Bridging Past and Future Climate across Paleoclimatic Reconstructions, Observations, and Models: A Hydroclimate Case Study*

JASON E. SMERDON

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

BENJAMIN I. COOK

NASA Goddard Institute for Space Studies, New York, and Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

EDWARD R. COOK AND RICHARD SEAGER

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

(Manuscript received 13 June 2014, in final form 7 December 2014)

ABSTRACT

Potential biases in tree-ring reconstructed Palmer drought severity index (PDSI) are evaluated using Thornthwaite (TH), Penman–Monteith (PM), and self-calibrating Penman–Monteith (SC) PDSI in three diverse regions of the United States and tree-ring chronologies from the North American drought atlas (NADA). Minimal differences are found between the three PDSI reconstructions and all compare favorably to independently reconstructed Thornthwaite-based PDSI from the NADA. Reconstructions are bridged with model-derived PDSI_TH and PDSI_PM, which both closely track modeled soil moisture (near surface and full column) during the twentieth century. Differences between modeled moisture-balance metrics only emerge in twenty-first-century projections. These differences confirm the tendency of PDSI_TH to overestimate drying when temperatures exceed the range of the normalization interval; the more physical accounting of PDSI_PM compares well with modeled soil moisture in the projection interval. Remaining regional differences in the secular behavior of projected soil moisture and PDSI_PM are interpreted in terms of underlying physical processes and temporal sampling. Results demonstrate the continued utility of PDSI as a metric of surface moisture balance while additionally providing two recommendations for future work: 1) PDSI_PM (or similar moisture-balance metrics) compare well to modeled soil moisture and are an appropriate means of representing soil-moisture balance in model simulations and 2) although PDSI_PM is more physically appropriate than PDSI_TH, the latter metric does not bias tree-ring reconstructions of past hydroclimate variability and, as such, reconstructions targeting PDSI_TH can be used with confidence in data–model comparisons. These recommendations and the collective results of this study thus provide a framework for comparing hydroclimate variability within paleoclimatic, observational, and modeled data.

1. Introduction

The increasing availability of forced-transient simulations over the last millennium (e.g., Fernández-Donado et al. 2013; Masson-Delmotte et al. 2014) from fully coupled general circulation models (GCMs) has drastically improved our ability to compare paleoclimate reconstructions with model output and to investigate multidecadal to centennial-scale climate dynamics (Schmidt et al. 2014). Among the currently available collection of simulations, the phase 5 of the Coupled Model Intercomparison Project (CMIP5) and phase 3 of the Paleoclimate Modelling Intercomparison Project (PMIP3) have for the first time produced multiple last-millennium, historical, and future simulations using the same model configurations and resolutions (Taylor et al. 2012). This important development makes comparisons between

Denotes Open Access content.

* Lamont-Doherty Earth Observatory Publication Number 7873.

Corresponding author address: Jason E. Smerdon, Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W, P.O. Box 1000, Palisades, NY 10964.
E-mail: jsmerdon@ldeo.columbia.edu

DOI: 10.1175/JCLI-D-14-00417.1

paleoclimatic data and last-millennium simulations directly applicable to historical simulations and future projections. Coincident with these advances in modeling efforts, the number of gridded or regional mean proxy reconstructions of multiple climatic variables is also increasing (e.g., Mann et al. 2009; Cook et al. 2010; PAGES 2k Consortium 2013; Wahl and Smerdon 2012; Neukom et al. 2010, 2011; Trouet et al. 2009), as is our understanding of the methods used to perform these reconstructions (e.g., Jones et al. 2009; Smerdon 2012; Tingley et al. 2012), expanding the detail and accuracy with which the actual climate of the last several millennia is characterized. This collective progress opens the possibility that comparisons between paleoclimatic reconstructions and model simulations can be used to improve understanding of decadal to centennial climate dynamics and to constrain model projections of twenty-first-century climate change in truly direct and quantitative ways (e.g., Schmidt 2010; Ault et al. 2014).

Many paleoclimate data–model comparison studies are emerging that both interpret comparison results and work to refine the methods by which the comparisons are made (e.g., Phipps et al. 2013; Schmidt et al. 2014; Hind et al. 2012; Hind and Moberg 2013; Coats et al. 2013a,b, 2015b; Anchukaitis et al. 2010; Seager et al. 2008; Fernández-Donado et al. 2013; Ault et al. 2013a,b; Lehner et al. 2012; Goosse et al. 2010, 2012; Sundberg et al. 2012). This collection of studies, inter alia, has highlighted numerous considerations for how comparisons should accommodate the unique strengths, weaknesses, and uncertainties of paleoclimatic reconstructions and model simulations. Each new comparison requires attention to the type of proxies used, the climatic variables considered, the means by which different models and model experiments are incorporated, and ultimately how to statistically characterize an ensemble of comparison results.

Among the collection of studies, paleoclimate data–model comparisons of hydroclimatic variability over the last millennium (e.g., Seager et al. 2008; Coats et al. 2013b, 2015b; Ault et al. 2013a; Tierney et al. 2013) is one emerging and important area of focus, given the critical social and ecological implications of hydroclimate variability and change (e.g., Allen and Ingram 2002; Hoerling et al. 2013, 2014; Ding et al. 2011; Headey 2011; Li et al. 2011; Lobell et al. 2011; Peng et al. 2011; Seager et al. 2013). Robust, high-resolution, and gridded drought atlas reconstructions that span most, if not all, of the last millennium and part of the first millennium of the Common Era are one critical tool for efforts on these time scales (Cook et al. 2007, 2010; Cook et al. 2014b). Comparisons between models and drought atlases are nevertheless complicated by the fact that the reconstructions have targeted the Palmer drought severity index (PDSI; Palmer 1965), an integrated estimate of soil-moisture balance that is not a simulated state variable in model integrations. In addition to the fact that model data must therefore be used in offline calculations that estimate PDSI for comparison to hydroclimate reconstructions [emerging capabilities using process-based tree-growth models (e.g., Anchukaitis et al. 2006; Evans et al. 2006) may provide an alternative approach in which model data are used to directly estimate tree growth chronologies that are not calibrated on PDSI or other climate variables], a growing debate has emerged around the efficacy of PDSI as a metric of soil-moisture balance in observations and model simulations (e.g., Guttman 1998; Vicente-Serrano et al. 2010b; Burke et al. 2006; Burke and Brown 2008; Burke 2011; Dai 2011a,b, 2013; van der Schrier et al. 2011, 2013; Seneviratne 2012; Hoerling et al. 2012; Sheffield et al. 2012; Trenberth et al. 2014; Cook et al. 2014a).

Much of the discussion and criticism of PDSI has hinged on its different formulations and more specifically on the means by which potential evapotranspiration (PET) is estimated within the PDSI calculation. A common method for estimating PET, the Thornthwaite formulation (Thornthwaite 1948), scales PET as a function of temperature and latitude only, and associated Thornthwaite-based PDSI estimates (PDSI_TH) can consequently overestimate drying when temperatures exceed the range of variability spanned by the PDSI normalization interval (Hoerling et al. 2012; Sheffield et al. 2012; Dai 2013). This effect has been shown to be significant for PDSI_TH estimates at the end of the twentieth century in observational records (Sheffield et al. 2012) and in the use of PDSI_TH as a moisture-balance metric in twenty-first-century model projections (Hoerling et al. 2012; Dai 2013; Schmidt et al. 2014). PDSI estimates that use the Penman–Monteith (PM) formulation (PDSI_PM) for PET alternatively have been proposed as more physically appropriate (e.g., van der Schrier et al. 2011), along with additional modifications using self-calibrating PDSI (PDSI_SC) that employ both the Penman–Monteith PET formulation and regionally estimated soil and vegetation properties (Wells et al. 2004; van der Schrier et al. 2013). Additional metrics, such as the standardized precipitation–evaporation index, which can incorporate Penman–Monteith estimated PET, also have been developed (e.g., Vicente-Serrano et al. 2010a; Hernandez and Uddameri 2014) and compare well with PDSI_PM in model projections (Cook et al. 2014a).

Despite the above discussions, PDSI remains a useful metric of soil-moisture balance for several reasons. While soil moisture is ultimately the applicable state variable for evaluating model-simulated changes in hydroclimate and drought, the land surface models in

coupled GCMs vary widely in their sophistication (e.g., soil depth, number of layers), tunings, and parameterizations (e.g., soil texture, rooting depths, vegetation types), thus complicating the meaningful comparison of direct soil-moisture variables across models. Offline metrics like PDSI therefore serve to homogenize accounting of soil-moisture balance and rely principally on atmospheric variables in their computation. Well-distributed records of observed soil moisture also are not widely available over many decades, making soil moisture a difficult quantity on which to validate simulated hydroclimate variability over the twentieth century. Recent work additionally has demonstrated the benefit of separating PET or PDSI into the constituent influences on these variables (Scheff and Frierson 2014; Cook et al. 2014a). The separation of such influences in coupled GCMs is not easily accomplished for model-simulated soil moisture in which variables such as temperature or precipitation cannot be held constant independent of other coupled variables. With regard to paleoclimatic data–model comparisons specifically, the currently available collection of drought atlases also have used PDSI as the reconstructed target variable, making it necessary to use model-derived PDSI for comparisons between simulations and reconstructions.

Concerns about paleoclimatic drought atlases nevertheless have been raised because tree-ring-derived products have traditionally targeted PDSI_TH (e.g., Cook et al. 2007, 2010). Sheffield et al. (2012) most recently noted that "paleoclimate reconstructions of drought may be particularly susceptible because they are often developed by scaling tree-ring data to match the calculated [PDSI] for their overlap period" (p. 437). The authors go on to surmise that such reconstructions may overestimate past changes while underestimating "recent trends in the context of the past" (p. 437). Such concerns are indeed serious, given the degree to which tree-ring-derived drought atlases have been used to characterize previous droughts and pluvials (e.g., Cook et al. 2007, 2010) and the extent to which such information has helped define potential multidecadal risk factors associated with hydroclimatic variability and change.

Herein we evaluate the dependence of regional reconstructions targeting different PDSI formulations using the same dendroclimatic chronologies used in the North American drought atlas (NADA) and observationally estimated PDSI_TH, PDSI_PM, and PDSI_SC. We derive new reconstructions in three diverse regions of the United States for each of these PDSI formulations, analyze their respective similarities and differences, and compare them to previous NADA estimates of PDSI_TH over the same regions, the latter of which are derived using a different reconstruction method, different predictor processing steps, and a spatially explicit PDSI_TH target on a $0.5° \times 0.5°$ latitude–longitude grid (in contrast to the regional indices that are targeted herein). We subsequently turn to bridging paleoclimatic estimates of hydroclimate variability with modeled twentieth-century climate simulations and twenty-first-century climate projections. This challenge is addressed by analyzing model-derived estimates of PDSI_TH; PDSI_PM; and two soil-moisture estimates from the CanESM2 and CCSM4 GCMs, which are both available through the CMIP5 data archive. In addition to comparing simulations and reconstructions, we characterize where and how the simulated PDSI and soil-moisture estimates agree within each model simulation. It is critically noted that the parameterized impacts of simulated CO_2 fertilization have implications for secular soil-moisture trends and may be an important source of uncertainty in comparisons between multiple PDSI and soil-moisture variables in twenty-first-century climate projections. Our investigation culminates with guidance on how to interpret our results as a framework for comparing hydroclimate variability across overlapping observational and modeling intervals and to ultimately use this framework to place future hydroclimate projections into a longer paleoclimatic context.

