Evaluation of a Regional Climate Modeling Effort for the Western United States Using a Superensemble from Weather@home*

SIHAN LI, PHILIP W. MOTE, DAVID E. RUPP, AND DEAN VICKERS

Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

ROBERTO MERA

Union of Concerned Scientists, Washington, D.C.

MYLES ALLEN

Smith School of Enterprise and the Environment, and Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

ABSTRACT

Simulations from a regional climate model (RCM) as part of a superensemble experiment were compared with observations of surface meteorological variables over the western United States. The RCM is the Hadley Centre Regional Climate Model, version 3, with improved physics parameterizations (HadRM3P) run at 25-km resolution and nested within the Hadley Centre Atmosphere Model, version 3 (HadAM3P). Overall, the means of seasonal temperature were well represented in the simulations; 95% of grid points were within 2.7°, 2.4°, and 3.6°C of observations in winter, spring, and summer, respectively. The model was too warm over most of the domain in summer except central California and southern Nevada. HadRM3P produced more extreme temperatures than observed. The overall magnitude and spatial pattern of precipitation were well characterized, though HadRM3P exaggerated the orographic enhancement along the coastal mountains, Cascade Range, and Sierra Nevada. HadRM3P produced warm/dry northwest, cool/wet southwest U.S. patterns associated with El Niño. However, there were notable differences, including the locations of the transition from warm (dry) to cool (wet) in the anomaly fields when compared with observations, though there was disagreement among observations. HadRM3P simulated the observed spatial pattern of mean annual temperature more faithfully than any of the RCM–GCM pairings in the North American Regional Climate Change Assessment Program (NARCCAP). Errors in mean annual precipitation from HadRM3P fell within the range of errors of the NARCCAP models. Last, this paper provided examples of the size of an ensemble required to detect changes at the local level and demonstrated the effect of parameter perturbation on regional precipitation.

1. Introduction

Although global climate models (GCMs) are important tools for investigating climate variability and

---

Corresponding author address: Sihan Li, Oregon Climate Change Research Institute, College of Earth, Ocean, and Atmospheric Sciences, CEOAS Admin Building 104, Oregon State University, Corvallis, OR 97330.
E-mail: sli@coas.oregonstate.edu

change on large scales, their coarse spatial resolution to date (typically 100–300 km) inhibits their ability to represent the interactions of synoptic-scale weather systems with local terrain and mesoscale processes. These interactions not only determine the mesoscale detail of climate patterns and variability but may also profoundly influence the magnitude of variability and the response to forced change at and above the mesoscale. Given current technology, the standard approach to incorporating mesoscale meteorology over a region is to nest a higher-resolution climate model [known as a regional climate model (RCM)] within a GCM. RCMs, with their increased resolution of local terrain, are able to represent finescale features

such as land–sea breeze, rain shadows, and windstorms (Leung et al. 2003a,b).

The western United States is an excellent test bed for investigating climate patterns emerging from interactions between atmospheric circulation and terrain because of the strong topographic gradients and wide range of climate zones over a relatively small area of the planet. In fact, the first use of a nested RCM was focused on the western United States (Dickinson et al. 1989). Since then there have been numerous studies simulating climate in the western United States using RCMs (e.g., Duffy et al. 2006; Salathé et al. 2008, 2010; Zhang et al. 2009; Dulière et al. 2011). Duffy et al. (2006) used four RCM–GCM combinations, run at 36–60-km spatial resolution, to simulate present and future climates to examine the intermodel variability of the response to increased atmospheric greenhouse gases (GHGs). Salathé et al. (2008) nested one 15-km-resolution RCM in a GCM to examine mesoscale feedbacks and localized responses to increased GHGs. By performing two regional climate simulations using one RCM driven by two different GCMs, Salathé et al. (2010) showed that mesoscale simulations produced regional changes substantially different from the GCMs or statistical downscaling.

In addition to the limitations associated with spatial resolution, modeling studies are subject to challenges associated with small sample sizes. Most studies using RCMs, and also many using GCMs, have performed a single simulation, which forms the basis for comparison with observations or for comparisons between simulations with differing forcings and/or boundary conditions. The disadvantage of using a single simulation is that it certainly undersamples the space of possible climate states, and while some uncertainty can be reduced by averaging over time (e.g., 30 years), even such averages can be substantially different between simulations whose only difference is a small perturbation to initial conditions (e.g., Deser et al. 2014). This problem is even more acute when comparing extremes. Adequately large ensembles of simulations are needed for determining the statistical properties of the model more accurately. For example, a 40-member ensemble generated with the National Center for Atmospheric Research (NCAR) Community Climate System Model, version 3 (CCSM3), has been used by many studies to investigate uncertainty due to internal variability, signal-to-noise ratio,[1] minimum ensemble size required to detect a forced signal, and time of emergence of the forced signals (Deser et al. 2012a,b; Oshima et al. 2012;

Kang et al. 2013; Hu and Deser 2013; Wettstein and Deser 2014; Wallace et al. 2015). These studies imply that even 40 members are sometimes insufficient to separate signal from noise, depending on the signal being sought, the domain, and the time of emergence of the forced signal. Several coordinated ensemble modeling experiments have been conducted to better quantify uncertainty on a regional level, such as the Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE; Christensen and Christensen 2007). Also, the North American Regional Climate Change Assessment Program (NARCCAP; 50-km resolution; Mearns et al. 2009, 2012) was implemented to explore the separate and combined uncertainties in regional climate change simulations that result from the use of different atmosphere–ocean general circulation models (AOGCMs) to provide boundary conditions for different RCMs.

A third challenge to modeling regional climate concerns is the dependence of the model simulation on physical parameterizations. Processes that occur at scales finer than the resolution of the climate model are necessarily left unresolved. These processes must be simulated using empirical representations of their aggregated behavior at resolved scales as functions of the resolved-scale variables such as temperature, humidity, wind, and pressure. While observational studies are usually used to set values for parameters, parameterizations could still introduce uncertainty into the modeling process when there is a range of parameter values that are physically possible. The uncertainties in the values of the parameters lead to errors in the simulated climate and uncertainty in the response to forcing.

We designed a study to address three challenges: 1) achieving relatively high spatial resolution to reproduce important features of climate in the western United States, 2) accounting for the internal variations associated with initial conditions, and 3) accounting for variations in simulated climate associated with parameter choices. To address the first challenge, we nested the Hadley Centre Regional Climate Model, version 3, with improved physics parameterizations (HadRM3P; Jones et al. 2004) at 25-km resolution within the Hadley Centre Atmosphere Model, version 3 (HadAM3P), which is a higher-resolution ($1.875° \times 1.25°$) version of the atmospheric component of HadCM3, the atmosphere–ocean coupled general circulation model (Gordon et al. 2000). To address the second and third challenges, we generated a "superensemble" of simulations (a total of over 130 000 model years) for the historical time period 1960–2009 using the HadISST, version 1.1, dataset (Rayner et al. 2003) to specify the SSTs and sea ice

---

[1] Defined here as delta mean divided by uncertainty.

fractions for each month. The overall experiment, some initial results, and the strengths and weaknesses of our approach in comparison with other studies are discussed in Mote et al. (2015), with further details of the modeling given in Massey et al. (2015).

