

Projections of Future Drought in the Continental United States and Mexico

MICHAEL WEHNER

Lawrence Berkeley National Laboratory, Berkeley, California

DAVID R. EASTERLING, JAY H. LAWRIKORE, RICHARD R. HEIM JR., AND RUSSELL S. VOSE

NOAA/National Climatic Data Center, Asheville, North Carolina

BENJAMIN D. SANTER

Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

(Manuscript received 23 August 2010, in final form 21 April 2011)

ABSTRACT

Using the Palmer drought severity index, the ability of 19 state-of-the-art climate models to reproduce observed statistics of drought over North America is examined. It is found that correction of substantial biases in the models' surface air temperature and precipitation fields is necessary. However, even after a bias correction, there are significant differences in the models' ability to reproduce observations. Using metrics based on the ability to reproduce observed temporal and spatial patterns of drought, the relationship between model performance in simulating present-day drought characteristics and their differences in projections of future drought changes is investigated. It is found that all models project increases in future drought frequency and severity. However, using the metrics presented here to increase confidence in the multimodel projection is complicated by a correlation between models' drought metric skill and climate sensitivity. The effect of this sampling error can be removed by changing how the projection is presented, from a projection based on a specific time interval to a projection based on a specified temperature change. This modified class of projections has reduced intermodel uncertainty and could be suitable for a wide range of climate change impacts projections.

1. Introduction

Sustained periods of drought can be disruptive to both human and natural systems. Drought is a relative condition, and is best quantified by considering local climatological factors. In general, drought is characterized by a lack of available water. There are different perspectives on what constitutes drought, and a variety of definitions are available (Heim 2002). The balance between evaporation and precipitation determines the amount of soil moisture—a quantity critical to agriculture. The Palmer drought severity index (PDSI) is a widely recognized measure of droughts extending several months or more (Heim 2002). In this paper, we examine the ability of the climate models used in the

Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) to reproduce observed PDSI statistics over North America (Solomon et al. 2007). We rank the models' performance by a variety of measures, and explore whether performance in simulating current climate is functionally related to the scatter in projections of future changes in drought.

The PDSI code used in this study is the same as that used for the National Oceanic and Atmospheric Administration's National Climatic Data Center (NOAA/NCDC) operational product and requires monthly precipitation and mean temperature as input. Precipitation is used as a measure of moisture supply, while temperature is used to estimate evapotranspiration or moisture demand. The use of temperature to estimate moisture demand makes the Palmer model well suited for temperature-sensitivity drought studies. Drought is classified into the following categories: incipient ($-0.5 \geq \text{PDSI} > -1.0$), mild ($-1.0 \geq \text{PDSI} > -2.0$), moderate ($-2.0 \geq \text{PDSI} > -3.0$), severe ($-3.0 \geq \text{PDSI} > -4.0$), and extreme

Corresponding author address: Michael Wehner, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., MS50F, Berkeley, CA 94720.
E-mail: mfwehner@lbl.gov

($-4.0 \geq \text{PDSI}$). Although not specifically designed to diagnose soil moisture, the NCDC PDSI algorithm contains its own, albeit simple, soil moisture scheme. In a multimodel intercomparison of drought statistics, a model-independent soil moisture treatment removes the complexities of directly comparing soil moisture schemes of varying sophistication. In an early study, Rind et al. (1990) found that the then-current version of the Goddard Institute for Space Studies General Circulation Model (GISS-GCM) potentially underestimated future drought in the United States because of land surface model component deficiencies. Land surface models have improved significantly in the past 20 years, however not all the modern climate models are state of the art in that category. Soil moisture feedback is still a factor through its influence on precipitation variability. Furthermore, details in the treatment of vegetation and its transpiration differ greatly among climate models, which may cause different stomatal responses to climate change that would eventually be reflected in the response of temperature and precipitation. A more detailed description of the Palmer index, including its soil moisture treatment, is given in appendix A.

In previous studies, Schubert et al. (2004) found that only one-third of the total low frequency variability in Great Plains precipitation is due to Pacific sea surface temperature fluctuations, with the remainder generated locally and driven by soil moisture feedback. Seager et al. (2005) obtained similar results, and traced persistent wet and dry periods in the western United States to persistent tropical Pacific SST variations. Easterling et al. (2007) showed that an increase in precipitation has masked the severity of recent drought in the western United States caused by increases in air temperature. Relevant to the current investigation is the study by Burke et al. (2006), which found a detectible contribution from anthropogenic emissions of greenhouse gases to the global drying that has been observed since 1980. The Burke et al. (2006) study relied on the PDSI to measure drought, using one of the 19 climate models considered here. Because the PDSI is generally more sensitive to observed temperature changes than observed precipitation, this detection result follows from the robust detectibility of twentieth century temperature changes.

Here, we consider simulations from 19 different climate models archived in the Coupled Model Intercomparison Project (CMIP3) database managed by the Program for Climate Model Diagnosis and Intercomparison. Monthly-mean surface air temperature and precipitation from simulations of the twentieth century (20c3m) and the twenty-first century were used to calculate monthly values of PDSI. The twenty-first century simulations were forced by anthropogenic atmospheric conditions prescribed by the IPCC Special Report on Emissions

Scenarios (SRES) A1B scenario. This protocol stabilizes atmospheric concentration of carbon dioxide at 720 ppm at the end of the twenty-first century and is considered to be a moderate greenhouse gas emission reduction scenario.

To reduce the “noise” of the natural internal variability of the climate system, some of the 19 models were integrated over these historical and future periods as ensembles of independent realizations. Ensembles were generated by perturbing the initial conditions of the atmosphere and/or ocean. Each realization contains some climate response to the imposed forcing changes (the “signal”) plus a specific sequence of climate noise. Since the noise is uncorrelated from one realization to the next, its amplitude is reduced by averaging over realizations, thus providing a better estimate of the underlying signal. After first calculating the ensemble average PDSI over each individual model’s 20c3m or A1B realizations, we then form a multimodel ensemble, equally weighted by all models considered.

The models are formulated at a variety of spatial resolutions. To ease intercomparison of the models’ ability to simulate PDSI statistics, modeled and observed temperature and precipitation data were regridded prior to the PDSI calculation to a spectral resolution of T42 (a Gaussian grid of about 400 km at the equator). This is the resolution of the coarsest model in the study. All models and observations were used in identical ways to drive the PDSI code. An examination of simulated PDSI quality as a function of model resolution is deferred to a later study. A list of the CMIP3 models and the number of realizations of each century used in this paper is presented in Table 1.

The U.S. and Canadian temperature and precipitation data were from their respective national archives (the National Climatic Data Center and Environment Canada) while the Mexican data were provided by A. Douglas (Creighton University, 2008, personal communication). Observations include monthly-averaged total precipitation and temperature. These datasets have been subject to extensive quality control. Temperature data have been adjusted for inhomogeneities arising from station moves, instrument changes, and other factors that can cause artificial discontinuities in the time series. Homogenization relies on the method documented in Menne and Williams (2009). Our analysis is restricted to North America. See the appendices for further details.

2. Simulated and observed drought statistics of North America

Accurate simulation of any measure of drought requires a realistic representation of the surface air temperature and precipitation climatology. Since drought conditions represent a significant departure from the average,

TABLE 1. The CMIP3 climate models used in this study. Full details about model formulation and resolution may be found at the CMIP3 Web site (http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php).

Model acronym	Institution	Country	Atmospheric component resolution
Bjerknes Center for Climate Research Bergen Climate Model version 2 (BCCR-BCM2.0)	Bjerknes Center for Climate Research	Norway	T63L31
CCCma CGCM3.1	Canadian Centre for Climate Modeling and Analysis	Canada	T47L31
CCCma CGCM3.1_t63	Canadian Centre for Climate Modeling and Analysis	Canada	T63L31
CCSM3	National Center for Atmospheric Research	United States	T85L26
CNRM-CM3	Centre National de Recherches Meteorologiques	France	T42L45
CSIRO Mk3.0	Commonwealth Scientific and Industrial Research Organization	Australia	T63L18
ECHAM5	Max Planck Institute for Meteorology	Germany	T63L32
GFDL CM2.0	Geophysical Fluid Dynamics Laboratory	United States	$\sim 2.5^\circ \times \sim 2^\circ \text{L}24$
GFDL CM2.1	Geophysical Fluid Dynamics Laboratory	United States	$\sim 2.5^\circ \times \sim 2^\circ \text{L}24$
Third climate configuration of the Met Office Unified Model (HadCM3)	Met Office	United Kingdom	$2.5^\circ \times 3.75^\circ \text{L}$
Hadley Centre Global Environmental Model version 1 (HadGEM1)	Met Office	United Kingdom	N96L38
Flexible Global Ocean–Atmosphere–Land System Model gridpoint version 1.0 (FGOALS-g1.0)	Institute of Atmospheric Physics	China	
INMCM3	Institute for Numerical Mathematics	Russia	$4^\circ \times 5^\circ \text{L}21$
IPSL CM4	L’Institut Pierre-Simon Laplace	France	$2.5^\circ \times 3.75^\circ \text{L}19$
MIROC(hires)	Center for Climate System Research–National Institute for Environmental Studies–Frontier Research Center for Global Change	Japan	T105L56
MIROC(medres)	Center for Climate System Research–National Institute for Environmental Studies–Frontier Research Center for Global Change	Japan	T42L19
ECHAM and the global Hamburg Ocean Primitive Equation (ECHO-G)	Meteorological Institute of the University of Bonn	Germany	T30L19
Meteorological Research Institute Coupled General Circulation Model, version 2.3.2a (MRI CGCM2.3.2a)	Meteorological Research Institute	Japan	T42L30
PCM	National Center for Atmospheric Research	United States	T42L17

accurate simulation of the climate variability should be integral to successful reproduction of key features of observed drought behavior. For accurate representation of a soil moisture-based metric like the PDSI, it is also crucial to maintain the correct balance between precipitation and evaporation. Too high an average temperature or insufficient average precipitation will bias the soil to be overly dry. Likewise, too low a temperature or too much precipitation will bias the soil to be excessively moist. The nonlinear dependence of soil moisture on temperature and precipitation can cause such systematic errors to adversely affect the simulation of PDSI statistics.