2. Data and methods

a. PDSI calculations

We focus on three areas of the United States that are representative of diverse hydroclimatic and vegetation regimes: the Four Corners (4C), northern plains (NP), and southeast (SE) regions as indicated in Fig. 1. These areas also contain well-documented and abundant tree-ring chronologies that will be used to derive regional reconstructions. Three different formulations of observational PDSI over the three regions are used: PDSI_TH, PDSI_PM, and PDSI_SC. The first two have been calculated using standard formulations (Thornthwaite 1948; Penman 1948; Xu and Singh 2002), while the third is taken from the self-calibrating PDSI dataset of van der Schrier et al. (2013). PET in PDSI_TH has the advantage of only requiring temperature and latitude data but is effectively a rescaling of these variables as an estimate of PET [see, e.g., the discussion in van der Schrier et al. (2011)],

$$\mathrm{PET} = 16\left(\frac{10T}{I}\right)^{a}, \tag{1}$$

where PET is in millimeters per month, T is the monthly mean of daily averaged temperature (°C), I is the heat index, and a is a third-order polynomial of the heat index

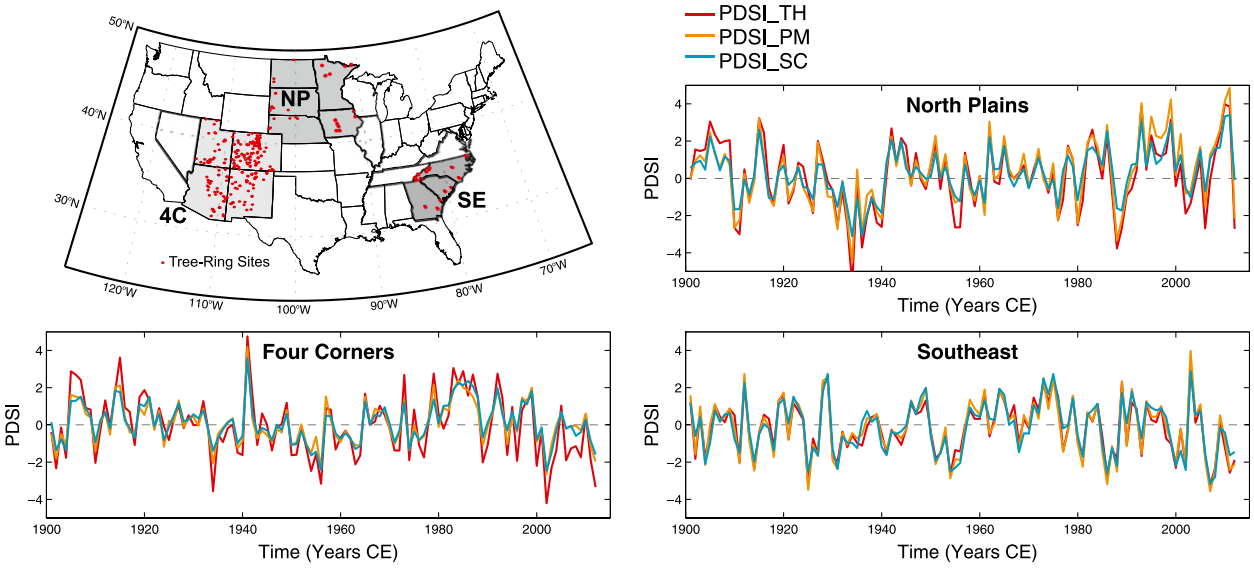

FIG. 1. Map of the contiguous United States and those states chosen to represent the Four Corners, Northern Plains, and Southeast (designated with gray shading). The tree-ring sites used in the regional reconstructions are shown as red dots. Plotted time series are the area-weighted mean PDSI estimates for each of the three regions estimated from observational data and three PDSI formulations: PDSI Thornthwaite, PDSI Penman–Monteith, and self-calibrating PDSI with Penman–Monteith.

(Thornthwaite 1948). This expression is scaled in the Thornthwaite calculation of PDSI to account for latitudinal variations in the length of months and days as

$$\mathrm{PET_{TH}} = \mathrm{PET}(\theta/30)(h/12), \tag{2}$$

where θ is the length of the month (in days) and h is the duration of daylight (in hours) on the 15th day of the month (e.g., van der Schrier et al. 2011; Willmott et al. 1985). The dependence of PET on T in Eq. (1) is thus the origin of the excessive drying estimated by PDSI_TH when temperatures are significantly outside the range of variability defined by the baseline normalization period.

The Penman–Monteith PDSI formulation (Penman 1948; Xu and Singh 2002) has been suggested as a physically appropriate alternative method for calculating PET in twenty-first-century projections (Dai 2013; Hoerling et al. 2012; van der Schrier et al. 2013; Sheffield et al. 2012; Cook et al. 2014a). The Penman–Monteith formulation is based on surface moisture and energy-balance considerations (Xu and Singh 2002) and a commonly used formulation is defined by the Food and Agricultural Organization (FAO) of the United Nations (Allen et al. 1998). In the FAO formulation, PET in millimeters per day is given by

$$\mathrm{PET} = \frac{0.408\Delta(R_n - G) + \gamma\dfrac{900}{T_a + 273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}, \tag{3}$$

where Δ is the slope of the vapor pressure curve (kPa °C^{-1}), R_n is the surface net radiation (MJ m^{-2} day^{-1}), G is the soil heat flux density (MJ m^{-2} day^{-1}), γ is the psychometric constant (kPa °C^{-1}), T_a is the air temperature at 2 m (°C), u_2 is the wind speed at 2 m (m s^{-1}), e_s is the saturation vapor pressure (kPa), and e_a is the actual vapor pressure (kPa). Although there have been some suggestions that PDSI_PM also may be susceptible to overestimated drying trends (Hoerling et al. 2012), this has not been widely observed in multiple studies (e.g., Sheffield et al. 2012; van der Schrier et al. 2011; Cook et al. 2014a) and reported differences in observationally based PDSI_PM estimates are more likely related to differences in the precipitation datasets used as inputs for various observational PDSI_PM calculations (Trenberth et al. 2014).

For both PDSI_TH and PDSI_PM calculations presented herein, the normalization interval is 1901–2012. Soil-moisture capacities for the top and bottom soil layers are set to the standard values of 25.4 mm (1 in.) and 127 mm (5 in.). Our observational PSDI_PM calculations use the estimate of PET from the updated version (3.21) of the latest CRU TS dataset, which has been derived using the FAO formulation of PET (Harris et al. 2014) given in Eq. (3). The Harris et al. (2014) CRU TS dataset is similarly the source for all other input variables used to calculate PDSI_TH and PDSI_PM. The PDSI_SC dataset of van der Schrier et al. (2013) also uses the FAO PET formulation and differs from our calculated version of PDSI_PM only in its incorporation

of regionally specific soil and vegetation properties and a snow model (van der Schrier et al. 2013); similar to PDSI_PM, the PDSI_SC estimates use climate data from the updated version (3.21) of the latest CRU TS dataset for observational forcing variables (Harris et al. 2014).

Area-weighted regional indices were calculated from the 0.5° × 0.5° gridded datasets of PDSI_TH, PDSI_PM, and PDSI_SC during the summer season (JJA) over the 4C, NP, and SE regions; all of the regional PDSI series begin in the year 1901 and extend to 2012. These regional series thus comprise the JJA PDSI calibration targets for the reconstructions discussed in the following subsections.

For calculations of PDSI_TH and PDSI_PM from the model simulations over the historical and projection intervals, we use the same conventions described above, with several additional applied choices. Modeled PET in the Penman–Monteith formulation is again calculated using the FAO version. Relative to changes in energy availability and the vapor pressure deficit, Penman–Monteith PET is relatively insensitive to near-surface wind speed (Scheff and Frierson 2014; Cook et al. 2014a), which we set equal to a constant $1\,m\,s^{-1}$ in the PDSI_PM model calculations. Ground heat fluxes similarly are only a small fraction of the total surface energy budget, typically totaling about 10%–15% (Betts et al. 1996; Sellers et al. 1997). We therefore set the ground heat flux equal to $0\,W\,m^{-2}$ in the model-derived estimates of PDSI_PM, a choice that has been found to yield little impact (Cook et al. 2014a).

We do not calculate model-derived estimates of PDSI_SC. Although the PDSI_SC calculation can be done in principle, it additionally includes a snow model and regionally varying vegetation and soil parameters. These considerations further remove the PDSI calculation from model-derived quantities (in the case of the snow model) and involve poorly constrained or unavailable model fields (vegetation and soil properties). In light of these factors and the good agreement that is later shown for PDSI_PM and PDSI_SC, we do not additionally calculate PDSI_SC fields for the models.

b. *Tree-ring chronologies and the NADA*

We use tree-ring width chronologies from the regional collection of states shown in Fig. 1. These chronologies are taken from the same network used in the NADA database to derive version 2a of the drought atlas (Cook et al. 2007). A total of 283, 36, and 26 chronologies are employed in the 4C, NP, and SE regions, respectively. The start and end dates vary across the collection of chronologies in each region, several of which extend to the first century of the first millennium of the common era, whereas many only begin in the eighteenth or nineteenth centuries. For our purposes herein, we only employ chronologies back to the year 1000; all chronologies extend to at least 1979. These constraints define calibration-interval selections and the nesting procedure for the reconstruction method detailed in section 2c.

The NADA version 2a is also used for comparison to the new regional reconstructions. Version 2a of the NADA is a gridded reconstruction on a 0.5° × 0.5° latitude–longitude grid of JJA average PDSI_TH values over much of North America. The grid has been reconstructed using a point-by-point regression scheme that employs a principal component ordinary least squares regression method, although multiple steps of predictor processing and ensemble estimation have been included [see Cook et al. (2007) and references therein for further details].

c. *Reconstruction method*

The new regional reconstructions are derived using a nested ''composite plus scale'' (CPS) method (e.g., Jones et al. 2009) with an ensemble correlation-weighting scheme (PAGES 2k Consortium 2013); 50-yr nests were used in each regional reconstruction that employed all chronologies available back to the beginning of each nest period. A CPS reconstruction was computed for each nest by standardizing (normalizing and centering) the available tree-ring chronologies over the calibration interval and subsequently calculating a weighted composite in which the relative weight of each chronology was determined by the strength of the correlation with the target index. Each composite was finally centered and scaled to have the same mean and variance as the target index during the calibration interval.

The CPS methodology was implemented using a resampling scheme for validation and calibration that uses 50- and 29-yr blocks for calibration and validation, respectively (the last year of uniformly available predictor series is 1979, providing 79 yr of overlap with the target indices that all start in 1901). The initial calibration period extends from 1901 to 1950 and was incremented by 1 yr until reaching the final period of 1930–79, yielding a total of 29 reconstructions for each nest. Within each calibration step, the 29 years excluded from calibration were used for cross validation. For each nest, the final CPS reconstruction was taken as the median reconstructed value in each year within the 29-member reconstruction ensemble. Uncertainties were estimated from the mean variance of the residuals across all of the validation intervals; 1.96 times the square root of the estimated mean variance was added to the median ensemble values in each year to derive the 95% confidence intervals. The final nested reconstruction was

combined by splicing the median reconstruction and estimated uncertainties of each nest such that every reconstructed year was derived from the nest with the maximum number of chronologies.

d. Model data

We use publicly available GCM output from the CMIP5 archive, the suite of model experiments organized and contributed in support of the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC). Output from the historical and RCP8.5 twenty-first-century projection experiments is used, the latter of which is the high-emission, business-as-usual scenario that has been justified in many studies by the current lack of international action to limit greenhouse gas emissions. Our analyses are nevertheless not dependent on the employed emissions scenario beyond the fact that the RCP8.5 simulations represent the most extreme warming scenario and therefore embody the maximum impact of temperature changes on PDSI comparisons. Historical CMIP5 experiments are run for the years 1850–2005 and are forced with observations of transient climate forcings [e.g., solar variability, land-use change, and greenhouse gas (GHG) concentrations]. These experiments are initialized in 1850 using output from long, unforced control runs with fixed preindustrial boundary conditions. The RCP8.5 projection scenario is one of a suite of future GHG forcing scenarios spanning the 2006–99 period; RCP8.5 is designed so that the top of the atmosphere radiative imbalance will equal approximately $+8.5\,\mathrm{W\,m^{-2}}$ by the end of the twenty-first century, relative to preindustrial conditions. The RCP8.5 projections are initialized using the end of the historical runs. Our analysis is restricted to two models (CanESM2 and CCSM4) that have available five continuous ensemble members spanning the historical through projection time intervals. These models and associated ensemble members also have been selected based on the availability of their layered soil-moisture output, which extends to ~4 and ~43 m over 3 and 15 layers in the CanESM2 and CCSM4 models, respectively. For both models, we employ a near-surface JJA soil-moisture estimate that is taken from approximately the first 30 cm in each model and full-column soil moisture taken over the total depth of each modeled soil column. For comparison to PDSI, both soil-moisture variables spanning the historical to projection interval are centered and scaled to match the corresponding PDSI_PM variance from 1901 to 2005, the same normalization window used to calculate the PDSI variables.