In the first phase of this project, two studies (Zhang et al. 2009; Dulière et al. 2011) used the HadRM3P at 25-km resolution over the western United States, just as in this paper. They compared HadRM3P driven by reanalysis data compared with station data and with simulations using the Weather Research and Forecasting (WRF) Model at 12- and 36-km resolution for the period 2003–07; these simulations were run in-house. Simulations for surface temperature in HadRM3P were about as skillful as for WRF at both resolutions, while for precipitation, the 12-km WRF simulation was better than the 36-km WRF and 25-km HadRM3P. The next paper (Mote et al. 2015) described the superensemble but did not thoroughly evaluate the model, a task left to this paper.

As discussed in more detail by Massey et al. (2015), our experiment covers one of six regions of the globe using the weather@home system to simulate regional climate. Weather@home uses a volunteer computing network (http://www.climateprediction.net; Allen 1999) to generate large ensembles of HadRM3P driven by HadAM3P. To generate the superensemble, we varied initial conditions and parameter values as described in section 2b below.

The primary purpose of this paper is to compare the climate as simulated by the superensemble against the observed climate over the western United States. We note that errors in the RCM results may be due to problems in the RCM itself or may reflect errors in the lateral boundary conditions supplied by the GCM. Thus, in our analysis of simulated historical climate, we evaluate not only the RCM alone but rather the coupled RCM–GCM model. We evaluate, primarily, the RCM's skill in reproducing spatial details of the regional climate. Moreover, given the importance of teleconnections between El Niño–Southern Oscillation (ENSO) and seasonal climate over the western United States (e.g., Ropelewski and Halpert 1986; Wallace et al. 1992; Gershunov 1998; Cayan et al. 1999), we also evaluate how well HadRM3P–HadAM3P reproduces the regional teleconnections to ENSO. Our evaluation is mainly focused on temperature and precipitation. Net downward solar radiation is also examined here, mainly as a diagnostic for understanding errors in temperature.

We describe the regional simulations and datasets used in this study in section 2. Section 3 examines climatological statistics such as seasonal mean states, spatial correlation, and temperature/precipitation–topography relationships, while section 4 discusses simulated ENSO

teleconnections in the western United States. An assessment of added value from a high-resolution superensemble is provided in section 5, and conclusions are given in section 6.

2. Datasets and simulations

a. Data

We compared simulations to six datasets of observed, or observation-based, meteorological variables. The first dataset was monthly mean temperature from the U.S. Historical Climatology Network (USHCN) monthly data, version 2 (Menne et al. 2009). The fully adjusted temperature series were used for stations in Oregon, Idaho, Washington, California, and Nevada. Stations that moved or have joined records were excluded, leaving a total of 147 stations. The USHCN temperature data were compared with the closest model grid point from the regional model. Differences in elevation are expected to influence the temperature comparisons because elevations of stations may be as much as 500 m above or below the elevation of the model grid cell, so we applied an elevation adjustment to the simulated temperature for this comparison. We used simple linear regression to estimate how much of the differences in temperature could be explained by the differences in elevation alone. We regressed the errors (calculated as simulated minus observed temperature) against the differences in elevation of the grid cells and the stations. The predictions from the linear regression were then subtracted from the simulated temperature to achieve the elevation-adjusted temperatures.

The second dataset, providing monthly mean maximum, minimum, and average temperature, precipitation rate, and net downward solar radiation, was the 32-km resolution National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR; Mesinger et al. 2006). NARR begins in 1979 and thus does not span our regional model simulations, which begin in 1960. When comparing temperature from NARR with HadRM3P, we used a standard lapse rate of $4.5°C\,km^{-1}$ for the Cascade Range (i.e., Cascades) and Sierra Nevada and $6.5°C\,km^{-1}$ elsewhere to account for differences in elevation between the datasets.

The third dataset consisted of monthly means of near-surface daily maximum and minimum temperature and precipitation rate from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; $2.5' \times 2.5'$, Daly et al. 2008).

The fourth dataset, providing the monthly gridded ($0.5° \times 0.5°$) maximum and minimum temperature and

TABLE 1. RCM–GCM combinations from NARCCAP used in this paper. (ECP2 is Experimental Climate Prediction Center Regional Spectral Model version 2; for additional acronym expansions, see http://www.ametsoc.org/PubsAcronymList.)

RCM	Driving GCM
CRCM	CCSM
CRCM	CGCM3
ECP2	GFDL
HadRM3	GFDL
HadRM3	HadCM3
MM5–Iowa State University (MM5I)	CCSM
MM5I	HadCM3
RegCM3	CGCM3
RegCM3	GFDL
WRF–Grell (WRFG)	CCSM
WRFG	CGCM3

precipitation rate, was from Climate Research Unit time series (CRU TS), version 3.10 (Harris et al. 2014).

The fifth dataset was the Climate Prediction Center (CPC) 0.25° × 0.25° daily U.S. unified gauge-based analysis of precipitation (CPC U.S. unified precipitation data provided by the NOAA/OAR/ESRL Physical Sciences Division (PSD) from their website at http://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html).

The sixth dataset was the Global Precipitation Climatology Centre (GPCC) 0.5° × 0.5° monthly precipitation dataset calculated from global station data (GPCC Precipitation data provided by the NOAA/OAR/ESRL PSD from their website at http://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html).

NARR datasets are reanalysis products, which are model simulations, and contain features of both the constraining observations and the underlying model. Thus, NARR itself may produce a pattern of precipitation that is different from observations and therefore may not be a fair test for another model; consequently, in this paper, we used PRISM, CRU TS, CPC, and GPCC precipitation as supplementary data sources against which to compare our simulations.

We also examined simulations from 11 RCM–GCM combinations (Table 1) from NARCCAP (Mearns et al. 2014). The historical simulations span the years 1969 to 1999.