To assess the models’ simulated PDSI statistics, we use the 1950–99 period as the base climatology from which

PDSI is calculated. We selected this period for two reasons: 1) because of the generally high quality of available observations, and 2) to avoid usage of the major drought of the 1930s in the calculation of “normal” conditions. The disadvantage of this choice is the significant anthropogenic trend in surface air temperature in the later part of the period. Additionally, a less severe although still major drought did occur in the 1950s. In the PDSI calculation shown below, each individual model’s drought measure is derived relative to its own climatology, not the actual or mean model climatology.

Figure 1 shows the difference between the multimodel mean temperature and precipitation (averaged over 1950–99) and the NCDC observational dataset. Large biases are revealed, such as the cold and wet model

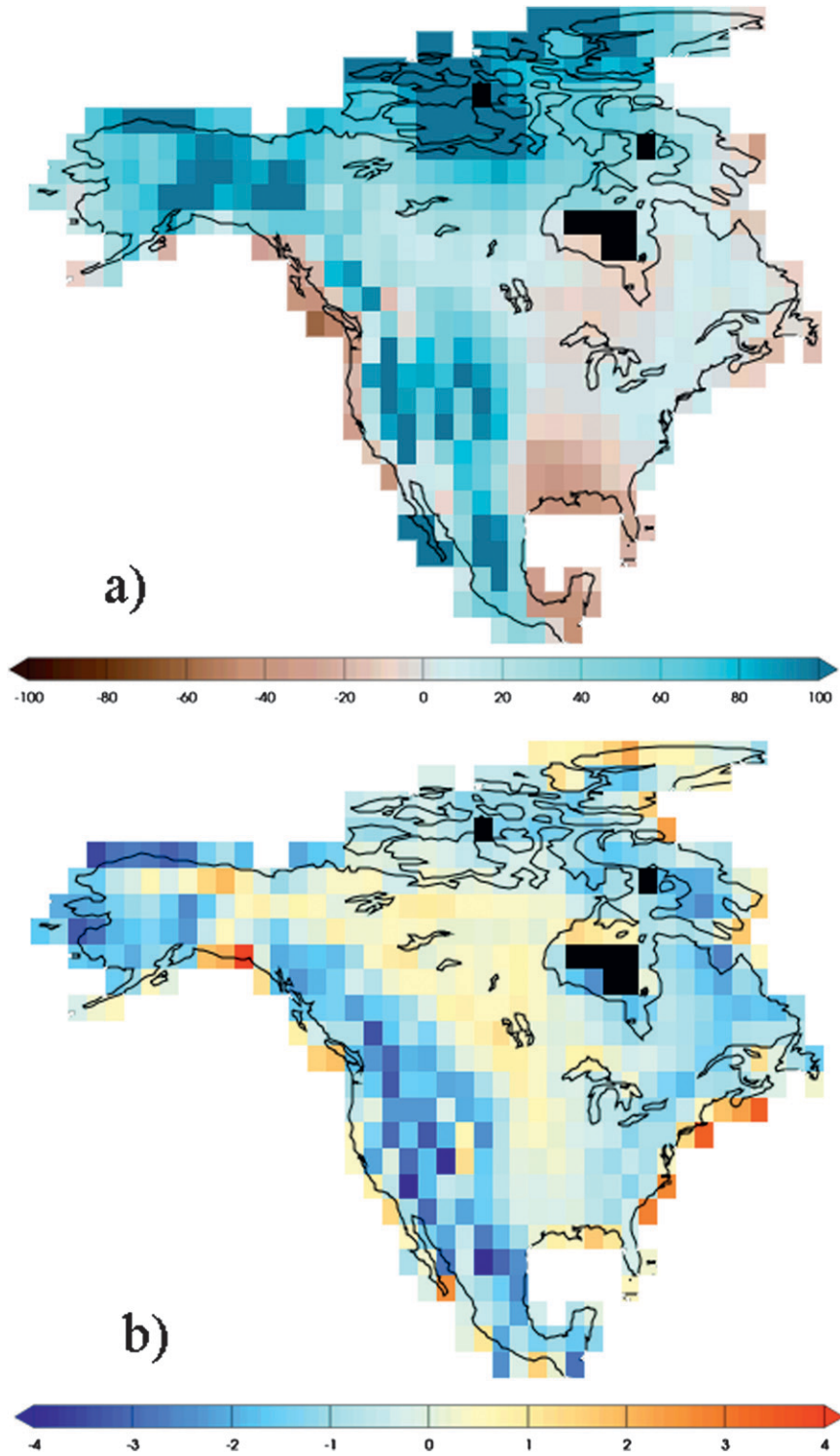


FIG. 1. The 1950–99 North American climatologies of the NCDC observations minus the mean model average. (a) Precipitation bias (%) and (b) temperature bias (K).

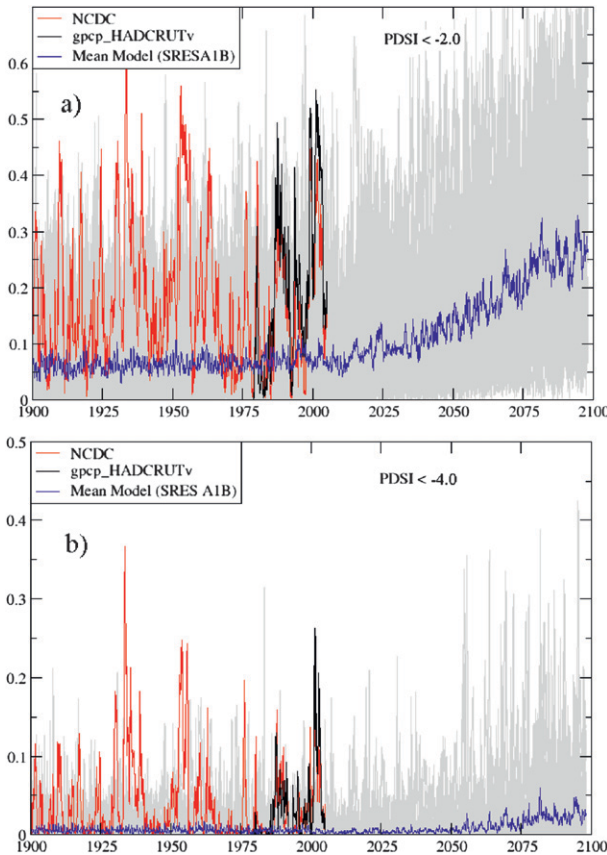


FIG. 2. (a) Area fraction of the continental United States and Mexico experiencing drought conditions (PDSI < -2) calculated from observations and raw model output. The light gray curves are individual model realizations. (b) As in (a) except for extreme drought conditions (PDSI < -4).

bias in the western United States and Mexico. These biases affect the calculation of drought metrics like the PDSI.

Use of the raw, uncorrected model output to calculate the PDSI leads to systematic underprediction of the occurrence of drought, especially in the arid southwest United States and Mexico. Figure 2a shows the percent area of the continental United States and Mexico in moderate drought conditions (PDSI < -2) for each month over the period 1900–2098. In this section, we focus attention on an analysis of model errors over the twentieth century. The discussion of future drought projections is deferred until section 3.

The results from the NCDC observations (in red) clearly show the “Dust Bowl” conditions of the mid-1930s. At the peak of these conditions, nearly 60% of the United States and Mexico is in moderate drought. There is some suggestion of a trend toward increasing drought beginning around 1980, but natural variability is very large. Our study does not explicitly examine whether an

anthropogenic fingerprint pattern is identifiable in observed patterns of changes in the PDSI over North America. As noted earlier, one optimal detection study found a detectable human influence on observed patterns of global-scale changes in the PDSI (Burke et al. 2006).

A second observational PDSI estimate is also shown (in black). This was constructed from the Hadley Centre Climate Research Unit global temperature dataset, version 3 (HadCRUT3v) surface temperature (Jones et al. 1999; data available online at <http://www.cru.uea.ac.uk/cru/data/temperature/>) and Global Precipitation Climatology Project (GPCP) precipitation datasets (Adler et al. 2003). The latter dataset spans 1979–2006 only, so this period is used to calculate the baseline climatology for this particular PDSI calculation only. There is close agreement with the NCDC-based estimate of changes in the PDSI. The multimodel mean result (shown in blue) reveals that relative to observations, the average area undergoing moderate drought conditions is systematically underpredicted in the simulations.

Because the model results shown by the blue line in Fig. 2 are averaged over both realizations and models, natural variability is considerably damped with respect to observations. Individual realizations are shown by the background gray lines, and reveal that some models can produce significant areas of moderate drought conditions. However, the frequency of these events is too low in all of the models considered in this study. Figure 2b shows the percent area in extreme drought conditions (PDSI < -4) for the same North American region. Again, the models employed here consistently underpredict the average area experiencing these conditions, as well as the frequency of occurrence of large drought events.

The results shown in Fig. 2 provide strong motivation for objective correction of the models’ temperature and precipitation fields. The intent of this correction is to improve the portrayal of future drought conditions. The assumption in our correction procedure is that the underlying bias between observations and in any given model simulation varies over the seasonal cycle, but not from year to year. Error fields constructed from the difference between the 1950–99 climatological average for each model and for the observations form the basis of this correction factor. To preserve the positive definiteness of precipitation, we apply the following multiplicative correction at each grid point to restore the simulated 1950–99 average to observations:

$$\text{correction}(m) = \frac{\sum_y \text{observed}(y, m)}{\sum_y \text{simulated}(y, m)},$$

$$\text{corrected}(y, m) = \text{correction}(m) \times \text{simulated}(y, m)$$

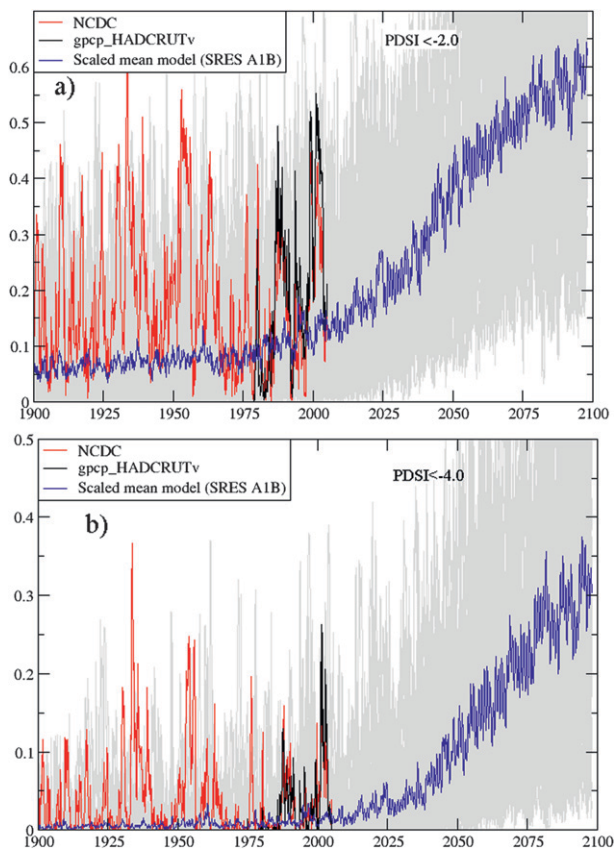


FIG. 3. Area fraction of the continental United States and Mexico experiencing drought conditions ($PDSI < -2$) calculated from observations and corrected model output. (b) As in (a) except for extreme drought conditions ($PDSI < -4$).

where m represents each of the 12 months and y denotes an index over years (from 1950 to 1999). We note that this correction technique also alters the simulated precipitation variability, although not necessarily in any preferential manner. We apply the correction to each individual month of each realization at each individual grid point. To be consistent, we also apply a similar correction to temperature. The effect on variability is minimal in this case as the units are Kelvins, keeping the temperature correction factor small in magnitude.