3. Analyses of target series and reconstructions

a. Observational estimates of PDSI

The area-weighted regional time series for the three observational PDSI estimates are shown for the 4C, NP, and SE regions in Fig. 1; all time series are centered to have a mean of zero over the reconstruction calibration/validation window (1901–79). The square of Pearson's correlation coefficient calculated between the series in each region is 0.83 or higher. The largest shared variances (r^2) occur over the SE region (0.96, 0.86, and 0.90), whereas the NP (0.88, 0.85, and 0.90) and 4C (0.85, 0.83, and 0.92) regions rank in descending order for parings of PDSI_TH versus PDSI_PM, PDSI_TH versus PDSI_SC, and PDSI_PM versus PDSI_SC, respectively. These relative comparisons are qualitatively consistent with the correlations between the Thornthwaite and Penman–Monteith versions of PDSI_SC reported by van der Schrier et al. (2011) for similar regions in North America.

The standard deviations of the observational PDSI series indicate regionally dependent differences in the variability of the PDSI estimates in the 4C (TH: 1.65; PM: 1.13; SC: 1.07), NP (TH: 1.84; PM: 1.73; SC: 1.27), and SE (TH: 1.38; PM: 1.50; SC: 1.27) regions. These estimates are consistent with previously reported comparisons between Thornthwaite- and Penman–Monteith-derived PDSI estimates (e.g., Sheffield et al. 2012, van der Schrier et al. 2011) in which PDSI_TH displays larger variance relative to the PDSI_PM formulation because of its greater sensitivity to temperature variations. Nevertheless, the results also indicate that in some regions, such as the SE, PDSI_PM has a larger variance than PDSI_TH; therefore, not all areas follow the generally described behavior.

b. Reconstructed estimates of PDSI

The median PDSI reconstructions and associated uncertainties for each region are compared to the corresponding observational target series in Fig. 2. Each reconstruction compares well with the targets, although there are some differences among the collection of validation statistics. The PDSI_TH reconstructions are generally the most skillful across all regions and nests (Fig. 3), while the PDSI_SC reconstruction is the second most skillful in the 4C region and the PDSI_PM reconstruction is second or first most skillful in the NP and SE regions. In all cases, the reconstructions yield validation-interval coefficients of determination over all nests that are skillful above the 99% ($p < 0.01$) significance level. Assuming the traditional threshold of zero for the reduction of error (RE) coefficient, all regional reconstructions are skillful across all nests, except for the

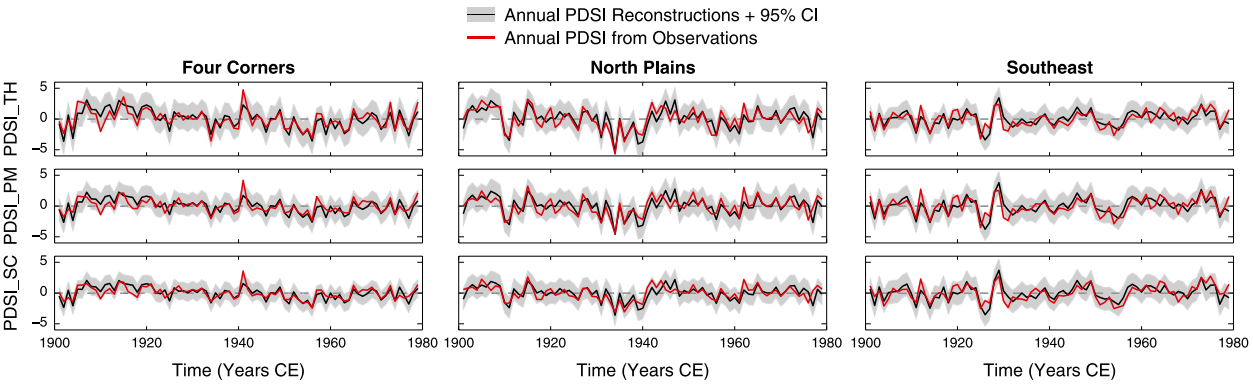

FIG. 2. (left to right) Targeted regional PDSI time series calculated from observational data for the PDSI_TH, PDSI_PM, and PDSI_SC formulations and the associated CPS reconstructions and 95% confidence intervals. Reconstructions (black line with associated gray shading) and observationally based estimates (red line) are shown during their common period of overlap from 1901 to 1979, which also comprises the calibration/validation interval for the reconstructions.

NP reconstructions, for which the PDSI_TH, PDSI_PM, and PDSI_SC reconstructions are only skillful back to 1500, 1500, and 1600, respectively. By the more stringent coefficient of efficiency (CE) statistic, which also uses a skill threshold of zero, the 4C and SE reconstructions are skillful across all nests, except the PDSI_SC reconstruction in the SE region during several nests in the early part of the millennium (Fig. 3). In the NP, all of the reconstructions yield negative CE values prior to 1600 and the PDSI_SC yields an additional negative nest from 1650 to 1700. Although not all nests in the NP region are skillful back to 1500 across all of the validation metrics, the reconstructed time series are truncated at that year to reflect the bulk validation performance described above.

Regional comparisons between each of the three different PDSI reconstructions yield consistent and very similar results, despite the small differences in the validation statistics discussed above. All regional pairings of reconstructions indicate shared variances of 0.99 or higher during the reconstruction intervals (1000–1900 for the 4C and SE regions and 1500–1900 for the NP). The only notable, though modest, differences between the reconstructions are the standard deviations for 4C (TH: 1.45; PM: 1.05; SC: 1.07), NP (TH: 1.64; PM: 1.39; SC: 1.10), and SE (TH: 1.26; PM: 1.39; SC: 1.32). With the exception of the SE, a region in which the three reconstructions are virtually identical, the PDSI_TH reconstructions exceed the standard deviations of the other two reconstructions by 20%–50%. These variance differences are further reflected in the right-hand panels of Fig. 3, in which the linear relationship between each combination of the reconstructions deviate from the one-to-one line by varying amounts in each region. Given the stated differences between the variances of the target indices, these differences in the standard deviations are expected based on the CPS methodology that matches variances between the composite reconstruction and target indices during the calibration interval. In other words, the standard deviations of each reconstruction reflect the relative differences between the variance of the target time series (although not perfectly so). Importantly, these differences in variance are merely a scaling of the overall reconstructions and do not reveal temporal variance dependencies that would bias interpretations of hydroclimate variability during the twentieth century relative to the past.

c. *Comparisons between NADA and the CPS reconstructions*

It is worth investigating whether the presented CPS reconstructions are representative of the NADA product, thus making our findings more applicable to interpretations of the latter. In Fig. 4 we compare the derived CPS PDSI_TH reconstructions to area-weighted regional averages from the NADA that have been sampled from gridded regions approximating the CPS reconstruction areas shown in Fig. 1 (see Figs. 11 and 12 for the specific sample regions extracted from the NADA grid). We limit the comparisons in Fig. 4 to PDSI_TH, given that the NADA also has targeted PDSI_TH. The shared variances between the NADA time series and the CPS reconstructions of PDSI_PM (4C: 0.94; NP: 0.65; SE: 0.68) and PDSI_SC (4C: 0.94; NP: 0.65; SE: 0.68) are nevertheless virtually identical to the shared variances reported for PDSI_TH in Fig. 4.

Comparisons between the NADA regional averages and the CPS PDSI_TH reconstructions are consistent, despite the different methods by which the reconstructions were derived. The CPS reconstructions target a single regional average and use a sliding calibration–validation scheme to weight chronologies based on correlation. The NADA index is the regional average of a gridded product that uses

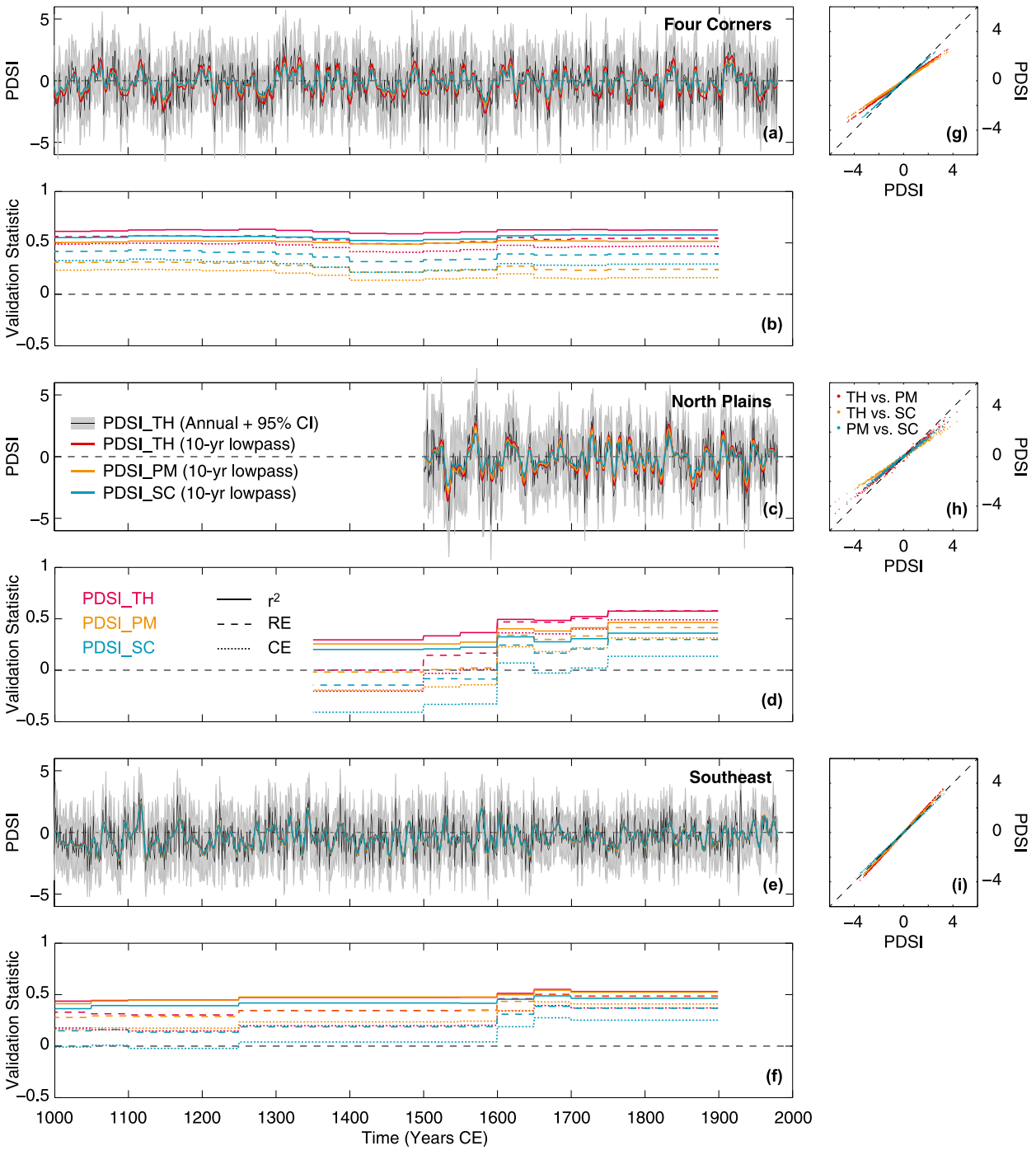

FIG. 3. (a),(c),(e) PDSI reconstructions for the three regions using the CPS method and three regional PDSI target series using the PDSI_TH, PDSI_PM, and PDSI_SC formulations. The 10-yr low-pass time series for each PDSI reconstruction (filtered using a 10-point Butterworth filter) are also shown; each panel also plots the annual PDSI_TH and associated 95% confidence intervals for reference. (b),(d),(f) Resolved variance, RE, and CE cross-validation statistics for each of the regional PDSI reconstructions as a function of each 50-yr nest. Figure legends for all panels are given in (c) and (d). (g)–(i) Comparisons between the annually reconstructed PDSI values are shown in scatterplots, in which each value of reconstructed PDSI is plotted against the other. Scatterplots do not include the calibration–validation interval from 1901 to 1979 and a dashed 1:1 line is plotted for reference.

