JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

AMERICAN WATER RESOURCES ASSOCIATION

ROBUSTNESS OF METEOROLOGICAL DROUGHTS IN DYNAMICALLY DOWNSCALED CLIMATE SIMULATIONS[1]

Poulomi Ganguli and Auroop R. Ganguly[2]

ABSTRACT: We examine the robustness of a suite of regional climate models (RCMs) in simulating meteorological droughts and associated metrics in present-day climate (1971-2003) over the conterminous United States (U.S.). The RCMs that are part of North American Regional Climate Change Assessment Program (NARCCAP) simulations are compared with multiple observations over the climatologically homogeneous regions of the U.S. The seasonal precipitation, climatology, drought attributes, and trends have been assessed. The reanalysis-based multi-model median RCM reasonably simulates observed statistical attributes of drought and the regional detail due to topographic forcing. However, models fail to simulate significant drying trend over the Southwest and West. Further, reanalysis-based NARCCAP runs underestimate the observed drought frequency overall, with the exception of the Southwest; whereas they underestimate persistence in the drought-affected areas over the Southwest and West-North Central regions. However, global climate model-driven NARCCAP ensembles tend to overestimate regional drought frequencies. Models exhibit considerable uncertainties while reproducing meteorological drought statistics, as evidenced by a general lack of agreement in the Hurst exponent, which in turn controls drought persistence. Water resources managers need to be aware of the limitations of current climate models, while regional climate modelers may want to fine-tune their parameters to address impact-relevant metrics.

(KEY TERMS: drought; sustainability; precipitation; stochastic models; geospatial analysis; time series analysis.)

Ganguli, Poulomi and Auroop R. Ganguly, 2015. Robustness of Meteorological Droughts in Dynamically Downscaled Climate Simulations. *Journal of the American Water Resources Association* (JAWRA) 1-30. DOI: 10.1111/1752-1688.12374

INTRODUCTION

The impacts of climate change on drought attributes continue to be debated in the scientific community, even as multiple regions, globally and in the United States (U.S.), experience severe droughts. Drought is a recurrent problem in many parts of the Conterminous United States (CONUS). Heat waves and droughts alone caused damage of around \$210.1 billion dollars during 1980-2011 in the U.S. and ranked second highest after tropical cyclones in terms of financial losses (Smith and Katz, 2013). Droughts are difficult to characterize because of complex interdependence among various drought attributes, such as severity (magnitude), duration, spatial coverage, frequency, and persistence. In a design context, assumption of complete independence or

[1]Paper No. JAWRA-15-0020-P of the *Journal of the American Water Resources Association* (JAWRA). Received February 16, 2015; accepted September 30, 2015. **Discussions are open until six months from issue publication**.

[2]Postdoctoral Research Associate (Ganguli) and Associate Professor (Ganguly), Sustainability and Data Sciences Laboratory, Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115 (E-Mail/Ganguli: p.ganguli@neu.edu).

dependence among drought attributes may lead to over- or underestimation of reservoir sizing (Salvadori and De Michele, 2004; Salvadori *et al.*, 2013). Similarly, the long-range persistence or the Hurst phenomena (Hurst, 1951), is one of the fundamental attributes of drought. An example of Hurst phenomena is the persistent drought condition in the Southwest of the CONUS (Stine, 1994; Woodhouse *et al.*, 2010). Understanding temporal scaling and long-term persistence within hydrologic variables is important for the design of water infrastructures. The uncertainty in the conventional statistical analysis may considerably increase due to the presence of long-term persistence in hydrologic time series (Koutsoyiannis and Montanari, 2007). Not considering persistence in the time series may lead to underestimation in return period, resulting in an inappropriate reservoir design (Douglas *et al.*, 2002).

Precipitation simulations from global climate models (GCMs) are derived variables and hence less robust than GCM-simulated state variables (such as temperature) and often fail to adequately capture important statistical characteristics, such as persistence (Johnson *et al.*, 2011; Rocheta *et al.*, 2014). Moreover, in order to make reliable decisions and ensure regional resilience in response to future climate change, water resources managers and planners can use climate projections at fine-scale resolution. The latest generation (Coupled Modeling Intercomparison Project phase 5, or "CMIP5") GCMs run at a spatial resolution of 150-300 km and are unable to resolve fine-scale features, such as clouds and topography explicitly. Assessments of environmental impacts typically require information at higher resolutions, for example, at resolutions of 50 km or higher. A perspective article (Bonnin, 2013) from the National Weather Service, the organization which develops precipitation frequency atlas of the U.S., mentions that insights from climate science "do not discuss frequencies and durations required for civil infrastructures." High resolution climate information is essential in impact assessment in hydrology (such as for the construction of *Intensity-Duration-Frequency* curves of precipitation extremes [Aron *et al.*, 1987; Yarnell, 1935], and *Severity-Duration-Frequency* curves for drought [Dalezios *et al.*, 2000; Halwatura *et al.*, 2014]) and agriculture (simulation of crop yield models [Olesen *et al.*, 2007; Xiong *et al.*, 2007]). Thus, to capture fine-scale regional information at stakeholder- (e.g., water resources managers and planners) relevant scales, different downscaling, such as statistical (Benestad *et al.*, 2008) or dynamical (Giorgi and Mearns, 1991) methods have been developed.

Dynamical downscaling is based on regional climate models (RCMs), where all vertical levels of the atmosphere, including the surface level are taken into account and relatively (compared to GCMs) fine-scale hydrometeorological processes are simulated (Leung *et al.*, 2004). In addition, due to enhanced resolution they are expected to provide added value in the frequency distribution of local weather anomalies, such as extreme daily precipitation (Laprise, 2008). On the other hand, statistical downscaling methods are based on finding statistical relationships between a set of predictors and predictands (Wilby *et al.*, 1998; Jeong *et al.*, 2012). Dynamically downscaled variables respond in physically consistent ways to external forcing (e.g., land-surface changes) and are therefore assumed to be less susceptible to non-stationarity (or fundamental changes in regional climate patterns owing to radiative forcing under global warming). Statistically downscaled variables are thought to be primarily a result of synoptic forcing and carefully selected large-scale climate parameters (Wilby *et al.*, 2004; Jeong *et al.*, 2013), which might not guarantee physical consistency under non-stationarity (Hayhoe *et al.*, 2008; Torma *et al.*, 2015). For the Northeast U.S., Hayhoe *et al.* (2008) reports superior performance of dynamically downscaled RCMs as compared to statistically downscaled daily precipitation extremes, especially along the coast. However, the value added by the RCMs, while not questioned for hypothesis testing, have been debated for downscaling (Racherla *et al.*, 2012; Kerr, 2013). Uncertainties in RCMs can result from parameterizations and resolutions, initial and lateral boundary conditions of the driving GCMs, and inter-model variability, all of which constrains their accuracy (Foley, 2010). RCMs are computationally demanding, i.e., they need more computer time to reproduce equivalent dynamical scenarios as compared to the statistical downscaling. The issue is further exacerbated by the combinatorics, since running an exhaustive set of GCM-RCM combinations is a substantial computational challenge, while an arbitrary sub-selection underestimates the variability. The suite of dynamically downscaled climate models that are part of the North American Regional Climate Change Assessment Program (NARCCAP) computes a selected subset of all possible combinations. A comparative analysis of statistical *vs.* dynamically downscaled daily precipitation suggests both methods exhibit considerable uncertainty to regional climate simulations, especially in simulating summer precipitation (Schmidli *et al.*, 2007).

To date, very few studies have attempted to evaluate the credibility of RCMs in the context of meteorological droughts (e.g., Gao *et al.*, 2012), although there is considerable prior literature on climate extremes (Bukovsky, 2012; Di Luca *et al.*, 2012; Mishra *et al.*, 2012; Singh *et al.*, 2013; Wehner, 2013). Jeong *et al.* (2014) compared future projected changes of meteorological drought duration and severity based on the Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) and showed the role of temperature in future drought changes. However, comparisons do exist for observed droughts with NARR (North American Regional Reanalysis; Mo and Chelliah, 2006; Karnauskas *et al.*, 2008; Weaver *et al.*, 2009; Sheffield *et al.*, 2012a, b) and the land surface models (LSMs) (e.g., the Noah model; Chen *et al.*, 1997 and the variable infiltration capacity, VIC; Liang *et al.*, 1994). NARR is based on reanalysis, which does not offer information about future climate conditions. The prior studies have mostly used the Palmer Drought Severity Index (PDSI) as the drought index, which in turn is based on water balance calculations. There have been concerns that the PDSI lacks multi-scale features and hence may not capture droughts on time scales less than about 12 months (Dai, 2014), thus being unsuitable for capturing seasonal droughts. The LSMs assess long-term hydrological drought involving soil moisture condition and overlooks seasonality. Physically distributed hydrological models, such as the VIC (Andreadis and Lettenmaier, 2006; Sheffield *et al.*, 2009), are expected to be better at handling surface and subsurface hydrological processes but may contribute larger uncertainties owing to the inherent parameterizations involved in the model.

Understanding the ability of climate models to reproduce regional drought (seasonal to decadal time scale) patterns and accurately reproduce historical observations at relevant (higher than GCMs) spatial scales is crucial for water resources planning. The NARCCAP is a plausible choice to study meteorological drought over the CONUS. NARCCAP offers an archive of simulated data at a horizontal resolution of 50 km (Mearns *et al.*, 2009, 2012) based on runs of six RCMs at three hourly time steps and produced in two phases. Phase I (which runs from 1979-2004 with usable period 1980-2004 excluding spin-up data) dynamically downscales retrospective atmospheric reanalysis and utilizes perfect boundary forcing. Phase II (runs from 1968-2000 with usable period 1971-2000) downscales data from free running coupled atmosphere-ocean general circulation models. The prior literature has not examined the performance of NARCCAP in simulating observed meteorological droughts with precipitation as a sole forcing input.

We examine meteorological droughts in present-day climate (1971-2003) in NARCCAP ensembles. Studying meteorological droughts is important primarily for two reasons. First, prolonged meteorological droughts often act as a catalyst for more damaging other drought categories, such as agricultural and hydrological droughts (Wilhite *et al.*, 2014). Second, the ability to reproduce observed meteorological drought trends can provide stakeholder confidence in model skills relevant for impact assessments. The analysis of present-day climate simulations allows an identification of systematic model errors, which in turn helps in developing a better understanding of climate change signals in projected time periods (Giorgi *et al.*, 2004). In this study, we examine the ability of both GCM-driven and NCEP-driven RCM runs to simulate observed drought attributes. Since reanalysis effectively encapsulates weather prediction model analysis fields, it is appropriate to compare the RCM output with observations on an individual event basis. In GCM-driven runs, GCM outputs are used to provide boundary conditions for both historical and future climate runs. However, for historical runs, model performance cannot be evaluated against individual events, and comparison with observations is only possible for statistical attributes independent of time steps. Hence, in the latter case, only those statistical properties of droughts, which are temporally independent, have been analyzed. We examine meteorological droughts with most commonly used indices, specifically, the SPI, which is a measure of water availability relative to the baseline condition (McKee *et al.*, 1993) and captures the multi-scale nature of drought.

The present study evaluates following primary research question for the CONUS:

1. How good are the Phase I and Phase II simulations of the NARCCAP RCMs in providing credible predictive insights for meteorological droughts and associated drought statistics?

This in turn leads to a few related ancillary questions, which directly relate to data and analysis methodology and how it translates to overall meteorological drought trends:

1. Do observational datasets obtained from different sources consistently simulate trends in precipitation, one of the major drivers of meteorological droughts?
2. How sensitive are the drought metrics and associated statistics at different temporal scales?
3. Do RCMs add substantial value in simulating observed precipitation as compared to “raw” precipitation directly obtained from GCMs?

Our study adds to existing literature in several aspects. First, to date very few studies attempted to evaluate the credibility of RCMs in general and the NARCCAP in particular in replicating meteorological droughts and associated attributes, although equivalent analyses have been performed for temperature (Jeong *et al.*, 2014), precipitation (Wehner, 2013; Singh *et al.*, 2013), and wind extremes (Pryor *et al.*, 2013a, b). Second, while previous work mostly evaluated RCM skills against observations based on reanalysis-driven RCM experiments or projection of extremes based on GCM-driven experiments, this study is one of the first kind that investigates the credibility of Phase I (NCEP-driven) and Phase II (GCM-driven) NARCCAP runs in simulating meteorological drought in the present-day (1971-2003) climate. Third, we use different quantitative metrics to assess RCM skills against multiple observational datasets in simulating wide ranges of both temporally dependent (such as trends, temporal variability in drought area) and independent (frequency and persistence) regional drought statistics, which are not yet investigated in any of the existing literature. In addition, we evaluate GCM-driven RCMs' skills with respect to their host GCMs. This is because in most cases the performance of the models is debatable when RCMs are driven by GCM fields due to the potential low quality of the forced GCM climatologies. We hope our analysis will assist the modeling community in recognizing systematic model errors, which provides an opportunity for further model development and improvement. Further, the analysis presented herein will be helpful for the stakeholder community in identifying model limitations before using these models in Impact, Adaptation and Vulnerability (IAV) studies.

Here the objectives are to evaluate the credibility of NARCCAP RCMs to reproduce observed statistical attributes of meteorological droughts over the CONUS. Robustness is typically examined at decadal to multi-decadal timescales based on maximum data availability of the NARCCAP model runs, specifically, 1980-2003 for NCEP-driven runs, and 1971-1999 for GCM-driven simulations. A 25-30 years' timescale is typically used to determine a climatological average by climate scientists since it is long enough to filter out natural variability (e.g., the impact of oceanic oscillators) in the climate systems. Coincidentally, the typical planning horizon of water resources planners and infrastructure managers is usually about 30 years in the future. Thus, insights based on 25-30-years' average can help to understand projection skills, quantify uncertainty, and assess regional impacts of meteorological droughts.

DATA AND METHODS

Study Region

Our study focuses on the CONUS (20°N-50°N, 125°W-60°W). We consider nine climatologically homogeneous regions across the CONUS as suggested in the literature (Karl and Koscielny, 1982; Karl and Koss, 1984). Figure 1 shows climatologically homogeneous regions and topography map of the CONUS. Delineation of these regions is performed using principle component analysis of gridded PDSI values. These regional classifications have been used by many researchers earlier in the context of drought (Soulé and Yin, 1995; Easterling *et al.*, 2007).

Observational Data

We used three different precipitation datasets available at monthly time steps for validation. The spatial resolutions of the observed datasets are close to that of NARCCAP simulated models (0.5°). The use of multiple observations, while not a norm in model evaluation studies, needs to be considered for droughts, specifically because this can address the issue of uncertainty in the observed drought patterns. As discussed by Trenberth *et al.* (2014), different observational datasets may generate considerably different insights regarding drought trends (e.g., see the diametrically opposite insights in Dai, 2013 and Sheffield *et al.*, 2012a, b). Thus, rather than recommending one observational dataset for validation, we believe a better strategy may be to examine multiple observational datasets and use agreement about these datasets as one measure of credibility for any insight (including diagnosis or prognosis) of droughts. In other words, we would assess credibility of models by examining those drought patterns that exhibit similarity across multiple observations.

The first dataset is produced by the Climate Research Unit (CRU TS3.22) of the University of the East Anglia (Harris *et al.*, 2014). This includes gridded precipitation data over land at 0.5° spatial resolution for the time period 1901-2013, out of which we extracted data for the period 1971-2003 for the analysis. Harris *et al.* (2014) presents a detailed comparison of CRU precipitation climatology against other available observed precipitation climatologies and found the dataset compares favorably; however, the major deviations are mostly in regions or time periods with sparser observational datasets.

The second dataset is from the Global Precipitation Climatology Center (GPCC) at Deutscher Wetterdienst, hereafter referred to as GPCC v.6 (Schneider *et al.*, 2014). This dataset contains global land surface precipitation based on 67,200 stations worldwide, which have record durations of 10 years or longer. The dataset contains monthly precipitation records at a regular spatial resolution of 0.5°, 1°, and 2.5° with temporal coverage ranges from 1901 to 2010. For the present

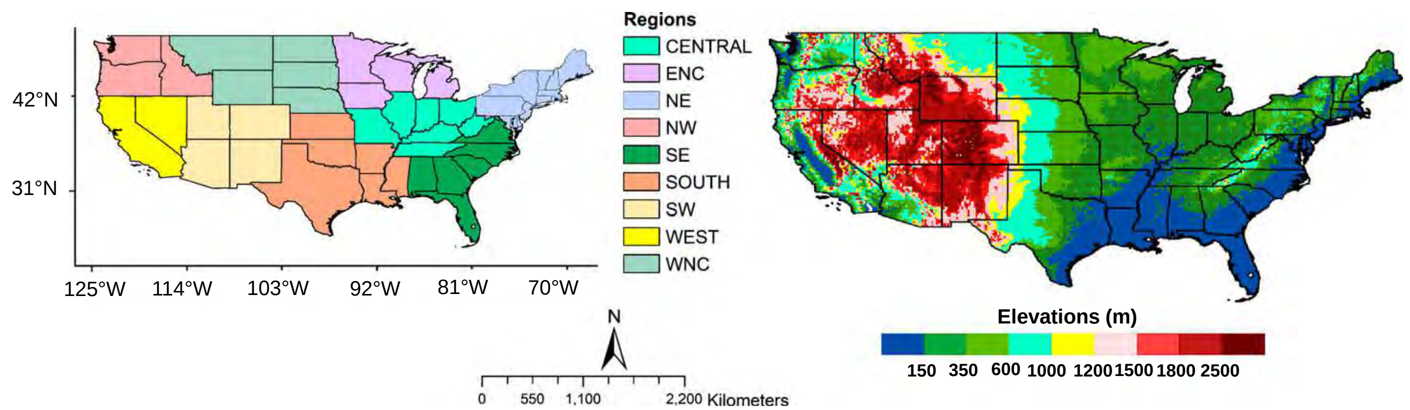

FIGURE 1. Climatologically Homogeneous Regions in U.S. The abbreviations ENC, NE, NW, SE, SW, and WNC denote East-North Central, Northeast, Northwest, Southeast, Southwest, and West-North Central regions, respectively (left); topography distribution (right).

study, we extracted data for the common period of 1971-2003 at a spatial resolution of 0.5°.

The third dataset is the University of Delaware v.3.01 (UDel v.3.01; Willmott and Matsuura, 2001). The dataset is available at a monthly temporal resolution over global land areas spanning from 1900-2010.

Regional Climate Model Data

We used archive simulated data from six RCMs driven by NCEP and GCM boundary conditions. Details of all NARCCAP models considered in this study are listed in Table 1. All operate at a spatial resolution of 50 km over landmasses of the 48 contiguous U.S., most of Canada to 60°N, and Northern Mexico. Simulations with these models are produced for the current and the mid-21st Century (2041-2070) under the SRES A2 emissions scenario. The models differ in structure and parameterization schemes. Two of the regional models, CRCM and ECP2 include the "spectral nudging" technique, which imposes time-variable large-scale atmospheric states in the integration area of the RCM domain (von Storch *et al.*, 2000; Wehner, 2013). The remaining regional models were unconstrained inside the integration area. To perform comparative analyses, the NARCCAP models with three hourly temporal resolutions are aggregated to a monthly time scale. To avoid missing data near the end of the simulations and to maintain consistency throughout the analyses while trying to include as much data as possible, all NCEP-driven runs are analyzed during 1980 to 2003, while GCM-driven RCMs are analyzed for the time frame of 1971-1999.

TABLE 1. Details of RCM Models Used in This Study.