Figure 3a shows the percent area of the continental United States and Mexico in moderate drought conditions from the corrected model temperature and precipitation fields (c.f. Figure 2). Even with the bias correction, the models continue to systematically underpredict drought extent and severity. However, the correction does increase both the average fraction of area experiencing drought conditions and the severity of the most intense events. Inspection of the twentieth century portion of Fig. 3b reveals that a number of 20c3m realizations

produced extreme drought events of magnitude similar to that of the 1930s Dust Bowl.

The effect of this bias correction on the simulated PDSI statistics varies significantly between models. In Table 2, we show the fractional area of the continental United States and Mexico experiencing moderate drought ($PDSI < -2$) and extreme drought ($PDSI < -4$) conditions. Results are averaged over the 1950–99 period. The highlighted row labeled “mean model” corresponds to the blue lines in Figs. 2 and 3. Note we have listed the models in their rank order, based on the amount of area experiencing moderate drought ($PDSI < -2$) conditions. None of the model datasets, either uncorrected or corrected, are able to exceed the observed drought areas. In all but one of the models [Institute for Numerical Mathematics Coupled Model Version 3.0 (INMCM3)], the fractional area experiencing drought conditions was increased by application of the correction. After correction, 11 of the models are able to reproduce more than half the observed fractional area experiencing moderate drought, with the best model [the Commonwealth Scientific and Industrial Research Organisation Mark version 3.0 (CSIRO Mk3.0)] producing 87% of the observational result. Thirteen of the 19 models also show increased extreme drought conditions ($PDSI < -4$) after bias correction. The CSIRO Mk3.0 model also best reproduced the observed fractional area experiencing extreme drought, but only three models could reproduce more than half of the observational estimate. The models that did not improve their simulation of extreme drought are preferentially clustered at the bottom of this rank-ordered list. A brief analysis of changing the reference period to the cooler interval of 1950–70 had no significant effect on model error. Given the improvements shown in Table 2, the remainder of the discussion in this study focuses on PDSI statistics calculated from the corrected model temperature and precipitation fields.

The results in Fig. 3 are an area integrated metric, and do not reveal any information about the models’ ability to reproduce drought conditions in the correct places. To gain some insight into the spatial structure of simulated drought, we count at each grid point the number of months experiencing moderate drought ($PDSI < -2$) and extreme drought ($PDSI < -4$) over the 600-month period from January 1950 through to December 1999. Figure 4a shows the observed percentage of moderate drought months for all of North America. Figure 4b shows the same percentage for extreme drought. A similar analysis is performed with each of the model realizations. Results are then averaged for models with multiple realizations. The substantial structure in Fig. 4 reveals that the base climatology is an important factor in determining the statistics of PDSI and reinforces the

TABLE 2. Fractional area of the continental United States and Mexico experiencing moderate drought ($PDSI < -2$) and extreme drought ($PDSI < -4$) conditions over the 1950–99 period calculated from the NCDC observations and uncorrected and corrected climate models.

NCDC observations Model name	PDSI < -2		PDSI < -4	
	16.9%		3.4%	
	Uncorrected	Corrected	Uncorrected	Corrected
CSIRO Mk3.0	8.9%	14.8%	0.6%	2.3%
PCM	8.8%	13.7%	0.6%	1.9%
GFDL CM2.0	7.7%	12.1%	0.4%	1.0%
GFDL CM2.1	8.4%	11.7%	1.2%	1.9%
CCCma CGCM3.1_t63	9.6%	11.5%	0.7%	1.4%
ECHAM5	7.1%	10.8%	0.4%	1.5%
CCSM3	8.7%	10.8%	1.2%	1.7%
MRI CGCM2.3.2a	6.0%	10.6%	0.5%	0.8%
FGOALS-g1.0	5.6%	10.2%	0.2%	0.9%
CCCma CGCM3.1	8.9%	9.9%	0.7%	1.4%
BCCR-BCM2.0	6.2%	9.1%	0.3%	0.7%
Mean model	6.3%	9.1%	0.6%	1.0%
CNRM-CM3	5.3%	8.1%	0.6%	0.7%
HadCM3	4.7%	7.8%	0.5%	0.5%
ECHO-G	6.1%	7.3%	0.7%	0.5%
HadGEM1	4.6%	7.2%	0.3%	0.6%
INMCM3	6.6%	5.9%	0.8%	0.2%
IPSL CM4	2.5%	5.1%	0.1%	0.2%
MIROC(medres)	2.0%	3.4%	0.4%	0.1%
MIROC(hires)	1.5%	2.4%	0.2%	0.1%

need to correct the model biases discussed earlier. Figures 4c,d show the same results averaged over all the models with equal weighting. These figures reveal that the models underpredict the amount of time spent in both moderate and extreme drought across almost all of North America.

We can use the values contained in Figs. 4a–d to compute the pattern correlation (Houghton et al. 2001) between the observed pattern of drought structure and each individual model’s drought pattern. Our correlation analysis is also limited to the land areas of the continental United States and Mexico (12° – 85° N, 180° – 310° W). As with our previous “drought fraction” results, we find that the models used here vary considerably in their ability to simulate the spatial distribution of drought. The left panels of Fig. 5 show the relationship between the centered (mean value removed) pattern correlation of simulated moderate drought and extreme drought with observations and the fractional area metric discussed in Fig. 3. There is some evidence of a functional relationship between these two metrics of model skill for extreme drought conditions but not for the weaker drought conditions. The right panels of Fig. 5 show this relationship when the mean value is retained (uncentered) in the correlation calculations. In these calculations, the four models least well simulating drought conditions are more clearly identified but no clear relationship exists between the metrics for the remaining models.

Interestingly, the mean model’s drought pattern exhibits better spatial correlation with the observations than most of the individual models although it remains rather low, as is evident from a visual inspection of Fig. 4. Similar results have been obtained in an analysis of patterns of ENSO variability (Pierce et al. 2009). Although the superiority of the mean model is not fully understood, it may be related to the combined effects of spatial smoothing and quasi-random distribution of model errors in drought patterns.

A similar analysis of the quality of the spatial distribution and areal extent of extreme drought reveals a clearer relationship between the two metrics than for the less dry case. Although most of the models produce no more than half the observed extreme drought area, as the extreme drought area increases the pattern correlation metric also generally increases.

The poor performance of all the models after bias correction of the mean temperature and precipitation must be rooted in a defective simulation of variability in one or both of these fields. We examined the effect of short term variability by removing the climatological annual cycle to calculate the monthly-mean anomaly over 1950–99 for each model at each grid point. We then compared the temporal standard deviation of the bias-corrected temperature and precipitation, as well as the local covariance of these two variables for each model against the NCDC observations. Noting that every