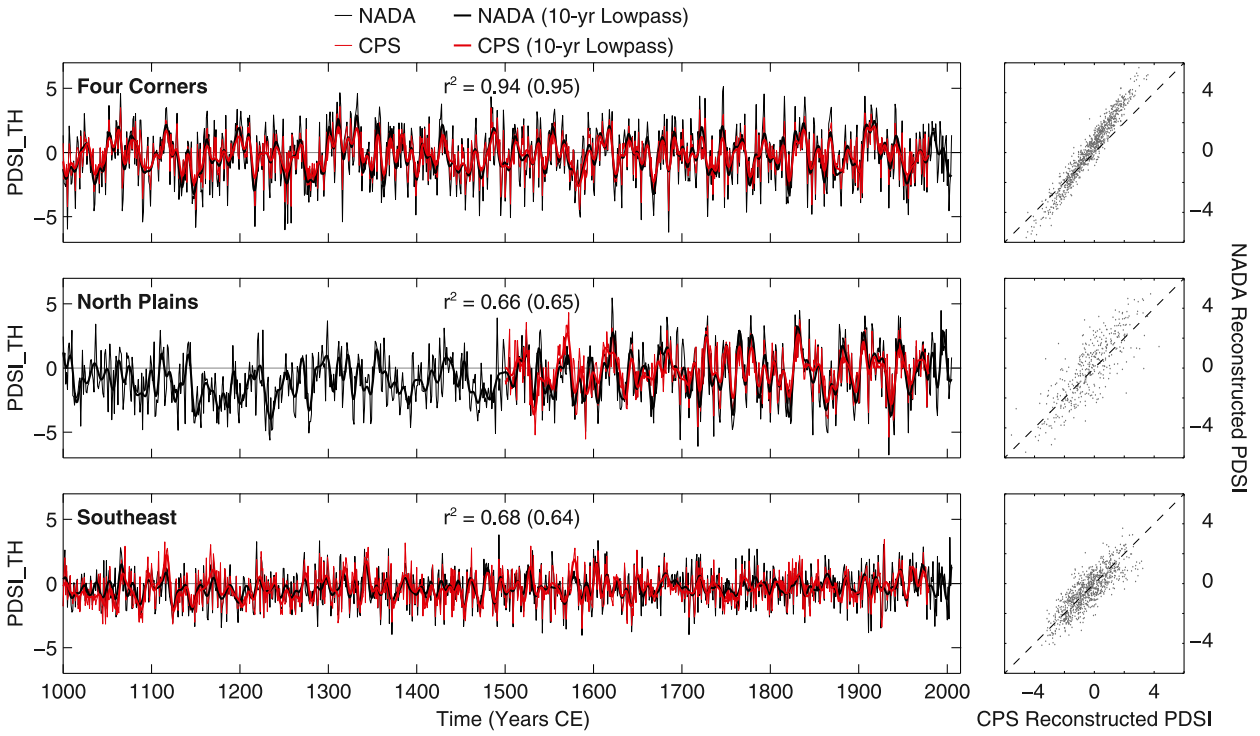

FIG. 4. (top to bottom) For each region, comparisons of reconstructed PDSI_TH from this study using CPS and the regional averages of the gridded NADA product (black lines) over the targeted CPS (red lines) reconstruction domains. Time series plot both annual values (thin lines) and the 10-yr low-pass-filtered time series using a 10-point Butterworth filter (thicker lines). (left) The shared variance (r^2) between the annual (filtered) CPS and NADA time series over their periods of overlap are shown within each plot. (right) Scatterplots for the annual CPS and NADA values are shown with 1:1 dashed lines for reference.

principal component regression to target PDSI values at each grid point on a 0.5° × 0.5° spatial grid, uses a static calibration–validation interval, and involves various predictor processing steps not included in the CPS methodology used herein. The reconstructions therefore have targeted different characteristics of the instrumental PDSI fields and used contrasting methods to derive the respective reconstructions. Despite these dissimilarities, the CPS PDSI_TH and NADA reconstructions compare extremely well (Fig. 4); the reported shared variances for PDSI_PM and PDSI_SC similarly support a strong agreement between the NADA and the two additional sets of regional reconstructions. These results not only support the idea that analyses of the new CPS reconstructions are relevant to interpretations of the NADA, but they more generally support the robustness and consistency of the employed methods and derived reconstructions.

4. Bridging paleoclimatic, observational, and model data

We now turn specifically to the challenge of bridging PDSI reconstructions with historical model simulations and future projections. We do not take up the related and important endeavor of comparing last-millennium PDSI reconstructions with forced-transient model simulations targeting the same interval. Such comparisons involve additional considerations regarding the interpretation of internal and forced variability and the realism with which the reconstructed exogenous forcings are estimated and employed in the last-millennium simulations. Detailed examples of such work can be found in Coats et al. (2013b, 2015a,b), who specifically perform paleoclimate data–model comparisons focused on hydroclimate in North America. Alternative to these specific comparisons of data and models over the last millennium, we endeavor to address how to bridge PDSI reconstructions with historical and future climate projections in the context of the noted differences between multiple PDSI formulations. We analyze four estimates of modeled soil-moisture balance: PDSI_TH, PDSI_PM, soil moisture in approximately the first 30 cm of the soil column (30cmSM), and full-column soil moisture (FCSM), the latter two of which are centered and scaled to match the PDSI_PM variance during the 1901–2012 interval for comparison.

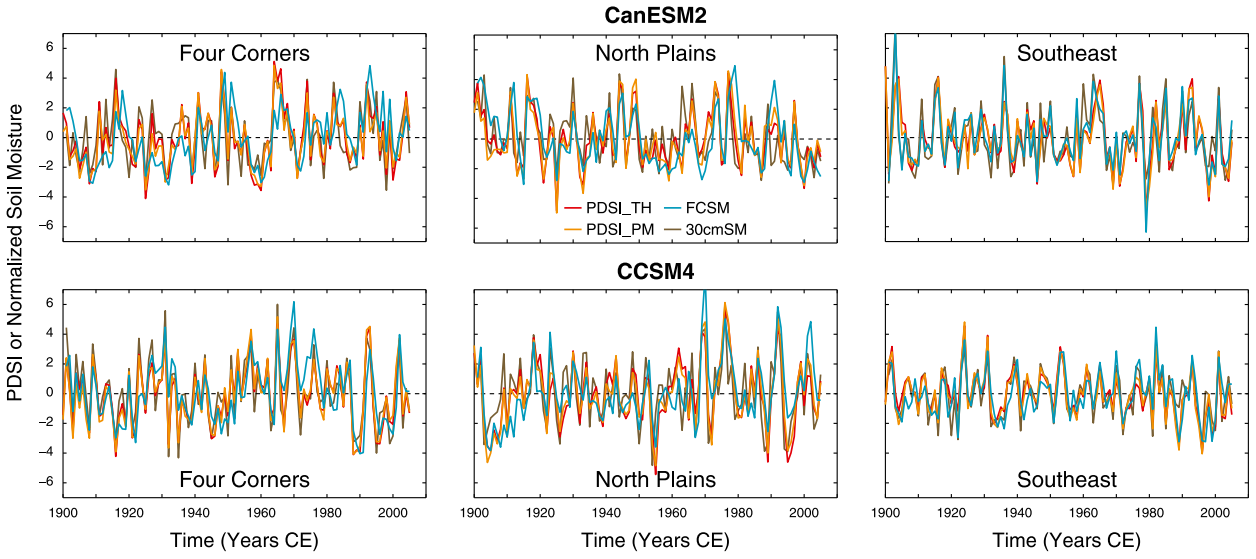

FIG. 5. (left to right) Regional model estimates of PDSI_TH, PDSI_PM, normalized 30cmSM, and normalized FCSM during the historical interval (1901–2005) in (top) the CanESM2 and (bottom) CCSM4 models. The first ensemble member of each model is plotted (interannual r^2 estimates across all five ensemble members are shown in Figs. 6 and 7). In all cases, PDSI or soil-moisture normalizations used the 1901–2005 intervals as a baseline, but time series are recentered from 1901 to 1979 to match the calibration/validation interval of the PDSI reconstructions.

a. Historical interval

The four characterizations of modeled soil-moisture balance in each region over the historical interval (1901–2005) are shown in Fig. 5 for the CanESM2 and CCSM4 models. Consistent with the observation-based estimates, these four measures of soil-moisture balance yield internally consistent results over the twentieth century, and both formulations of PDSI reproduce modeled soil moisture with high fidelity. Agreement between the two PDSI calculations is expected given that they are both calibrated over the 1901–2012 period and therefore any unrealistic temperature-driven differences in the PDSI_TH calculation are minimized. PDSI_TH and PDSI_PM in fact reveal a large amount of shared variance over the historical interval, matching or exceeding r^2 values of 0.9 in all regions over all ensemble members in both models (Figs. 6 and 7).

Comparisons between the two PDSI estimates and modeled soil moisture during the historical interval indicate weaker but still high levels of shared variance (Figs. 6 and 7). In most cases, PDSI_TH and PDSI_PM compare best with 30cmSM. The shared variance is strongest in CCSM4 for the 4C and SE regions, where r^2 values exceed 0.7 for comparisons between the two PDSI variables and 30cmSM. These numbers reduce slightly for the NP where some ensemble members yield values between 0.5 and 0.6. Comparisons weaken in CanESM2, in which values of shared variance between the two PDSI variables and 30cmSM range between about 0.5 and 0.75 across all regions. Comparisons between the two PDSI variables and FCSM generally indicate less shared variance than with 30cmSM because FCSM incorporates longer-scale variations and time lags that exceed the time scales that PDSI and near-surface soil moisture more strongly sample. The depth of sampling is not, however, the only factor, as indicated by the fact that shared variances between PDSI estimates and FCSM are larger than for 30cmSM in the SE region in the CanESM2 model and for some variable pairings and ensemble members in both models in the NP. Moreover, FCSM spans a much greater depth in the CCSM4 model than in CanESM2, but the r^2 values between the PDSI variables and FCSM are comparable or larger in CCSM4 than in CanESM2. Depth effects are therefore not the only determining factor in the comparison.

Comparisons between the two direct soil-moisture estimates (30cmSM and FCSM) are similar or worse within each model than the comparisons between the two PDSI estimates and either of the soil-moisture variables, with the one exception being in the CanESM2 model in the SE. This highlights the fact that it is not straightforward to determine which metric is most appropriate as a measure of modeled soil-moisture variability or even which soil-moisture target is the most appropriate analog to compare against PDSI. Even within models, the agreement between direct

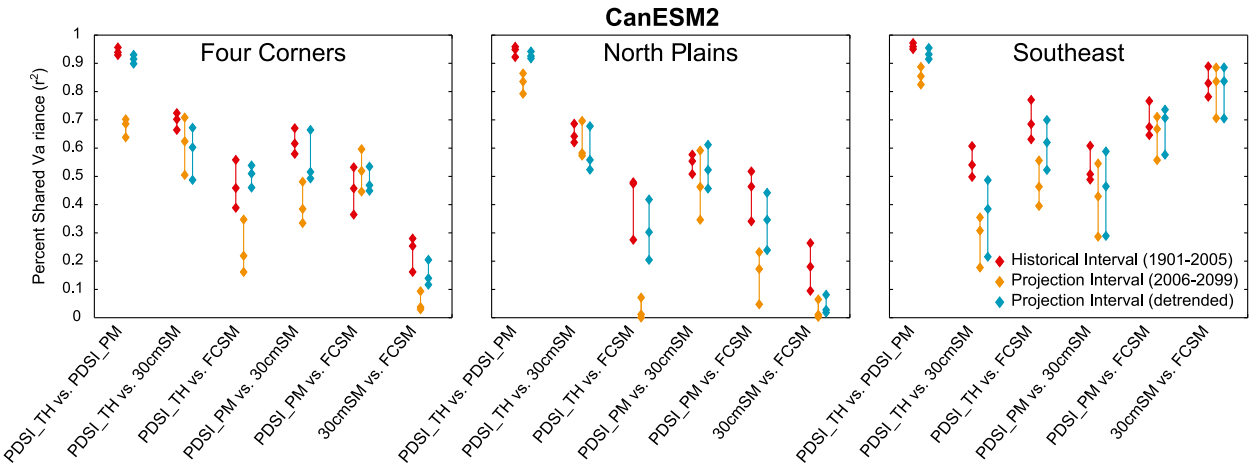

FIG. 6. (left to right) Regional shared variances (r^2) among the collection of soil-moisture metrics in the CanESM2 model simulations (x axis) during the historical (1901–2005) and projection (2006–99) intervals. Vertical diamond triplets correspond to the maximum, minimum, and median shared variance across the five members of the ensemble. Detrended results have been computed for the projection interval after removing a linear trend over the same period.

soil-moisture metrics is variable and depends strongly on the selected regions. The differences among these variables are likely associated with the specific tunings, parameterizations, and other modeling choices in the land surface components of the GCMs and further highlights the challenges of even direct comparisons between soil-moisture variables in the current generation of coupled GCMs.