Model Acronym	RCM with Institution/Model Origin and Reference	Boundary GCM(s)
CRCM	Canadian RCM (Caya and Laprise, 1999)	CCSM, CGCM3
ECP2	Scripps Experimental Climate Prediction Center Regional Spectral Model (Juang *et al.*, 1997)	GFDL
HRM3	Third-generation Hadley Center RCM (Jones *et al.*, 2003)	HADCM3, GFDL
MM5I	Fifth-generation Pennsylvania State University — National Center for Atmospheric Research Mesoscale Model (Grell *et al.*, 1994) run by Iowa State University	CCSM
RCM3	Regional Climate Model version 3 (Giorgi *et al.*, 1993a, b), run by UC Santa Cruz	GFDL, CGCM3
WRFG	Weather Research and Forecasting Model, run by Pacific Northwest National Laboratory (Skamarock *et al.*, 2005)	CCSM, CGCM3

In general, as compared to the single model, the multi-model ensembles increase the overall skill, reliability, and consistency of the model performance (Tebaldi and Knutti, 2007) while characterizing model uncertainty from the ensemble spread (Sanderson and Knutti, 2012). Hence, apart from individual model performance, the performance is also evaluated on multi-model ensemble members (multi-model median and bounds).

Global Climate Model Data

To understand value added by RCMs, we compare the performance of NARCCAP ensembles with precipitation simulations of the 20th Century (20C3M) scenarios from their host (or driving) global coupled atmospheric ocean general circulation models (AOGCMs) archived at monthly time steps: CCSM3.0, CGCM3.1, and GFDL-CM2.0 and HadCM3. These AOGCMs are made available by the World Climate Research Program (WCRP) Coupled Model Intercomparison Project Phase 3 (CMIP3; Meehl *et al.*, 2007). The HadCM3 run for NARCCAP was different from that in CMIP3 archive, therefore, the outputs of this GCM simulation were obtained by contacting NARCCAP team directly. To maintain simplicity in the analysis the initial condition biases associated with the GCM simulations are assumed to be insignificant and only the first realization was used when multiple ensemble runs were available for each of the driving GCMs (Rocheta *et al.*, 2014).

All climate model outputs are interpolated to a common grid of 0.5° latitude/longitude resolution using the bilinear interpolation technique in the Climate Data Operators software (CDO, https://code.zmaw.de/projects/cdo). To compare with observations, land/sea mask at 0.5° spatial resolutions are obtained from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) and are applied to GCM and NARCCAP simulated fields.

Meteorological Drought Attributes

Meteorological drought is referred to as a precipitation deficiency, in comparison to normal or baseline conditions. We use Standardized Precipitation index (SPI-n, where n = 3-, 6-, 9-, and 12-month accumulation period) as an index of meteorological drought. SPI represents the number of standard deviations above or below that an event is from the long-run mean (Sims *et al.*, 2002). To estimate SPI at an "n-month" time scale (hence, SPI-n), an accumulation window of n-months is applied to a given monthly precipitation time series, following which a statistical distribution is fitted. In this article, as in the original work of McKee *et al.* (1993), we used Gamma distribution to fit precipitation time series aggregated at n = 6 months (McKee *et al.*, 1993; Sims *et al.*, 2002). SPI is spatially invariant and probabilistic in nature and able to capture different drought states ranging from short-, medium-, and long-term drought conditions depending on the length of the accumulation period. SPI has a number of advantages (Lloyd-Hughes and Saunders, 2002), such as: (1) The SPI is based on precipitation and requires computation of only two parameters, compared to multiple computational terms needed to compute PDSI. (2) By avoiding dependence on soil moisture conditions, it can be effectively used both in summer and winter seasons and is not adversely affected by topography. (3) It can be tailored to specific needs for impact assessment. For example, its variable time scales are useful for modeling a wide range of meteorological, agricultural, and hydrological applications. The temporal nature of the index facilitates understanding drought dynamics, such as onset and cessation, which is difficult to be tracked by other indices. (4) Standardized nature ensures that the frequency of extreme events at any location and on any time scale is consistent. Conversely, application of SPI has a few potential disadvantages: (1) the quantity and reliability of the data used to fit a suitable probability distribution, (2) due to its standardized nature, SPI is incapable of identifying regions that are more "drought prone" than others, and (3) employing SPI at shorter time scales (such as 1, 2, and 3 months) to the regions with low seasonal precipitation, which results in erroneous large positive or negative SPI values.

The drought properties are derived using threshold methods, which are based on statistical theory of runs (Yevjevich, 1983) for analyzing sequential time series. A drought event is identified when an uninterrupted sequence of SPI values (at monthly time scales) remains equal to or below the 20th percentile of the SPI distribution over the period analyzed at a specific site (Svoboda *et al.*, 2002). We characterize meteorological drought to the following properties (McKee *et al.*, 1993):

1. **Duration**: Number of consecutive months when SPI remains equal to or below the threshold value.
2. **Severity**: Cumulative values of SPI within the drought duration. In general, for convenience severity of drought events at a particular time scale is taken to be positive and expressed as (McKee *et al.*, 1993)

$$S_i = -\sum_{t=1}^{D} \mathrm{SPI}_{i,t} \quad i = 1,\ldots,n \tag{1}$$

3. **Percentage Area under Drought (PAUD)**: The fraction of the area (in percentage) is considered under drought if the SPI values for the grid cells reach below the specified threshold limit in line with the former study (Sheffield and Wood, 2008). Accordingly, PAUD (A_t) at a time step t is computed using the expression

$$A_t = \frac{\sum_{i=1}^{N_{\mathrm{grid}}} 1\{Z_{i,t} \le Z_{\mathrm{thr}}\} A_i}{\sum_{i=1}^{N_{\mathrm{grid}}} A_i} \forall t = 1,2,\ldots,n \tag{2}$$

where $1\{\psi\}$ is a logical indicator function of set ψ, taking the value of either 0 (if ψ is false) or 1 (if ψ is true), $Z_{i,t}$ is the SPI value of month t, Z_{thr} is the threshold limit of SPI for identifying drought in the grid, A_i denotes influence area of the grid i and is computed by area of the grid cell i weighted by the cosine of the grid latitude and N_{grid} is the total number of grids in the region.

4. **Persistence**: The persistence in hydroclimatic time series represents temporal grouping of non-periodic similar events, such as occurrence of similar conditions such as dry (or wet) spells in a cluster of time frames (Outcalt *et al.*, 1997; Mesa *et al.*, 2012; Kumar *et al.*, 2013). Drought persistency is quantified by the Hurst exponent (index, H). The Hurst index, $0.5 < H < 1$ ($H = 0.5$, the data is independent, which is expected in a random series and due to the absence of long-term temporal correlation) represents positive persistency in the time series and reinforces the trend. This implies, if the series is showing a downward (upward) trend of its long-term average in the previous period, it is likely to follow the same trend in the subsequent time period (Kumar *et al.*, 2013). For many geophysical time series, H values range between 0.6 and 0.9 (Outcalt *et al.*, 1997). We use detrended fluctuation analysis (Peng *et al.*, 1994; Weron, 2002) to compute the Hurst exponent.

Methodology and Evaluation Metrics

First, we evaluate uncertainty in different observational datasets during 1971-2003 in simulating precipitation. For this, we compare the first and second moment properties (mean and standard deviation) and lag-1 autocorrelation in the datasets. Then we investigate the sensitivity of SPI at different time scales by comparing following metrics over the land grids: (1) spatial pattern during notable drought years, and regional distributions of (2) cross correlation fields, (3) weighted average drought severity (weighted by the drought duration), and (4) frequency.

After investigating uncertainty in different observations and the evolution of meteorological drought at multiple time scales, we analyzed skills of the GCM and reanalysis-driven NARCCAP ensembles in reproducing observed drought statistics. We adopted three general criteria to assess the robustness of NARCCAP RCMs to simulate historical drought climatology. The first criterion assesses the value added by the multi-model (hereafter referred to as MME) median NARCCAP GCM-RCM ensemble against the simulation of precipitation fields from the MME median raw "host" GCM during 29 years (1971-1999) of the simulation time period. The second criterion assesses the robustness of the NARCCAP, driven by NCEP boundary conditions to emulate observed drought trends (1980-2003). Here we used individual NARCCAP RCMs and their multi-model ensembles, MME median and its bounds (MME minimum and MME maximum). The MME minimum and MME maximum ensembles of the RCMs are computed at 10th and 90th percentile levels of the model runs. The third criterion assesses the ability of GCM-driven RCMs in simulating statistical metrics of temporally independent observed drought properties over 1971-1999. Brief descriptions of each of the assessment methods and metrics used are summarized here.

In the first assessment, we evaluate the added value of GCM-RCM NARCCAP run against the host GCM by comparing spatial patterns of climatology and variability across the CONUS. We also compare the variability in seasonal precipitation in GCM-driven NARCCAP RCMs and their host GCMs against observations. We use these metrics to assess the impact of climate change; it is necessary to examine, as a minimum, the ability of the models to simulate mean and variance reasonably with respect to the observations. The second assessment is based on the number of test statistics between observations and models during the analysis period (1980-2003):

1. The Taylor diagrams (Taylor, 2001) of regionally averaged SPI time series to assess pattern error over the nine regions.
2. Trends in SPI time series using non-parametric Mann-Kendall trend statistics with correction for the ties and autocorrelation (Hamed and Ramachandra Rao, 1998; Reddy and Ganguli, 2013). The slope of the trend is estimated using Theil-Sen estimator.
3. Pattern correlation analysis of drought climatology using non-parametric Spearman's rank correlation.
4. Spatiotemporal variability of average (median) percentage area under droughts (PAUD) using box-plots and spatial autocorrelation plots.
5. Distribution of drought frequency using box plots.
6. Box plots and scatter plots of observed *vs.* absolute bias (modulus difference between model and observations) in median drought severity depicting associated uncertainty.

In the third assessment that compares baseline simulations (1971-1999) of GCM-driven NARCCAP runs relative to observations, only those statistical metrics of droughts are considered that are time independent. In this case, we consider two drought properties, i.e., frequency and drought persistence. We analyze the distribution of drought frequency over nine regions using box plots. The persistence in SPI time series is computed using Hurst exponent. Agreement in Hurst exponent between multiple observations and the RCMs is analyzed using pattern correlations.

ANALYSIS

Comparison of Precipitation Datasets

Figure 2 depicts the mean annual precipitation, standard deviation and lag-1 autocorrelation for the CRU TS3.22, GPCC v.6, and UDel v.3.01 precipitation datasets over the 33-year time period (1971-2003). We found an overall agreement among observations in simulating statistical properties and general spatial patterns of annual precipitation. The agreement among datasets is the highest in the West-North Central regions and lowest in the Northeast region as analyzed by pattern correlation metrics of annual average precipitation. In addition, pattern correlation analysis of mean annual precipitation over different meteorological subdivisions shows the correlation between GPCC v.6 and UDel v.3.01 datasets are higher as compared to the correlation between GPCC v.6 and CRU TS3.22 dataset. The pattern correlation of mean precipitation between GPCC v.6 and UDel v.3.01 datasets ranges from 0.89 (Northeast) to 0.98 (West-North Central), whereas correlation between GPCC v.6 and CRU TS3.22 datasets ranges from 0.76 (Northeast) to 0.96 (West-North Central). On the other hand, the correlation of annual precipitation between CRU TS3.22 and UDel v.3.01 datasets varies between 0.72 (Northeast) and 0.97 (West-North Central). Spatial analysis of annual average precipitation suggests CRU TS3.22 tends to overestimate annual precipitation in the West as compared to the other two datasets (Figure 2; top panel). CRU TS3.22 dataset shows less than 11% of grid points with an average precipitation of 300 mm or less per year in contrast to other two datasets, in which 13% and more grid points have annual average precipitation of 300 mm or less. Further, CRU TS3.22

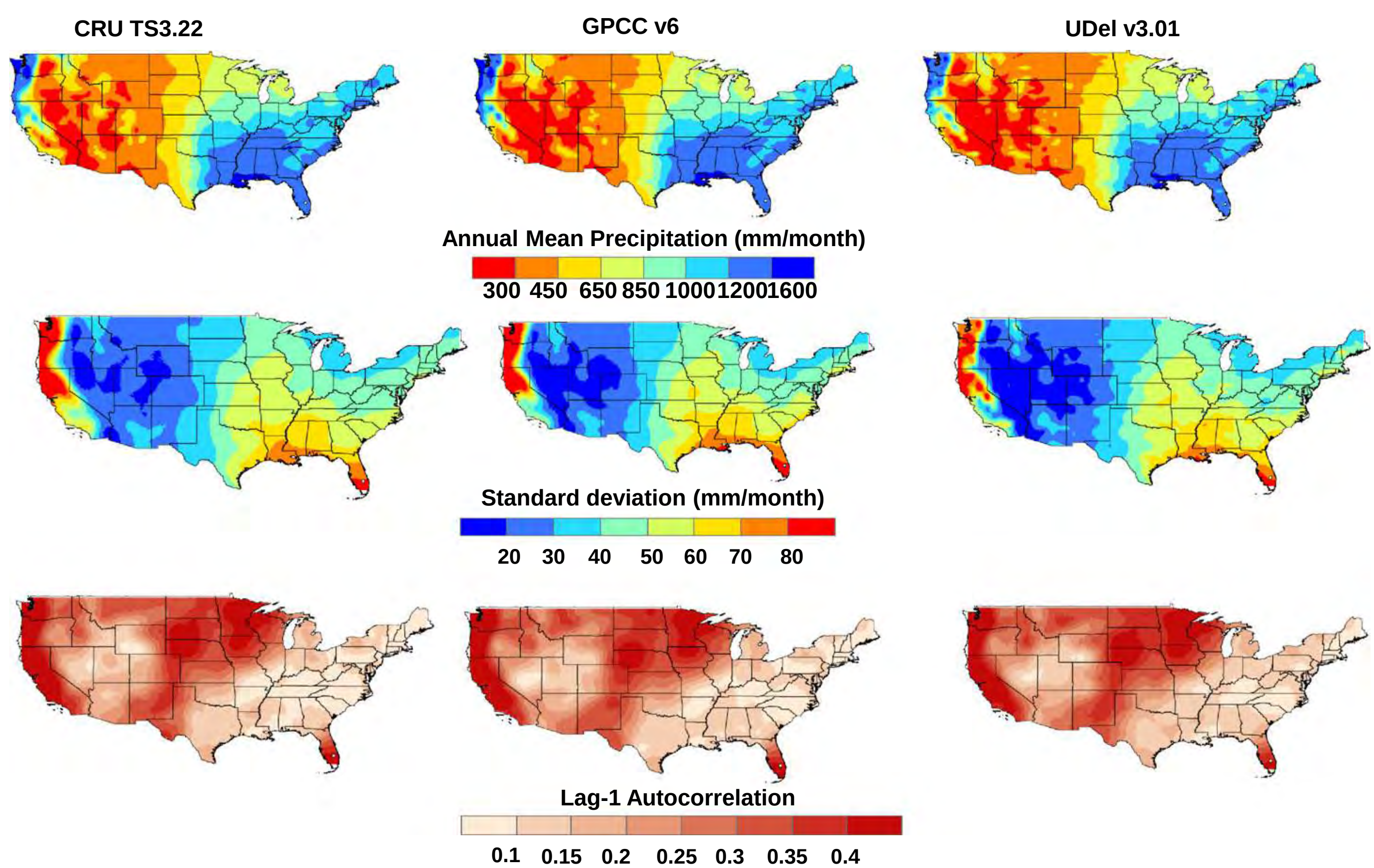

FIGURE 2. Spatial Distribution of Observed U.S. Precipitation Climatology (top row), Standard Deviation (middle row), and Lag-1 Autocorrelation (bottom row) for Different Datasets for the Common Period 1971-2003.

dataset underestimates the standard deviation for low values of precipitation, especially over the Southwest as compared to the other two datasets (Figure 2; bottom panel). The CRU TS3.22 datasets reportedly have the wet bias with respect to the other datasets since around 1996 as noted in earlier studies (Fekete *et al.*, 2004; Trenberth *et al.*, 2014).

Sensitivity in Spatial Patterns of Droughts at Different Time Scales

In the following analysis, we investigate the sensitivity of SPI at different time scales. Figure 3 shows spatial distributions of SPI calculated at 3-, 6-, 9-, and 12-month time scales at the end of July during three notable drought years between 1971 and 2003. For instance, at the end of July 1976 a 3-month SPI uses precipitation total of May, June, and July 1976, while the 12-month SPI uses the precipitation total from August 1975 through July 1976. These time scales reflect the impact of drought on the availability of different water resources. For example, soil moisture responds to precipitation anomalies on the relatively small time scale, therefore a three-month SPI can be used to monitor soil moisture conditions in different stages of plant development. On the other hand, streamflow, reservoir storage, and groundwater respond to long-term precipitation anomalies, therefore a 12-month SPI reflects hydrological drought condition.

In general, on a three-month time scale most of the regions have patches of the dry and wet pattern, and are characterized by near-normal conditions (i.e., $-0.8 \leq \mathrm{PI} \leq 0.8$). The Midwestern and coastal California regions during 1976, Midwest and Southeast regions during 1988, and part of West-North Central, Southwest, and West regions during 2002 are in moderate to extreme dry conditions. During 1976, on shorter (three months) and medium (six and nine months) time scales Midwest region is characterized by severe drought conditions, whereas on a longer time scale (12 months) the region is affected by near-normal to medium drought state. Conversely, on shorter and medium time scales, Southwest and West regions are mostly characterized by near-normal to wet state. However, on a longer time scale, the regions are in a moderate drought state, indicating the probability of hydrological droughts with a consequent loss of water resources. A similar trend was also noted in 1988, in which on a shorter time scale Midwest region was affected under severe drought

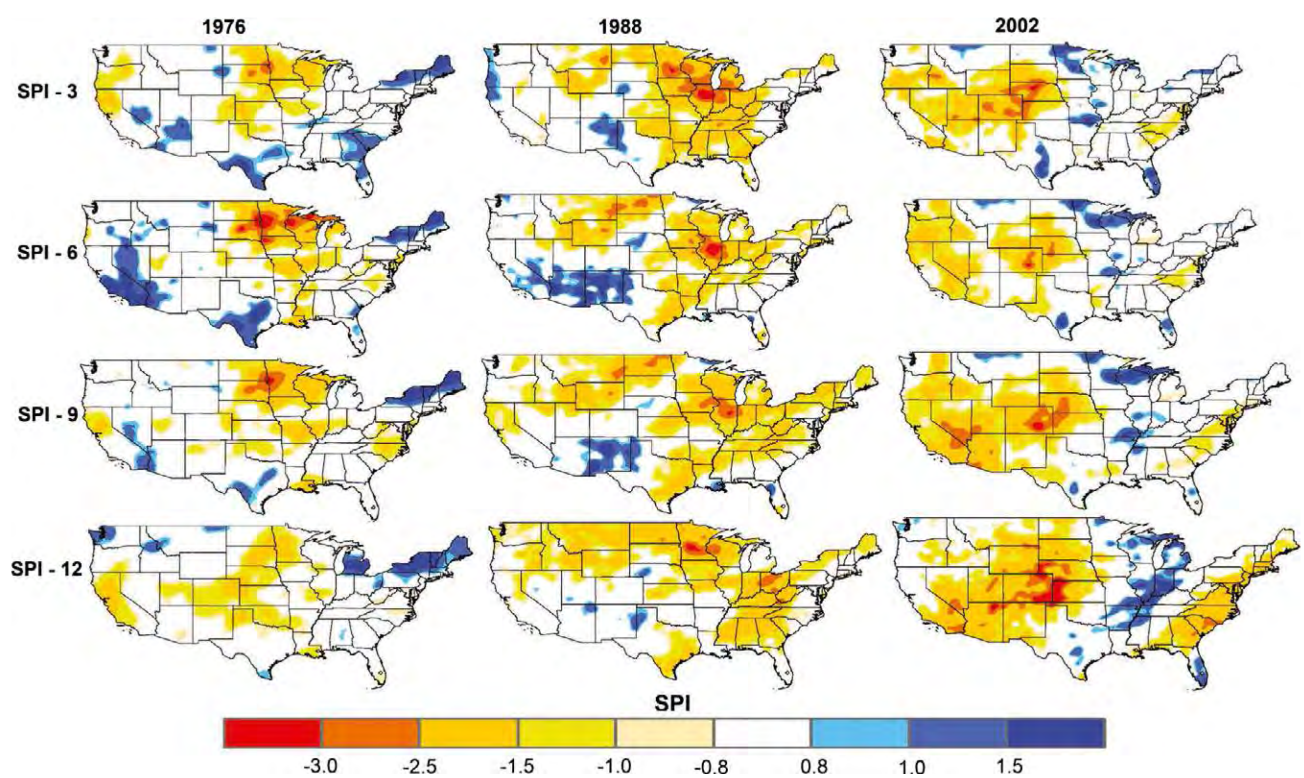

FIGURE 3. Spatial Maps of SPI at Time Scales of 3, 6, 9, and 12 Months in GPCC v.6 Data. Shades of blue from dark to light on top of the scale show wetter conditions while light green, yellow, orange, red, and dark red on the bottom show drier conditions; white indicates near-normal condition.

condition, whereas on longer time scale the region was characterized by medium to severe dry state with moderate to severe drought conditions extended toward West-North Central regions. Further evidence of moderate drought state is prominent over South and Southeast using SPI-12. However, at time scales of six and nine months, Southeast is found to be in near-normal condition. In 2002, the percentage grid points in the near normal state at time scales 3-, 6-, 9-, and 12-months are found to be around 54.1%, 53.8%, 48%, and 39.3%, respectively, indicating that with an increase in accumulation time scales, there is an increase in spatial extent of dry and wet pattern. The percentage grid points under extreme drought state (SPI < -2) are found to be around 10% in SPI-3 and 17% in SPI-12, showing signs of long-term hydrological drought especially over the Southwest, West, and parts of Southeast regions during 2002. Previous studies (US Drought Monitor; http://droughtmonitor.unl.edu/) suggest in summer 2002, more than 50% of the contiguous U.S. was under moderate to severe drought conditions, whereas the western part of the country has been in the grip of severe droughts since late 1999. The three-month SPI may be misleading in the Southwest and West. Since these regions are characterized by little rain, the corresponding historical totals will be small leading to relatively small deviations on either side of the mean, which could result in large negative or positive SPIs (WMO, 2012).