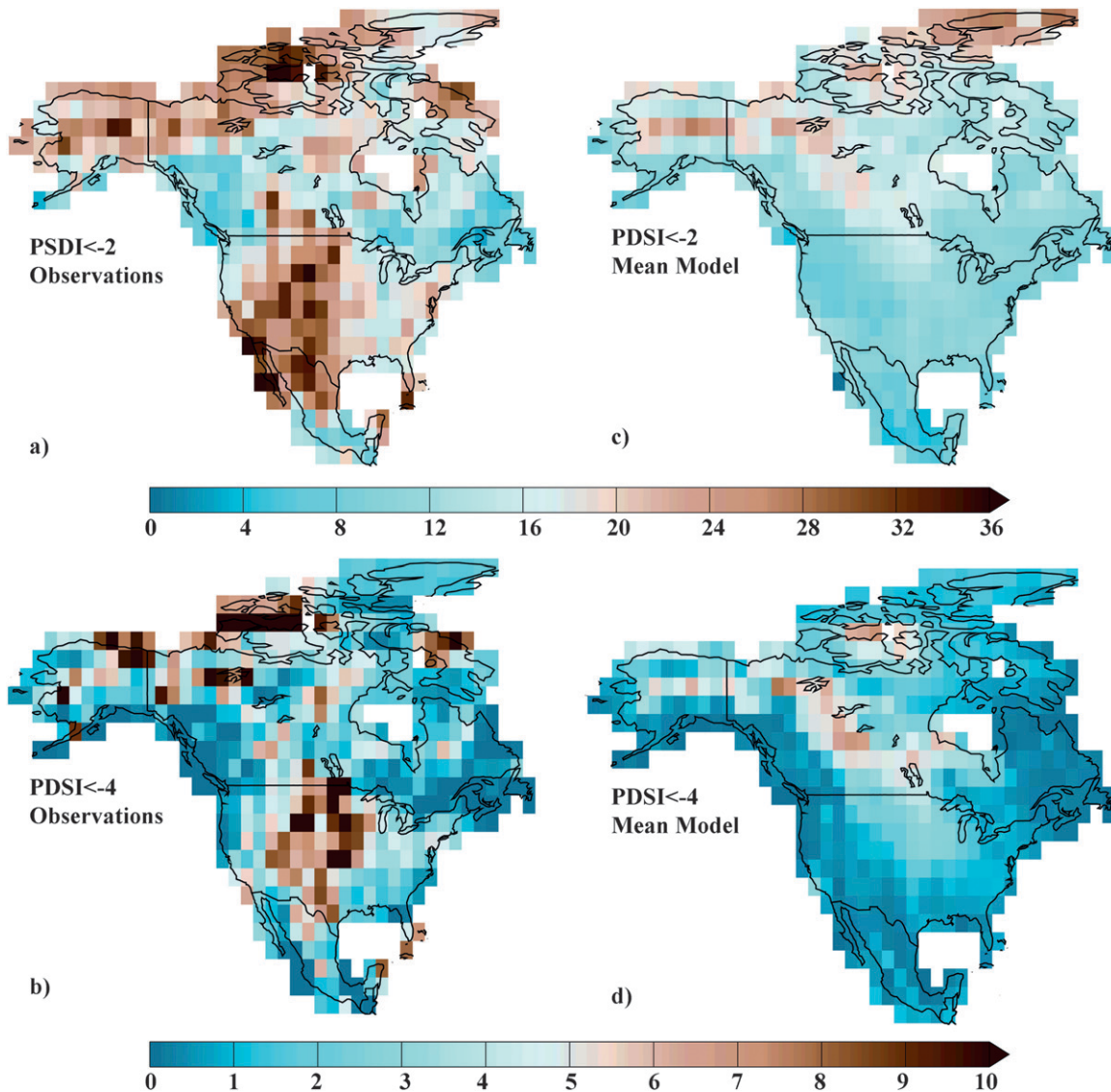


FIG. 4. (a) Observed percentage of time spent in drought conditions ($\text{PDSI} < -2$) during the 1950–99 period. (b) As in (a) except for observed extreme drought conditions ($\text{PDSI} < -4$). (c) As in (a) except showing the multimodel average result. (d) As in (b) except showing the multimodel average result. These fields form the basis for the pattern correlation metrics described in the text.

corrected model underpredicts current observed drought area extent, we find that there is no relationship between these second-order measures of short time variability and any of the four drought measures. Model errors of variance and covariance are of both signs with some models exhibiting too much variability while others too little. Neither of the two superior models [CSIRO Mk3.0 and the National Center for Atmospheric Research (NCAR) Parallel Climate Model (PCM)] appears to stand out in this regard. For instance, CSIRO Mk3.0 poorly simulates the covariance but is superior in its simulation of the individual variances relative to the other models. Conversely, PCM simulates the covariance well but performs

poorly in its simulation of the individual variances. Nor do the worst PDSI models systematically underperform. For instance, the Model for Interdisciplinary Research on Climate, high-resolution version [MIROC(hires)], ranked near the bottom for the drought statistics, is one of the best at reproducing the covariance of temperature and precipitation.

We also constructed multimonth averages of the simulated bias-corrected fields to examine model performance of variability over longer periods. Results from averages six months long or less were as inconclusive as that described above. This may not be surprising as a more characteristic time scale for the PDSI is about

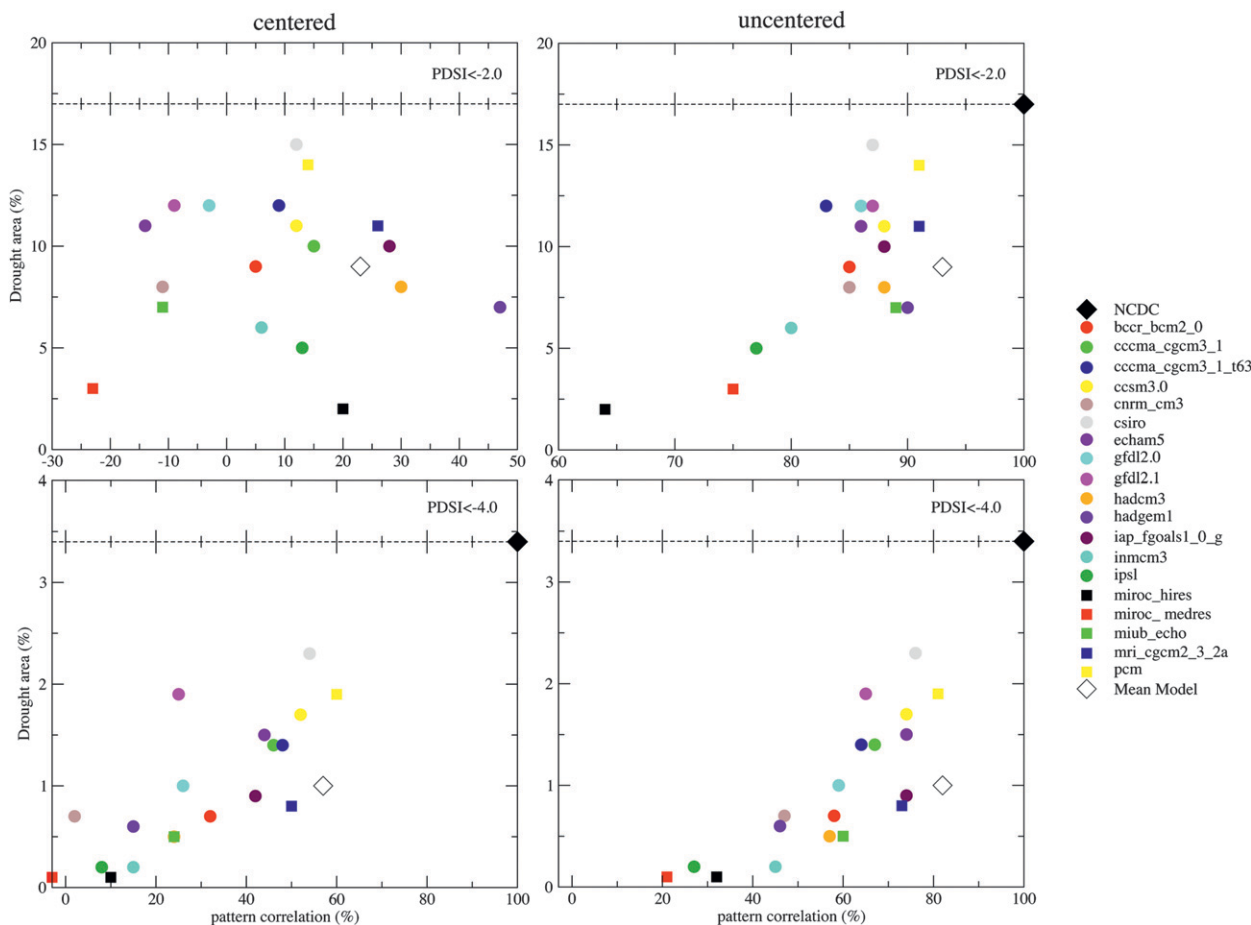


FIG. 5. Scatterplot showing the relationship between the average area of the continental United States and Mexico experiencing drought and the pattern correlation between the observed and simulated spatial distribution of the percentage of time experiencing drought. Both (left) centered and (right) uncentered are shown. The analysis period is 1950–99. (top) Drought conditions ($PDSI < -2$) and (bottom) extreme drought conditions ($PDSI < -4$).

nine months (Guttman 1998). The total (co)variance mixes wet, dry, warm, and cold excursions from the mean state. As the PDSI is a measure of the balance between evaporation and precipitation, drought conditions must occur when conditions are either warmer and/or drier than the mean state. To investigate this further, we separated out periods of the observations and each of the simulations into three sets of data points based on whether a period is warmer than average, drier than average, or both of these conditions. We calculated the total number of warm periods, the total number of dry periods, the total number of periods when conditions were both warm and dry, and the average and integrated temperature and precipitation during these three periods for a total of nine potentially relevant performance metrics. Again, for averaging periods six months or less, comparison of these nine metrics between models and observations is inconclusive. Even on the nine-month

time scale, there is no definitive bias that explains why the all of the simulations underpredict drought extent in the United States and Mexico. On the nine-month time scale, most of the simulations have too many periods that are both warmer and drier than the average. If this were the controlling factor, the simulations would exceed the observed drought area. The amount of the precipitation deficit during these warm, dry nine-month periods is not systematically lower for the simulations than for the observations, although it is for some of them. Similarly, the simulated excess temperature during these periods can be both higher and lower than observed. Nonetheless, it is instructive to examine these metrics in further detail. In Table 3, a relative rank of each model is shown ordered such that the best model on the drought area index described above is first and the worst is last. The rank for the average nine-month precipitation deficit is defined such that the first model has

TABLE 3. Relative order of 20 bias-corrected climate models ranked by their ability to simulate observed PDSI metrics and precipitation deficits and temperature surpluses during warm and dry periods over the 1950–99 period. One is the best model and 20 is the worst model. The measured region is confined to the land areas of the continental United States and Mexico. Warmer means warmer than observed; similarly for colder, drier, and wetter.

	Surface air temperature deficit rank		Precipitation deficit rank		Area metric moderate drought	Area metric extreme drought	Correlation metric moderate drought	Correlation metric extreme drought
CSIRO Mk3.0	9	Warmer	3	Drier	1	1	10	3
PCM	13	Warmer	4	Drier	2	3	3	1
GFDL CM2.0	10	Warmer	7	Drier	3	8	13	11
GFDL CM2.1	2	Warmer	1	Drier	4	2	11	12
CCCma	8	Warmer	14	Wetter	5	6	16	6
CGCM3.1_t63								
ECHAM5	6	Warmer	6	Drier	6	5	12	8
CCSM3	4	Warmer	17	Wetter	7	4	8	4
MRI CGCM2.3.2a	19	Colder	8	Drier	8	11	4	5
FGOALS-g1.0	1	Warmer	13	Wetter	9	10	2	9
CCCma CGCM3.1	7	Warmer	18	Wetter	10	7	9	7
BCCR-BCM2.0	17	Colder	16	Wetter	11	13	14	10
CNRM-CM3	14	Warmer	15	Wetter	13	12	15	19
HadCM3	3	Warmer	11	Wetter	14	15	7	14
ECHO-G	11	Warmer	20	Wetter	15	16	6	13
HadGEM1	5	Warmer	10	Wetter	16	14	5	15
INMCM3	16	Warmer	9	Wetter	17	17	17	16
IPSL CM4	15	Warmer	5	Drier	18	18	18	18
MIROC(medres)	18	Colder	2	Drier	19	19	19	20
MIROC(hires)	20	Colder	19	Wetter	20	20	20	17

the lowest amount of precipitation during the wet and dry periods. Additionally, we note by the word “drier” or “wetter” if a particular model’s deficit is lower or higher than the observations. Similarly for the temperature surplus, the warmest models are assigned the lowest ranks and the words “warmer” or “cooler” are used to describe each model relative to the observations during these periods. In addition to being warm and dry more often than the observation, nearly all the models are warmer on average during these periods than the observations. We note that the 5 of the best 6 models ranked by the drought area index are both warmer and drier than the observations. We also note that the worst performers tend to be biased wet with 9 of the bottom 11 models described accordingly. But there are some notable inconsistencies in this ranking. For instance, the Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (CGCM3.1_t63) is biased both wet (and warm), yet performs well in three of the four drought metrics relative to other models. And L’Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4), which is both warmer and drier than the observations, performs very poorly in all four drought metrics. Additionally, interpretation of the relative ranks for the warm or dry metrics in the context of model performance in simulating drought is not very informative. We are left to

conclude that the nine-month warm and dry period average precipitation deficit and temperature surplus are only suggestive of model performance in simulating the PDSI and that some other, perhaps nonlinear, combination of precipitation and temperature variations controls the balance between evaporation and precipitation.