All gridded datasets were regridded to a common 0.25° × 0.25° grid using bilinear interpolation.

b. Simulations

The HadRM3P domain covers the western United States, a portion of Canada, and the northwestern Pacific Ocean. This analysis, however, focused on the western United States west of 110°W longitude (see Fig. 1) largely because one of the primary observational datasets we used covers the United States only. Details
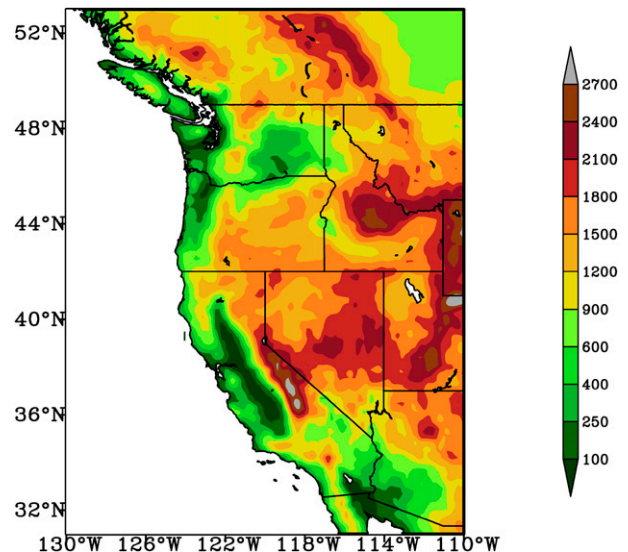

FIG. 1. Approximate domain and elevation (m) of terrain used in this research.

of the model configuration can be found in Massey et al. (2015) but with the western U.S. region replacing the European region. Briefly, HadAM3P runs first for one full model day, providing the lateral boundary conditions to HadRM3P, which also runs for one full model day; there is no feedback from HadRM3P to HadAM3P. HadRM3P defines a four-point buffer zone (100 km) around the perimeter of the region, which we exclude from our analysis. The main variables comprising the lateral boundary conditions are relaxed across the buffer zone to values temporally interpolated from 6-hourly output from HadAM3P.

Each simulation is for a single year, but simulations can be connected to make a longer time series. Initially, a pool of 1-year "work units" are created at 5-yearly intervals, each with the same starting conditions of the state of the model after nine years (12 January 1960–30 November 1968) of integration under observed climate forcing. These work units are distributed to client computers, and the results from the integration are returned, along with the final state of the model. This final state is then incorporated into a new work unit describing the next year of the climate scenario, using this final state as the starting condition. This process is repeated for the subsequent integrations, enabling strings of several-year runs to be built from the single-year runs (Massey et al. 2015).

Two experiments were run for the years 1960–2009: 1) a perturbed initial conditions experiment with standard, or default, model parameters [standard physics (SP)] and 2) a perturbed physics (PP) experiment. In the SP experiment, the initial condition perturbation is drawn from a large set of possible perturbations defined as deltas in potential temperature and is calculated as a

TABLE 2. Perturbed parameters in the new perturbed physics experiment used in this analysis. The upper and lower bounds for the values of the parameters were specified from expert solicitation.

Parameters	Description/process affected	Lower/upper bound
ENTCOEF	Scales rate of mixing between environmental (clear) air and convective plume (cloud)	[0.6, 9]
CT	How quickly cloud droplets convert to rain	$[0.5, 4] \times 10^{-4}$
VF1	Ice fall speed through clouds—important for the development of clouds and determining type (rain, sleet, hail, and snow) and amount of precipitation	[0.5, 2]

fully 3D field by first taking 348 next-day differences from a 1-yr-long integration of the GCM, with a scaling function applied in the vertical to ensure that there is no perturbation at the top of the atmosphere, and then applying five global scaling factors to the perturbations to generate a set of 1740 initial condition perturbations [for details see Massey et al. (2015)]. In the results of the SP experiment presented below (sections 3 through 5 and the first half of section 6), 20–500 runs per year, each with a unique set of initial condition, were used for the years 1979–98, depending on availability, and 500 runs were used per year for 1999–2009.

The motivation behind the PP experiment is that there may be many variants of the climate model that are as good as if not "better" than the standard version, and their response to a given climate forcing may be different from the standard version. Our ultimate goal (beyond the scope of this paper) is to cast a wide net and explore parameter space to find regions of parameter space where the model performs well and to assess the implications for climate response to changes in GHGs.

In the historical PP experiment, the perturbations of 12 parameters appear to have been conservative and did not notably alter the probability distributions of regional temperature and precipitation. Therefore, we ran a subsequent PP experiment where we isolated and per-turbed only three parameters (listed in Table 2) while keeping the rest of the parameters at their respective default value. The three parameters were chosen be-cause they were shown by previous studies to have bearing on the vertical moisture profile and energy balance. Findings of previous studies (Stainforth et al. 2005; Sanderson and Piani 2007; Knight et al. 2007; Sanderson et al. 2008a,b) suggested the dominant in-fluence of the entrainment coefficient (ENTCOEF) in establishing different relative humidity profiles that lead to different climate sensitivities. Other investigations (Grabowski 2000; Wu 2002; Sanderson and Piani 2007; Sanderson et al. 2008a,b) indicated that a low ice fall speed (VF1) would lead to a warm, moist, cloudy at-mospheric profile with less precipitation. Sanderson et al. (2010) showed that the accretion constant (CT) affects water vapor, cloud, and lapse-rate feedbacks.

The range of parameter values was increased over the previous PP experiment based on previous studies (e.g., Stainforth et al. 2005; Sanderson et al. 2008a,b) done on the predecessor ensemble of weather@home; Latin hypercube sampling was used to create 200 distinct parameter sets, with the same 15 initial condition perturbations applied to each parameter set. Results from the 3000 runs were used to identify parameter combinations that lead to warmer and drier, or cooler and wetter, conditions. Then a new set of runs was sent out with three parameter sets P1, P2, and P3—P1 is the default setting and P2 and P3 corresponding to warmer and drier and cooler and wetter conditions, respectively. We applied 1000 initial condition perturbations to each parameter set. By the time of this manuscript, an en-semble of 110 simulations for each parameter set had been completed for year 2011; some initial results of this PP experiment were presented here.

3. Climatological mean statistics

To establish how well the regional climate simula-tions reproduce the observed climate of the western United States, first we compared seasonal mean simu-lations to gridded observations averaged for the period December 1979 through November 2009, in a similar manner to Leung et al. (2003a,b). We attempt, through averaging over a period of 31 years, to reduce the in-ternal atmospheric variability about the mean state so that differences between the simulations and gridded observations are primarily the result of model de-ficiencies and differences in grid resolutions. We sepa-rated our analysis into winter [December–February (DJF)], spring [March–May (MAM)], and summer [June–August (JJA)]. For brevity we omit autumn, al-though we have analyzed autumn results as well; adding panels for autumn would overcrowd the figures without adding substantially meaningful information, and be-sides, few impacts of climate change are connected with changes in autumn.