Next we examine the sensitivity of drought statistics using distributions of drought properties at different time scales. Figure 4 shows distributions of the spatial cross-correlation of SPI time series over the nine regions. In these figures, the interquartile range of the box-plots shows a measure of spatial variability across the regions. While point statistics such as SPI time series, may be described by a map, spatial properties such as cross-correlation vary between a pair of grid points, and should be available for every possible grid location. The cross-correlations of SPI time series between a pair of grid points are computed using non-parametric Kendall's τ correlation over the land grids of the CONUS. Figure 4 suggests regional cross-correlations are positively correlated over most of the grid points. The drought indices tend to be closer and less uncertain (shown by 5th and 95th percentile whisker plots) at smaller time scales; however, at a longer time scale the indices, in general, differed more. We found that the uncertainty bounds (25th and 75th percentiles) in spatial cross-correlation grow with the increase in SPI-time scales over most of the regions. The median cross-correlation is found to be least in the Southwest and ranges from 0.29 (SPI-3) to 0.33 (SPI-12) and highest in the Northwest region and varies between 0.44 (SPI-3) and 0.5 (SPI-12). Figure 5 shows distributions of weighted average severity and number of droughts. The regional distributions of weighted average severity (Figure 5, top panel) show the increase in drought severity with the increase in the time scales. In both shorter (3 months) and longer (12 months) time scales Midwest (East-North Central) is characterized by the highest median drought severity (around 16.65). On medium term (six and nine month time scales), the West-North Central

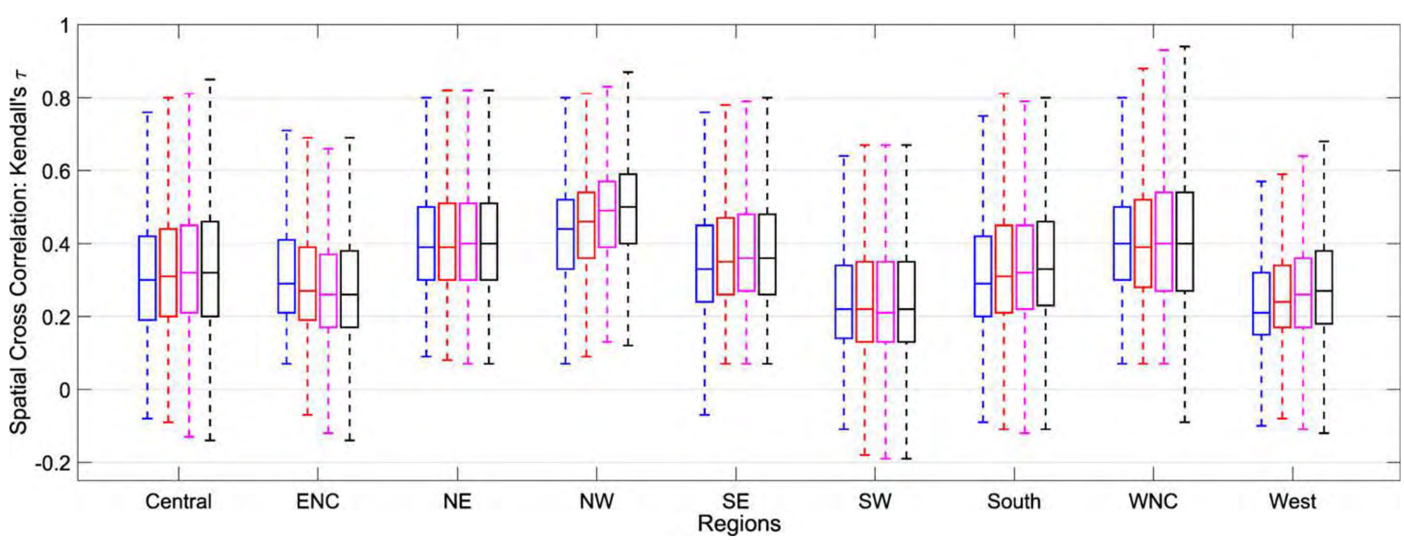

FIGURE 4. Box Plots of Spatial Cross-Correlation of Kendall's τ Statistics (significant at 5% significance level) of SPI Time Series between All Pairs of Grid Points at Different Time Scales for the Nine Regions. Distributions in blue, red, magenta, and black correspond to 3-, 6-, 9-, and 12-month time scales, respectively. Box-and-whisker plots show distribution of SPI with median (horizontal line), interquartile range (box), and 5th-95th percentiles (whiskers) of the data.

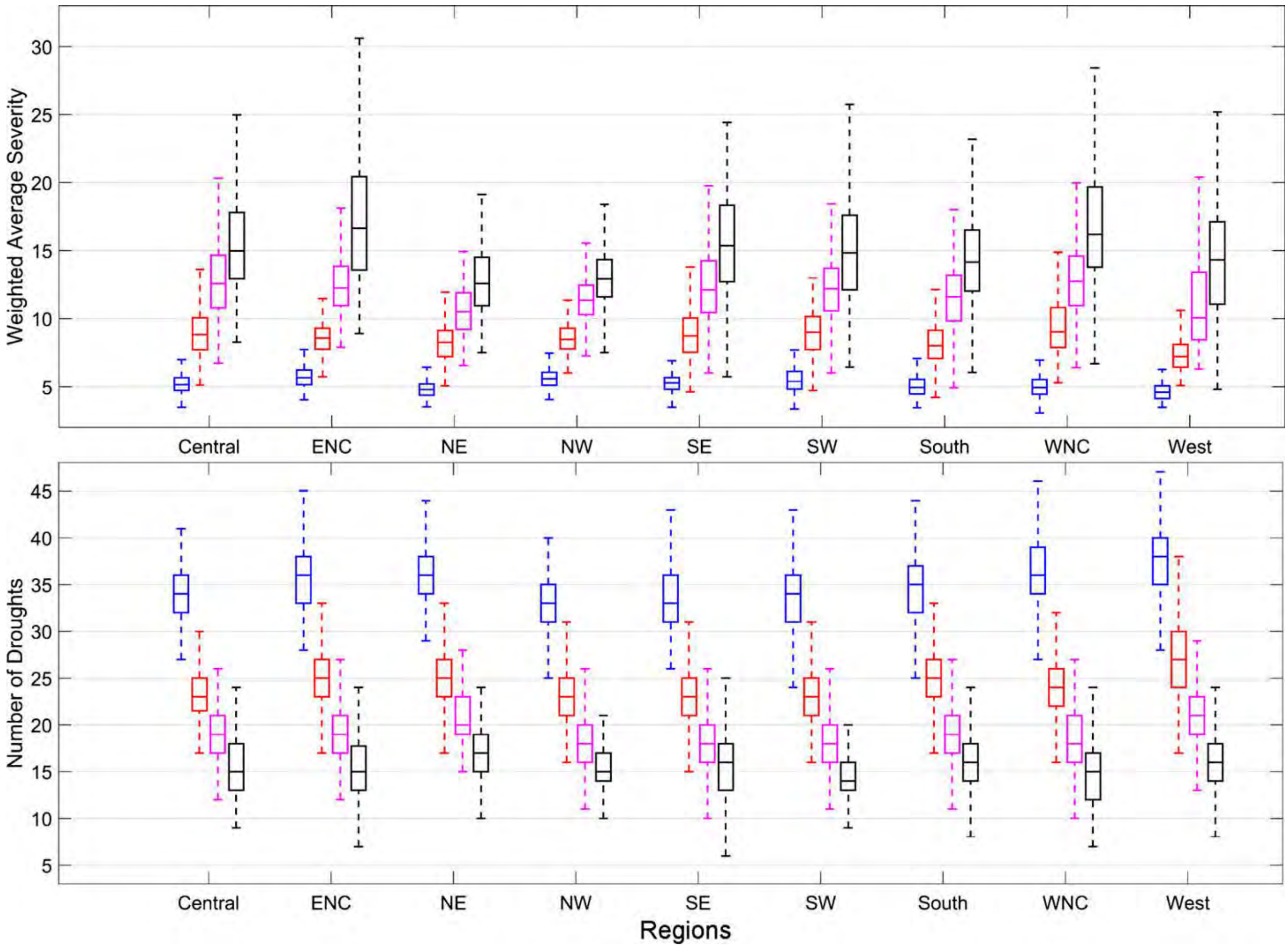

FIGURE 5. Box Plots of Weighted Average Severity (weighted by duration) and Drought Frequency at Different Time Scales for the Nine Regions. Distributions in blue, red, magenta, and black correspond to 3-, 6-, 9-, and 12-month time scales, respectively. Box-and-whisker plots are defined in Figure 4.

region is characterized by the highest average drought severity. At longer time scales, many regions, including Central, South, Southeast, and West-North Central show high median drought severity (exceeding 14). On the other hand, regional distributions of drought frequency (Figure 5, bottom panel) suggest decrease in the number of droughts with increase in the time scales. At short-to-medium time scales, the highest average drought frequency is noted over the West and ranges between around 21 (at a time scale of nine month SPI) and 38 (three month SPI) droughts on an average (per 33-year). At a 12-month time scale, the median drought frequency is found to be highest over the Northeast with around 17 droughts on an average (per 33-year) followed by the West, South, and Southeast with around 16 droughts on an average (per 33-year). Although at longer time scale, the Northeast has the highest average drought frequency, it is characterized by the least median severity (around 12.5). This implies higher average drought frequency in Northeast is counterbalanced by a lesser average severity. Recently, Hayhoe *et al.* (2007) reported frequent droughts in the Northeast in recent years with extended low-flow periods in summer.

In subsequent analysis, SPI at an accumulation period of six months is chosen over other time scales since it reflects seasonal to moderate trends in precipitation (WMO, 2012). SPI in this time scale is effective for the detection of agricultural drought conditions because it indicates the water content of vegetation and the soil moisture conditions (Sims *et al.*, 2002; Ji and Peters, 2003). Moreover, SPI at shorter time scales (such as at 1 or 3 months) may give erroneous results at dry regions, while at longer accumulation periods (such as 9 and 12 months) the uncertainty in drought conditions may increase due to the limited number of available records.

Robustness of GCM Forced NARCCAP RCMs in Simulating Regional Precipitation

In this section, a quantitative evaluation of the RCM skill over GCMs during 1971-1999 is presented. Figure 6 compares annual average precipitation and standard deviation of the ensemble median of ten GCM-RCM pairs (i.e., CRCM-CCSM, MM5I-CCSM, WRFG-CCSM, CRCM-CGCM3, RCM3-CGCM3, WRFG-CGCM3, ECP2-GFDL, HRM3-GFDL, RCM3-GFDL, and HRM3-HadCM3) with the ensemble median of their four host GCMs (CCSM3.0, CGCM3.1, GFDL-CM2.0, and HadCM3). In general, GCMs are able to capture broad features of time-averaged precipitation pattern reasonably well, however, they fail to simulate topographically induced features of precipitation due to inherent coarser horizontal resolutions. On the other hand, precipitation pattern simulated from RCMs shows a number of topographically induced fine-scale regional features and their variability, such as precipitation pattern over the Southwest, West-North Central, and Northwest regions, respectively, although their simulated intensity may differ from observations in many regions. In general, the annual average precipitation map shows a tendency of GCM-driven NARCCAP ensembles to produce larger precipitation over the Pacific Northwest and Northeast regions as compared to their host GCMs. However, they could be able to simulate the dry zone transition that arises from precipitation shadowing by the mountain ranges in the Intermountain region, which is not adequately represented by their host GCM field (Figure 6).

Next we compare seasonal precipitation fields simulated by the GCM and the GCM-driven NARCCAP RCMs against multiple observations over the nine regions. Figure 7 shows the distribution of seasonal precipitation in the three observations, multi-model median GCM and GCM-driven NARCCAP ensembles. Among observations, a close agreement is noted between CRU TS3.22 and GPCC v.6 data in simulating seasonal precipitation. However, UDel v.3.01 underestimates precipitation over the Northwest and West in all seasons. For the DJF (December-February) season, both climate models (the MME median of GCM-driven NARCCAP RCMs and the MME median of their host GCMs), especially NARCCAP overestimates winter precipitation variability over the Northwest and Southwest regions. Both climate models fail to capture the spatial variability of precipitation over Northeast, Southeast, and South regions. Over Southeast, the performance of MME median NARCCAP RCM is found to be superior to the host GCM. For MAM (March-May), climate models (MME median host GCM and NARCCAP RCM), overestimate precipitation over the Northeast, Northwest, Southwest, West, and West-North Central regions and underestimate in Southeast and South. Taken together, in both winter and spring seasons, the spatial variability of NARCCAP RCMs is found to be higher over the Northwest and Southwest regions relative to their host GCMs. The winter precipitation

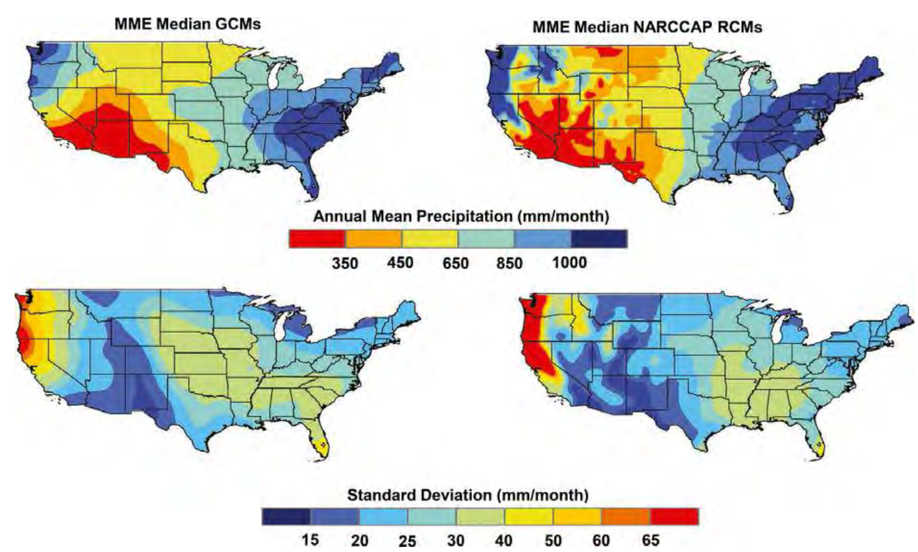

FIGURE 6. Spatial Distribution of Precipitation Climatology and Standard Deviation in Multi-Model Median of Regional Climate Models/ General Circulation Model (RCM/GCM) Combination (right) and Their Host GCMs (left) over the Period 1971-1999. Four GCMs: CCSM3.0, CGCM3.1, HadCM3, and GFDL-CM2.0 in 20C3M scenario and ten GCM-RCM NARCCAP models: CRCM-CCSM, WRFG-CCSM, MM5I-CCSM, CRCM-CGCM3, RCM3-CGCM3, WRFG-CGCM3, ECP2-GFDL, HRM3-GFDL, HRM3-HadCM3, and RCM3-GFDL are considered for the multi-model median computation.