We finally note that we avoid direct comparisons between the regional PDSI_TH reconstructions and the associated model variables beyond the collective plotting of these variables in Figs. 8–10. We have normalized and centered all of the moisture-balance variables over the common reference interval from 1901 to 2005, leaving any consistencies in means and variances of the reconstructed and modeled variables over the reference interval a product of construction. Furthermore, shared variance between the reconstructed and modeled moisture-balance metrics should not be expected for the CMIP5 historical model runs, in which the simulations are initialized from preindustrial control runs. These initializations do not constrain modes of internal variability that impact hydroclimate over North America, such as the El Niño–Southern Oscillation or the Pacific decadal oscillation, to be in phase with those that have dictated the actual climate states represented in the reconstructions. Unless all of the hydroclimate variability in the analyzed regions was forced over the twentieth century, which is certainly not the case, the interannual to decadal variance in the reconstructions and the

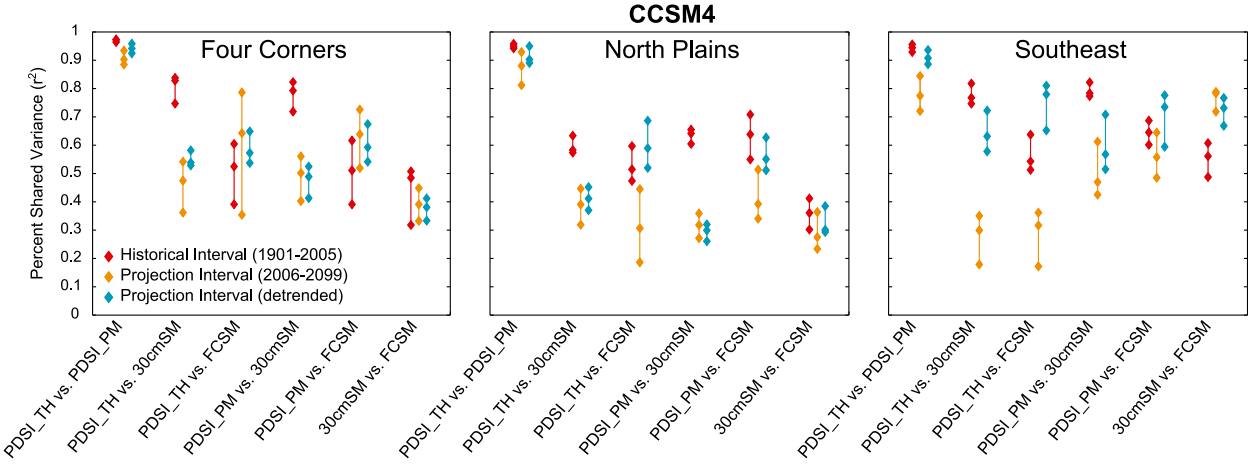

FIG. 7. As in Fig. 6, but for the CCSM4 model simulations. Maximum, minimum, and median results were similarly computed from an ensemble of five simulations.

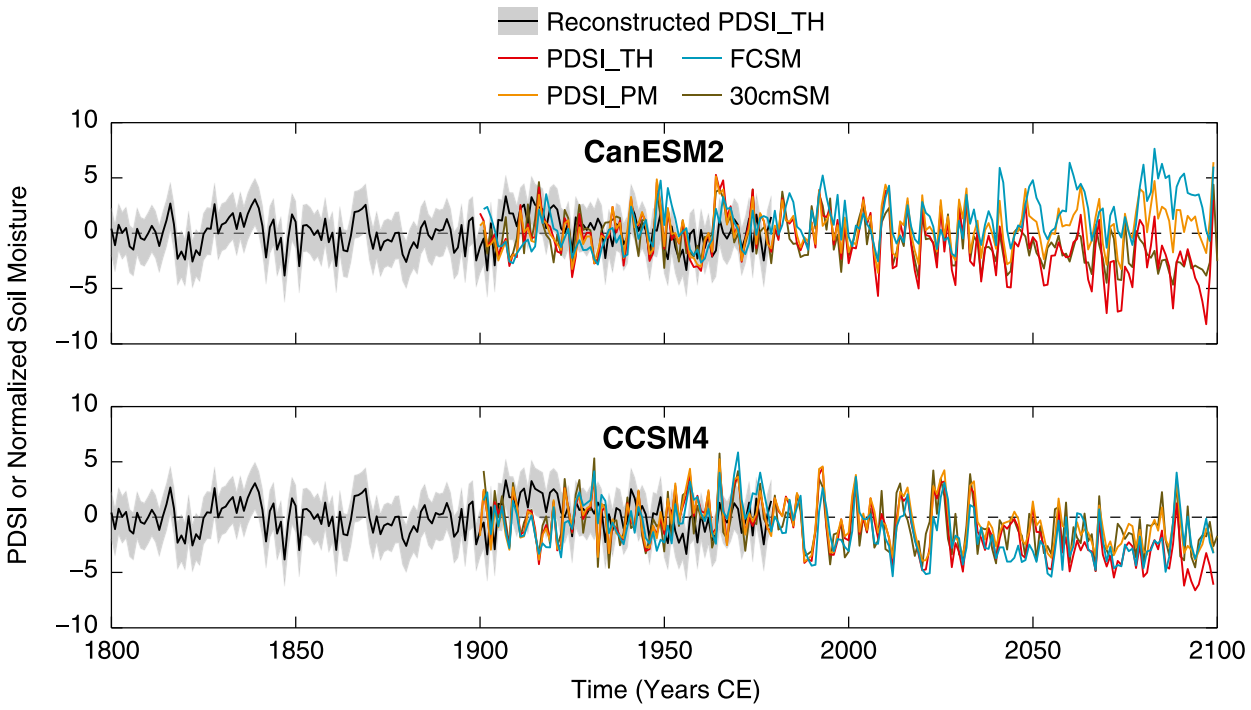

FIG. 8. A comparison of reconstructed PDSI_TH (black with gray shading) in the 4C region (see Fig. 1) with collocated model estimates of PDSI_TH (red), PDSI_PM (orange), normalized 30cmSM (black), and normalized FCSM (blue) during the historical and projection intervals. Results are shown for the first ensemble member of (top) the CanESM2 and (bottom) CCSM4 model simulations to allow direct visual comparisons between the reconstructed and model variance. All time series have been centered over the 1901–79 interval, the time period of common overlap. Reconstructions extend prior to 1800, but the 1800–2100 interval is chosen for ease of visual comparison.

models will not share consistently common features. These considerations therefore must inform attempts to compare reconstructions and model simulations directly over the interval of overlap and require more detailed and specific analyses that are beyond the scope of this investigation. The bridging approach that we have outlined herein nevertheless demonstrates the use of the twentieth century as a common interval for referencing reconstruction and model data in order to compare earlier paleoclimatic intervals and model projections of the twenty-first century.

b. *Projection interval*

Shared variance between the four soil-moisture variables during the projection interval (2005–99) are given in Figs. 6 and 7, whereas Figs. 8–10 provide comparisons between the four soil-moisture metrics for the first ensemble member from both models over the historical interval through projection intervals. Shared variances for detrended time series over the projection interval, in addition to the original time series, are also shown in Figs. 6 and 7. The shared variances among all of the detrended soil-moisture metrics in the projection interval in all regions are generally comparable to those of the historical interval (with some exceptions, particularly in the CCSM4 output). Larger differences are observed between the original time series, in which differences in twenty-first-century trends reduce the r^2 values among the different soil-moisture metrics (in cases where trends are comparable among variables, the shared variances in original and detrended time series are similar). Secular drying in PDSI_TH exceeds that of all other modeled soil-moisture metrics over the twenty-first century in all three regions, a clear demonstration of the tendency for PDSI_TH to overestimate drying during the projection interval. There are, however, regional differences in the relative comparisons among variables, as discussed below.

PDSI_TH includes secular trends in the 4C region that in some cases are larger in magnitude or opposite in sign relative to the three other metrics in both models (Fig. 8), despite the fact that some of the other metrics compare favorably to PSDI_TH over the full century or during specific intervals of time. PDSI_PM and 30cmSM compare well throughout the projection interval in the CCSM4 model, whereas 30cmSM projects enhanced drying, relative to PDSI_PM, in the CanESM2 model. FCSM again compares favorably to PDSI_PM and 30cmSM in the CCSM4 simulation, whereas it suggests a wetting trend in the CanESM2 simulation that is not reflected in any of the

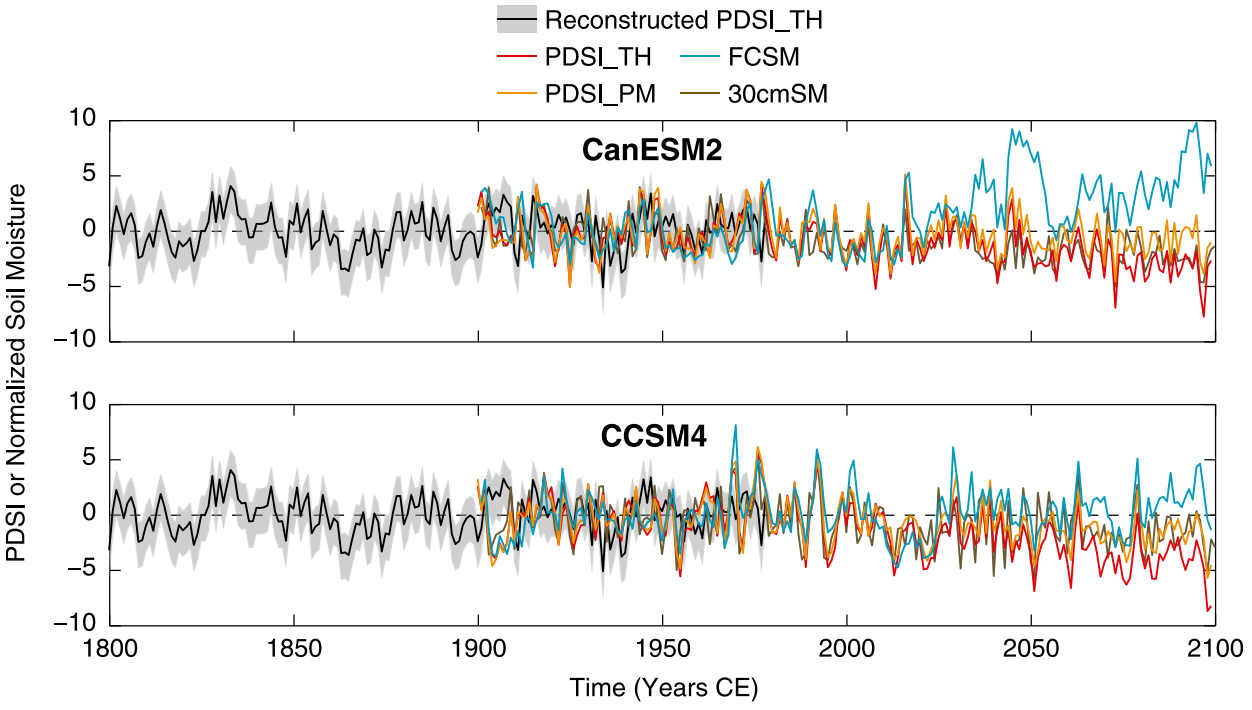

FIG. 9. As in Fig. 8, but for the North Plains region (see Fig. 1).

other estimates. This may again be representative of the deeper-column soil-moisture dynamics in the model (although no such trend is observed for the CCSM4 model with a much deeper soil column).