The spatial patterns of seasonal average temperature (tavg) for HadRM3P were very similar to those of PRISM and NARR (Fig. 2). Overall, temperature was
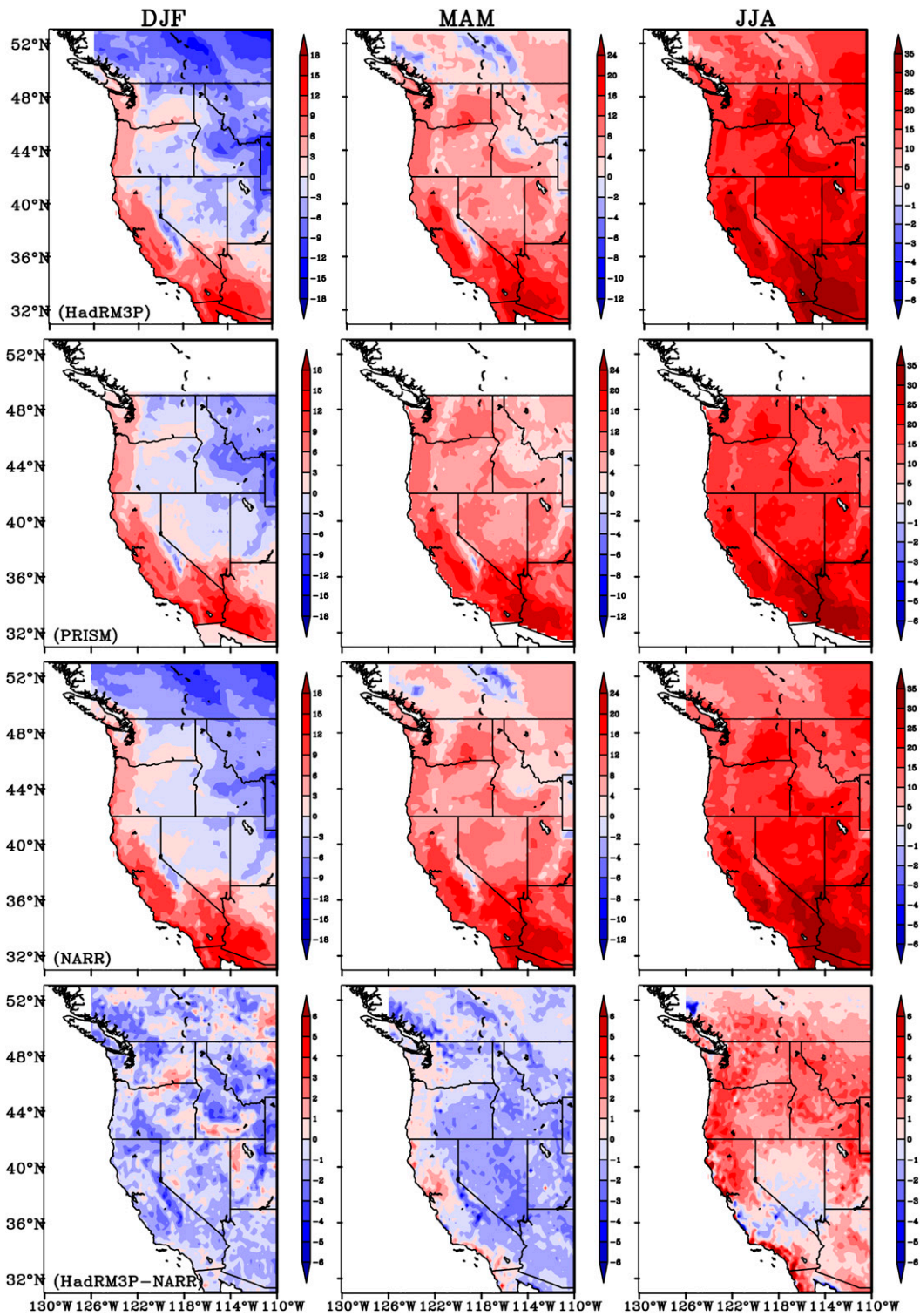
FIG. 2. Seasonal mean temperature (°C) in 1979–2009 for (left) DJF, (center) MAM, and (right) JJA from (top) HadRM3P, (middle top) PRISM, (middle bottom) NARR, and (bottom) the bias (HadRM3P minus NARR).
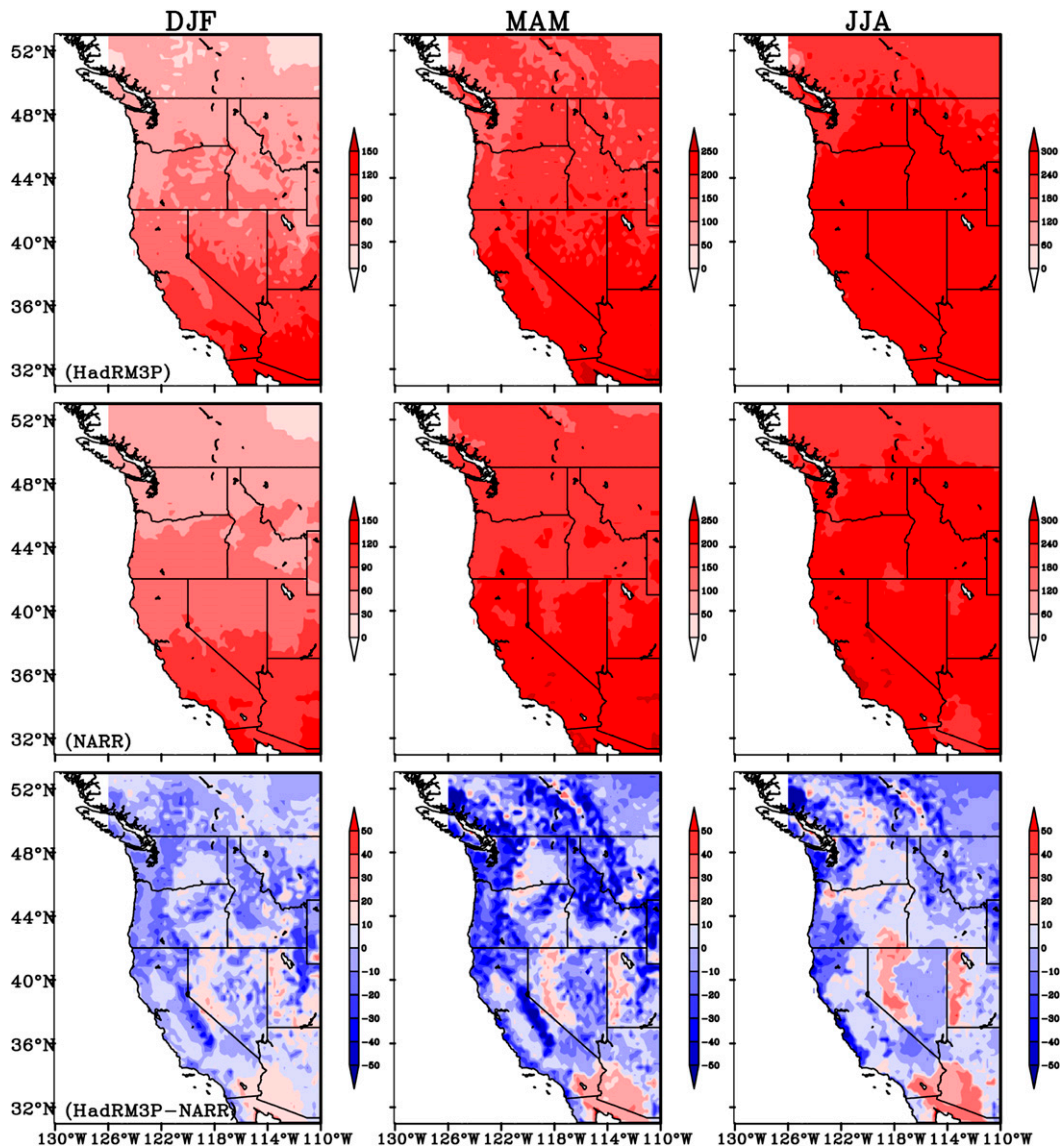
FIG. 3. Seasonal mean monthly net downward solar radiation (W m^{-2}) in 1979–2009 for (left) DJF, (center) MAM, and (right) JJA from (top) HadRM3P, (middle) NARR, and (bottom) the bias (HadRM3P minus NARR).