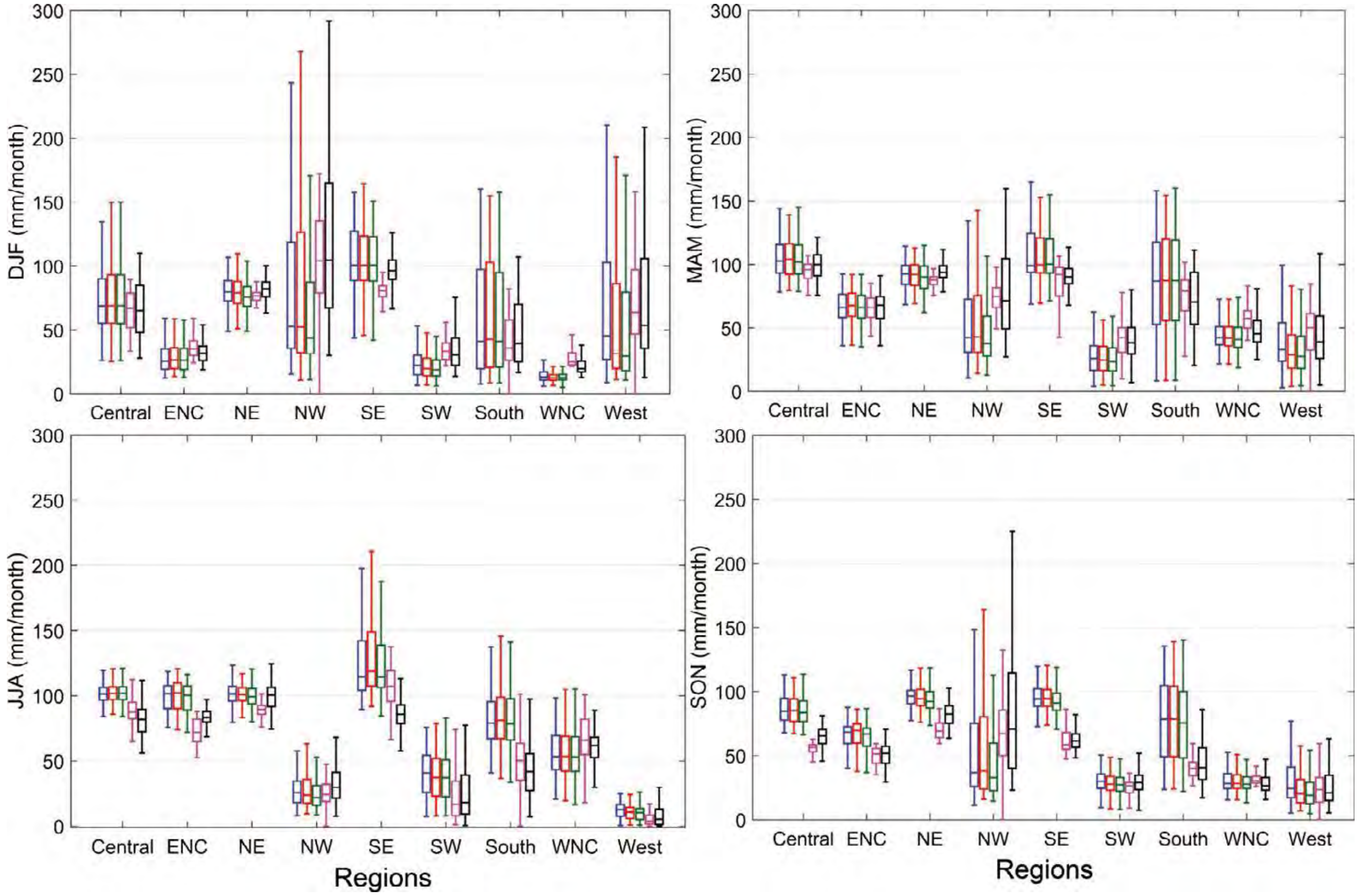

FIGURE 7. Seasonal Distributions of Observed (CRU TS3.22, GPCC v.6, and UDel v. 3.01) and Simulated (interpolated GCM and NARCCAP) Multi-Model Median Climate Model over the Nine Climatological Homogeneous Regions. Seasons included in the analysis are: DJF (December-February), MAM (March-May), JJA (June-August), and SON (September-November). Distributions of CRU, GPCC, and UDel data are shown in blue, red, and green colors, whereas those for GCM and NARCCAP are shown in magenta and black colors, respectively.

over the Southwest is produced by large-scale low pressure frontal systems generated from the upper level mid-latitude and subtropical jet streams, drawing necessary moisture from the Pacific Ocean (Woodhouse, 1997). Simulation of precipitation may be sensitive to the model resolution irrespective of the topographic forcing as shown by Giorgi and Marinucci (1996). In their experiments, the precipitation amount tended to increase at finer resolutions. Greater topographic factors at higher resolution further strengthen this effect. For JJA (June-August), in general, both MME median GCM-RCM pair and the host GCM underestimate seasonal precipitation trends except in the Northwest and West-North Central regions. In the Northwest, MME median RCM overestimates precipitation variability relative to the MME median host GCM and the observations. Conversely, in the West-North Central region, the MME median GCM simulates highest median precipitation. Neither the NARCCAP nor their host GCM could simulate the signature of the North American monsoon (NAM) over the Southwest. The inability to simulate precipitation in the Southwest is primarily due to the issue with downscaled simulation for the region (Wang *et al.*, 2009; Dominguez *et al.*, 2012) due to its complex topographical features. As noted by Bukovsky *et al.* (2013), the dry bias in RCMs, especially over Arizona is potentially due to inability of RCMs to develop low level onshore flow as well as Gulf of California low-level jet during monsoon season, needed for transporting necessary moisture for precipitation in the region, which causes very low precipitation amount during JJA months over the Southwest. Moreover, in their study, they found that the GFDL model lacked the skill of providing adequate boundary conditions for RCM to simulate summer precipitation climatology over the Southwest. These features and regional terrains are not well simulated by most of the GCMs as shown by earlier studies (Collier and Zhang, 2007; Lee *et al.*, 2007). For SON (September-November) season, MME median GCM largely underestimates seasonal precipitation followed by the MME median NARCCAP ensemble over Central, East-North Central, Northeast, Southeast, and South regions. Over Southwest and West seasonal precipitation is reasonably well simulated by the MME median NARCCAP GCM-RCM pair. Except, JJA in all seasons, variability or the "spread" of NARCCAP RCM is found to be much higher as compared to the host GCM and the observations. The MME median NARCCAP ensembles show modest skill in simulating winter precipitation over Southwest as compared to its host GCM, however, MAM is the season when the model overestimates observed precipitation amount the most while SON is closest to the observations.

Robustness of NCEP Forced NARCCAP RCMs in Simulating Regional Drought Statistics

In this section, we illustrate the quantitative evaluation of the Phase I NARCCAP runs with NCEP boundary conditions relative to multiple observational datasets during the 1980-2003 period. Since the Phase I simulations directly utilize "perfect" boundary forcing, therefore it is expected that the regional atmospheric model results, such as precipitation can be deterministically compared with the observations. We evaluate the robustness of NCEP-driven NARCCAP models with those statistical metrics that are time-varying. In this category, we include simultaneous spatial intercomparison between models and the observations using Taylor diagrams, trends, simulations of regional drought properties, and temporal variability in drought area over the nine regions.

Simulation of SPI Statistics and Regional Drought Trends. To make a simultaneous intercomparison of the spatial pattern between models and observation at regional levels, we employed Taylor diagrams. Taylor diagrams provide the concise statistical summary of how well patterns match each other in terms of their correlation, root-mean-square difference and the ratio of their variances. Figure 8 shows the Taylor diagrams for individual NCEP-driven NARCCAP models and their multi-model ensembles for regional median SPI-6 time series. Each model is compared with respect to GPCC v.6 data using centered root-mean-square-difference (RMSD), Pearson's correlation coefficient and standard deviation (SD). To investigate observational data uncertainty, CRU TS3.22 and UDel v. 3.01 are also compared to GPCC v.6 and plotted on the same figure. The observation is shown on the x-axis of the figure as a reference point. The distance from this reference point from the origin is proportional to the standard deviation of the spatial pattern for each region. Standard deviation contours from the origin are shown in black. Contours showing the RMS differences between the NARCCAP ensembles and the observation are shown in green. Model results are then plotted in the azimuthal position based on the centered RMS difference and correlation with (reference) observation and shown in blue contours, representing the spatial correlation between the models and the observation.

In examining these figures, we note the agreement between observational datasets is high in most of the subregions. In general, the agreement between CRU TS3.22 and the reference observation data (GPCC v.6) is higher as compared to between UDel v.3.01 and GPCC v.6 data. The spatial agreement between CRU TS3.22 and GPCC v.6 is highest over the Central region with RMSD error of 0.064 and pattern correlation of 0.9954. However, over the Northwest and West both CRU TS3.22 and UDel v.3.01 are in close agreement with each other. The agreement between UDel v.3.01 and GPCC v.6 is low over the Northeast and East-North Central regions, which may be attributed to local differences within regions across the gridded datasets. Consistent with previous findings, we find multi-model ensembles perform relatively better as compared to single individual models. For example, over the Central region, the standard deviation of MME maximum is close to GPCC v.6 with a centered RMSD of 0.55 and pattern correlation of 0.68 (statistical significance level at 5% level), whereas MME median achieves a correlation with GPCC v.6 of 0.70 with RMSD of 0.60. The pattern correlation statistics for MME median ranges from 0.27 (Southwest) to 0.58 (West-North Central). On an individual model basis, the two RCMs that use spectral nudging, the CRCM and the ECP2 perform better in all subregions. The spatial pattern correlations of CRCM vary between 0.35 (Northeast) and 0.66 (West-North Central). All models exhibit the weakest spatial correlation over the Northeast (ranges between 0.15 and 0.35) and highest over the West (ranges between 0.55 and 0.69). Further, except Central and East-North Central regions, all models underestimate the spatial variance over the rest of the regions. Over the East-North Central, the MME median and MME minimum overestimate spatial variance. On the other hand, over the Central region, all models including their multi-model ensembles overestimate spatial variance of the SPI field with the highest deviation noted by the MME median followed by the MME minimum and the CRCM. There is little spread among models in simulating regional SPI statistics over the West. Among individual models, HRM3 performs poorly over the Central, East-North Central, Northeast, West, West-North Central,

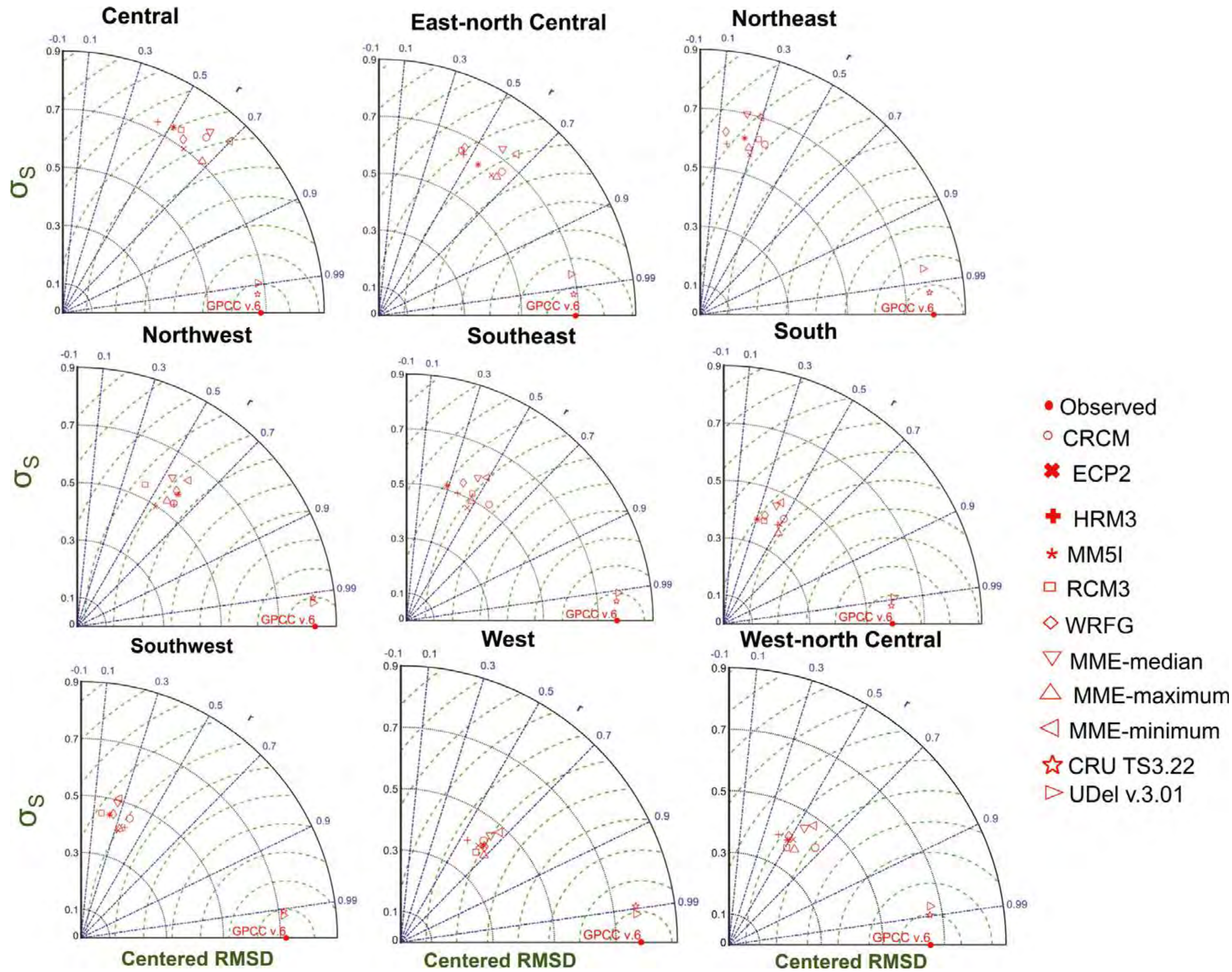

FIGURE 8. Taylor Diagrams of Observed *vs.* Individual NARCCAP RCMs and Their Multi-Model Ensembles in Simulating Spatial Patterns of SPI Time Series over the Nine Regions. NCEP runs are driven by NCEP-2 reanalysis and time period for the analysis is 1980-2003. Averaged spatial variation (σ_S) is shown by the radial distance from the origin, which is proportional to spatial standard deviation normalized by observations. The spatial correlation (r) between observation and models are shown in azimuthal direction.

and Southeast regions; likewise RCM3 over the Northwest and MM5I over the South and Southeast regions.

Next, we analyze SPI trends simulated by models against observations during 1980-2003. The ability of RCMs to reproduce observed drought trends or patterns may be considered a necessary (but not sufficient) condition for these RCMs to credibly simulate signatures of anthropogenic climate changes under future emission scenarios. However, anticipated changes in radiative forcing imply that historical skills may not necessarily be adequate to infer future performance. At the very least, the agreement among multiple model ensembles, or model consensus, needs to be considered. Furthermore, the inability of RCMs to reproduce observed drought patterns may not necessarily be the failure of the RCMs exclusively, rather they may be inherited from GCMs forcing. Thus, prior literature (Seager *et al.*, 2009; Schiermeier, 2013; Ault *et al.*, 2014) suggests that GCM projections may not always be able to simulate historical mega-drought events, which in turn are related to large-scale dynamical patterns. Nevertheless, over this short period of time, the results of trend estimates may be uncertain due to a number of factors (e.g., large-scale climatic oscillations, intrinsic climate variability, and anthropogenic changes) as noted by Bukovsky (2012). This can provide predictive insight about models' credibility to simulate regional climate in the projected time period (Giorgi *et al.*, 2004).

Moreover, analyzing trends is helpful in identifying causes and characteristics of the model bias. Figure 9 shows the spatial pattern of slopes (expressed in mm month^{-1}decade^{-1}) in the observations (CRU TS3.22, GPCC v.6, and UDel v.3.01), and simulations from six RCMs (CRCM, ECP2, HRM3, MM5I, RCM3, and WRFG) along with their multi-model ensembles (MME minimum, MME median, and MME maximum) driven by NCEP boundary conditions. The slope of SPI at each grid point is computed using nonparametric Theil-Sen estimator. Statistical significance of trend is examined using Mann-Kendall test statistics at 5% significance level. Grid points with significant trends are marked with asterisks. We notice from Figure 9 that the two observation datasets — CRU TS3.22 and GPCC v.6 — are in close agreement with each other in simulating SPI trends; however, the trends in UDel v.3.01 are noisier as compared to the other two datasets especially over the Northwest. Further, the UDel v.3.01 dataset overestimates drying trends in East-North Central (around 51% grid points show significant drying against ~23% grid points in CRU TS3.22 and 20.2% grid points in GPCC v.6) and Central (around 25.6% grid points show significant drying pattern against around 10% grid points in CRU TS3.22 and around 7% grid points in GPCC v.6) regions, and shows a wet trend in part of the Northwest region (around 2% grid points with significant upward/wet pattern while other two data show overall significant drying pattern). The disagreement in the pattern of trends in UDel is mainly due to the differences in the interpolation methodology used in the dataset and the number of station observations used as noted in the earlier studies (Nickl *et al.*, 2010; Trenberth *et al.*, 2014).

Both CRU and GPCC datasets show presence of a significant drying (negative slope of SPI) trend in the observed SPI time series over most of North America including West, Northwest, part of Rocky Mountain, and Midwest regions. Conversely, few regions also exhibit wet or positive trends, which include the Northeast, West-North Central, part of the Southeast (coastal Gulf Coast and interior regions, such as Alabama, Georgia, North and South Carolina) and South (Northeast of Texas) regions. The two observed datasets (CRU and GPCC) show statistically significant drying and wet trend over around 70-76% and 24-30% of the domain, respectively. Among models, the MME median performs best and simulates a 72% drying and 29% wet trend (significant). However, it fails to capture a significant drying trend over Rocky Mountain regions in the Southwest. Among individual RCMs, as a whole MM5I performs the best and simulates a 76% drying and 24% wet trend (significant) akin to CRU dataset. The other RCMs, the CRCM and RCM3 exhibit a widespread significant wet pattern (over around 61% and 74% of the

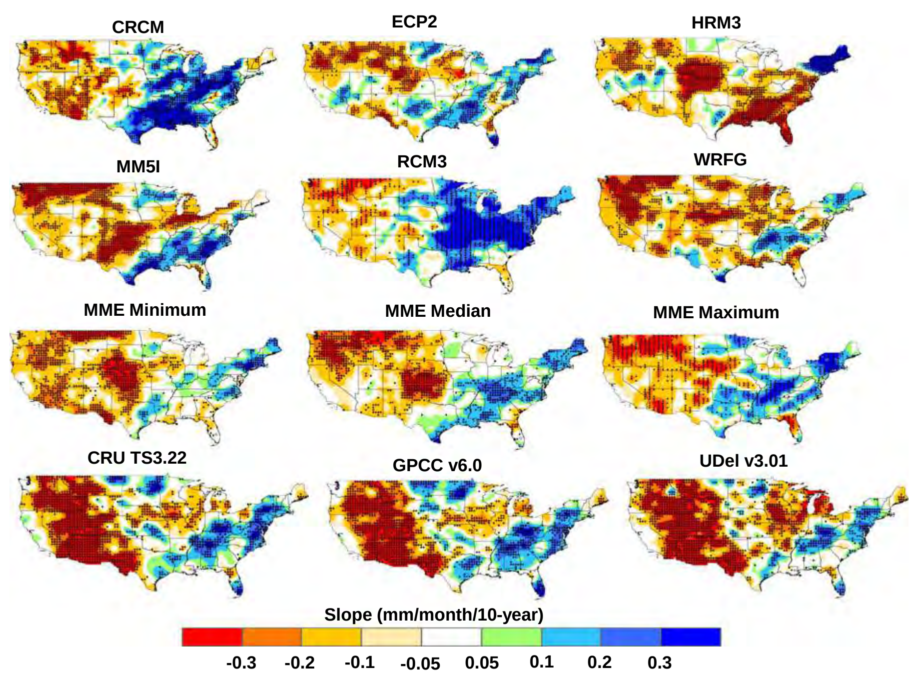

FIGURE 9. Spatial Distributions of Slope (mm-month^{-1}-decade^{-1}) of the Trends in SPI Time Series during 1980-2003 from Six NCEP-Driven NARCCAP RCMs, CRU v. 3.22, GPCC v. 6, and UDel v.3.01 Datasets. The slopes are estimated with the Theil-Sen estimator, which is robust against outliers. Stippling indicates trends are statistically significant at 5% significance level obtained using non-parametric Mann-Kendall two sided test. Drought is modeled using SPI at time scale of six months (SPI-6).

domain, respectively); the HRM3 and WRFG show a significant drying pattern (over around 87% of the domain in both models). Most of the models simulate a drying trend over the Northwest. However, none of them simulates widespread significant drying trends over the West and Southwest. While GPCC v.6, CRU TS3.22, and UDel v.3.01 all indicate a statistically significant drying over 72-79% of the domain in the Southwest and 44-46% of the domain in the West, most of the models cover only 14% (i.e., ECP2) to 39% (i.e., CRCM) of the domain in the Southwest and 1.4% (i.e., ECP2) to 34% (i.e., multi-model minimum) of the domain in the West.