In the NP, the behavior of PDSI_TH is again consistent with previously discussed expectations (Fig. 9). The variable estimates drier twenty-first-century mean conditions than any of the other soil-moisture metrics and projects secular drying trends in both models that are either larger or not present in the other variables. Similar to the 4C region, PDSI_PM compares favorably to 30cmSM throughout the projection interval in the CCSM4 model, whereas 30cmSM projects slightly enhanced drying, relative to PDSI_PM, in the CanESM2 model. The FCSM estimate in CanESM2 contains a pronounced wetting trend and multidecadal variability in the NP that, in addition to contrasting with PDSI_PM, is also not present in the 30cmSM output. The FCSM is similarly wetter in the CCSM4 projection for the NP but not by nearly as much as the CanESM2 projection. In the SE, the behavior of PDSI_TH is the same as discussed for the other two regions in both models (Fig. 10). Comparisons between 30cmSM and FCSM variables are more consistent in both models in the SE, which both project wetter conditions or a wetting trend (CCSM4) in the SE that is not present in the PDSI_PM estimate.

Given the three regional analyses presented above, it is evident that all of the soil-moisture metrics compare well over the twentieth century in both models over all three regions. Regional differences are evident, however, in comparisons between the variables over the twenty-first-century, including some large disparities between the two modeled soil-moisture metrics themselves. This is particularly true of the secular trends estimated by the various metrics. While PDSI_TH projects varying degrees of exaggerated drying over all regions and models, PDSI_PM either matches 30cmSM well, falls between 30cmSM and FCSM, or projects relatively constant moisture-balance conditions when 30cmSM and FCSM indicate wetting trends. These regional observations are given broader context in Figs. 11 and 12, which plot the projected mean values of PDSI_PM, 30cmSM and FCSM for the last two decades (2080–99) of the twenty-first-century using the first ensemble member of the CanESM2 and CCSM4 simulation; the agreement between the three variables in the direction of either wetting or drying is also plotted. Inspections of Figs. 11 and 12 indicate that it is difficult to draw consistent characterizations of how the variables compare across models or regions. In the CanESM2 simulation, for instance, PDSI_PM and FCSM project wetting throughout much of western North America, whereas 30cmSM projects drying over those regions [note that this wetting projection is anomalous over the CMIP5 ensemble (Cook et al. 2014a) but the analyzed models have been selected based on the availability of their layered

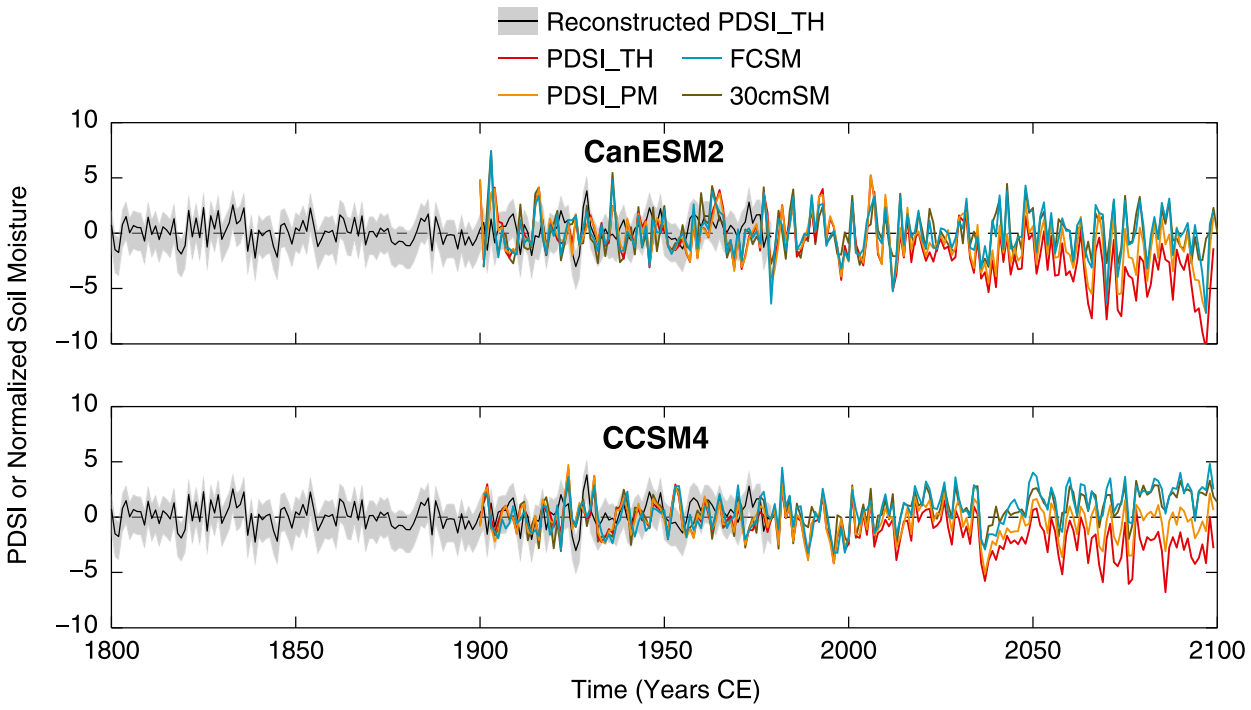

FIG. 10. As in Fig. 8, but for the Southeast region (see Fig. 1).

soil-moisture output and ensemble members that are continuous across the historical to projection intervals]. In contrast to CanESM2, the PDSI_PM projection from CCSM4 is almost uniformly toward drying, whereas a heterogeneous pattern of wetting emerges in 30cmSM and is further enhanced in FCSM. A full diagnosis of the physical underpinnings of these differences is beyond the scope of this paper, but we discuss various explanations for the regional differences in the following section and suggest that parsing the differences between these soil-moisture metrics is an important area of further research.

5. Discussion

This study has been motivated by two considerations: 1) the growing interest in performing paleoclimatic model–data comparisons and 2) the ongoing debate about how to properly represent hydroclimatic variability and change from a host of possible soil-moisture metrics. Our results demonstrate the robustness of PDSI as a metric of near-surface moisture variability in observations, reconstructions, and twentieth-century model simulations. Reconstructions derived for three diverse regions of the United States compare favorably, regardless of the reconstruction target (PDSI_TH, PDSI_PM, or PDSI_SC) or technique (CPS versus the NADA-based approach). The variances of the derived reconstructions are the principal, though modest, differences among them and are consistent with the expected character of PDSI formulations estimated from observations over the historical interval. Importantly, these results indicate that previous concerns about biases in tree-ring reconstructions due to the use of PDSI_TH as a calibration target (Sheffield et al. 2012) are unfounded over North America. Similarly, model estimates of soil moisture and PDSI are all consistent during the historical interval: PDSI_TH, PDSI_PM, and two normalized soil-moisture estimates all compare well in the CanESM2 and CCSM4 historical model simulations. Together, these results support the continued use of PDSI as a valuable tool for empirical and model-based investigations of drought and hydroclimate.

Principal differences emerge only in model-derived estimates of PDSI and soil moisture during the twenty-first-century projection interval. Specifically, the secular behavior of the various metrics diverges in the projections, whereas the interannual variability remains relatively consistent across all variables throughout the twenty-first-century projection interval. In the case of PDSI_TH, enhanced secular drying is now a well-understood consequence of unrealistically scaling PET as a function of temperature when values exceed the range defined by the normalization interval. It is less clear, however, how and why modeled near-surface and full-column soil moisture would compare well with the more physically based PDSI_PM during the twentieth

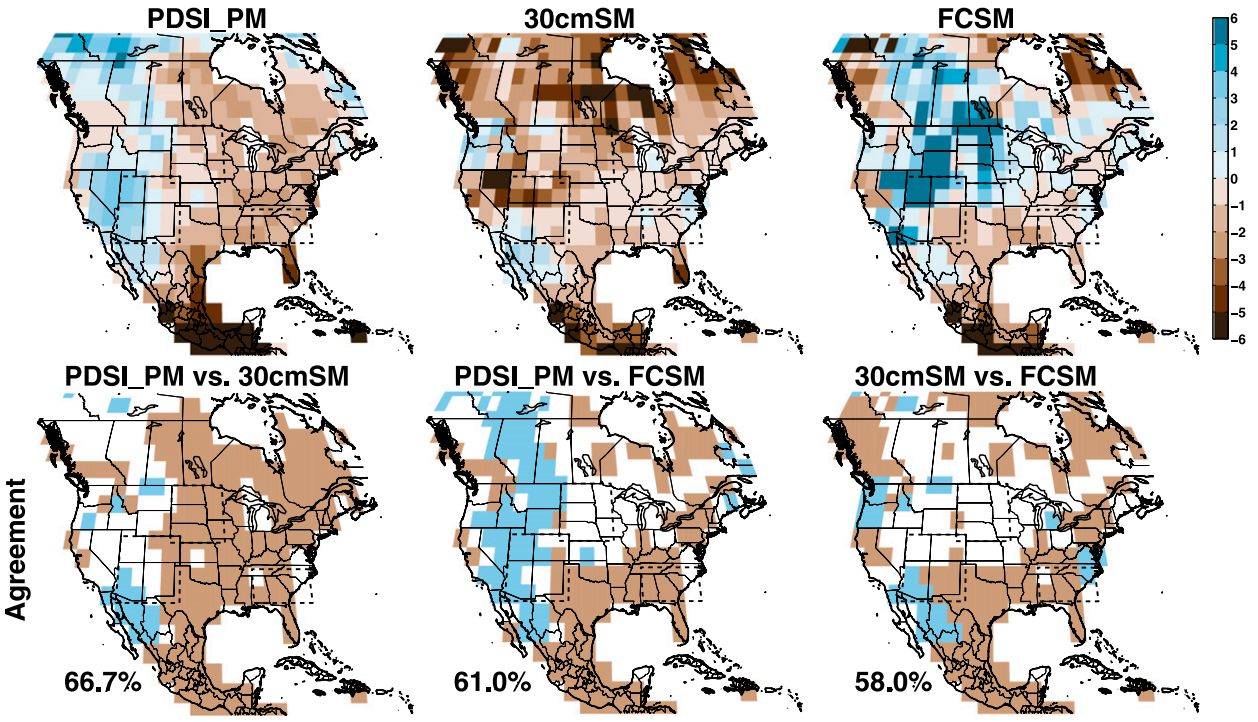

FIG. 11. (top) (left to right) Mean PDSI_PM, normalized 30cmSM, and normalized FCSM for the last two decades (2080–99) of the twenty-first-century projection interval (ensemble member 1) from the CanESM2 model. (bottom) Agreement between the sign of the wetting (blue) or drying (brown) as projected by the three variables in the 2080–99 interval. The total percentage of grid cells that agree on the wetting or drying trends are provided in the bottom left of the comparison maps. The boxes defined by dashed lines indicate the regions extracted from the 0.5° × 0.5°NADA, observation, and model grids as representative of the 4C, NP, and SE regions.

century and in the twenty-first-century on interannual time scales, whereas their twenty-first-century secular trends would diverge.

Several possibilities may explain some of the different secular behavior in the model-based metrics. The differences may arise, in part, from the difficulty in identifying the most appropriate modeled soil-moisture variable for comparison with the model-derived PDSI. We have demonstrated various situations in which PDSI most closely reflects near-surface soil moisture, others where PDSI was a better indicator of full-column soil moisture, or places where PDSI resolved both equally well. Importantly, we also illustrated ample instances where even the two soil-moisture metrics diverge, sometimes to quite extreme effect (e.g., the northern plains in CanESM2). Additional differences between PDSI and modeled soil moisture also are likely to arise through the typically more sophisticated treatment of processes (e.g., vegetation, soil physics) in the land surface components of GCMs. These parameterizations and tunings vary across models, meaning that any concerns in comparing PDSI and soil moisture within models must also be extended to comparing soil-moisture trends across models.

One major issue that is often discussed in the context of future drought projections is the CO_2 fertilization effect. The direct physiological effect of elevated atmospheric CO_2 concentrations (eCO2) is to reduce water loss during photosynthesis by lowering stomatal conductance. This effect is typically incorporated into the physics of GCM land surface and vegetation models but not in standard PDSI calculations, including those employed herein. Two expectations are therefore associated with eCO2: 1) drought stress on model vegetation will be reduced, translating into lower rates of evapotranspiration and increases in soil moisture and runoff, and 2) standard PDSI projections, which do not include the impact of eCO2, will not reflect any associated wetting tied to the effect. This may explain areas in the presented projections where PDSI indicates drying while model soil moisture indicates wetting or little change.