well represented in the simulations; the influences of the major geographical features like mountain ranges were evident, and the seasonal cycle was reproduced. The simulated spatial patterns during winter and spring were very close to those observed, except in winter the cold bias was almost everywhere, while the spring simulated temperature showed a warm bias of ~1°C along western Washington, Oregon, and California, with the rest of the domain showing cold bias. During summer, HadRM3P was warmer than observations over most of the western U.S. domain except central California and southern Nevada, and the warm bias was larger than in winter and spring. In winter, spring,

and summer, 95% of grid points were within 2.7°, 2.4°, and 3.6°C of NARR, respectively.

To investigate the possible reasons for the temperature biases, we compared the seasonal mean monthly net downward solar radiation at the surface from HadRM3P simulations and NARR (Fig. 3). Overall, the spatial pattern and seasonal cycle were well represented in the simulations. Negative biases in net downward solar radiation were present in all seasons, which ruled out solar radiation as the reason for the warm bias in summer. The biases were largest in spring and were mainly associated with mountain ranges such as the Cascades, Sierra Nevada, and Rockies, where large
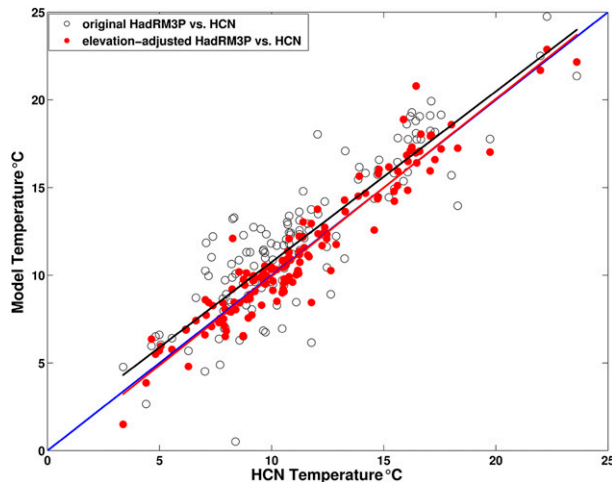
FIG. 4. The 50-yr average (1960–2009) temperatures (°C) from HadRM3P and the USHCN for 147 USHCN observing stations in a five-state region (Oregon, Idaho, Washington, California, and Nevada). Each dot denotes a different USHCN station. The open black circles show the original HadRM3P vs USHCN temperature, and the solid black line is the corresponding linear regression. The filled red circles show the elevation-adjusted HadRM3P vs USHCN temperature, and the solid red line is the corresponding linear regression. The 1:1 line is shown in blue.

snowpacks are present, and biases may be related to the snow–albedo feedback represented in the model. The specific mechanisms behind these biases still need further investigation.

For average temperature, good agreement was found with USHCN station data. Regression of the model tavg on USHCN tavg yielded a slope of 0.97 and an R-squared value of 0.76 (Fig. 4). Most of the scatter in the relationship can be explained by the difference in elevation between the 25-km model resolution and the actual elevation of the station; after removing the influence of elevation, the regression slope and R-squared value were improved to 1.01 and 0.89, respectively.

The magnitude of precipitation and its spatial pattern were characterized reasonably well by the regional simulations for all seasons (Fig. 5). During winter, the simulation showed a distinct spatial pattern that was strongly influenced by topography, just as in PRISM. The two precipitation bands along the U.S. West Coast corresponded to orographic precipitation associated with the coastal mountain ranges and, a little inland, Cascades and Sierra Nevada. Further inland, precipitation decreased in the basins and the intermountain zone and increased again as the prevailing westerly flow encounters the Rockies. With its coarser resolution, NARR precipitation showed less detailed change with respect to terrain. The HadRM3P-simulated precipitation exaggerated the orographic enhancement across

the coastal mountains, Cascades, and Sierra Nevada relative to both NARR and PRISM.

Seasonal regionwide model biases for mean temperature and precipitation and correlation coefficients between observed and simulated time-averaged spatial fields are presented in Tables 3 and 4. The spatial correlation between HadRM3P and NARR was higher for temperature than for precipitation in all seasons. For temperature, the spatial correlations were all above 0.95, with winter and spring season being as high as 0.98. The regional simulations showed cold bias in winter and spring and warm bias in summer and fall. The magnitude of summer bias was about 2–3 times the magnitude in other seasons. The spatial correlations for precipitation were above 0.7 for all seasons.

To further illustrate the influence of topography on simulated and observed temperature and precipitation, we examined winter and summer temperature and precipitation cross sections along 47.75°N (across the Olympic Mountains, northern Cascades, and Rockies) as in Salathé et al. (2010). Temperature gradients from HadRM3P along the transect were consistent with the observations (Fig. 6), though summer temperatures were too warm, echoing Fig. 2. Precipitation on the mountain ranges, including the westward shifts in the precipitation peak relative to the crest and the rapid drop in the lee (i.e., the rain shadow effect), were simulated realistically. However, HadRM3P exaggerated the orographic enhancement across the coastal mountains and Cascades relative to the observed. This pattern was seen in other transects (e.g., Sierra Nevada; shown in Fig. S1 of the supplemental information).