Simulation of Regional Drought Properties. Figure 10 shows the spatial distributions of maximum drought severity. The spatial maps reveal wide variations among the models, for instance, HRM3 and WRFG overestimate drying trends over the Northeast and Southeast, respectively. Among multi-model ensembles, MME median performs best in reproducing a drying trend over the Southwest; however, overestimates drying over the Central and East-North Central regions. Moreover, most of the models fail to simulate maximum drought properties over the Great Plains and the Rocky Mountain States satisfactorily. The spatial pattern of maximum drought duration is similar to those of severity (hence, not shown here). For quantitative evaluations of the model skill, we present heat maps of pattern correlation analysis of maximum drought severity and duration in Figure 11. GPCC v.6 is chosen over the other two datasets as a baseline for comparison because it is not affected by wet bias unlike CRU TS3.22 and contains relatively smoother trend field unlike UDel v.3.01. To investigate the relative agreement between different observational data, CRU TS3.22 and UDel v.3.01 are also compared in the heat maps. Pattern correlation statistics are analyzed using non-parametric rank-based Spearman's correlation (ρ), which is robust against outliers. As observed from Figure 11, weak positive correlation (ranges from 0.2 to 0.4) exist between CRU TS3.22 and GPCC v.6 data over the Northeast region. The correlations

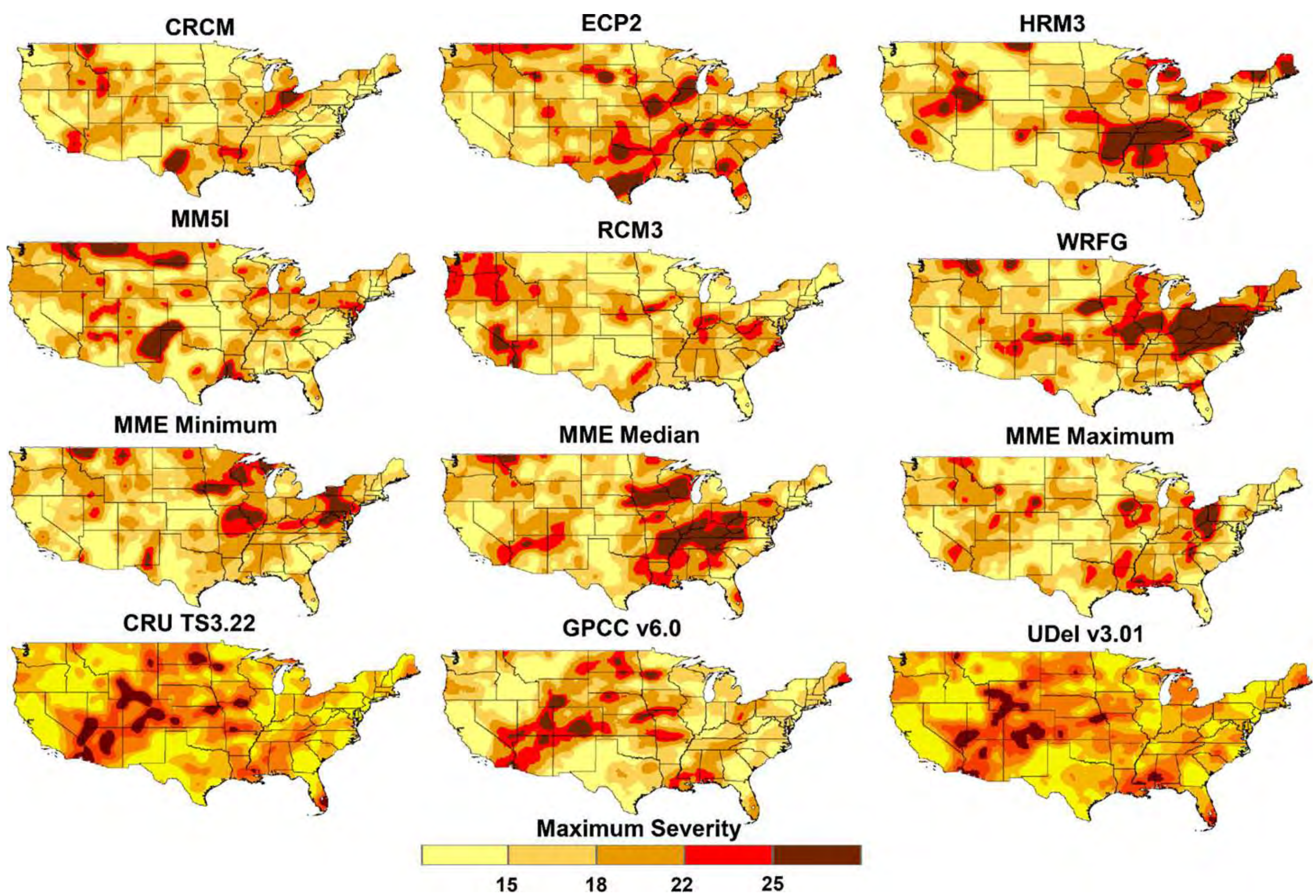

FIGURE 10. Maps of Maximum Drought Severity in NCEP-Driven NARCCAP Simulations and Observation during 1980-2003.

between GPCC and the other two datasets are strong (more than 0.6) over Central, West, South, and West-North Central regions. In simulating maximum drought severity, MME median NARCCAP RCM performs best as compared to the individual RCMs. Among individual models, RCM3 performed best over the West and showed highest pattern correlation ranges between 0.6 and 0.8 (Figure 11). Many models show a modest positive rank correlation (between 0.2 and 0.4) with reference observation (GPCC v.6); for example, CRCM over the West, Southwest; HRM3 over the South; RCM3 over the Southwest; and WRFG over the West. In contrast, many of the models also exhibit negative correlation coefficients, such as CRCM and MME maximum over the Northwest, ECP2 over the South and Southeast, and RCM3 and MME median over the Northeast. None of the models including their multi-model ensembles satisfactorily simulates maximum drought severity and duration over the Central, East-North Central, Northwest, and West-North Central regions, respectively. In simulating maximum drought duration, we note a modest positive correlation (ρ = 0.2-0.4) between models and the GPCC in a few regions; such as, CRCM over the West and Southwest, HRM3 over the South, MM5I over the Northeast, RCM3 and MME median over the West and Southwest, MME maximum over the West, respectively. We also note the presence of negative correlation between RCM3 and GPCC over the Southeast, Northwest, and Northeast regions.

In general, no single model stands out as superior as compared to its peers, hence we employ regional bias plots to assess model performance over the regions as a whole. We compute median drought severity and absolute bias (absolute difference between observed and modeled drought severity) relative to GPCC data at individual grid points over the nine regions. Figure 12 shows the relation between regional median drought severity and associated bias. The median drought severity is found to be highest over the Southeast followed by the Northwest during the analysis period (1980-2003). The uncertainties in the median biases over different regions are small and in the ranges between 0.77 and 1.36; however, considering individual grid points the model bias is found to be highest over the Southeast (with highest magnitude 8.0, a location in North Carolina) and lowest over the West (with highest magnitude 5.2, a location in Nevada). The correlation between observed (GPCC) regional median severity and the absolute model bias is found to be positive and statistically significant at 1% significance level with Kendall's τ dependence 0.67 (Figure 12; right). This implies with an increase in severity the model bias grows higher, which indicates model performance drops in simulating extreme drought statistics. The relation between observed *vs.* modeled median drought severity is found to be negative; however statistically insignificant.

Next, we examine the performance of NCEP-driven RCMs in simulating the observed drought frequency. Figure 13 shows the distribution of drought frequency in the nine regions. Among observations, the agreement between GPCC v.6 and CRU TS3.22 data is high, with the exception of the Northwest region. Overall, the spread of UDel v3.01 data is high as compared to the other two datasets except in the Northwest, South, and West regions. Observational datasets suggest highest average drought frequency over the West (no. of droughts per 24 years: 20) and

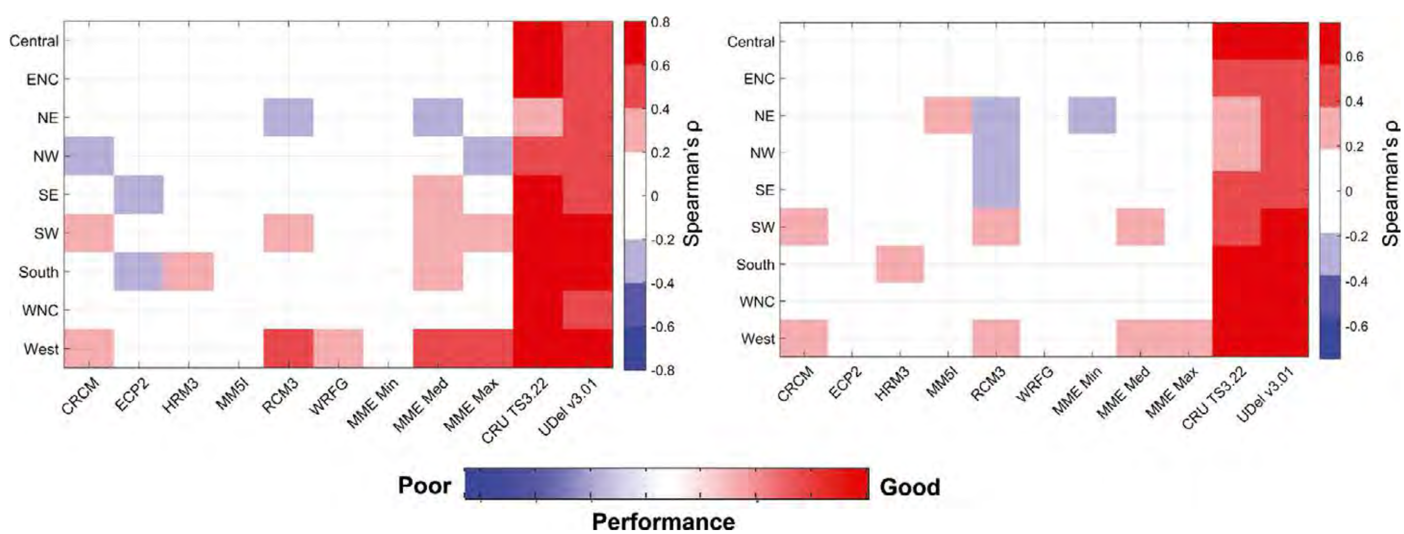

FIGURE 11. Heat Maps of Regional Pattern Correlation Statistics for Maximum Drought (left) Severity and (right) Duration. GPCC v.6 is chosen as reference observation.

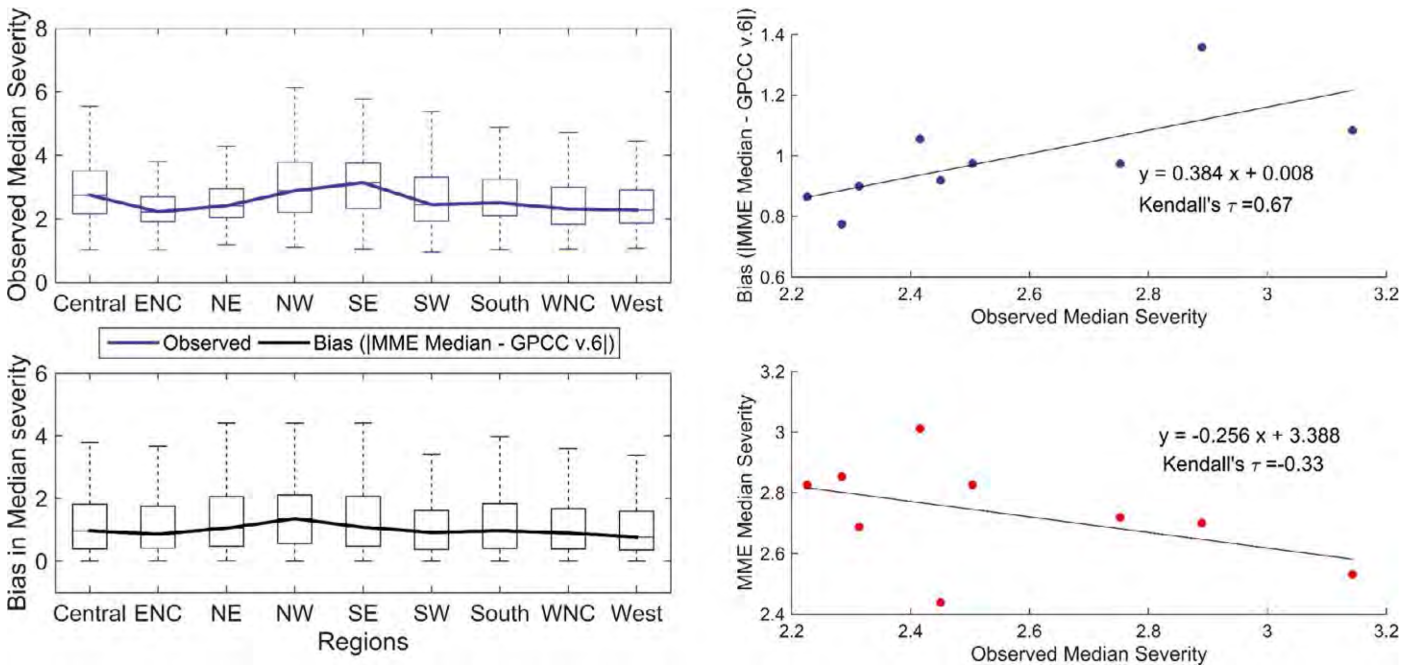

FIGURE 12. Relation between Observed *vs.* Absolute Bias (difference between multi-model median drought severity and observed drought severity) of Median Drought Severity Averaged over Nine Regions in NCEP-Driven NARCCAP Simulations (left); Scatter Plots of (top right) Regional Bias *vs.* Observed (bottom right), and Modeled *vs.* Observed Median Severity.

lowest over the Southwest (no. of droughts per 24 years: 16). Among the models, the individual RCMs including MME median NARCCAP underestimate median drought frequency over all regions with respect to observational datasets, however, they satisfactorily simulate regional spread. Few regional exceptions exist. For example, ECP2 in the Northeast, RCM3 in the East- and West-North Central, MM5I in the Central and Southeast, and WRFG in the Southeast overestimate average drought frequency. Further, in the Southwest all models including their multi-model ensembles overestimate the observed drought frequency. The overestimation of drought frequency by the model ensembles concurrent with the inability to capture the significant drying trends in SPI time series in the Southwest (Figure 9) suggests that RCMs may generate more frequent meteorological droughts compared to observations but fail to capture the intensity of the event. Over the Northwest, the performance of the two models with spectral nudging, the CRCM and ECP2 is comparable to that of GPCC in simulating the median drought frequency; however, extremes are not adequately simulated by the models as indicated by the relatively short length of the whiskers as compared to the GPCC datasets.

In the following sections, we evaluate the skills of NCEP forced NARCCAP models in simulating spatial extent of drought. Figure 14 shows temporal variability of average (median) PAUD for the nine regions. The maximum PAUD is found to be ~81% during 2002 in the Southwest region, followed by ~68% during 1985 in the East-North Central regions, respectively. The mean PAUD time series is reasonably well simulated by the MME median NARCCAP ensembles driven by the NCEP boundary condition. However, a few regional exceptions exist in some of the years. For example, NARCCAP RCMs overestimate PAUD over the Southern U.S. during 1986 and underestimate over the Southwest, Northwest, and West-North Central regions during 2002. In particular, NCEP-driven RCMs overestimate variability of PAUD time series over the West in most of the years. The box plots in Figure 14 show interannual variability of the models, which further confirms the discrepancy between observed and model-simulated PAUD time series.

Next, we evaluate the robustness of RCMs in simulating persistence in the PAUD time series. This helps to identify the inconsistency between RCMs and multiple observations in simulating spatial extent and timings of drought. Persistence in the hydrologic event results from the presence of memory in the system, such as prolonged duration of a drought event. A high frequency of drought often results from low persistence in the hydrologic system, which in turn links to low autocorrelation in the drought time series, both at spatial and temporal scale levels (Tallaksen and Stahl, 2014). Hence, we developed an autocorrelation function (ACF;

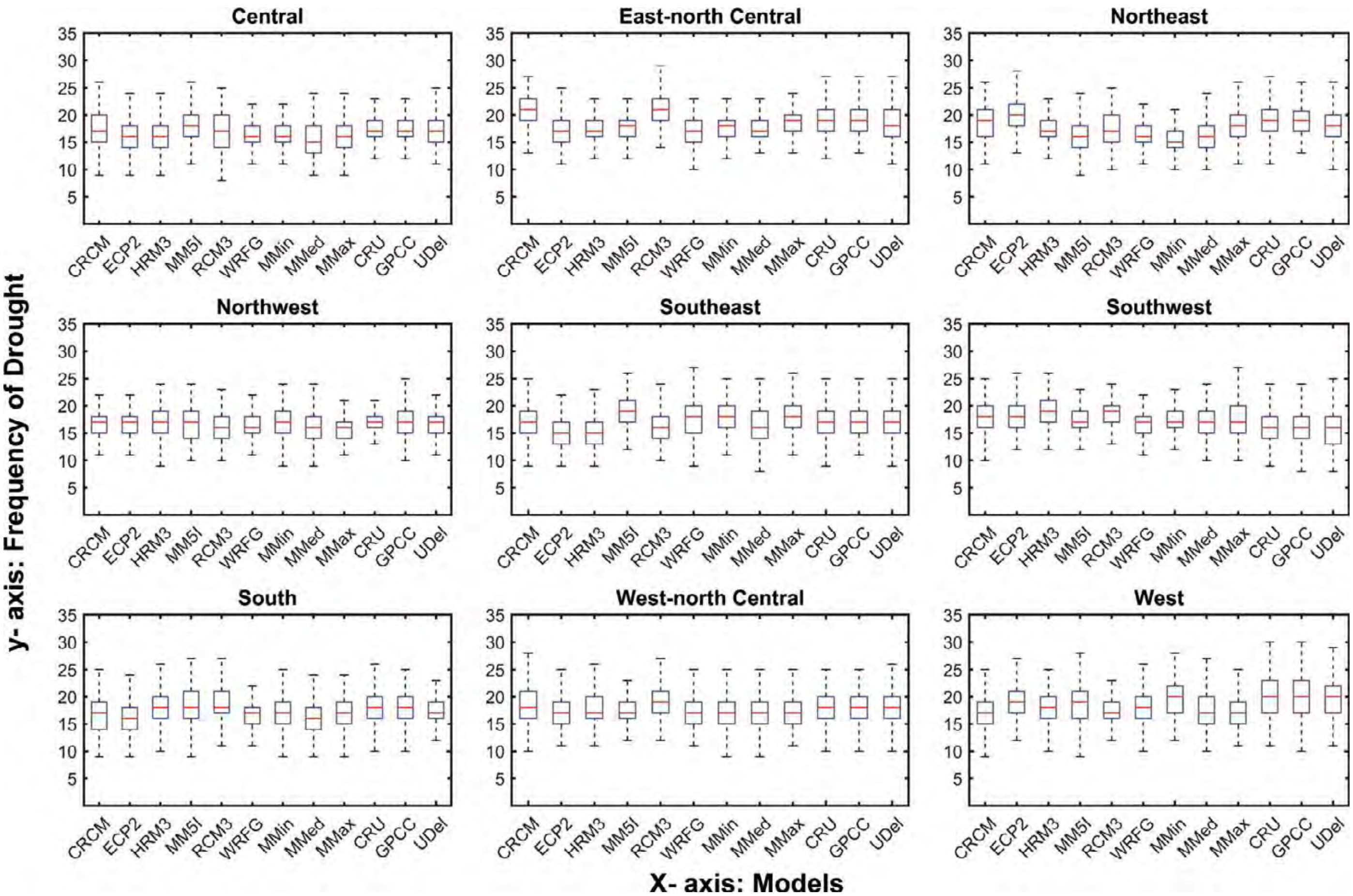

FIGURE 13. Box Plots of Drought Frequency in Individual NCEP-Driven NARCCAP Models, Their Multi-Model Ensembles, and Multiple Observational Datasets (CRU TS3.22, GPCC v.6, and UDel v.3.01).