3. Drought metrics and projections of the future

Projections of future drought statistics through the PDSI can be expressed in a number of ways. A change in the mean climatology is reflected as a change in the time-averaged value of PDSI. The variations of PDSI around such an altered climatology are also important. The PDSI metrics described in the previous section are useful to describe these altered drought statistics. The time-dependent metrics depicted in Figs. 3a,b, the average area experiencing moderate drought or extreme drought, provides one measure of projected drought severity. The spatially dependent map of the time spent in moderate drought or extreme drought, as depicted for the present-day observations in Figs. 4a,b, is another such measure.

The correction factors applied to the simulated precipitation and temperature generally improve the models’ ability to simulate the four PDSI metrics defined in the previous section averaged over the reference period 1950–99. The correction also significantly amplifies the

late twentieth century and entire twenty-first century trends. In Fig. 3, a positive trend in the multimodel mean moderate drought areal extent begins around 1975. For extreme drought areal extent, a trend in the multimodel mean begins to appear around 2000. Detection and attribution of a human influence on these trends at this spatial scale is unlikely at the present time because of a large natural variability. In fact, inspection of the results in Fig. 3 considering only the observed data does not suggest anything particularly unusual about the recent past compared to the early portion of the twentieth century. However, if the future real world bears any resemblance to this ensemble of projections, detection and attribution of a human influence on this regional measure of drought could be expected by the middle of the twenty-first century, if not considerably earlier. Figure 6a shows the projected multimodel average value of PDSI over North America at the end of the twenty-first century under the SRES A1B forcing scenario. In this projection, what is considered severe to extreme drought conditions would be the normal climatological state over much of the continental United States and Mexico. In fact, roughly two-thirds of the area in this region's normal state would be considered moderate drought conditions today and a tenth would be considered extreme drought conditions. These fractions refer to a change in the drought climatology, and the average fractional area actually in moderate drought or extreme drought conditions as pictured in Figs. 3a,b could be different. Interestingly, this fraction is also about two-thirds for moderate drought conditions but is about a quarter for extreme drought conditions at the end of this century. We interpret that for those regions where drought conditions become the norm, not having drought is the unusual event. Likewise, extreme drought is not as rare an event as it is currently. Even in much of Canada, where precipitation is projected to increase by all models (Karl et al. 2009), moderate drought or mild drought conditions are projected to be the normal state. Here and perhaps elsewhere, the increased precipitation does not offset the increase in evapotranspiration due to warmer surface temperature. This leads to a reduction in soil moisture that is reflected in negative values of PDSI.

Multimodel projections of the future can be made more credible if objective measures of model performance can be used to weight each model's contribution to the multimodel mean. The variations in model rank shown in Table 3 reveal that a single performance metric does not suffice to sort out the good from the bad models. Any weighting scheme chosen to construct a multimodel mean is arbitrary and different schemes would certainly differ in the details. Santer et al. (2009) ranked climate models by

70 different metrics both parametrically and nonparametrically for a detection and attribution study of atmospheric moisture trends. Although they found it impossible to identify any model "best" in all categories, it was possible to identify certain models that could be judged superior when considering the entire set of metrics. For projection studies, consideration of metrics based on the quantities to be predicted is fair game. Hence, the area and correlation metrics in Table 3 can be used to rank models for their usefulness in projecting future PDSI drought statistics. As the uncentered correlation provides marginally more information for drought conditions than the centered correlation, we compute the model ranking based on the correlations with the mean retained. The reader should regard this choice as an arbitrary one to illustrate the implementation of ranking for climate change projection purposes. Following Santer et al. (2009), we normalize the metrics to use them as the basis for both parametric and nonparametric ranks in Table 4. The nonparametric score for a given model is determined by simply averaging the model rank for each of the four metrics shown in columns 2–5 of Table 3. The parametric score is determined by the average of the four normalized metrics themselves shown in columns 3–6 of Table 4.

Normalization of the errors allows us to consider the different categories of errors with equal weighting in the construction of the parametric rank. Since the errors are not necessarily regularly distributed, a model is strongly penalized in a parametric ranking if it performs poorly on a single metric. The nonparametric ranking also penalizes such a model, but with a different weight. In fact, the parametric and nonparametric rank shown in the first two columns of Table 4 are very similar. The top five best performing models according to parametric rank are 1) CSIRO Mk3.0, 2) PCM, 3) Community Climate System Model, version 3 (CCSM3), 4) Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL CM2.1), and 5) ECHAM5. (Actually, the mean model would be fourth on the list but we will exclude it from this part of the discussion.) The worst five models are 19) MIROC(hires), 18) MIROC, medium-resolution version [MIROC(medres)], 17) IPSL CM4, 16) INMCM3, and 15) Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3 (CNRM-CM3). Nonparametric ranking changes the order only slightly; see Table 4.

A selection of only the top five models in the construction of a mean model can provide a significant improvement in these metrics over the inclusion of all of the models in a mean model. For the top five (corrected) models, using the parametric ranking, the 1950–99 average continental U.S. and Mexico moderate drought

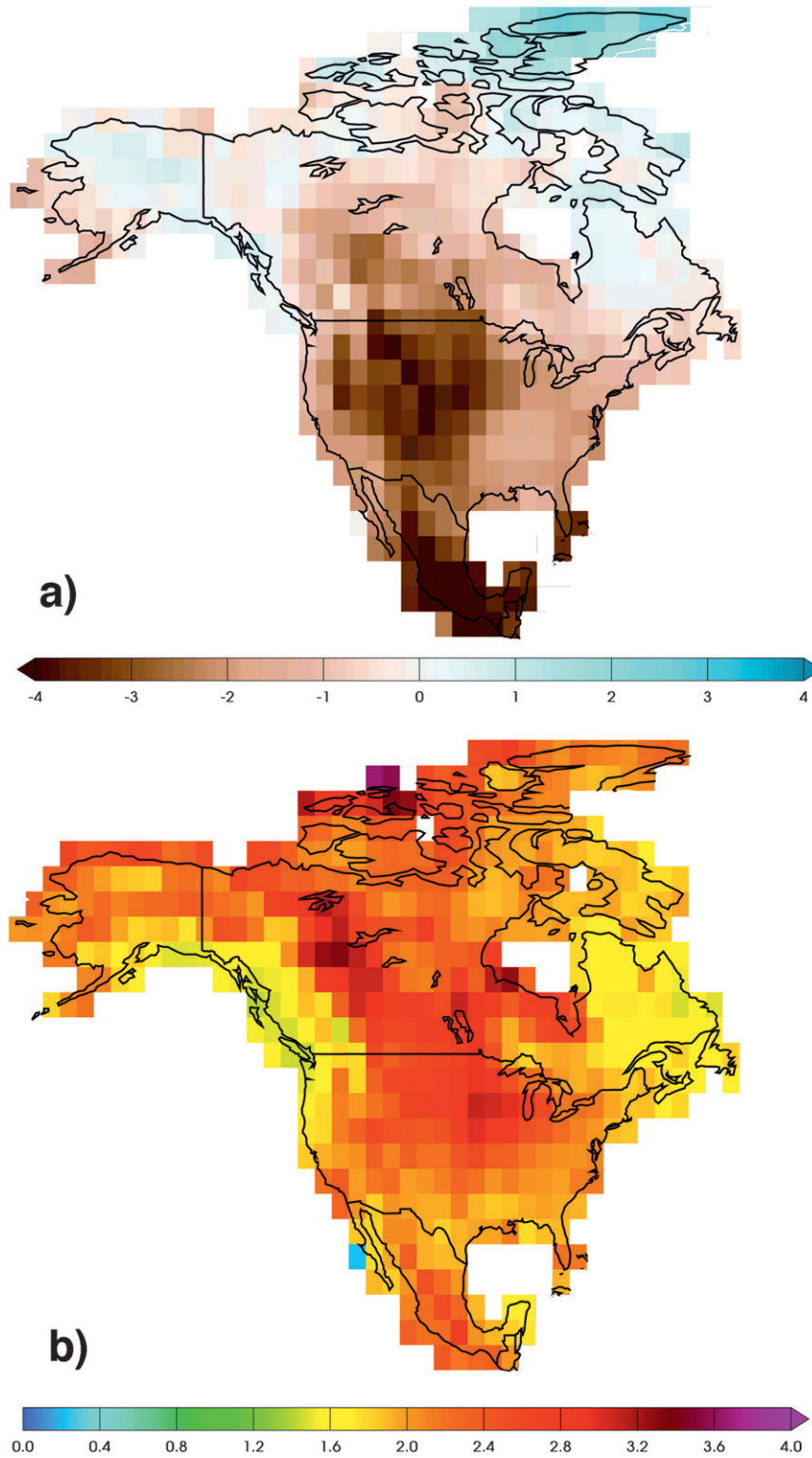


FIG. 6. (a) Projected multimodel mean PDSI averaged over the period 2089–98 for North America from 19 CMIP3 models driven by the SRES A1B forcing scenario. (b) Intermodel standard deviation of PDSI for this same period and forcing scenario.

TABLE 4. Normalized drought metrics with parametric and nonparametric scores for the 19 models in this study and the mean model. Associated parametric and nonparametric relative ranking are shown in parenthesis. The mean model is unranked.