Comparing the seasonal mean diurnal temperature range (DTR) simulated by HadRM3P with NARR, we found the spatial pattern and seasonal cycle were relatively well represented in the simulations (Fig. 7). The model produced larger DTR in all three seasons and the largest DTR in spring, which suggests that the model produced less cloud cover than NARR. This larger DTR occurred because when there is less cloud cover, there is more incoming solar radiation reaching the surface during the day (more heating) and more longwave radiation escaping during nighttime (more cooling), producing a larger diurnal temperature range. A more complete diagnosis is beyond the scope of this paper, though this points to a need for further research.

Because of the interest in extreme weather events, we also compared observed and simulated winter minimum temperature (Tmin1; formed by averaging the coldest day for each of the three winter months) and summer maximum temperature (Tmax1; average of the hottest day for each of the summer months) (Fig. 8). The spatial

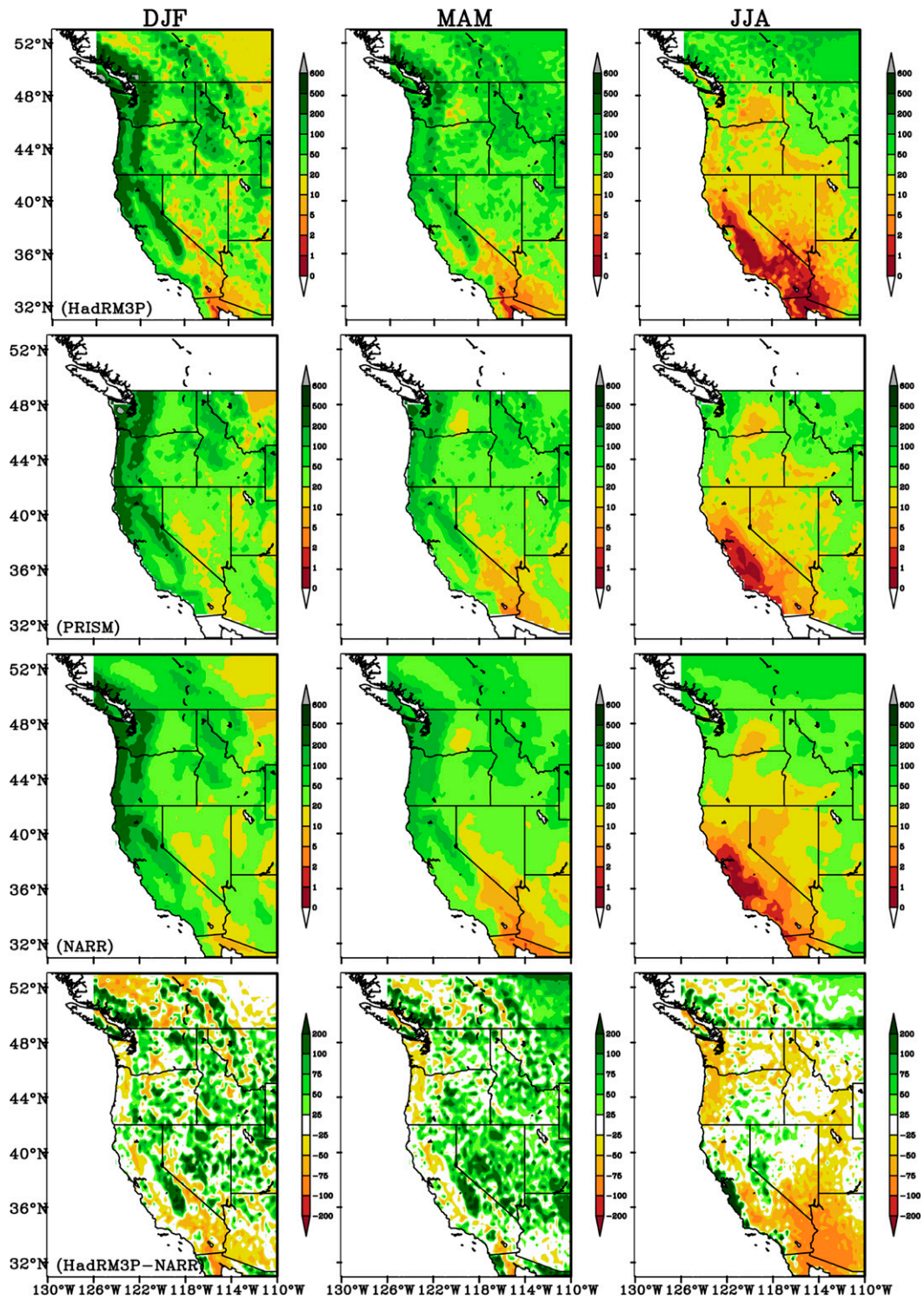FIG. 5. Seasonal mean precipitation (mm month^{-1}) in 1979–2009 for (left) DJF, (center) MAM, and (right) JJA from (top) HadRM3P, (middle top) PRISM, (middle bottom) NARR, and (bottom) the bias in percent (HadRM3P minus NARR).

TABLE 3. The 31-yr average (1979–2009) seasonal temperature (°C) from the HadRM3P SP experiment and from NARR, the bias (model minus observed), and the spatial correlation R over the western United States. (SON denotes September–November.)

Season	HadRM3P	NARR	Bias	R
DJF	0.20	1.16	−0.96	0.98
MAM	8.03	9.08	−1.04	0.98
JJA	21.96	20.26	1.70	0.96
SON	11.48	11.18	0.30	0.99

TABLE 4. The 31-yr average (1979–2009) seasonal precipitation (mm month^{-1}) from the HadRM3P SP experiment and from NARR, the bias (model minus observed), and the spatial correlation R over the western United States.

Season	Model	NARR	Bias	R
DJF	81.4	66.8	14.6	0.82
MAM	58.3	45.7	12.6	0.74
JJA	17.2	21.7	−4.4	0.75
SON	48.5	41.8	6.7	0.83

patterns of Tmin1 and Tmax1 simulated by HadRM3P resembled those of NARR. However, HadRM3P produced more extreme temperatures than NARR: simulated winter Tmin1 is much colder than observations, with the 10th percentile showing a cold bias of −10°C; simulated summer Tmax1 is warmer than observations, with the 90th percentile showing a warm bias of +6.40°C. It is worth noting here that tavg (Fig. 2) also showed cold bias in winter and warm bias in summer; therefore, some of the bias in Tmin1 and Tmax1 simply reflected the bias in the mean. When the bias in the mean was removed, the magnitude of the bias was reduced in both winter Tmin1 and summer Tmax1 (shown

in Fig. S2 of the supplemental material) while the spatial patterns of biases stayed the same in general.