Figure 15) for the PAUD time series up to 1 year lags (i = 1, 2, ..., 12 months) for each of the nine regions. The drop in the ACFs at lag-6 is an inherent property of the data because the SPI time series is computed at six month accumulation running window. The ACF plots show a declining autocorrelation pattern with increasing time lags, but the nature differs regionally and among multiple datasets. The temporal variability of annual PAUD time series shows high-frequency variability characterized by low autocorrelation (i.e., less persistence) over the West. On the other hand, we observe relatively low-frequency variability and consequently high persistence in drought area over the Southwest and West-North Central regions, respectively. This implies the likelihood of hydrological droughts in these regions. Our finding is consistent with the earlier study on the Southwest drought (Cayan *et al.*, 2010). The early 21st Century drought over the Southwest started during 2000 with exceptionally warm temperature and low precipitation (30th percentile or below) over the interior Pacific Coast, which again spread over Colorado, Utah, Arizona, and Southern Nevada by 2002, with monthly precipitation percentiles dropped to 20th percentile or below (Cayan *et al.*, 2010). Further, our analysis in the previous few paragraphs suggests during 1980-2003, Southwest is characterized by lowest average drought frequency as compared to the other regions. This implies evidence of spatially (i.e., drought affected area) and temporally (i.e., longer duration) persistent hydrological droughts over the Southwest. Except for the Southwest and West-North Central regions, the regional ACFs are well simulated by the RCMs and their multi-model ensembles.

Robustness of GCM Forced NARCCAP RCMs in Simulating Regional Drought Statistics

In this section, we evaluate robustness of NARCCAP RCMs driven by GCM boundary conditions in simulating statistical properties of droughts that are temporally independent, as GCMs do not contain the

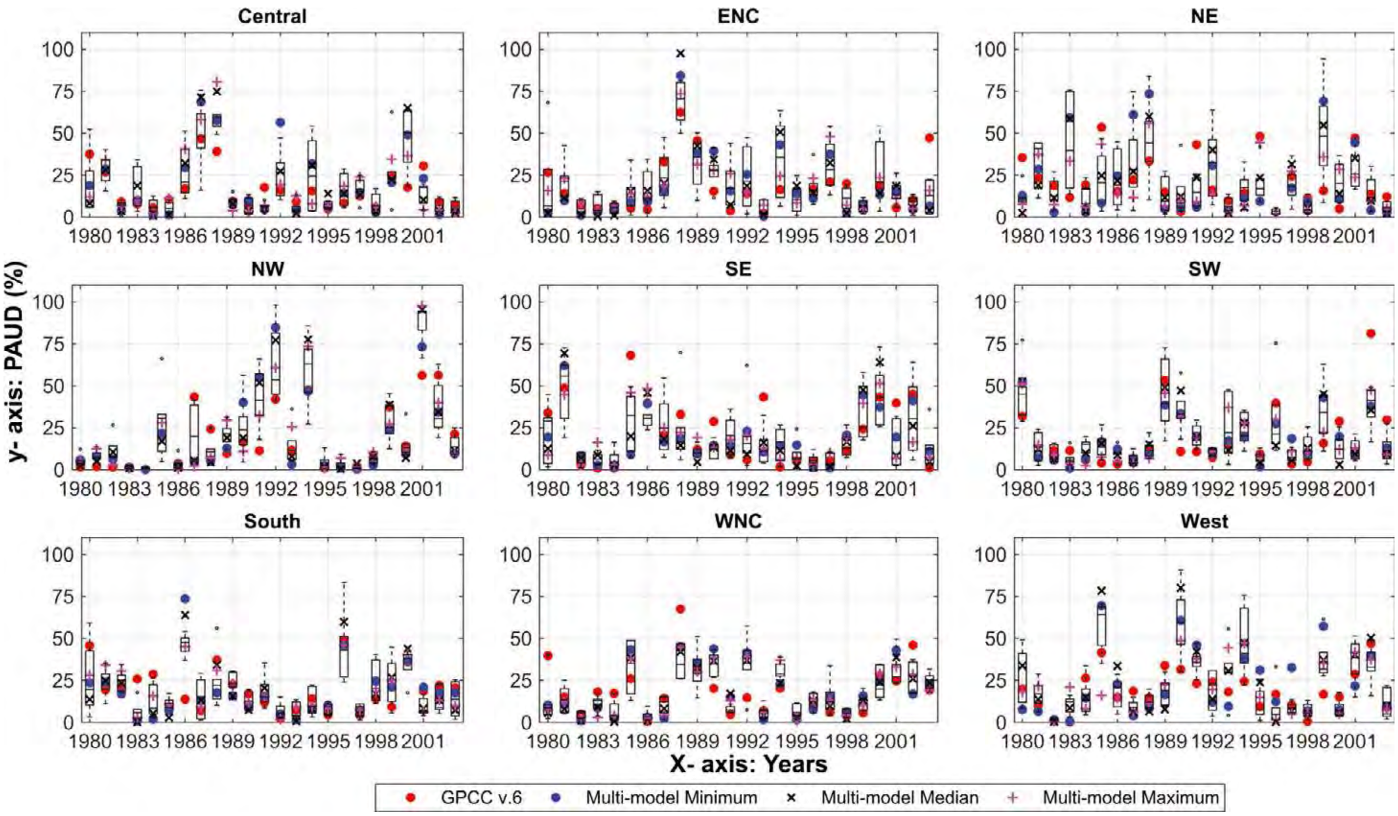

FIGURE 14. Time Series of Percentage Area under Drought, PAUD (regional median) in Observation (GPCC v.6) and Models. The simulation is performed from the set of six NCEP-driven RCMs over the period 1980-2003. Box plots show distributions of PAUD from individual RCMs.

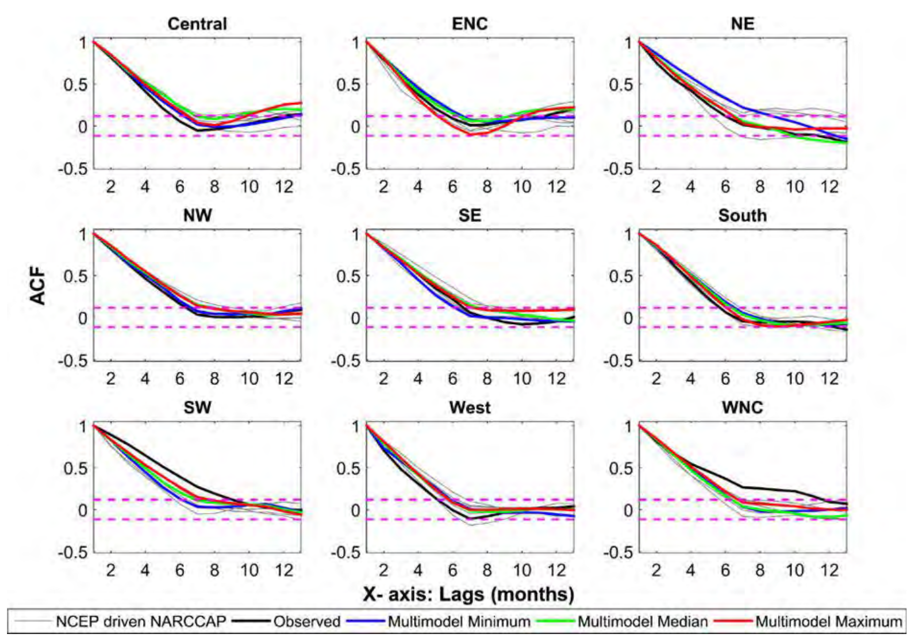

FIGURE 15. Autocorrelation Function (ACF) of PAUD (up to 12 months lag) in NCEP-Driven Individual NARCCAP Simulations and Their Multi-Model Ensembles over the Nine Regions During 1980-2003. The two horizontal lines indicate lower and upper confidence bounds at 5% significance level.

same sequence of sea surface temperature variability and associated signals at same temporal phasing as that in the observations. In this category, we examine frequency and persistence as the statistical metrics. The time slice for comparison is from 1971 to 1999, the span of maximum GCM-forced NARCCAP data availability.

Simulation of Drought Frequency: 1971-1999. We identify the meteorological drought episodes from historical SPI time series and compare the regional drought frequency in observations and in GCM-forced NARCCAP RCMs (Figure 16). We calculate the number of drought events and their durations at each grid and present spatial distribution of drought frequency for the nine regions. Figure 16 shows discrepancy among the observations in simulating the regional drought frequency. Both CRU TS3.22 and UDel v.3.01 underestimate drought frequency as compared to the GPCC v.6. The disparity between GPCC v.6 and CRU TS3.22 datasets is clearly due to the markedly different (high) precipitation value simulated by CRU TS3.22 dataset (Figure 2). Further, the discrepancy in UDel v.3.01 data relative to the other two datasets is attributed to the differences in interpolation methodologies used in the dataset as discussed earlier. The average (median) observed drought frequency is highest over the West and the Northeast regions, whereas the frequency is observed over the Southwest. The intercomparison of individual observational datasets reveals substantial numerical differences in simulating drought frequency; however, the overall pattern is identical across all datasets. The MME-median GCM-RCM pair overestimates observed drought frequency over most of the regions; however, it underestimates drought frequency over the Northeast and the Southeast relative to GPCC data (Figure 16). Likewise in most of the regions, individual NARCCAP members overestimate the observed drought frequency.

Simulation of Drought Persistence. The persistence in the SPI time series is analyzed using the Hurst index. Figure 17 compares the distribution of the Hurst index and the observed median severity in the three observational datasets. Figure 17 shows wide variations among the observations in simulating regional SPI persistence. In general, UDel v.3.01 followed by CRU TS3.22 overestimates regional distribution of Hurst index in all regions relative to GPCC v.6 data. In Northeast and Southwest, UDel v.3.01 overestimates average drought severity, whereas it underestimates in the West relative to other two datasets. While comparing the relation between persistence in SPI time series and average drought severity, we find the Southwest region is characterized by the high SPI persistence with high median severity, indicating evidence of hydrological drought in this region. In contrast, although the West is characterized by high values of Hurst index in SPI time series, the average severity is relatively less as compared to the other regions. The persistence in SPI

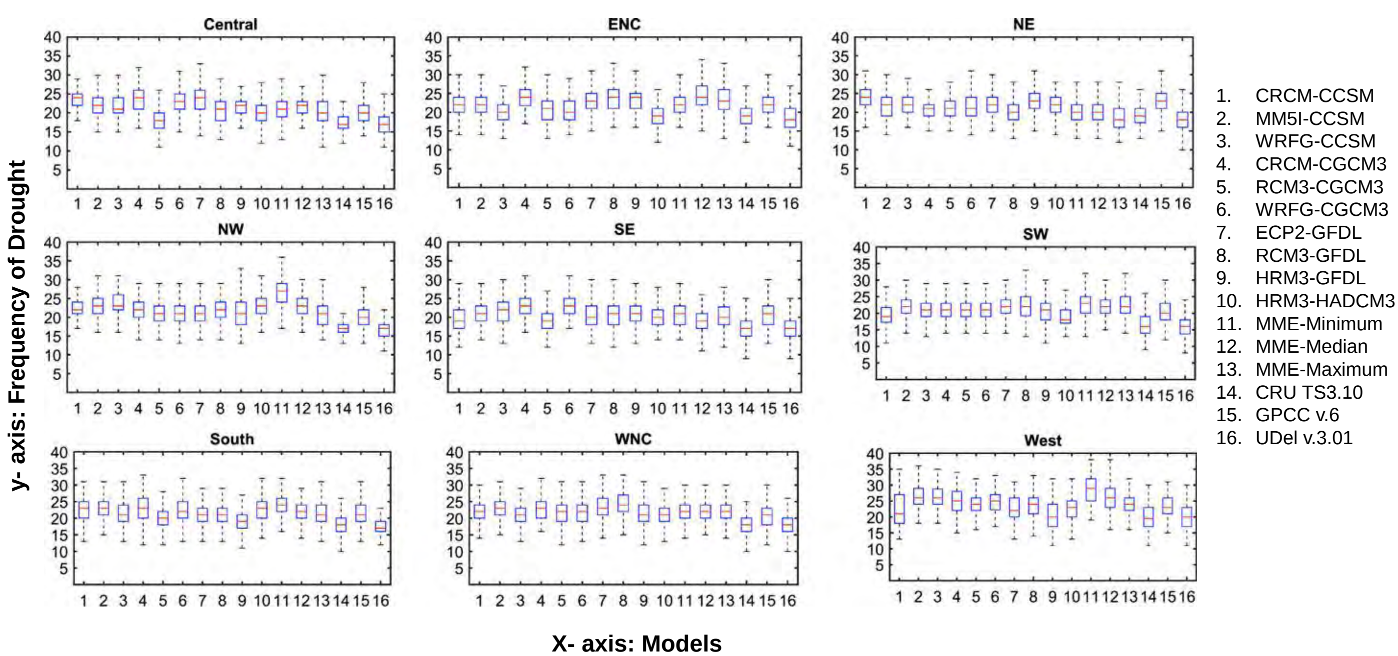

FIGURE 16. Box Plots of Drought Frequency in Individual GCM-Driven NARCCAP Models, Their Multi-Model Ensembles, and Multiple Observational Datasets (CRU TS3.22, GPCC v.6, and UDel v.3.01).

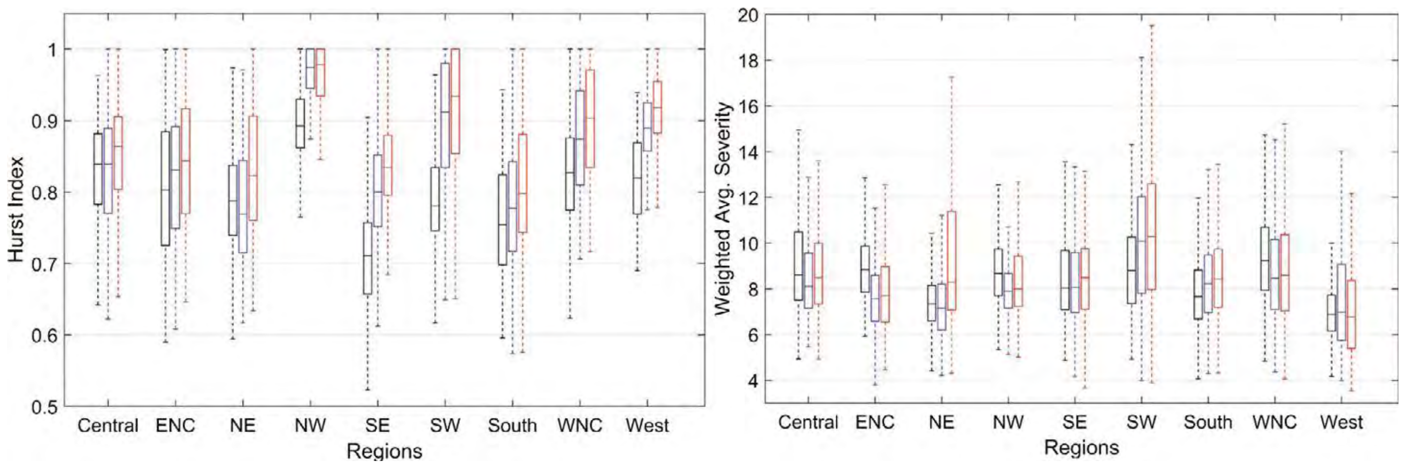

FIGURE 17. Box Plots of (left) Hurst Exponent of SPI Time Series and (right) Weighted Average Severity of Drought in GPCC v.6 (black), CRU TS3.22 (blue), and UDel v.3.01 (red) Datasets.

time series over the Southeast appears to be least using GPCC v.6 data. Barring few exceptions (such as, between the years 2005-2007 and 1986-1987), in general drought persistence in the Southeast is relatively rare as compared to the other regions of the U.S. as shown previously (Mo and Schemm, 2008a, b; Ford and Labosier, 2014). Further, average severity of drought over the Northeast appears to be less severe and characterized by low values of persistence in the SPI time series.

We compare individual GCM-driven NARCCAP models and their multi-model ensembles in simulating regional SPI persistence with respect to reference observations using pattern correlation analysis. Figure 18 shows a heat map of model performance against GPCC v.6. Pattern correlation analysis suggests multi-model ensembles do not concur well with the observations in simulating regional Hurst index. Among individual models, CRCM-CCSM and WRFG-CGCM3 over the East-North Central, CRCM-CGCM3, RCM3-CGCM3 and ECP2-GFDL over the Southeast, and HRM3-GFDL and HRM3-HadCM3 over Northeast show high correlation value (Spearman's $\rho \geq 0.6$). Over West, spatial patterns of many models in particular CCSM group (CRCM-CCSM, MM5I-CCSM, and WRFG-CCSM) are not in phase with observed Hurst index.

DISCUSSIONS

We evaluate robustness of NARCCAP in two phases: Phase I NARCCAP simulations driven by NCEP boundary conditions compare individual drought events and associated properties with observations while the Phase II simulations forced by GCM boundary conditions test only those statistical metrics of drought that are temporally independent.

To assess how NCEP-driven NARCCAP RCMs are able to simulate regional SPI statistics, we compare the spatial pattern between models and the reference observation using Taylor diagrams. Consistent with previous studies (Arritt, 2008; Bukovsky *et al.*, 2013), our results suggest spatial pattern correlation of models are high over the West and correlation value gradually decreases as we move from west to east. The deterioration of model performance from the west (inflow boundary) to east (outflow boundary) is due to the incorporation of large-scale information in the model solution at lateral boundaries (Arritt, 2008). As pointed out by Arritt (2008), the deterioration of model performance with distance from the inflow boundary has improved to some extent in the models that include time-variant large-scale atmospheric states in the model solution, such as inclusion of spectral nudging. In this context, we find CRCM and ECP2 the two RCMs that include spectral nudging perform reasonably well in simulating regional SPI statistics as compared to the other RCMs. The results of the regional trend analysis suggest MME median RCM simulates regional trends satisfactorily relative to observations; however, it fails to simulate widespread significant drying over the West and Southwest. Most of the RCMs fail to simulate drying trend over the Southwest, which appears to be a problem with downscaled simulation for the region (Wang *et al.*, 2009; Dominguez *et al.*, 2012) due to its complex terrain. The CRCM and RCM3 exhibit excessive wetness over the eastern domain, whereas WRFG shows an overall drying. However, the two RCMs, CRCM and WRFG, are able to reproduce signs of drying (significant) trend over the Southwest. This is

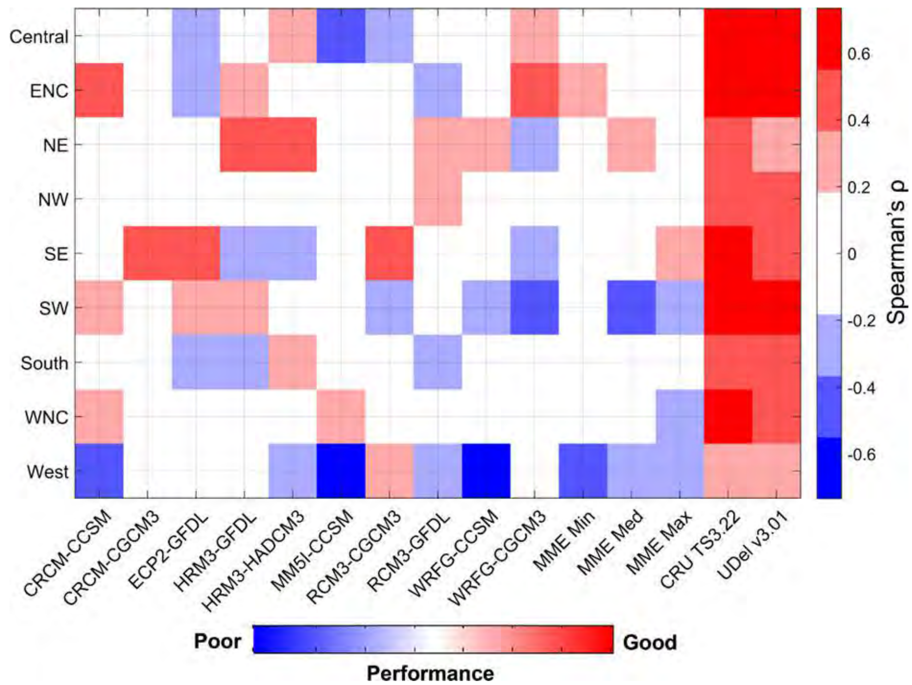

FIGURE 18. Heat Map of Regional Pattern Correlation Statistics for SPI Persistence. GPCC v.6 is chosen as baseline.

because these two models are able to develop onshore low level monsoon flow over the Southwest unlike other RCMs as noted in a previous study (Bukovsky *et al.*, 2013). On the other hand, the excessive wet trend in CRCM over the eastern domain is due to its greater moisture convergence at the near surface level (Bukovsky *et al.*, 2013). The overall drying trend in WRFG is likely the result of weaker orographic lift simulated by the model leading to reduced precipitation over the domain (Liang *et al.*, 2012). In agreement with earlier findings (Bukovsky *et al.*, 2013; Cruz, 2014), our analysis suggests ensemble median NARCCAP RCM forced with GCM boundary conditions fail to simulate signature of NAM satisfactorily. One hypothesis that may be examined by future studies is that the smaller precipitation variability in NARCCAP RCMs during JJA and the underestimation of the NAM summer precipitation is related to RCM warm season convection initiation.