While the physiological effect of eCO2 is well understood at the leaf level (Ainsworth and Rogers 2007), there are large uncertainties associated with the scaling of this effect on hydrology at larger scales that may lead to an overestimate of the eCO2 effect in model projections. For example, many experiments in which plants are exposed to elevated levels of CO_2 show only modest

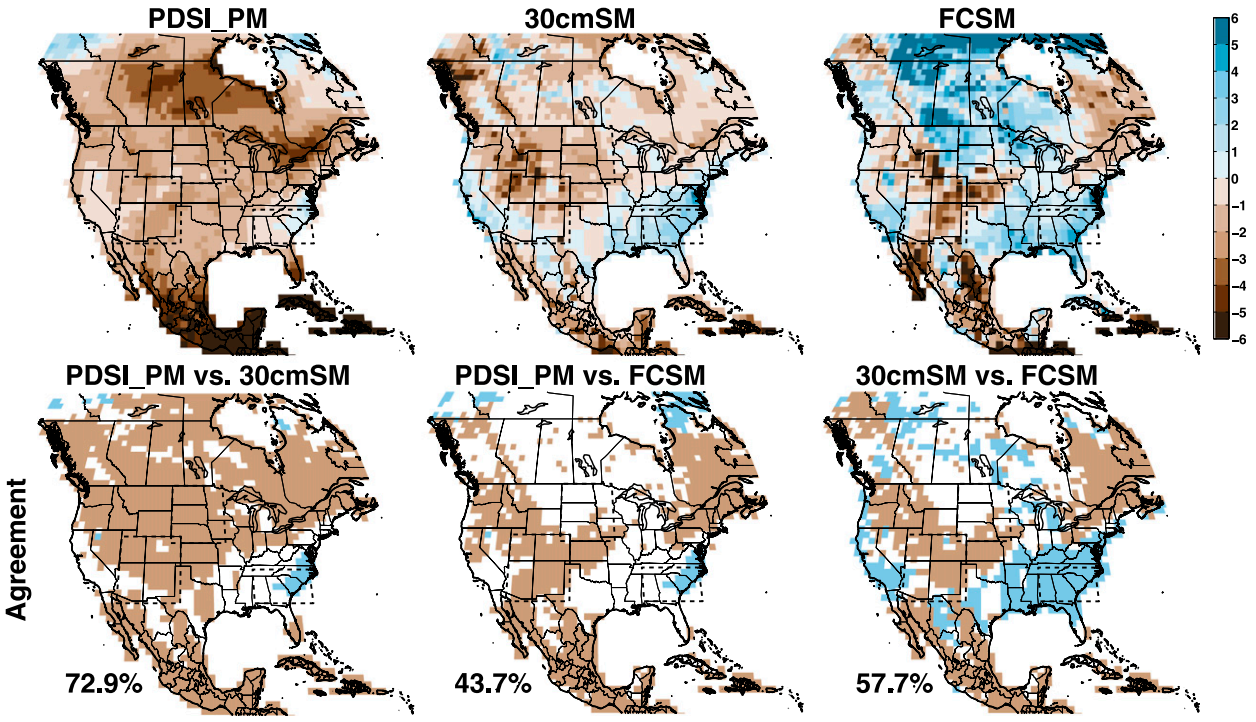

FIG. 12. As in Fig. 11, but for the CCSM4 model.

and often insignificant changes in transpiration and soil moisture (e.g., Domec et al. 2009; Hussain et al. 2013; Inauen et al. 2013; Naudts et al. 2013; Stocker et al. 1997). Additionally, few eCO_2 studies simultaneously incorporate ambient warming into their experiments, an accompanying element of climate change that is expected to increase ET and potentially counteract any moisture gains from eCO2. Despite this ambiguity in the empirical evidence, however, most land surface and vegetation models, including those models typically incorporated into GCMs, substantially reduce transpiration when exposed to eCO_2 and, in some cases, dramatically overestimate the transpiration response relative to observations (De Kauwe et al. 2013). This suggests that the modeled response to eCO_2 may be oversimplified and the eCO_2 effect on hydrology overestimated. In fact, this host of uncertainties in both the modeled and empirical eCO2 responses led Working Group II of the IPCC to conclude in AR5 (Gerten et al. 2014) that the net effect of eCO_2 on runoff and transpiration is still ''poorly constrained'' and that ''precipitation and temperature effects are likely to remain the prime influence on global runoff'' (p. 158). Notably, eCO_2 does not appear to have a strong influence on soil moisture in other CMIP5 simulations before the rapid increases in the RCP scenarios, as evidenced by the tight coupling between PDSI and modeled soil moisture in the historical simulations presented herein.

Clearly, there remain significant uncertainties in the interpretation of modeled soil-moisture and PDSI responses in twenty-first-century GCM projections. Further characterizing and resolving the impact of these modeling choices will be important as a means of resolving differences between PDSI_PM and the soil-moisture estimates discussed herein but even more importantly to constrain the range of realistic soil-moisture conditions into the future. In the context of paleoclimate data–model comparisons, however, it appears that projections using PDSI_PM or normalized near-surface soil moisture are the most appropriate variables for characterizing future projections and for comparing projections to modern and paleoclimatic observations (e.g., Dai 2013; Cook et al. 2014a).

6. Conclusions

Given our principal observations, a framework emerges for comparisons between PDSI reconstructions and model simulations. It is first of all evident that tree-ring-based PDSI reconstructions are robust to the selection of PDSI target variables and the current generation of PDSI_TH drought atlases can be confidently used as estimates of hydroclimatic variability over the last millennium. With regard to comparing these drought atlases to last-millennium simulations, our results suggests that model-derived PDSI or soil moisture

should all estimate similar hydroclimate histories given that modeled temperature variations over the last millennium are typically within the range of modeled twentieth-century climate (e.g., Fernández-Donado et al. 2013; Masson-Delmotte et al. 2014). Indeed, the similarity between multiple hydroclimate metrics over the last millennium has been demonstrated specifically in the U.S. Southwest using multiple simulations in the context of assessing the occurrence of multidecadal drought events in the region (Coats et al. 2013b, 2015b). Similarly, we have shown that comparisons among observational or modeled soil-moisture metrics during the twentieth century are largely insensitive to the choice of metric. While it is most consistent to compare the same metric across reconstructions, observations, and models, such consistency does not appear essential and is sometimes not possible over the last millennium and through the twentieth century. In lieu of homogenized products, it therefore is suggested that comparisons simply reference and scale means and variances of all variables over a common period of overlap.

Overall, the current collection of tools can provide robust characterizations of hydroclimate variability and change during the last millennium, which in turn can be meaningfully compared to observations and model simulations of twentieth-century hydroclimate and model projections of twenty-first-century change. Within these comparisons the most important interval of consideration is the twenty-first century, in which model projections should be characterized using either PDSI_PM (or similarly a Penman–Monteith version of PDSI_SC or standardized precipitation–evaporation index) or a near-surface normalized soil-moisture output. In the case of the latter, however, these results must be carefully interpreted in terms of their applicability across a collection of model simulations and in terms of the processes that they include. The impact of CO_2 fertilization on twenty-first-century hydroclimate projections is a particularly important process to evaluate across models, given the poorly constrained nature of the process on large-scale vegetation response and because it appears to have potentially important implications for projected hydroclimatic trends. Although these details complicate comparisons and leave open important research questions, future model–data comparisons that span the last millennium through the twenty-first-century projection interval will be vital for assessing and characterizing future risks associated with hydroclimate variability and change. As we have demonstrated, these comparisons are possible and appropriate given the current collection of data and tools, which should all be used for comparisons within the framework outlined herein.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and leads development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Haibo Liu and Naomi Henderson provided computational and data processing support at LDEO. RS and JES were supported in part by the NOAA Award Global Decadal Hydroclimate Variability and Change (NA10OAR4310137). RS was also supported by DOE Award DE-SC0005107. Further support came from NSF Award AGS-1243204. BIC was supported by NASA.

REFERENCES

Ainsworth, E. A., and A. Rogers, 2007: The response of photosynthesis and stomatal conductance to rising [CO_2]: Mechanisms and environmental interactions. *Plant Cell Environ.,* **30,** 258–270, doi:10.1111/j.1365-3040.2007.01641.x.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature,* **419,** 224–232, doi:10.1038/nature01092.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, 1998: Crop evapotranspiration-guidelines for computing crop water requirements. Food and Agriculture Organization of the United Nations Irrigation and Drainage Paper 56, 326 pp.

Anchukaitis, K. J., M. N. Evans, A. Kaplan, E. A. Vaganov, H. D. Grissino-Mayer, M. K. Hughes, and M. A. Cane, 2006: Forward modeling of regional-scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophys. Res. Lett.,* **33,** L04705, doi:10.1029/2005GL025050.

——, B. M. Buckley, E. R. Cook, B. I. Cook, R. D. D'Arrigo, and C. M. Ammann, 2010: Influence of volcanic eruptions on the climate of the Asian monsoon region. *Geophys. Res. Lett.,* **37,** L22703, doi:10.1029/2010GL044843.

Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser, 2013a: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate,* **26,** 5863–5878, doi:10.1175/JCLI-D-11-00732.1.

——, C. Deser, M. Newman, and J. Emile-Geay, 2013b: Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium. *Geophys. Res. Lett.,* **40,** 3450–3456, doi:10.1002/grl.50647.

——, J. E. Cole, J. T. Overpeck, G. T. Pederson, and D. M. Meko, 2014: Assessing the risk of persistent drought using climate model simulations and paleoclimate data. *J. Climate,* **27,** 7529–7549, doi:10.1175/JCLI-D-12-00282.1.

Betts, A. K., J. H. Ball, A. C. M. Beljaars, M. J. Miller, and P. A. Viterbo, 1996: The land surface-atmosphere interaction: A review based on observational and global modeling perspectives. *J. Geophys. Res.,* **101,** 7209–7225, doi:10.1029/95JD02135.

Burke, E. J., 2011: Understanding the sensitivity of different drought metrics to the drivers of drought under increased atmospheric CO_2. *J. Hydrometeor.,* **12,** 1378–1394, doi:10.1175/2011JHM1386.1.

——, and S. J. Brown, 2008: Evaluating uncertainties in the projection of future drought. *J. Hydrometeor.,* **9,** 292–299, doi:10.1175/2007JHM929.1.

——, ——, 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. Hydrometeor.,* **7,** 1113–1125, doi:10.1175/JHM544.1.

Coats, S., J. E. Smerdon, B. I. Cook, and R. Seager, 2013a: Stationarity of the tropical Pacific teleconnection to North America in CMIP5/PMIP3 model simulations. *Geophys. Res. Lett.,* **40,** 4927–4932, doi:10.1002/grl.50938.

——, ——, R. Seager, B. I. Cook, and J. F. González-Rouco, 2013b: Megadroughts in southwestern North America in ECHO-G millennial simulations and their comparison to proxy drought reconstructions. *J. Climate,* **26,** 7635–7649, doi:10.1175/JCLI-D-12-00603.1.

——, B. I. Cook, J. E. Smerdon, and R. Seager, 2015a: North American pancontinental droughts in model simulations of the last millennium. *J. Climate,* doi:10.1175/JCLI-D-14-00634.1, in press.

——, J. E. Smerdon, B. I. Cook, and R. Seager, 2015b: Are simulated megadroughts in the North American southwest forced? *J. Climate,* **28,** 124–142, doi:10.1175/JCLI-D-14-00071.1.

Cook, B. I., J. E. Smerdon, R. Seager, and S. Coats, 2014a: Global warming and 21st century drying. *Climate Dyn.,***43,** 2607–2627, doi:10.1007/s00382-014-2075-y.

——, ——, ——, and E. R. Cook, 2014b: Pan-continental droughts in North America over the last millennium. *J. Climate,* **27,** 383–397, doi:10.1175/JCLI-D-13-00100.1.

Cook, E. R., R. Seager, M. A. Cane, and D. W. Stahle, 2007: North American drought: Reconstructions, causes and consequences. *Earth-Sci. Rev.,* **81,** 93–134, doi:10.1016/j.earscirev.2006.12.002.

——, ——, R. R. Heim Jr., R. S. Vose, C. Herweijer, and C. Woodhouse, 2010: Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *J. Quat. Sci.,* **25,** 48–61, doi:10.1002/jqs.1303.

Dai, A., 2011a: Characteristics and trends in various forms of the Palmer drought severity index during 1900-2008. *J. Geophys. Res.,* **116,** D12115, doi:10.1029/2010JD015541.