	Nonparametric metric	Parametric metric	Area metric 2	Area metric 4	Correlation metric 2	Correlation metric 4
PCM	2.25(1)	0.87(2)	0.95	2.30	0.23	0.00
CSIRO Mk3.0	3.75(2)	0.86(1)	0.62	1.69	0.79	0.34
CCSM3	5.75(3)	1.39(3)	1.87	2.61	0.65	0.45
Mean model	6.00	1.56	2.39	3.71	0.00	0.16
MRI CGCM2.3.2a	7.00(4)	1.66(6)	1.92	3.90	0.30	0.52
GFDL CM2.1	7.25(5)	1.63(4)	1.59	2.28	0.85	1.80
FGOALS-g1.0	7.50(6)	1.72(9)	2.03	3.79	0.14	0.93
ECHAM5	7.75(7)	1.63(5)	1.86	2.89	0.94	0.84
CCCma CGCM3.1	8.25(8)	1.66(7)	2.15	3.02	0.72	0.75
CCCma CGCM3.1_t63	8.25(9)	1.68(8)	1.64	3.00	1.46	0.62
GFDL CM2.0	8.75(10)	1.96(10)	1.45	3.67	0.96	1.77
BCCR-BCM2.0	12.00(11)	2.28(11)	2.39	4.15	1.12	1.45
HadCM3	12.50(12)	2.42(12)	2.78	4.41	0.61	1.86
ECHO-G	12.50(13)	2.47(13)	2.94	4.48	0.60	1.84
HadGEM1	12.50(14)	2.49(14)	2.98	4.30	0.39	2.31
CNRM-CM3	15.00(15)	2.74(15)	2.69	4.14	1.17	2.97
INMCM3	16.75(16)	3.08(16)	3.38	4.81	1.81	2.33
IPSL CM4	18.00(17)	3.35(17)	3.64	4.82	2.26	2.69
MIROC(medres)	19.00(18)	3.75(18)	4.14	5.07	2.54	3.24
MIROC(hires)	19.25(19)	4.08(19)	4.45	5.08	4.24	2.55

area is 12.4% and extreme drought area is 1.9%. From Table 2, this compares to values of 9.1% and 1.0%, respectively, when all models are included in the mean. The observed values, also from Table 2, are 16.9% and 3.4%.

A principal motivation for comparing the models against these metrics of simulating the past is to improve confidence in projections of future drought statistics. In fact, all models project increases in the severity of future North American drought, however the amount and distribution are very sensitive to the particular selection of models used in the projection. For instance, an end-of-century SRES A1B forcing projection of PDSI using all models indicates significantly more severe drought conditions than a projection using only the top 5 PDSI models described above. However, the reduction in projected future drought severity by choosing these top 5 PDSI models can be misleading as there is a strong correlation between models' climate sensitivity to CO₂ and their ability to replicate the four drought metrics.

Figure 7 shows the annual mean surface air temperature over the continental United States and Mexico subject to the IPCC 13-point filter averaged over all models, the top five models, and bottom five models, as well as the observations. The top five PDSI models, which happen to project less future drought area, also project significantly lower future temperatures over the region of interest than the average of all models, while the bottom five models project temperatures higher than the average of all models. Table 5 shows the

temperature increases over the continental United States and Mexico for each model compared against their PDSI metric rankings. From this table we note that the average surface air temperature increase of the top five PDSI models is a 0.8 K lower-than-average increase of all models for the end of the twenty-first century relative to the beginning of the twentieth century. We also note that the average surface air temperature increase of the bottom five PDSI models is 0.9 K higher than the average of all models. If this relationship between model

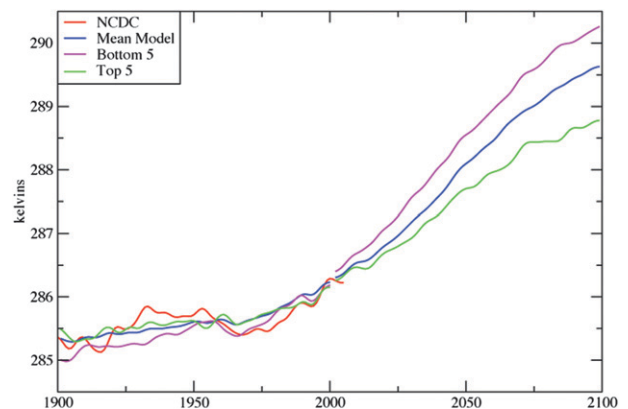


FIG. 7. Observed and corrected model surface air temperatures used in the calculation of PDSI averaged over the continental United States and Mexico. The annual average was further subject to the 13-point filter used in the IPCC AR4. Note that the 1950–99 average is the same for all four curves because of the correction period described in the text.

TABLE 5. Intermodel comparison of projected temperature changes (K) over the continental United States and Mexico to combined nonparametric ranking of four drought metrics. The third column shows the projected temperature change is difference between the 2089–98 and the 1900–09 averages. The numbers in parentheses are the order of the models in terms of amount of temperature change from lowest to highest. The last column shows the date when each model's running decadal mean globally averaged surface air temperature change first reaches +2.5 K over its 1900–09 value.

	Nonparametric rank	Surface air temperature difference (K)	Year that surface air temperature increases 2.5 K
PCM	1	2.8 (2)	2085
CSIRO Mk3.0	2	2.4 (1)	2110
CCCma CGCM3.1_t63	3	4.4 (11)	2038
MRI CGCM2.3.2a	4	3.3 (4)	2065
GFDL CM2.1	5	4.4 (10)	2057
FGOALS-g1.0	6	3.7 (5)	2069
ECHAM5	7	4.2 (9)	2057
CCSM3	8	3.8 (7)	2046
CCCma CGCM3.1_t47	9	4.4 (12)	2048
GFDL CM2.0	10	3.3 (3)	2089
BCCR-BCM2.0	11	3.7 (6)	2078
HadCM3	12	4.8 (15)	2059
HadGEM1	13	5.0 (16)	2054
ECHO-G	14	4.8 (14)	2059
CNRM-CM3	15	4.5 (13)	2045
INMCM3	16	4.0 (8)	2050
IPSL CM4	17	5.6 (18)	2049
MIROC(medres)	18	5.3 (17)	2054
MIROC(hires)	19	6.0 (19)	2029
Mean model		4.2	
Top five models		3.4	
Bottom five models		5.1	

climate sensitivity and drought metric skill is real, a relationship between long term climate sensitivity and variability on the PDSI time scale must also be real. A relationship between temperature or precipitation variability on the PDSI time scale (approximately nine months) and drought metric skill (see Table 3) is not definitely supported, although these numbers are suggestive for certain models. We are left to cautiously conclude that this correlation between climate sensitivity and model performance may be a sampling error due to the small size of the ensemble of CMIP3 models and should be considered coincidental.

Because of the strong relationship between evapotranspiration and air temperature noted earlier, comparing PDSI statistics between models of different average temperatures increases projection uncertainty. In a stationary climate, drought statistics are determined only by climate variability. In the nonstationary climate considered in projections, drought is determined both by climate variability and by the difference in the mean climate from that of the reference period. There is a wide range of temperature responses illustrated in Table 5. The intermodel standard deviation of projected PDSI from all 19 CMIP3 models at the end of the twenty-first century under SRES A1B forcing is shown in Fig. 6b and ranges from 1.5 to 3.5 over North America. Much of this

variation is due to mean temperature differences between models at the end of the century.

The surface air temperature sampling error incurred by choosing a subset of models based on performance on the four PDSI metrics could bias the projection for similar reasons. For the average of the top five PDSI models, the projected future drought is less at the end of the twenty-first century than when choosing all models because the average temperature is so much lower. This source of bias and uncertainty in highly derived quantities like PDSI can be reduced by slightly changing the nature of the projection. Typically, projections are framed by a time interval—that is, “How does some aspect of the climate change by the end of the twenty-first century under a specified forcing?” However, a related question may also be posed: “How does that aspect of the climate change when the temperature reaches a specified level?” This may be a more appropriate question in climate change impacts applications such as the present study (Clark et al. 2010). The practical difference is that models of differing climate sensitivities are analyzed over different periods rather than over the same one.

Figure 8a shows the PDSI from all 19 CMIP3 models averaged over the decade when each model's global mean surface air temperature first increases 2.5 K relative to the 1900–09 average. (Note: the choice of +2.5 K