On evaluating maximum drought severity using pattern correlation analysis, we find wide spatial variations among the models. The spatial pattern of maximum observed drought properties (severity and duration) shows a sharp gradient stretching from the Southwest towards the East-North Central regions. In particular, none of the models forced with NCEP boundary conditions simulates this spatial pattern satisfactorily. Some of the models overestimate drought only to a few specific regions. For example, HRM3 and WRFG overestimate drought severity and duration over the Southeast and in the portion of the Northeast and Central regions, respectively. The discrepancy in simulating regional extremes motivates us to investigate further if the mean behavior of drought properties is well simulated by the models. Therefore, we analyze averaged (represented by median as this measure is robust against outliers) severity *vs.* absolute model bias for each region using scatter plot. A statistically significant positive relationship exists between severity *vs.* the model bias. This implies with larger severity, the model skill grows worse. Specifically, with larger severity, the uncertainty in model projections grows higher. The worst plausible case is even higher because the upper bound of the uncertainty is larger for extreme droughts, leading adaptation difficult because of larger variability.

Observed and NCEP-driven NARCCAP simulated multi-model ensembles of PAUD time series show large spatiotemporal variability. NARCCAP models overestimate PAUD time series in the West in most of the years, whereas they underestimate in the Northeast, Southeast, Southwest, and West-North Central regions in some of the years. Further, to investigate the behavior of spatial persistence simulated by the NARCCAP RCMs, we analyze ACF plots of observations and models up to one year time lag. The ACF plots of observation at different lags show evidence of high-frequency variability in PAUD time series over the West and relatively low over the Southwest and West-North Central regions, respectively. The low-frequency variability in PAUD time series corresponds to high persistence in the system, which in turn can be linked to the likelihood of hydrological droughts in these regions (Southwest and West-North Central). The reanalysis-based NARCCAP RCMs fail to simulate high persistence in the PAUD time series satisfactorily. In a modeling framework, spatiotemporal continuity of drought not only depends on the model ability to reproduce mean precipitation throughout the annual cycle, but also on the variability of precipitation to maintain precipitation deficit over a sustained period. The lack of agreement among different models in simulating droughts is due to different parameterizations involved in modeling framework and persistence in the hydrological system is not adequately addressed by the models (Blenkinsop and Fowler, 2007; Wang *et al.*, 2009; Tallaksen and Stahl, 2014).

In Phase II assessment, we evaluate robustness of the GCM-forced NARCCAP ensembles. The assessments include analysis of regional drought frequency and persistence in the SPI time series. Overall models overestimate the regional drought frequency. Discrepancy in simulating regional drought properties, particularly the severe events, is related to the convective parameterization schemes, which are not properly resolved at fine-scale RCM grid cells as shown in many studies (Fowler and Ekström, 2009; Tripathi and Dominguez, 2013). This in turn can be linked to the failure of RCMs to simulate persistently low regional precipitation (Blenkinsop and Fowler, 2007).

The regional trend patterns in observed SPI time series show statistically significant drying trends, especially over the West and Southwest, which further motivates the analysis of drought persistence. The persistence in the time series often leads to underestimation of variance and subsequently overestimation of the statistical significance of trends (Hamed and Ramachandra Rao, 1998; Koutsoyiannis, 2003). Our analysis suggests the MME median GCM-RCM pair does not concur well with observations in simulating the regional Hurst index. One plausible reason for the inability of RCMs to simulate drought persistence is that climatic oscillations may not always be well simulated by climate models. This opens up possibilities for future research in model improvements, both GCMs and RCMs.

One of the potential caveats of our analysis is the relatively short record (25-29 years of monthly time series data) to characterize the space-time nature of drought persistence. Although a few studies (Wang *et al.*, 2009; Tallaksen and Stahl, 2014) have attempted to check performance of LSMs to capture hydrological droughts over the CONUS and Europe, they did not consider temporal scaling properties and focused on the autoregressive nature of the associated time series. Our study analyzes temporal scaling behavior of the meteorological drought index and its spatial coverage in RCM-simulated climate models and compares them with multiple observations to check the credibility of the climate models to simulate temporal and spatial persistence.

CONCLUSIONS

Precise projections of regional drought properties are essential to mitigate the impact of droughts on the water supply system (Shiau and Shen, 2001). Although a limited number of studies (Blenkinsop and Fowler, 2007; Sheffield *et al.*, 2012a, b) have attempted to evaluate robustness of high-resolution climate models in the context of meteorological droughts, the potential of dynamically downscaled NARCCAP models is not fully explored yet (Kerr, 2013). The dynamically downscaled climate variables are assumed to be physically more consistent as compared to their statistically downscaled counterparts (Laprise, 2008). However, the value added by the RCMs is still uncertain (Racherla *et al.*, 2012; Kerr, 2013) due to propagation of systematic biases from coarser resolution global models to regional models (Giorgi and Mearns, 1991). Therefore, it is important to know how well these climate models are able to simulate drought properties consistent with observations and offer predictive insights into drought. Hence, to bridge the existing gap in current understanding of NARCCAP RCMs in the context of drought, we evaluate the robustness of NARCCAP RCMs in simulating meteorological droughts over the CONUS against multiple observational datasets. We analyze meteorological drought as it often translates into potentially damaging other drought categories (Wilhite *et al.*, 2014). Meteorological drought is quantified using the SPI (SPI-6) due to its multi-scalar nature and its ability to capture seasonality, as opposed to the PDSI and soil moisture-based indices. Our primary and secondary findings are as follows:

Robustness analysis of Phase I NARCCAP simulations (1980-2003) *vs.* multiple observations yields following set of insights:

1. In general, multi-model median NARCCAP RCM outperforms individual models in simulating regional drought statistics, such as median SPI, associated trends, the maximum drought properties, and mean PAUD. However, a few regional exceptions exist, where extremes are

over/underestimated by the models. For example, models fail to simulate widespread drying trend over the Southwest and the West.
2. Among individual models, the RCMs with spectral nudging techniques, the CRCM and the ECP2 simulate regional SPI time series satisfactorily. Further, we note that the HRM3 and the RCM3 are in good agreement with observations to simulate regional maximum drought properties.
3. A statistically significant positive correlation between the model bias and the observed drought severity indicates with larger severity, the model skill grows worse. This implies with larger severity, the uncertainty in model projections grow higher. The worst plausible case is even higher because the upper bound of uncertainty is larger on the most severe cases that render adaptation difficult because of larger variability.
4. Overall, MME median NARCCAP ensembles underestimate median drought frequency with the exception of the Southwest. In the Southwest, all models including their multi-model ensembles overestimate observed drought frequency.
5. Over the Southwest, the overestimation of drought frequency by the model ensembles concurrent with the inability to capture the widespread drying trend suggests RCMs may generate more frequent meteorological droughts compared to observations but are unable to capture the intensity of the events.
6. The regional ACF plots suggest persistent droughts over the Southwest and West-North Central is underestimated by the RCMs. In addition, none of the models simulates widespread drying trend over the Southwest.

Robustness analysis of Phase II NARCCAP RCMs *vs.* observations suggests the following insights:

1. Individual NARCCAP RCMs and their multi-model ensembles overestimate the regional drought frequency over most of the regions.
2. Multi-model ensembles do not concur well with observations in simulating the regional Hurst Index; however, few individual models, for example, models with CGCM3 as the boundary conditions perform reasonably well.

Analysis of observational data across multiple sources, suggests the following insights:

1. Differences exist among three datasets for simulating annual average precipitation: CRU TS 3.22 shows the tendency toward wetter trend and underestimates variance especially over the Southwest.
2. Both CRU TS 3.22 and GPCC v.6 are in good agreement with each other in simulating regional trends in SPI. In contrast, UDel v.3.01 is relatively noisier and overestimates the drying trend over the East-North Central and the Central regions.
3. Simulation of the regional drought during 1971-1999 shows both CRU TS 3.22 and UDel v.3.01 datasets underestimate drought frequency as compared to the GPCC v.6. On the other hand, the agreement between CRU TS3.22 and GPCC v.6 data is found to be high during 1980-2003 in simulating the regional drought frequency.
4. UDel v.3.01 followed by CRU TS 3.22 overestimates regional distribution of the Hurst Index as compared to the GPCC v.6 dataset.
5. Analysis of drought persistence in PAUD time series reveals the Southwest and the West-North Central regions have higher drought persistence, whereas persistence analysis of the SPI time series (using Hurst index) shows the Southeast is characterized by least persistence.

Analysis of SPI sensitivity across multiple time scales shows the following insights:

1. Distributions of spatial cross-correlations over different regions show in general, drought indices tend to be closer and less uncertain at smaller time scales, whereas spatial variability increases with the increase in accumulation time scales. Further median cross-correlation is least over the Southwest and highest over the Northwest.
2. Regional distributions of weighted average severity show an increase in drought severity with the increase in the accumulation time scales, whereas opposite trends are noted for the spatial distribution of drought frequency. The highest average drought severity is found over the Midwest (East-North Central) at 3- and 12-month time scales, and also in the West-North Central at 6- and 9-month time scales. At the 12-month accumulation time scale, the Northeast is characterized by the highest average drought frequency with least median severity.

Examination of the added value of ensemble median GCM-forced NARCCAP RCMs against their host GCM shows the following insights:

1. Precipitation pattern simulated by the RCMs shows topographically induced fine-scale regional

features and their variabilities, such as regional precipitation patterns over the Southwest, West-North Central, and the Northwest regions.

2. Analysis of mean annual precipitation in both datasets shows that RCM produces larger precipitation over the Pacific Northwest and Northeast as compared to their host GCM. Analysis of seasonal precipitation over different subregions shows, except JJA, in all seasons, the variability of RCMs are larger as compared to their host GCM.
3. Neither NARCCAP nor their host GCMs are able to simulate signals of the North American Monsoon effectively and underestimate JJA precipitation over the Southwest.

Under non-stationary climate conditions, a credible projection of drought at fine-scale resolution is crucial for early warning, mitigation, and forming adaptation strategies (Duncan *et al.*, 2015). Although gridded precipitation and temperature data are routinely available at finer resolutions in observations and climate models, the higher resolution drought indices have limited availability through operational systems. In the U.S., specifically, many states lack indicator data at spatial and temporal scales needed for effective monitoring (Fontaine *et al.*, 2014). Further, uncertainty in drought quantifications and associated projections stem from a vast array of datasets from multiple sources that often limit our ability to frame appropriate mitigation strategies (Bishop and Beier, 2013). The difference in performance between the models and the observations primarily arises due to different initial conditions, structural dissimilarity, parameterization schemes, and the limited skills of the RCMs to simulate large-scale atmospheric pattern. To provide better regional assessment, the modeling community should continuously evaluate RCM output before it is used by the stakeholders for planning purposes. We hope the analysis will assist modelers to identify model deficiencies and further improve model performance, which will be helpful in providing credible drought projections. A proper coordination between NARCCAP modelers and the stakeholder community is needed to improve parameterization schemes of the models, so that the model is able to capture hydrologically relevant metrics (such as long-range persistence) useful for end user applications. Data-driven methods such as enhanced statistical downscaling schemes may help to achieve this goal.

ACKNOWLEDGMENTS

The work was funded by the United States (US) National Science Foundation (NSF) Expeditions in Computing Award # 1029711 and the NSF BIGDATA Award # 1447587. The Northeastern University provided partial funding. We thank the North American Regional Climate Change Assessment Program (NARCCAP, http://www.narccap.ucar.edu) for providing the data used in this article. NARCCAP is funded by the National Science Foundation (NSF), the U.S. Department of Energy (DoE), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency Office of Research and Development (EPA). We thank Dr. Seth McGinnis at Computational and Information Systems Laboratory, UCAR for providing access to HadCM3 GCM boundary condition data.

LITERATURE CITED

Andreadis, K.M. and D.P. Lettenmaier, 2006. Trends in 20th Century Drought over the Continental United States. Geophysical Research Letters 33, L10403:1-4, DOI: 10.1029/2006GL025711.

Aron, G., D.J. Wall, E.L. White, and C.N. Dunn, 1987. Regional Rainfall Intensity-Duration-Frequency Curves for Pennsylvania. Journal of the American Water Resources Association 23:479-485.

Arritt, R.W., 2008. Regional Drought in the North American Regional Climate Change Assessment Program (NARCCAP), in 33rd Annual Climate Diagnostics and Prediction Workshop/CLIVAR Drought Workshop (CDPW). Climate Prediction Center, NOAA. http://www.cpc.ncep.noaa.gov/products/outreach/workshops/CDPW33/abstracts/rarritt.shtml.

Ault, T.R., 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. Journal of Climate 27:7529-7549.

Benestad, R.E., I. Hanssen-Bauer, and D. Chen, 2008. Empirical-Statistical Downscaling. World Scientific, ISBN: 978-981-281-912-3.

Bishop, A.D. and M.C. Beier, 2013. Assessing Uncertainty in High-Resolution Spatial Climate Data across the US Northeast. PLoS ONE 8(8):e70260.

Blenkinsop, S. and H.J. Fowler, 2007. Changes in Drought Frequency, Severity and Duration for the British Isles Projected by the PRUDENCE Regional Climate Models. Journal of Hydrology 342:50-71.

Bonnin, G., 2013. Precipitation Frequency Estimates for the Nation and Extremes-A Perspective. Workshop on Probabilistic Flood Hazard Assessment (PFHA) January 29-31, 2013. pbadupws.nrc.gov/docs/ML1305/ML13059A419.pdf, *accessed* December 2013.

Bukovsky, M.S., 2012. Temperature Trends in the NARCCAP Regional Climate Models. Journal of Climate 25:3985-3991.

Bukovsky, M.S., D.J. Gochis, and L.O. Mearns, 2013. Towards Assessing NARCCAP Regional Climate Model Credibility for the North American Monsoon: Current Climate Simulations. Journal of Climate 26:8802-8826.

Caya, D. and R. Laprise, 1999. A Semi-Implicit Semi-Lagrangian Regional Climate Model: The Canadian RCM. Monthly Weather Review 127:341-362.

Cayan, D.R., T. Das, D.W. Pierce, T.P. Barnett, M. Tyree, and A. Gershunov, 2010. Future Dryness in the Southwest US and the Hydrology of the Early 21st Century Drought. Proceedings of the National Academy of Sciences 107(50):21271-21276.

Chen, F., Z. Janjić, and K. Mitchell, 1997. Impact of Atmospheric Surface-Layer Parameterizations in the New Land-Surface Scheme of the NCEP Mesoscale Eta Model. Boundary-Layer Meteorology 85:391-421.

Collier, J.C. and G.J. Zhang, 2007. Effects of Increased Horizontal Resolution on Simulation of the North American Monsoon in the NCAR CAM3: An Evaluation Based on Surface, Satellite, and Reanalysis Data. Journal of Climate 20:1843-1861.

Cruz, C.M.C., 2014. North American Monsoon Variability from Paleoclimate Era to Climate Change Projection: A Multiple Dataset Perspective. Ph.D. Dissertation, The University of Arizona, pp. 1-179.

Dai, A., 2013. Increasing Drought under Global Warming in Observations and Models. Nature Climate Change 3:52-58.

Dai, A., 2014. The Climate Data Guide: Palmer Drought Severity Index (PDSI). https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi, *accessed* September 2015.

Dalezios, N.R., A. Loukas, L. Vasiliades, and E. Liakopoulos, 2000. Severity-Duration-Frequency Analysis of Droughts and Wet Periods in Greece. Hydrological Sciences Journal 45:751-769.

Di Luca, A., R. de Elía, and R. Laprise, 2012. Potential for Added Value in Precipitation Simulated by High-Resolution Nested Regional Climate Models and Observations. Climate Dynamics 38:1229-1247.

Dominguez, F., E. Rivera, D.P. Lettenmaier, and C.L. Castro, 2012. Changes in Winter Precipitation Extremes for the Western United States under a Warmer Climate as Simulated by Regional Climate Models. Geophysical Research Letters 39, L05803:1-7, DOI: 10.1029/2011GL050762.

Douglas, E.M., R.M. Vogel, and C.N. Kroll, 2002. Impact of Streamflow Persistence on Hydrologic Design. Journal of Hydrologic Engineering 7:220-227.

Duncan, L.L., D. Perrone, J.H. Jacobi, and G.M. Hornberger, 2015. Drought Planning and Management: Using High Spatial Resolution as Part of the Solution. Environmental Science and Technology 49(5):2639-2647.

Easterling, D.R., T.W. Wallis, J.H. Lawrimore, and R.R. Heim, 2007. Effects of Temperature and Precipitation Trends on US Drought. Geophysical Research Letters 34, L20709:1-4, DOI: 10.1029/2007GL031541.

Fekete, B.M., C.J. Vörösmarty, J.O. Roads, and C.J. Willmott, 2004. Uncertainties in Precipitation and Their Impacts on Runoff Estimates. Journal of Climate 17(2):294-304.

Foley, A.M., 2010. Uncertainty in Regional Climate Modelling: A Review. Progress in Physical Geography 34(5):647-670.

Fontaine, M.M., A.C. Steinemann, and M.J. Hayes, 2014. State Drought Programs and Plans: Survey of the Western US. Natural Hazards Review 15(1):95-99.

Ford, T. and C.F. Labosier, 2014. Spatial Patterns of Drought Persistence in the Southeastern United States. International Journal of Climatology 34:2229-2240.

Fowler, H.J. and M. Ekström, 2009. Multi-Model Ensemble Estimates of Climate Change Impacts on UK Seasonal Precipitation Extremes. International Journal of Climatology 29:385-416.