——, 2011b: Drought under global warming: a review. *Wiley Interdiscip. Rev.: Climate Change,* **2,** 45–65, doi:10.1002/wcc.81.

——, 2013: Increasing drought under global warming in observations and models. *Nat. Climate Change,* **3,** 52–58, doi:10.1038/nclimate1633.

De Kauwe, M. G., and Coauthors, 2013: Forest water use and water use efficiency at elevated CO_2: A model-data intercomparison at two contrasting temperate forest FACE sites. *Global Change Biol.,* **19,** 1759–1779, doi:10.1111/gcb.12164.

Ding, Y., M. J. Hayes, and M. Widhalm, 2011: Measuring economic impacts of drought: A review and discussion. *Disaster Prev. Manage.,* **20,** 434–446, doi:10.1108/09653561111161752.

Domec, J. C., S. Palmroth, E. Ward, C. A. Maier, M. Therezien, and R. Oren, 2009: Acclimation of leaf hydraulic conductance and stomatal conductance of Pinus taeda (loblolly pine) to long-term growth in elevated CO_2 (free-air CO_2 enrichment) and N-fertilization. *Plant Cell Environ.,* **32,** 1500–1512, doi:10.1111/j.1365-3040.2009.02014.x.

Evans, M. N., B. K. Reichert, A. Kaplan, K. J. Anchukaitis, E. A. Vaganov, M. K. Hughes, and M. A. Cane, 2006: A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J. Geophys. Res.,* **111,** G03008, doi:10.1029/2006JG000166.

Fernández-Donado, L., and Coauthors, 2013: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Climate Past,* **9,** 393–421, doi:10.5194/cp-9-393-2013.

Gerten, D., R. Betts, and P. Döll, 2014: Cross-chapter box on the active role of vegetation in altering water flows under climate change. *Climate Change 2014: Impacts, Adaptation, and Vulnerability,* C. B. Field et al., Eds., Cambridge University Press, 157–161.

Goosse, H., E. Crespin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J. Geophys. Res.,* **115,** D09108, doi:10.1029/2009JD012737.

——, ——, S. Dubinkina, M.-F. Loutre, M. E. Mann, H. Renssen, Y. Sallaz-Damaz, and D. T. Shindell, 2012: The role of forcing and internal dynamics in explaining the ''Medieval Climate Anomaly.'' *Climate Dyn.,* **39,** 2847–2866, doi:10.1007/s00382-012-1297-0.

Guttman, N. B., 1998: Comparing the Palmer drought index and the standardized precipitation index. *J. Amer. Water Res. Assoc.,* **34,** 113–121, doi:10.1111/j.1752-1688.1998.tb05964.x.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.,* **34,** 623–642, doi:10.1002/joc.3711.

Headey, D., 2011: Rethinking the global food crisis: The role of trade shocks. *Food Policy,* **36,** 136–146, doi:10.1016/j.foodpol.2010.10.003.

Hernandez, E., and V. Uddameri, 2014: Standardized precipitation evaporation index (SPEI)-based drought assessment in semiarid south Texas. *Env. Earth Sci.,* **71,** 2491–2501, doi:10.1007/s12665-013-2897-7.

Hind, A., and A. Moberg, 2013: Past millennial solar forcing magnitude. A statistical hemispheric-scale climate model versus proxy data comparison. *Climate Dyn.,* **41,** 2527–2537, doi:10.1007/s00382-012-1526-6.

——, ——, and R. Sundberg, 2012: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium—Part 2: A pseudo-proxy study addressing the amplitude of solar forcing. *Climate Past,* **8,** 1355–1365, doi:10.5194/cp-8-1355-2012.

Hoerling, M., J. K. Eischeid, X. W. Quan, H. F. Diaz, R. S. Webb, R. M. Dole, and D. R. Easterling, 2012: Is a transition to semipermanent drought conditions imminent in the U.S. Great Plains? *J. Climate,* **25,** 8380–8386, doi:10.1175/JCLI-D-12-00449.1.

——, and Coauthors, 2013: Anatomy of an extreme event. *J. Climate,* **26,** 2811–2832, doi:10.1175/JCLI-D-12-00270.1.

——, J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains drought. *Bull. Amer. Meteor. Soc.,* **95,** 269–282, doi:10.1175/BAMS-D-13-00055.1.

Hussain, M. Z., A. VanLoocke, M. H. Siebers, U. M. Ruiz-Vera, R. J. Cody Markelz, A. D. B. Leakey, D. R. Ort, and C. J. Bernacchi, 2013: Future carbon dioxide concentration decreases canopy evapotranspiration and soil water depletion by field-grown maize. *Global Change Biol.,* **19,** 1572–1584, doi:10.1111/gcb.12155.

Inauen, N., C. Körner, and E. Hiltbrunner, 2013: Hydrological consequences of declining land use and elevated CO_2 in alpine grassland. *J. Ecol.,* **101,** 86–96, doi:10.1111/1365-2745.12029.

Jones, P. D., and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *Holocene,* **19,** 3–49, doi:10.1177/0959683608098952.

Lehner, F., C. C. Raible, and T. F. Stocker, 2012: Testing the robustness of a precipitation proxy-based North Atlantic Oscillation reconstruction. *Quat. Sci. Rev.,* **45,** 85–94, doi:10.1016/j.quascirev.2012.04.025.

Li, X., S. R. Waddington, J. Dixon, A. K. Joshi, and M. C. Vicente, 2011: The relative importance of drought and other water-related constraints for major food crops in South Asian farming systems. *Food Security,* **3,** 19–33, doi:10.1007/s12571-011-0111-x.

Lobell, D. B., W. Schlenker, and J. Costa-Roberts, 2011: Climate Trends and global crop production since 1980. *Science,* **333,** 616–620, doi:10.1126/science.1204531.

Mann, M. E., and Coauthors, 2009: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science,* **326,** 1256–1260, doi:10.1126/science.1177303.

Masson-Delmotte, V., and Coauthors, 2014: Information from paleoclimate archives. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 383–464.

Naudts, K., J. Berge, I. Janssens, I. Nijs, and R. Ceulemans, 2013: Combined effects of warming and elevated CO_2 on the impact of drought in grassland species. *Plant Soil,* **369** (1–2), 497–507, doi:10.1007/s11104-013-1595-2.

Neukom, R., J. Luterbacher, R. Villalba, M. Küttel, D. Frank, P. D. Jones, M. Grosjean, J. Esper, L. Lopez, and H. Wanner, 2010: Multi-centennial summer and winter precipitation variability in southern South America. *Geophys. Res. Lett.,* **37,** L14708, doi:10.1029/2010GL043680.

——, and Coauthors, 2011: Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Climate Dyn.,* **37** (1–2), 35–51, doi:10.1007/s00382-010-0793-3.

PAGES 2k Consortium, 2013: Continental-scale temperature variability over the last two millennia. *Nat. Geosci.,* **6,** 339–346, doi:10.1038/ngeo1797.

Palmer, W. C., 1965: Meteorological drought. U.S. Weather Bureau Research Paper 45, 58 pp.

Peng, S., and Coauthors, 2011: Recent change of vegetation growth trend in China, *Environ. Res. Lett.,* **6,** 044027, doi:10.1088/1748-9326/6/4/044027.

Penman, H. L., 1948: Natural evaporation from open water, bare soil and grass. *Proc. Roy. Soc. London,* **193,** 120–145, doi:10.1098/rspa.1948.0037.

Phipps, S. J., and Coauthors, 2013: Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years. *J. Climate,* **26,** 6915–6936, doi:10.1175/JCLI-D-12-00108.1.

Scheff, J., and D. M. W. Frierson, 2014: Scaling potential evapotranspiration with greenhouse warming. *J. Climate,* **27,** 1539–1558, doi:10.1175/JCLI-D-13-00233.1.

Schmidt, G. A., 2010: Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change. *J. Quat. Sci.,* **25,** 79–87, doi:10.1002/jqs.1314.

——, and Coauthors, 2014: Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate Past,* **10,** 221–250, doi:10.5194/cp-10-221-2014.

Seager, R., R. Burgman, Y. Kushnir, A. Clement, E. Cook, N. Naik, and J. Miller, 2008: Tropical Pacific forcing of North American medieval megadroughts: Testing the concept with an atmosphere model forced by coral-reconstructed SSTs. *J. Climate,* **21,** 6175–6190, doi:10.1175/2008JCLI2170.1.

——, M. Ting, C. Li, N. Naik, B. Cook, J. Nakamura, and H. Liu, 2013: Projections of declining surface-water availability for the southwestern United States. *Nat. Climate Change,* **3,** 482–486, doi:10.1038/nclimate1787.

Sellers, P. J., and Coauthors, 1997: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science,* **275,** 502–509, doi:10.1126/science.275.5299.502.

Seneviratne, S. I., 2012: Climate science: Historical drought trends revisited. *Nature,* **491,** 338–339, doi:10.1038/491338a.

Sheffield, J., E. F. Wood, and M. L. Roderick, 2012: Little change in global drought over the past 60 years. *Nature,* **491,** 435–438, doi:10.1038/nature11575.

Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *Wiley Interdiscip. Rev.: Climate Change,* **3,** 63–77, doi:10.1002/wcc.149.

Stocker, R., P. W. Leadley, and C. Körner, 1997: Carbon and water fluxes in a calcareous grassland under elevated CO_2. *Funct. Ecol.,* **11,** 222–230, doi:10.1046/j.1365-2435.1997.00071.x.

Sundberg, R., A. Moberg, and A. Hind, 2012: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium—Part 1: Theory. *Climate Past,* **8,** 1339–1353, doi:10.5194/cp-8-1339-2012.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thornthwaite, C., 1948: An approach toward a rational classification of climate. *Geogr. Rev.,* **38,** 55–94, doi:10.2307/210739.

Tierney, J. E., J. E. Smerdon, K. J. Anchukaitis, and R. Seager, 2013: Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. *Nature,* **493,** 389–392, doi:10.1038/nature11785.

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam, 2012: Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.,* **35,** 1–22, doi:10.1016/j.quascirev.2012.01.012.

Trenberth, K. E., A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield, 2014: Global warming and changes in drought. *Nat. Climate Change,* **4,** 17–22, doi:10.1038/nclimate2067.

Trouet, V., J. Esper, N. E. Graham, A. Baker, D. C. Frank, and J. D. Scourse, 2009: Persistent positive North Atlantic Oscillation mode dominated the Medieval Climate Anomaly. *Science,* **324,** 78–80, doi:10.1126/science.1166349.

van der Schrier, G., P. D. Jones, and K. R. Briffa, 2011: The sensitivity of the PDSI to the Thornthwaite and Penman-Monteith parameterizations for potential evapotranspiration. *J. Geophys. Res.,* **116,** D03106, doi:10.1029/2010JD015001.

——, J. Barichivich, K. R. Briffa, and P. D. Jones, 2013: A scPDSI-based global data set of dry and wet spells for 1901–2009. *J. Geophys. Res. Atmos.,* **118,** 4025–4048, doi:10.1002/jgrd.50355.

Vicente-Serrano, S. M., S. Beguera, and J. I. López-Moreno, 2010a: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate,* **23,** 1696–1718, doi:10.1175/2009JCLI2909.1.

——, ——, ——, M. Angulo, and A. El Kenawy, 2010b: A new global 0.5° gridded dataset (1901–2006) of a multiscalar drought index: Comparison with current drought index datasets based on the Palmer drought severity index. *J. Hydrometeor.,* **11,** 1033–1043, doi:10.1175/2010JHM1224.1.

Wahl, E. R., and J. E. Smerdon, 2012: Comparative performance of paleoclimatic field and index reconstructions derived from climate proxies and noise-only predictors. *Geophys. Res. Lett.,* **39,** L06703, doi:10.1029/2012GL051086.

Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating Palmer drought severity index. *J. Climate,* **17,** 2335–2351, doi:10.1175/1520-0442(2004)017<2335:ASPDSI>2.0.CO;2.

Willmott, C. J., C. M. Rowe, and Y. Mintz, 1985: Climatology of the terrestrial seasonal water cycle. *J. Climatol.,* **5,** 589–606.

Xu, C. Y., and V. P. Singh, 2002: Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. *Water Res. Manage.,* **16,** 197–219, doi:10.1023/A:1020282515975.