P.C.L., J.K., H.L., D.E.W., C.M.). Part of this research was funded by NOAA NA11OAR4310099 (J.D.N.). We thank Jonathan Hobbs for his valuable advice during this work. We also thank the NARCCAP team for production of the model simulations and archiving of the data.

REFERENCES

Beniston, M., 2004: The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations. *Geophys. Res. Lett.,* **31,** L02202, doi:10.1029/2003GL018857.

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.

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

Cavanaugh, N. R., and S. S. P. Shen, 2014: Northern Hemisphere climatology and trends of statistical moments documented from GHCN daily surface air temperature station data from 1950 to 2010. *J. Climate,* **27,** 5396–5410, doi:10.1175/JCLI-D-13-00470.1.

Caya, D., and R. LaPrise, 1999: A semi-Lagrangian semi-implicit regional climate model: The Canadian RCM. *Mon. Wea. Rev.,* **127,** 341–362, doi:10.1175/1520-0493(1999)127<0341:ASISLR>2.0.CO;2.

D'Odorico, P., and A. Porporato, 2004: Preferential states in soil moisture and climate dynamics. *Proc. Natl. Acad. Sci. USA,* **101,** 8848–8851, doi:10.1073/pnas.0401428101.

Dole, R., and Coauthors, 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.,* **38,** L06702, doi:10.1029/2010GL046582.

Donat, M. G., and L. Alexander, 2012: The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.,* **39,** L14707, doi:10.1029/2012GL052459.

Field, C. B., and Coauthors, Eds., 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* Cambridge University Press, 582 pp.

Fields Development Team, 2006: fields: Tools for Spatial Data. National Center for Atmospheric Research. [Available online at http://www.image.ucar.edu/GSP/Software/Fields/.]

Fischer, E. M., S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär, 2007: Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *J. Climate,* **20,** 5081–5099, doi:10.1175/JCLI4288.1.

Giorgi, F., M. R. Marinucci, and G. T. Bates, 1993a: Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. *Mon. Wea. Rev.,* **121,** 2794–2813, doi:10.1175/1520-0493(1993)121<2794:DOASGR>2.0.CO;2.

——, ——, ——, and G. De Canio, 1993b: Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions. *Mon. Wea. Rev.,* **121,** 2814–2832, doi:10.1175/1520-0493(1993)121<2814:DOASGR>2.0.CO;2.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.,* **113,** D06104, doi:10.1029/2007JD008972.

Gómez-Navarro, J. J., J. P. Montávez, S. Jerez, P. Jiménez-Guerrero, and E. Zorita, 2012: What is the role of the observational dataset in the evaluation and scoring of climate models? *Geophys. Res. Lett.,* **39,** L24701, doi:10.1029/2012GL054206.

Grell, G., J. Dudhia, and D. R. Stauffer, 1993: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398-STR, 107 pp.

Griffiths, G. M., and Coauthors, 2005: Change in mean temperature as a predictor of extreme temperature change in the Asia-Pacific region. *Int. J. Climatol.,* **25,** 1301–1330, doi:10.1002/joc.1194.

Hegerl, G. C., F. W. Zwiers, P. A. Stott, and V. V. Kharin, 2004: Detectability of anthropogenic changes in annual temperature and precipitation extremes. *J. Climate,* **17,** 3683–3700, doi:10.1175/1520-0442(2004)017<3683:DOACIA>2.0.CO;2.

Hong, S.-Y., and E. Kalnay, 2000: Role of sea surface temperature and soil-moisture feedback in the 1998 Oklahoma–Texas drought. *Nature,* **408,** 842–844, doi:10.1038/35048548.

Hughes, M., and A. Hall, 2010: Local and synoptic mechanisms causing Southern California's Santa Ana winds. *Climate Dyn.,* **34,** 847–857, doi:10.1007/s00382-009-0650-4.

Jones, R. G., D. C. Hassell, D. Hudson, S. S. Wilson, G. J. Jenkins, and J. F. B. Mitchess, 2004: Workbook on generating high resolution climate change scenarios using PRECIS. Hadley Centre for Climate Prediction and Research, 39 pp.

Juang, H., S. Hong, and M. Kanamitsu, 1997: The NCEP regional spectral model: An update. *Bull. Amer. Meteor. Soc.,* **78,** 2125–2143, doi:10.1175/1520-0477(1997)078<2125:TNRSMA>2.0.CO;2.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643, doi:10.1175/BAMS-83-11-1631.

Karl, T. R., and Coauthors, 2012: U.S. temperature and drought: Recent anomalies and trends. *Eos,* **93,** 473–475, doi:10.1029/2012EO470001.

Kenyon, J., and G. C. Hegerl, 2008: Influence of modes of climate variability on global temperature extremes. *J. Climate,* **21,** 3872–3889, doi:10.1175/2008JCLI2125.1.

Kim, J., and Coauthors, 2013: Evaluation of the surface climatology over the conterminous United States in the North American Regional Climate Change Assessment Program hindcast experiment using a regional climate model evaluation system. *J. Climate,* **26,** 5698–5715, doi:10.1175/JCLI-D-12-00452.1.