4. ENSO teleconnections

The dependence on ENSO of winter season precipitation and temperature anomaly patterns in the western United States has been well studied (e.g., Ropelewski and Halpert 1986; Wallace et al. 1992; Gershunov 1998; Cayan et al. 1999). Here we evaluated HadRM3P's ability to simulate ENSO teleconnections as one measure of how well the climate model simulates climate variability.
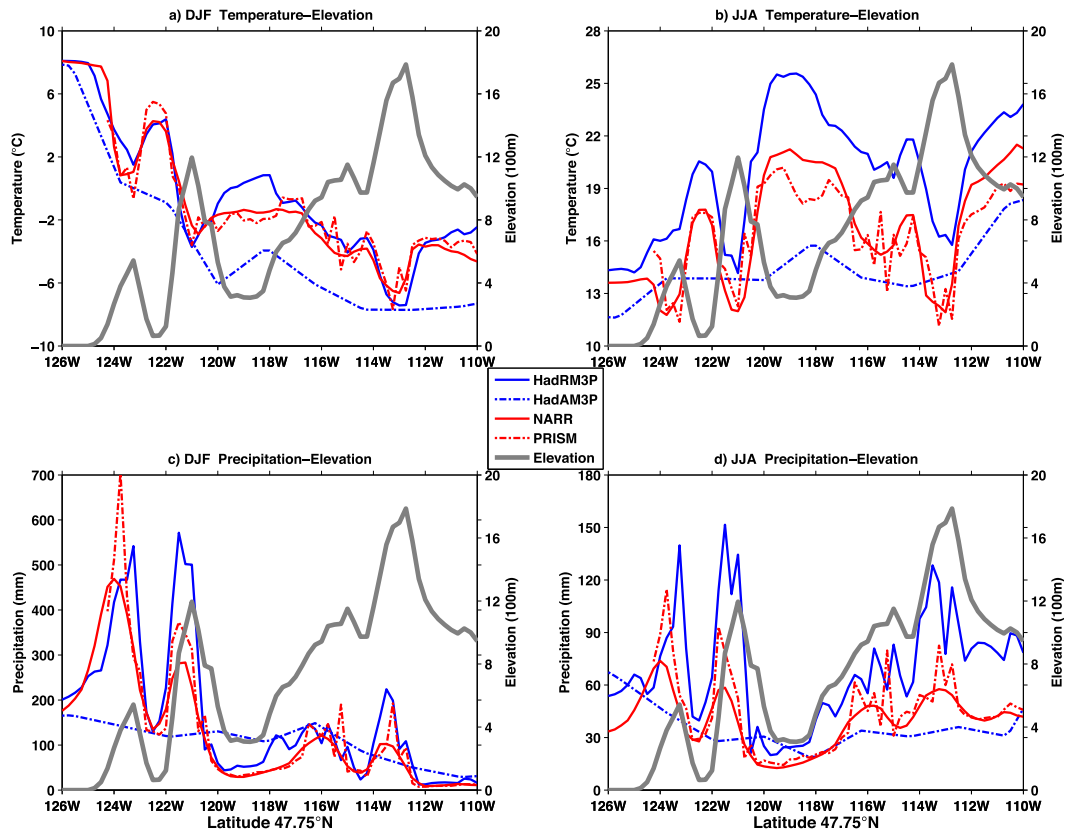

FIG. 6. Simulated and observed (a) DJF temperature, (b) JJA temperature, (c) DJF precipitation, and (d) JJA precipitation along a west–east transect at 47.75°N.
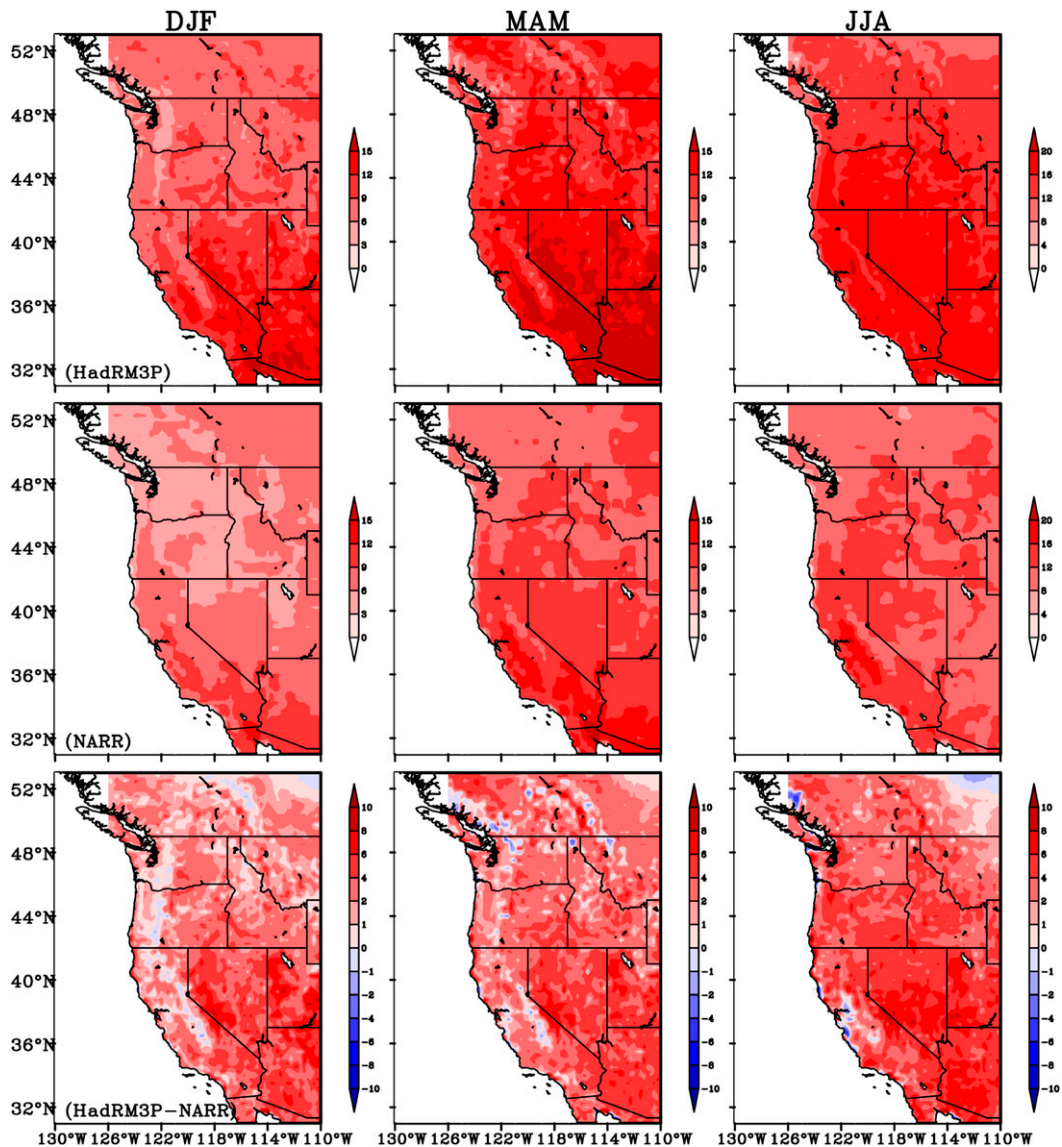
FIG. 7. Seasonal mean monthly DTR (°C) in 1979–2009 for (left) DJF, (center) MAM, and (right) JJA from (top) HadRM3P, (middle) NARR, and (bottom) the bias (HadRM3P minus NARR).