Gao, Y., L.R. Leung, E.P. Salathé, F. Dominguez, B. Nijssen, and D.P. Lettenmaier, 2012. Moisture Flux Convergence in Regional and Global Climate Models: Implications for Droughts in the Southwestern United States under Climate Change. Geophysical Research Letters 39, L09711:1-5, DOI: 10.1029/2012GL051560.

Giorgi, F., X. Bi, and J.S. Pal, 2004. Mean, Interannual Variability and Trends in a Regional Climate Change Experiment over Europe. I. Present-Day Climate (1961-1990). Climate Dynamics 22:733-756.

Giorgi, F. and M.R. Marinucci, 1996. An Investigation of the Sensitivity of Simulated Precipitation to Model Resolution and Its Implications for Climate Studies. Monthly Weather Review 124:148-166.

Giorgi, F., M.R. Marinucci, and G.T. Bates, 1993a. Development of a Second-Generation Regional Climate Model (Regcm2). Part I: Boundary-Layer and Radiative Transfer Processes. Monthly Weather Review 121:2794-2813.

Giorgi, F., M.R. Marinucci, G. de Canio, and G.T. Bates, 1993b. Development of a Second-Generation Regional Climate Model (Regcm2). Part II: Convective Processes and Assimilation of Lateral Boundary Conditions. Monthly Weather Review 121:2814-2832.

Giorgi, F. and L.O. Mearns, 1991. Approaches to the Simulation of Regional Climate Change: A Review. Reviews of Geophysics 29:191-216.

Grell, G.A., J. Dudhia, and D.R. Stauffer, 1994. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN- 398 + STR, 121 pp.

Halwatura, D., A.M. Lechner, and S. Arnold, 2014. Design Droughts as Planning Tool for Ecosystem Establishment in Post-Mining Landscapes. Hydrology and Earth System Sciences Discussions 11:4809-4849.

Hamed, K.H. and A. Ramachandra Rao, 1998. A Modified Mann-Kendall Trend Test for Autocorrelated Data. Journal of Hydrology 204:182-196.

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. International Journal of Climatology 34:623-642.

Hayhoe, K., C. Wake, B. Anderson, X.-Z. Liang, E. Maurer, J. Zhu, J. Bradbury, A. DeGaetano, A.M. Stoner, and D. Wuebbles, 2008. Regional Climate Change Projections for the Northeast USA. Mitigation and Adaptation Strategies for Global Change 13(5-6):425-436.

Hayhoe, K., C.P. Wake, T.G. Huntington, L. Luo, M.D. Schwartz, J. Sheffield, E. Wood, B. Anderson, J. Bradbury, A. DeGaetano, T.J. Troy, and D. Wolfe, 2007. Past and Future Changes in Climate and Hydrological Indicators in the US Northeast. Climate Dynamics 28(4):381-407.

Hurst, H.E., 1951. Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers 116:770-808.

Jeong, D.I., A. St-Hilaire, T.B.M.J. Ouarda, and P. Gachon, 2012. Multisite Statistical Downscaling Model for Daily Precipitation Combined by Multivariate Multiple Linear Regression and Stochastic Weather Generator. Climatic Change 114(3-4):567-591.

Jeong, D.I., A. St-Hilaire, T.B.M.J. Ouarda, and P. Gachon, 2013. Projection of Future Daily Precipitation Series and Extreme Events by Using a Multi-Site Statistical Downscaling Model over the Great Montréal Area, Québec. Canada Hydrology Research 44(1):147-168.

Jeong, D.I., L. Sushama, and M.N. Khaliq, 2014. The Role of Temperature in Drought Projections over North America. Climatic Change 127:289-303.

Ji, L. and A.J. Peters, 2003. Assessing Vegetation Response to Drought in the Northern Great Plains Using Vegetation and Drought Indices. Remote Sensing of Environment 87:85-98.

Johnson, F., S. Westra, A. Sharma, and A.J. Pitman, 2011. An Assessment of GCM Skill in Simulating Persistence across Multiple Time Scales. Journal of Climate 24:3609-3623.

Jones, R.G., D.C. Hassell, D. Hudson, S.S. Wilson, G.J. Jenkins, and J.F.B. Mitchell, 2003. Workbook on Generating High Resolution Climate Change Scenarios Using PRECIS. UNDP Technical Report, 32 pp.

Juang, H.M., S.Y. Hong, and M. Kanamitsu, 1997. The NCEP Regional Spectral Model: An Update. Bulletin of the American Meteorological Society 78:2125-2143.

Karl, T. and W.J. Koss, 1984. Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983. National Climatic Data Center, Asheville, North Carolina.

Karl, T.R. and A.J. Koscielny, 1982. Drought in the United States: 1895-1981. Journal of Climatology 2:313-329.

Karnauskas, K.B., A. Ruiz-Barradas, S. Nigam, and A.J. Busalacchi, 2008. North American Droughts in ERA-40 Global and

NCEP North American Regional Reanalyses: A Palmer Drought Severity Index Perspective. Journal of Climate 21:2102-2123.

Kerr, R.A., 2013. Forecasting Regional Climate Change Flunks Its First Test. Science 339:638.

Koutsoyiannis, D., 2003. Climate Change, the Hurst Phenomenon, and Hydrological Statistics. Hydrological Sciences Journal 48:3-24.

Koutsoyiannis, D. and A. Montanari, 2007. Statistical Analysis of Hydroclimatic Time Series: Uncertainty and Insights. Water Resources Research 43, W05429:1-9, DOI: 10.1029/2006WR005592.

Kumar, S., V. Merwade, J.L. Kinter, III, and D. Niyogi, 2013. Evaluation of Temperature and Precipitation Trends and Long-Term Persistence in CMIP5 Twentieth-Century Climate Simulations. Journal of Climate 26:4168-4185.

Laprise, R., 2008. Regional Climate Modeling. Journal of Computational Physics 227:3641-3666.

Lee, M.-I., S.D. Schubert, M.J. Suarez, I.M. Held, A. Kumar, T.L. Bell, J.-K.E. Schemm, N.-C. Lau, J.J. Ploshay, and H.-K. Kim, 2007. Sensitivity to Horizontal Resolution in the AGCM Simulations of Warm Season Diurnal Cycle of Precipitation over the United States and Northern Mexico. Journal of Climate 20:1862-1881.

Leung, L.R., Y. Qian, X. Bian, W.M. Washington, J. Han, and J.O. Roads, 2004. Mid-Century Ensemble Regional Climate Change Scenarios for the Western United States. Climatic Change 62:75-113.

Liang, X., D.P. Lettenmaier, E.F. Wood, and S.J. Burges, 1994. A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models. Journal of Geophysical Research: Atmospheres (1984-2012) 99:14415-14428.

Liang, X.-Z., M. Xu, X. Yuan, T. Ling, H.I. Choi, F. Zhang, L. Chen, S. Liu, S. Su, F. Qiao, Y. He, J.X.L. Wang, K.E. Kunkel, W. Gao, E. Joseph, V. Morris, T.-W. Yu, J. Dudhia, and J. Michalakes, 2012. Regional Climate-Weather Research and Forecasting Model. Bulletin of the American Meteorological Society 93:1363-1387.

Lloyd-Hughes, B. and M.A. Saunders, 2002. A Drought Climatology for Europe. International Journal of Climatology 22:1571-1592.

McKee, T.B., N.J. Doesken, and J. Kleist, 1993. The Relationship of Drought Frequency and Duration to Time Scales. Proceedings of the 8th Conference on Applied Climatology, pp. 179-183.

Mearns, L.O., R. Arritt, S. Biner, M.S. Bukovsky, S. McGinnis, S. Sain, D. Caya, J. Correia, Jr., D. Flory, and W. Gutowski, 2012. The North American Regional Climate Change Assessment Program: Overview of Phase I Results. Bulletin of the American Meteorological Society 93:1337-1362.

Mearns, L.O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes, and Y. Qian, 2009. A Regional Climate Change Assessment Program for North America. Eos, Transactions American Geophysical Union 90:311.

Meehl, G.A., C. Covey, K.E. Taylor, T. Delworth, R.J. Stouffer, M. Latif, B. McAvaney, and J.F.B. Mitchell, 2007. The WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. Bulletin of the American Meteorological Society 88:1383-1394.

Mesa, O.J., V.K. Gupta, and P.E. O'Connell, 2012. Dynamical System Exploration of the Hurst Phenomenon in Simple Climate Models. Geophysical Monograph Series 196:209-229.

Mishra, V., F. Dominguez, and D.P. Lettenmaier, 2012. Urban Precipitation Extremes: How Reliable Are Regional Climate Models? Geophysical Research Letters 39(3), L03407: 1-8, DOI: 10.1029/2011GL050658.

Mo, K.C. and M. Chelliah, 2006. The Modified Palmer Drought Severity Index Based on the NCEP North American Regional Reanalysis. Journal of Applied Meteorology and Climatology 45:1362-1375.

Mo, K.C. and J.E. Schemm, 2008a. Droughts and Persistent Wet Spells over the United States and Mexico. Journal of Climate 21:980-994.

Mo, K.C. and J.E. Schemm, 2008b. Relationships between ENSO and Drought over the Southeastern United States. Geophysical Research Letters 35, L15701:1-5, DOI: 10.1029/2008GL034656.

Nickl, E., C.J. Willmott, K. Matsuura, and S.M. Robeson, 2010. Changes in Annual Land-Surface Precipitation over the Twentieth and Early Twenty-First Century. Annals of the Association of American Geographers 100(4):729-739.

Olesen, J.E., T.R. Carter, C.H. Diaz-Ambrona, S. Fronzek, T. Heidmann, T. Hickler, T. Holt, M.I. Minguez, P. Morales, J.P. Palutikof, M. Quemada, M. Ruiz-Ramos, G.H. Rubæk, F. Sau, B. Smith, and M.T. Sykes, 2007. Uncertainties in Projected Impacts of Climate Change on European Agriculture and Terrestrial Ecosystems Based on Scenarios from Regional Climate Models. Climatic Change 81:123-143.

Outcalt, S.I., K.M. Hinkel, E. Meyer, and A.J. Brazel, 1997. Application of Hurst Resecaling to Geophysical Serial Data. Geographical Analysis 29:72-87.

Peng, C.-K., S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L. Goldberger, 1994. Mosaic Organization of DNA Nucleotides. Physical Review E 49:1685.

Pryor, S.C., R.J. Barthelmie, and J.T. Schoof, 2013a. High-Resolution Projections of Climate-Related Risks for the Midwestern USA. Climate Research 56:61-79.

Pryor, S.C., R.J. Barthelmie, and J.T. Schoof, 2013b. Past and Future Wind Climates Over the Contiguous USA Based on the NARCCAP Model Suite. Journal of Geophysical Research 117: D19119, DOI: 10.1029/2012JD017449.

Racherla, P.N., D.T. Shindell, and G.S. Faluvegi, 2012. The Added Value to Global Model Projections of Climate Change by Dynamical Downscaling: A Case Study over the Continental US Using the GISS-ModelE2 and WRF Models. Journal of Geophysical Research: Atmospheres D20118:1-8, DOI: 10.1029/2012JD018091.

Reddy, M.J. and P. Ganguli, 2013. Spatio-Temporal Analysis and Derivation of Copula-Based Intensity-Area-Frequency Curves for Droughts in Western Rajasthan (India). Stochastic Environmental Research and Risk Assessment 27(8):1975-1989.

Rocheta, E., M. Sugiyanto, F. Johnson, J. Evans, and A. Sharma, 2014. How Well Do General Circulation Models Represent Low-Frequency Rainfall Variability? Water Resources Research 50:2108-2123.

Salvadori, G. and C. De Michele, 2004. Frequency Analysis via Copulas: Theoretical Aspects and Applications to Hydrological Events. Water Resources Research 40, W12511:1-17, DOI: 10.1029/2004WR003133.

Salvadori, G., G.R. Tomasicchio, and F. D'Alessandro, 2013. Multivariate Approach to Design Coastal and Off-Shore Structures. Journal of Coastal Research 1(65):339-386.

Sanderson, B.M. and R. Knutti, 2012. On the Interpretation of Constrained Climate Model Ensembles. Geophysical Research Letters 39, L16708:1-6, DOI: 10.1029/2012GL052665.

Schiermeier, Q., 2013. Climate Models Fail to "Predict" US Droughts. Nature 496:284, DOI: 10.1038/496284a.

Schmidli, J., C.M. Goodess, C. Frei, M.R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith, 2007. Statistical and Dynamical Downscaling of Precipitation: An Evaluation and Comparison of Scenarios for the European Alps. Journal of Geophysical Research: Atmospheres D04105:1-20, DOI: 10.1029/2005JD007026.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014. GPCC's New Land Surface Precipitation Climatology Based on Quality-Controlled In Situ Data and Its Role in Quantifying The Global Water Cycle. Theoretical and Applied Climatology 115:15-40.

Seager, R., A. Tzanova, and J. Nakamura, 2009. Drought in the Southeastern United States: Causes, Variability over the Last Millennium, and the Potential for Future Hydroclimate Change. Journal of Climate 22:5021-5045.

Sheffield, J., K.M. Andreadis, E.F. Wood, and D.P. Lettenmaier, 2009. Global and Continental Drought in the Second Half of the Twentieth Century: Severity-Area-Duration Analysis and Temporal Variability of Large-Scale Events. Journal of Climate 22:1962-1981.

Sheffield, J., B. Livneh, and E.F. Wood, 2012a. Representation of Terrestrial Hydrology and Large-Scale Drought of the Continental United States from the North American Regional Reanalysis. Journal of Hydrometeorology 13:856-876.

Sheffield, J. and E.F. Wood, 2008. Projected Changes in Drought Occurrence under Future Global Warming from Multi-Model, Multi-Scenario, IPCC AR4 Simulations. Climate Dynamics 31:79-105.

Sheffield, J., E.F. Wood, and M.L. Roderick, 2012b. Little Change in Global Drought over the Past 60 Years. Nature 491:435-438.

Shiau, J.-T. and H.W. Shen, 2001. Recurrence Analysis of Hydrologic Droughts of Differing Severity. Journal of Water Resources Planning and Management 127:30-40.

Sims, A.P., D.S. Niyogi, and S. Raman, 2002. Adopting Drought Indices for Estimating Soil Moisture: A North Carolina Case Study. Geophysical Research Letters 29(8), 1183: 1-4, DOI: 10.1029/2001GL013343.

Singh, D., M. Tsiang, B. Rajaratnam, and N.S. Diffenbaugh, 2013. Precipitation Extremes over the Continental United States in a Transient, High-Resolution, Ensemble Climate Model Experiment. Journal of Geophysical Research: Atmospheres 118:7063-7086, DOI: 10.1002/jgrd.50543.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang, and J.G. Powers, 2005. A Description of the Advanced Research WRF Version 2. NCAR Technical Note NCAR/TN-4681STR, 88 pp.

Smith, A.B. and R.W. Katz, 2013. US Billion-Dollar Weather and Climate Disasters: Data Sources, Trends, Accuracy and Biases. Natural Hazards 67(2):387-410.

Soulé, P.T. and Z.Y. Yin, 1995. Short-to Long-Term Trends in Hydrologic Drought Conditions in the Contiguous United States. Climate Research 5:149-157.

Stine, S., 1994. Extreme and Persistent Drought in California and Patagonia During Mediaeval Time. Nature 369:546-549.

Svoboda, M.D., M. Lecomte, R. Hayes, K. Heim, J. Gleason, B. Angel, R. Rippey, M. Tinker, D. Palecki, D. Stooksbury, D. Miskus, and S. Stephens, 2002. The Drought Monitor, Bulletin of the American Meteorological Society 83:1181-1190.

Tallaksen, L.M. and K. Stahl, 2014. Spatial and Temporal Patterns of Large-Scale Droughts in Europe: Model Dispersion and Performance. Geophysical Research Letters 41:429-434, DOI: 10.1002/2013GL058573.

Taylor, K.E., 2001. Summarizing Multiple Aspects of Model Performance in a Single Diagram. Journal of Geophysical Research 106(D7):7183-7192.

Tebaldi, C. and R. Knutti, 2007. The Use of the Multi-Model Ensemble in Probabilistic Climate Projections. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365:2053-2075.

Torma, C., G. Filippo, and E. Coppola, 2015. Added Value of Regional Climate Modeling over Areas Characterized by Complex Terrain—Precipitation over Alps. Journal of Geophysical Research: Atmospheres 120:3957-3972, DOI: 10.1002/2014JD022781.

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. Nature Climate Change 4:17-22.

Tripathi, O.P. and F. Dominguez, 2013. Effects of Spatial Resolution in the Simulation of Daily and Subdaily Precipitation in the Southwestern US. Journal of Geophysical Research: Atmospheres 118:7591-7605.

von Storch, H., H. Langenberg, and F. Feser, 2000. A Spectral Nudging Technique for Dynamical Downscaling Purposes. Monthly Weather Review 128:3664-3673.

Wang, A., T.J. Bohn, S.P. Mahanama, R.D. Koster, and D.P. Lettenmaier, 2009. Multimodel Ensemble Reconstruction of Drought over the Continental United States. Journal of Climate 22:2694-2712.

Weaver, S.J., A. Ruiz-Barradas, and S. Nigam, 2009. Pentad Evolution of the 1988 Drought and 1993 Flood over the Great Plains: An NARR Perspective on the Atmospheric and Terrestrial Water Balance. Journal of Climate 22:5366-5384.

Wehner, M.F., 2013. Very Extreme Seasonal Precipitation in the NARCCAP Ensemble: Model Performance and Projections. Climate Dynamics 40:59-80.

Weron, R., 2002. Estimating Long-Range Dependence: Finite Sample Properties and Confidence Intervals. Physica A: Statistical Mechanics and Its Applications 312:285-299.

Wilby, R.L., S.P. Charles, E. Zorita, B. Timbal, P. Whetton, and L.O. Mearns, 2004. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods, Technical Report, Data Distribution Center Intergovernmental Panel on Climate Change, Norwich, UK.

Wilby, R.L., T.M.L. Wigley, D. Conway, P.D. Jones, B.C. Hewitson, J. Main, and D.S. Wilks, 1998. Statistical Downscaling of General Circulation Model Output: A Comparison of Methods. Water Resources Research 34(11):2995-3008.

Wilhite, A., M.V.K. Sivakumar, and R. Pulwarty, 2014. Managing Drought Risk in a Changing Climate: The Role of National Drought Policy. Weather and Climate Extremes 3:4-13.

Willmott, C.J. and K. Matsuura, 2001. Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950-1999), http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html, *accessed* July 2015.

WMO (World Meteorological Organization), 2012. Standardized Precipitation Index User Guide. Report No. WMO-No. 1090, Geneva.

Woodhouse, C.A., 1997. Winter Climate and Atmospheric Circulation Patterns in the Sonoran Desert Region, USA. International Journal of Climatology 17:859-873.

Woodhouse, C.A., D.M. Meko, G.M. MacDonald, D.W. Stahle, and E.R. Cook, 2010. A 1,200-Year Perspective of 21st Century Drought in Southwestern North America. Proceedings of the National Academy of Sciences 107:21283-21288.

Xiong, W., R. Matthews, I. Holman, E. Lin, and Y. Xu, 2007. Modelling China's Potential Maize Production at Regional Scale Under Climate Change. Climatic Change 85:433-451.

Yarnell, D.L., 1935. Rainfall Intensity-Frequency Data. U.S. Department of Agriculture, Publication No. 204, Washington, D.C.

Yevjevich, V., 1983. Methods for Determining Statistical Properties of Droughts. *In*: Coping with Droughts, V. Yevjevich, L. da Cunha, and E. Vlachos (Editors). Water Resources Publications, Fort Collins, Colorado, pp. 22-43.