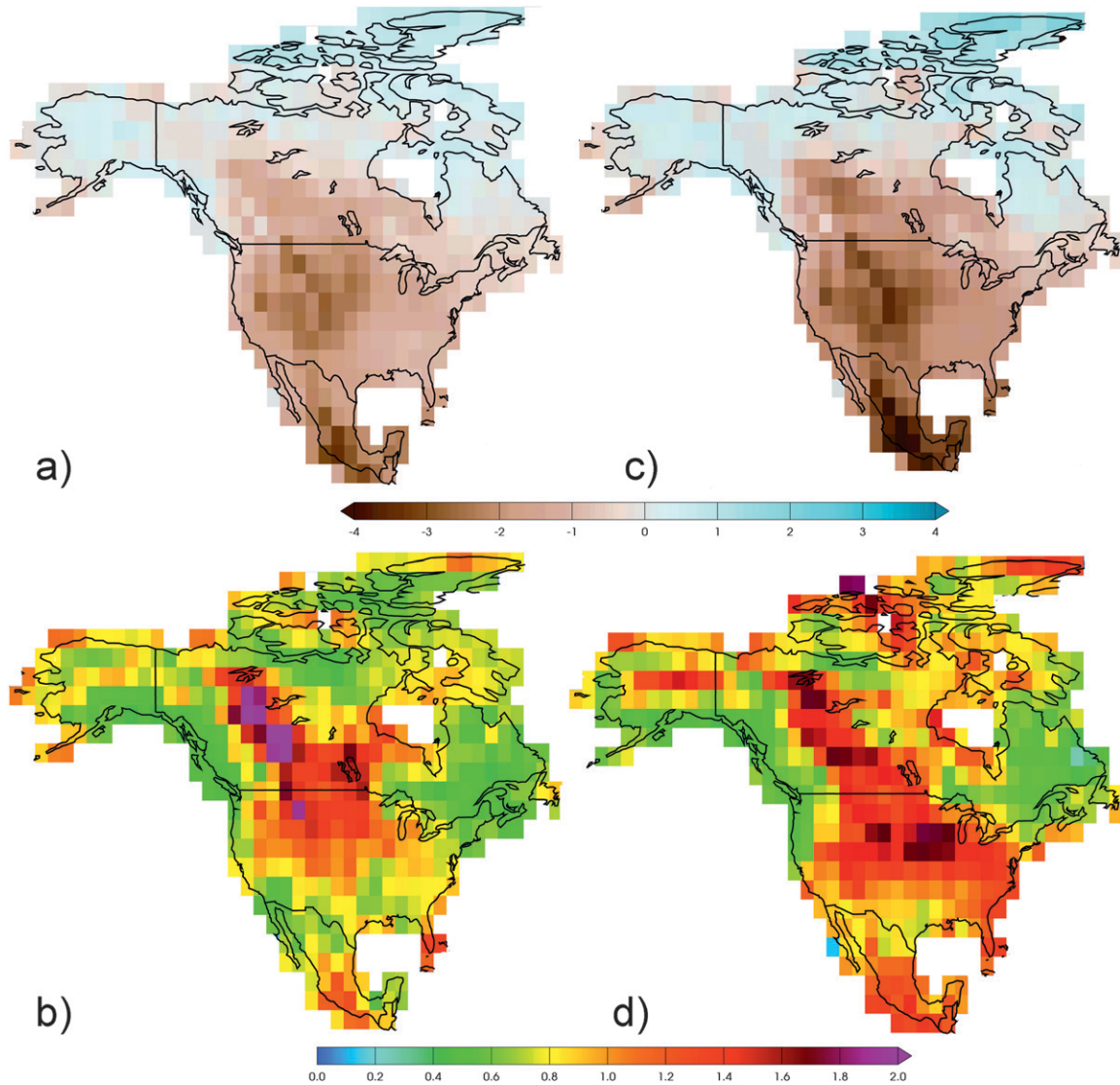


FIG. 8. (a) All-model average value of PDSI when the global average surface air temperature has increased 2.5 K over its 1900–09 mean value, (b) intermodel standard deviation of the values shown in (a), (c) all-model average value of PDSI for the decade centered at 2070, and (d) intermodel standard deviation of the values shown in (c).

is determined by the maximum warming of the model with the lowest climate sensitivity, which in the present study is CSIRO Mk3.0.) The date of this occurrence is listed in Table 5 for each model and ranges from 2038 to 2110. A running decadal mean was calculated for each model and compared to its 1900–09 mean temperature. For models with more than a single realization, the numbers in Table 5 are the average over realizations. However, each realization of each model was analyzed separately and the average performed over the +2.5 K date as a final step. As was the case when considering the end of the twenty-first century, widespread drying is projected over much of the continent when the global mean increases by 2.5 K. Severe drought conditions are

projected to be the normal state in southern Mexico while moderate drought conditions are projected for most of the western United States. For the continental United States and Mexico, about 35% of the region's climatology is moderate drought and about 5% is extreme drought in this projection. Figure 8b shows that the intermodel standard deviation of this projection of PDSI ranges from about 0.5 to 2. Two sources of intermodel variations contribute to this projection uncertainty. The first are the intermodel differences in the projected future changes in precipitation. The second, and perhaps smaller since the global mean changes are identical, are the differences between models in the spatial pattern of warming at the specified global surface air temperature

change of +2.5 K. As a comparison, Fig. 8c and shows the PDSI and its intermodel standard deviation from all 19 CMIP3 models averaged over the decade centered at 2070, which is when the mean model change is about +2.5 K (from Fig. 7). This estimate of projected PDSI is slightly higher, indicating that the individual models' changes do not combine in a linear fashion. The uncertainty in this estimate is slightly larger over most of the continent although actually a bit larger over the Canadian Rockies. Comparison with Figs. 6a,b, when the mean model global temperature is only 0.5 K higher, indicates both how rapidly multimodel PDSI might change toward the end of the century and how much more uncertainty in that estimate increases as due to the uncertainty in climate sensitivity.

The widespread drying projected by the all-model average in Fig. 8a is lessened when only the top five PDSI models are considered as in Fig. 9a. In this projection at +2.5 K, southern Mexico remains in severe drought but the drought in the western United States is reduced. PDSI is increased (indicating a reduction in drought) by about 0.5 in this region. Hence, the fractional area of the continental United States and Mexico where the normal state is moderate drought is about halved. The fractional area where the normal state is extreme drought does not change. Conversely, drought in this region is enhanced by considering the bottom five PDSI models as in Fig. 9b. However, these differences between the best and worst models are not statistically significant at the 90% confidence level according to a Student's *t* test using the intermodel variance calculated from all models (Fig. 8b) as estimates of the sample variances.

4. Discussion

We present an analysis of the ability of output from the IPCC AR4 climate models archived in the CMIP3 database to reproduce observed PDSI statistics over North America, particularly Mexico and the continental United States. Using the period 1950–99 as the base climatology, we find that none of the models are able to reproduce the frequency, severity, or extent of observed moderate drought conditions ($\text{PDSI} < -2$) or extreme drought conditions ($\text{PDSI} < -4$). Application of a multiplicative correction factor to the models' monthly temperature and precipitation fields to remove the temperature and precipitation bias over this period helps some of the models. We define six metrics of model performance based on drought severity as well as temporal and spatial measures applied over the base period. However, all models produce too little moderate and extreme drought in this region, even after correction of the biases. The difference in the ability to reproduce

observations varies greatly between the 19 models considered with two or three models performing nearly satisfactorily, four models failing spectacularly, and the remainder falling somewhat between these two extremes. Generally, the models simulate the amount of moderate drought conditions better than extreme drought conditions. However, the models tend to simulate the pattern of moderate drought conditions worse than for extreme drought conditions. A weak correlation between models' abilities to simulate the temporally based metrics and the spatially based metrics is found. However, there appears to be little or no correlation between model skill in reproducing these four drought metrics and model skill in reproducing the covariance of temperature and precipitation as well as the individual variances of temperature and precipitation. The absence of a connection between these second-order measures of variability and the higher-order variations quantified by PDSI is unexplained.

All models, regardless of their ability to simulate the base-period drought statistics, project significant future increases in drought frequency, severity, and extent over the course of the twenty-first century under the SRES A1B emissions scenario. Using all 19 models, the average state in the last decade of the twenty-first century is projected under the SRES A1B forcing scenario to be conditions currently considered severe drought ($\text{PDSI} < -3$) over much of continental United States and extreme drought ($\text{PDSI} < -4$) over much of Mexico. A significant amount of the intermodel uncertainty in this projection can be traced to differences in the models' climate sensitivity. The models with the largest temperature increase exhibit the largest increases in drought extent and severity because of the strong dependence of evapotranspiration on surface air temperature.

Periods of drought intensity comparable to the massive droughts of the 1930s or 1950s are replicated in the simulated twentieth century by the corrected models, albeit less frequently than observed. By the end of the twenty-first century, this condition becomes the normal one.

Using the four of the drought metrics, we have constructed both parametric and nonparametric model ranks to form a simple weighting scheme as a basis for improving confidence in multimodel projections. A projection using only the best models based on their performance in reproducing the four PDSI metrics leads to serious sampling errors because of an apparently coincidental relationship between model climate sensitivity and their ability to simulate these drought statistics. This bias may be removed by considering the PDSI change at a specified amount of temperature change rather than over a specific future time interval. Such a projection is less uncertain as local intermodel

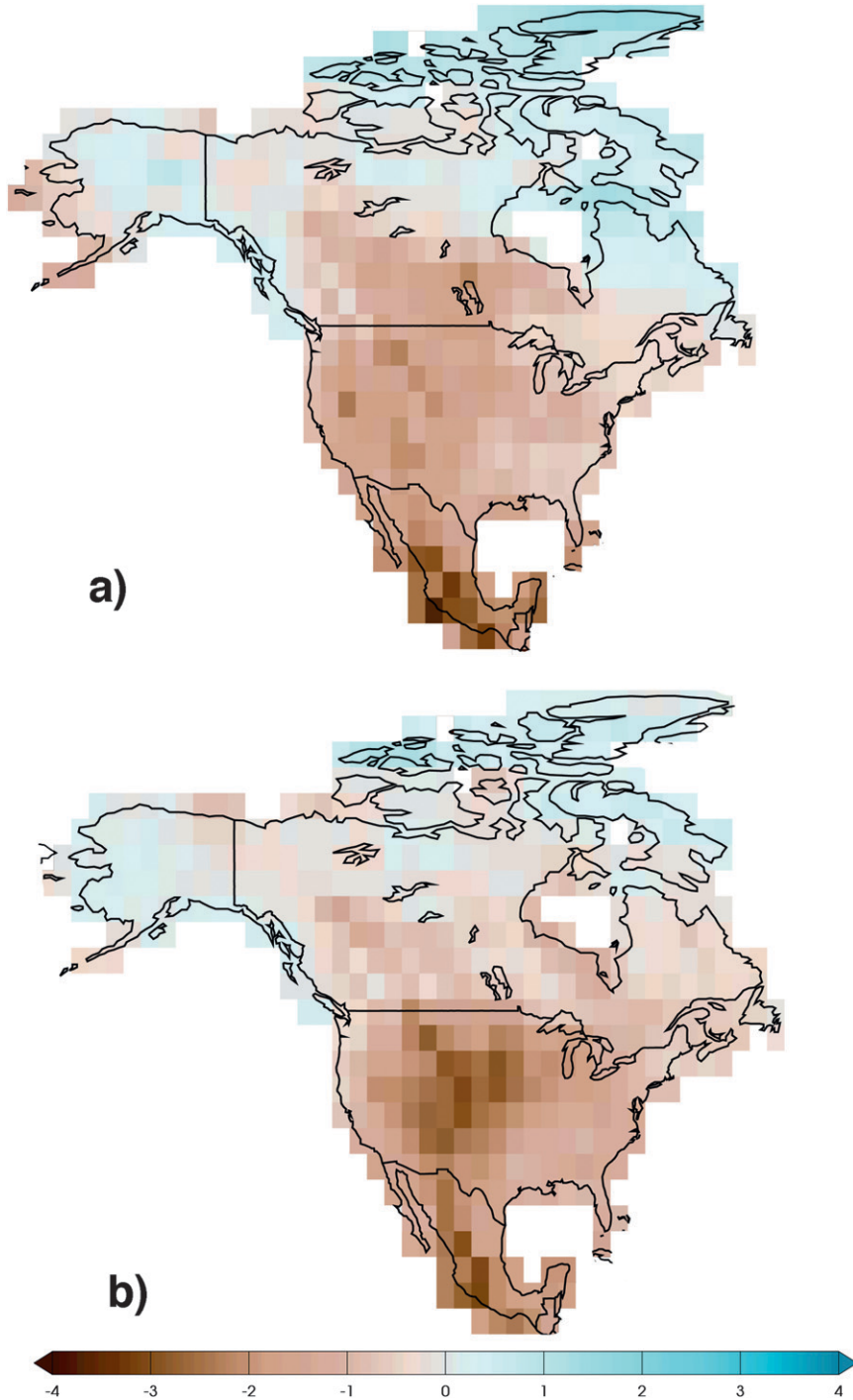


FIG. 9. (a) PDSI averaged over the top five PDSI models when the global average surface air temperature has increased 2.5 K over its 1900–09 mean value. (b) As in (a) except for the bottom five PDSI models.