Kjellström, E., F. Boberg, M. Castro, J. Hesselbjerg Christensen, G. Nikulin, and E. Sánchez, 2010: Daily and monthly temperature and precipitation statistics as indicators for regional climate models. *Climate Res.,* **44,** 135–150, doi:10.3354/cr00932.

——, G. Nikulin, U. Hansson, G. Strandberg, and A. Ullerstig, 2011: 21st century changes in the European climate: Uncertainties derived from an ensemble of regional climate model simulations. *Tellus,* **63A,** 24–40, doi:10.1111/j.1600-0870.2010.00475.x.

Lau, N.-C., and M. J. Nath, 2012: A model study of heat waves over North America: Meteorological aspects and projections for the twenty-first century. *J. Climate,* **25,** 4761–4784, doi:10.1175/JCLI-D-11-00575.1.

Lee, D. T., and B. J. Schachter, 1980: Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Inf. Sci.,* **9,** 219–242, doi:10.1007/BF00977785.

Loikith, P. C., and A. J. Broccoli, 2012: Characteristics of observed atmospheric circulation patterns associated with temperature extremes over North America. *J. Climate,* **25,** 7266–7281, doi:10.1175/JCLI-D-11-00709.1.

——, and ——, 2014: The influence of recurrent modes of climate variability on the occurrence of winter and summer extreme temperatures over North America. *J. Climate,* **27,** 1600–1618, doi:10.1175/JCLI-D-13-00068.1.

——, B. R. Lintner, J. Kim, H. Lee, J. D. Neelin, and D. E. Waliser, 2013: Classifying reanalysis surface temperature probability density functions (PDFs) over North America with cluster analysis. *Geophys. Res. Lett.,* **40,** 3710–3714, doi:10.1002/grl.50688.

Luber, G., and M. McGeehin, 2008: Climate change and extreme heat events. *Amer. J. Prev. Med.,* **35,** 429–435, doi:10.1016/j.amepre.2008.08.021.

Mearns, L. O., W. J. Gutowski, R. Jones, L.-Y. Leung, S. McGinnis, A. M. B. Nunes, and Y. Qian, 2009: A regional climate change assessment program for North America. *Eos,* **90,** 311–312, doi:10.1029/2009EO360002.

——, and Coauthors, 2012: The North American Regional Climate Change Assessment Program: Overview of phase I results. *Bull. Amer. Meteor. Soc.,* **93,** 1337–1362, doi:10.1175/BAMS-D-11-00223.1.

——, and Coauthors, cited 2013: The North American Regional Climate Change Assessment Program dataset. National Center for Atmospheric Research Earth System Grid, doi:10.5065/D6RN35ST.

Meehl, G. A., and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science,* **305,** 994–997, doi:10.1126/science.1098704.

——, and Coauthors, 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 747–845.

——, C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.,* **36,** L23701, doi:10.1029/2009GL040736.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360, doi:10.1175/BAMS-87-3-343.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.,* **25,** 693–712, doi:10.1002/joc.1181.

Otto, F. E. L., N. Massey, G. J. van Oldenborgh, R. G. Jones, and M. R. Allen, 2012: Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.,* **39,** L04702, doi:10.1029/2011GL050422.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate,* **20,** 4356–4376, doi:10.1175/JCLI4253.1.

Perron, M., and P. Sura, 2013: Climatology of non-Gaussian atmospheric statistics. *J. Climate,* **26,** 1063–1083, doi:10.1175/JCLI-D-11-00504.1.

Rahmstorf, S., and D. Coumou, 2011: Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci. USA,* **108,** 17 905–17 909, doi:10.1073/pnas.1101766108.

Rangwala, I., J. Barsugli, K. Cozzetto, J. Neff, and J. Prairie, 2012: Mid-21st century projections in temperature extremes in the southern Colorado Rocky Mountains from regional climate models. *Climate Dyn.,* **39,** 1823–1840, doi:10.1007/s00382-011-1282-z.

Rhines, A., and P. Huybers, 2013: Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl. Acad. Sci. USA,* **110,** E546, doi:10.1073/pnas.1218748110.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Ruff, T. W., and J. D. Neelin, 2012: Long tails in regional surface temperature probability distributions with implications for extremes under global warming. *Geophys. Res. Lett.,* **39,** L04704, doi:10.1029/2011GL050610.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature,* **427,** 332–336, doi:10.1038/nature02300.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRFG version 2. NCAR Tech. Note NCAR/TN-468+STR, 88 pp.

Sobolowski, S., and T. Pavelsky, 2012: Evaluation of present and future North American Regional Climate Change Assessment Program (NARCCAP) regional climate simulations over the southeast United States. *J. Geophys. Res.,* **117,** D01101, doi:10.1029/2011JD016430.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. M. B. Tignor, and H. L. Miller Jr., 2007: *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, 996 pp.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution for the European heatwave of 2003. *Nature,* **432,** 610–614, doi:10.1038/nature03089.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes, an intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change,* **79,** 185–211, doi:10.1007/s10584-006-9051-4.

Wang, A., and X. Zeng, 2013: Development of global hourly 0.5° land surface air temperature datasets. *J. Climate,* **26,** 7676–7691, doi:10.1175/JCLI-D-12-00682.1.

——, and ——, 2014: Global hourly 0.5-degree land surface air temperature datasets. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (accessed 24 Jun 2014), doi:10.5065/D6PR7SZF.