temperature differences are much reduced. At a 2.5 K global increase in surface air temperature relative to the 1900–09 average, an all-model projection exhibits moderate drought conditions over most of the western

United States and severe drought over southern Mexico as the mean climatological state. Using the best five models, as determined by a nonparametric ranking of the models against the four selected PDSI metrics, leads

to a projection with a moderate reduction in the western U.S. drought. Usage of the five worst models increases the projected drought severity in this region. However, these differences are not highly statistically significant.

The response of PDSI to future temperature increases is very robust and indicates that in many regions increased evapotranspiration will lead to decreases in soil moisture regardless of how mean precipitation changes. This sensitivity calls into question the usefulness of the four PDSI performance metrics presented here in making projections of future drought statistics. The PDSI metrics are a measure of how well the models reproduce current high-order variability and covariability of precipitation and temperature. Of more importance to future projections of PDSI and other measures of soil moisture is the amount of warming. The true climate sensitivity remains very uncertain and is not constrained by these metrics.

Acknowledgments. This work was performed under the auspices of the U.S. Department of Energy (DOE) by the Lawrence Berkeley National Laboratory (LBNL) under Contract DE-AC03-76SF00098 (LBNL) and with support from the DOE Regional and Global Climate Modeling Program. Support for the National Climatic Data Center was provided by the U.S. Department of Energy, Office of Biological and Environmental Sciences under Interagency Agreement DE-AI02-96ER62276, and the NOAA/Climate Program Office. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI), and the WCRP's Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

APPENDIX A

The Palmer Drought Severity Index

The Palmer drought severity index (PDSI) assesses the total environmental moisture status. It incorporates information on antecedent precipitation, moisture supply, moisture demand, and soil moisture into a hydrologic accounting system (Palmer 1965; Heim 2002). The PDSI is a dimensionless index that measures both drought and wet spell conditions. Moisture demand (evapotranspiration) is estimated from monthly mean temperature and solar zenith angle using a Thornthwaite model. The two-layer model used for soil moisture computations assumes moisture is not transferred to the bottom layer until the top layer is saturated, runoff does not occur until both soil layers are saturated, and all of the precipitation occurring in a month is utilized during that

month to meet evapotranspiration and soil moisture demand or be lost as runoff. The model, as originally designed by Palmer, does not account for precipitation that may fall as snow and, therefore, does not enter into the hydrologic computations during the month it occurs. It also assumes moisture can always freely move between the soil moisture layers—that is, the ground is never frozen. Alley (1984) discussed these limitations and others, including how the Palmer model and water balance models in general treat the distribution of precipitation and evapotranspiration within a month or week and how they fail to consider seasonal or annual changes in vegetation cover and root development. Several authors (referenced in Heim 2002) noted that the PDSI, as originally developed by Palmer, treats the drought problem in semiarid and dry subhumid climates where local precipitation is the sole or primary source of moisture and that it may not be applicable to other areas, that the drought severity classes were arbitrarily assigned, and that the normalization process Palmer used may not be adequate to allow direct spatial comparisons. The backstepping process used to determine the beginning and ending of droughts and wet spells is dependent on antecedent and future moisture conditions, which makes it best suited for nonoperational computations.

Some of these concerns have been addressed in subsequent modifications of the Palmer model. For example, the enhanced Palmer drought index (EPDI) was created for cold climate conditions in Alberta, Canada, and incorporates snowpack, mountain precipitation, stream flow, and soil moisture conditions (Agriculture and Agri-Food Canada 2002). The self-calibrating Palmer drought severity index (SC-PDSI) automatically calibrates the behavior of the index at any location by replacing empirical constants in the index computation with dynamically calculated values, thus making it more spatially consistent (Wells et al. 2004). A probability factor was incorporated into the backstepping process to enable a more accurate assessment of drought termination in operational environments (Heddinghaus and Sabol 1991). In spite of the criticisms, the PDSI remains a popular model because it addresses both sides of the drought equation (water supply and water demand) and is easily computed from readily available variables.

The operational PDSI utilized by NOAA/NCDC used in this study incorporates the probability factor developed by Heddinghaus and Sabol (1991) and requires monthly precipitation and mean temperature as input. Drought is classified into the following categories: incipient ($-0.5 \geq \text{PDSI} > -1.0$), mild ($-1.00 \geq \text{PDSI} > -2.00$), moderate ($-2.00 \geq \text{PDSI} > -3.00$), severe ($-3.00 \geq \text{PDSI} > -4.00$), and extreme ($-4.00 \geq \text{PDSI}$).

APPENDIX B

Temperature and Precipitation Observations

The observed data used here consist of 5360 temperature and 7544 precipitation measurements between 15° and 52°N, the study area extending northward across the Canada–United States border to eliminate “edge” effects. The U.S. and Canadian data were from their respective national archives (the National Climatic Data Center and Environment Canada) while the Mexican data were provided by A. Douglas (Creighton University, 2008, personal communication). All temperature series were adjusted using the pairwise approach of Menne and Williams (2009) to account for historical changes in station location, instrumentation, and observing practice. The irregularly spaced station data were interpolated to a half degree latitude–longitude grid using a three-step approach known as climatologically aided interpolation (Willmott and Robeson 1995). The first step involved gridding the 1961–90 base-period normals using trivariate thin plate smoothing splines that employed latitude, longitude, and elevation as predictors (as in Hutchinson et al. 2009). The second step involved gridding the temperature and precipitation anomalies in each year and month for the period 1895–2005 using the inverse distance weighting approach of Willmott et al. (1985). Finally, the gridded anomalies were added to the gridded climatologies in each year and month to create the actual temperature and precipitation grids in each year and month.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Agriculture and Agri-Food Canada, cited 2002: Agriculture drought risk management plan for Alberta: Strategic plan. [Available online at [http://www1.agric.gov.ab.ca/\\$department/deptdocs.nsf/all/ppe3883](http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/ppe3883).]
- Alley, W. M., 1984: The Palmer Drought Severity Index: Limitations and assumptions. *J. Climate Appl. Meteorol.*, **23**, 1100–1109.
- Burke, E. J., S. J. Brown, and N. Christidis, 2006: Modeling the recent evolution of global drought and projections for the twenty-first century with the Hadley Centre climate model. *J. Hydrometeorol.*, **7**, 1113–1125.
- Clark, R. T., J. M. Murphy, and S. J. Brown, 2010: Do global warming targets limit heatwave risk? *Geophys. Res. Lett.*, **37**, doi:10.1029/2010GL043898.
- Easterling, D. R., T. W. R. Wallis, J. H. Lawrimore, and R. R. Heim Jr., 2007: Effects of temperature and precipitation trends on U.S. drought. *Geophys. Res. Lett.*, **34**, L20709, doi:10.1029/2007GL031541.
- Guttman, N. B., 1998: Comparing the Palmer Drought Index and the Standardized Precipitation Index. *J. Amer. Water Resour. Assoc.*, **34**, 113–121.
- Heddinghaus, T. R., and P. Sabol, 1991: A review of the Palmer Drought Severity Index and where do we go from here? Preprints, *Seventh Conf. on Applied Climatology*, Salt Lake City, UT, Amer. Meteor. Soc., 242–246.
- Heim, R. R., Jr., 2002: A review of twentieth-century drought indices used in the United States. *Bull. Amer. Meteor. Soc.*, **83**, 1149–1165.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Hutchinson, M. F., D. W. McKenney, K. Lawrence, J. H. Pedlar, R. F. Hopkinson, E. Milewska, and P. Papadopol, 2009: Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. *J. Appl. Meteor. Climatol.*, **48**, 725–741.
- Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its variations over the last 150 years. *Rev. Geophys.*, **37**, 173–199.
- Karl, T. R., J. M. Mellilo, T. C. Peterson, Eds., 2009: Global climate change impacts in the United States: A state of knowledge report from the U.S. Global Change Research Program. Cambridge University Press, 188 pp. [Available online at <http://downloads.globalchange.gov/usimpacts/pdfs/climate-impacts-report.pdf>.]
- Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717.
- Palmer, W. C., 1965: Meteorological drought. U.S. Weather Bureau Research Paper 45, 58 pp. [Available online at www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf.]
- Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. USA*, **106**, 8441–8446.
- Rind, D., R. Goldberg, J. Hansen, C. Rosenzweig, and R. Ruedy, 1990: Potential evapotranspiration and the likelihood of future drought. *J. Geophys. Res.*, **95** (D7), 9983–10 004, doi:10.1029/JD095iD07p09983.
- Santer, B. D., and Coauthors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Natl. Acad. Sci. USA*, **106**, 14 778–14 783, doi:10.1073/pnas.0901736106.
- Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, 2004: Causes of long-term drought in the U.S. Great Plains. *J. Climate*, **17**, 485–503.
- Seager, R., Y. Kushnir, C. Herweijer, N. Naik, and J. Velez, 2005: Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856–2000. *J. Climate*, **18**, 4065–4088.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.
- Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating Palmer Drought Severity Index. *J. Climate*, **17**, 2335–2351.
- Willmott, C. J., and S. M. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.*, **15**, 221–229.
- , C. M. Rowe, and W. D. Philpot, 1985: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Amer. Cartographer*, **12**, 5–16.

Copyright of Journal of Hydrometeorology is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.