To test the ability of the regional model to simulate ENSO teleconnections, In this paper we showed the western U.S. winter season anomalies in temperature and precipitation associated with warm ENSO events during the 31-yr period of 1979–2009 from HadRM3P simulations and from various observational datasets (Figs. 9 and 10). The Niño-3.4 index (an average of sea surface temperature in the region bounded by 120°–170°W and 5°S–5°N) derived from the HadISST1 global sea surface temperature dataset (Rayner et al. 2003) was used to identify warm and cold ENSO events. Events were defined as five consecutive months [November–March (NDJFM)] at or above 1°C anomaly for warm phase and at or below −1°C for cold phase. Six warm

events (1983, 1987, 1992, 1995, 1998, and 2003) were identified and used in the following composite analysis. The climate anomalies were computed as the deviation of 3-month (December through February) means for the warm ENSO years from the 31-yr averages. As can be seen in Figs. 9 and 10, warm ENSO was characterized by warmer, drier conditions in the northwestern United States and cooler, wetter conditions in the southwestern United States in HadRM3P simulations. The spatial patterns of the HadRM3P-simulated and observed ENSO anomalies were not identical. For precipitation, the correlations between the HadRM3P-simulated anomaly pattern and those of NARR, PRISM, CRU TS, CPC, and GPCC are 0.85, 0.83, 0.82, 0.75, and 0.81,
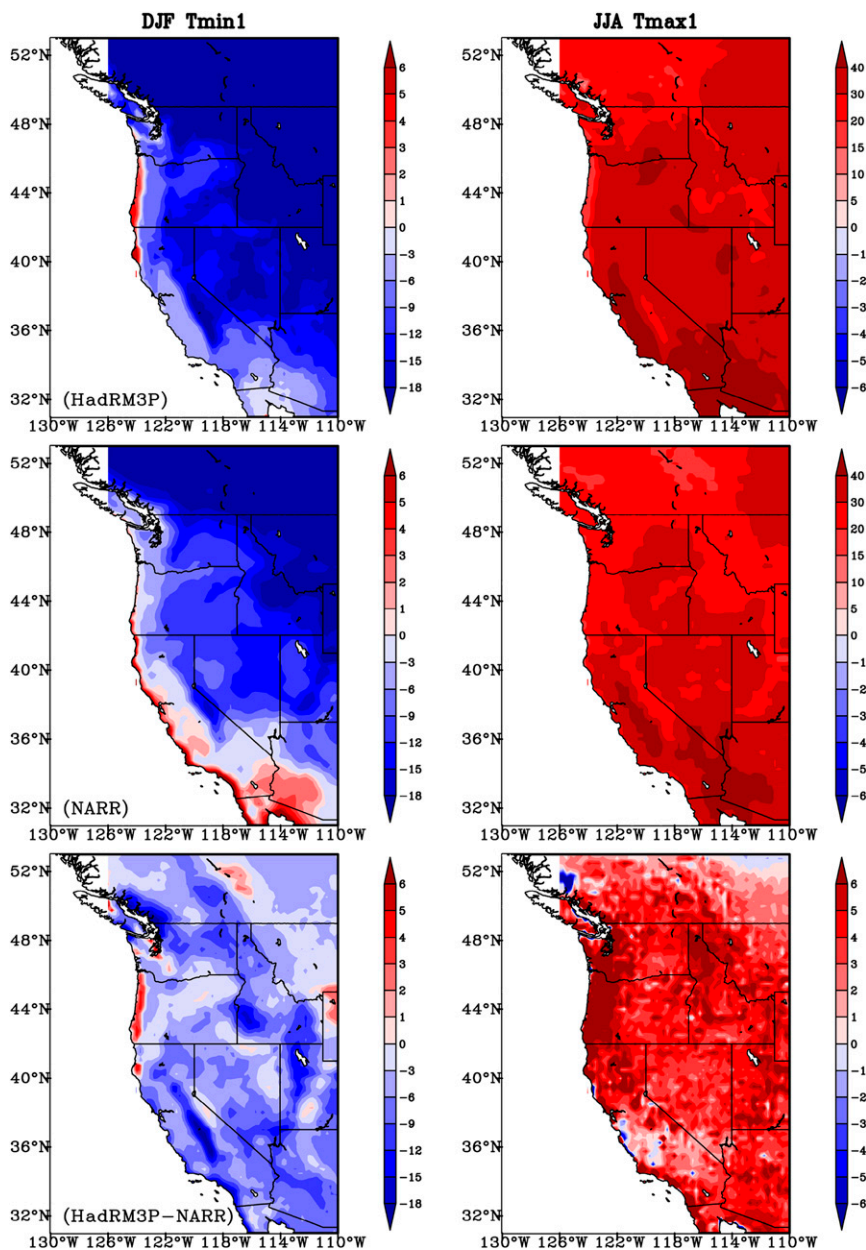
FIG. 8. Seasonal mean monthly Tmin1 (°C) in 1979–2009 for (left) DJF, and seasonal mean monthly Tmax1 (°C) in 1979–2009 for (right) JJA from (top) HadRM3P, (middle) NARR, and (bottom) the bias (HadRM3P minus NARR).

respectively; meanwhile, for temperature, the correlations between the HadRM3P-simulated anomaly pattern and those of NARR, PRISM, and CRU TS are 0.68, 0.69, and 0.76, respectively (the correlations for La Niña years are shown in Tables S1 and S2 in the supplemental material). For both temperature and precipitation, the transition of the sign of the anomalies occurred through Southern California, Nevada, and Utah in the HadRM3P simulations, while the location of the transition was not in complete agreement among observational datasets. For the

different observational datasets, the transition of the sign of the precipitation anomalies occurred roughly all along northeastern California, central Nevada, and southern Utah, while in the northern states the agreement was less strong than in the south. Observations showed wetter anomalies along the U.S. West Coast and much drier anomalies in northern Montana in warm ENSO events.

It is worth noting that in the results presented here, for each observation record, we had 6 examples of atmospheric response to El Niño forcing but 6 times ~500
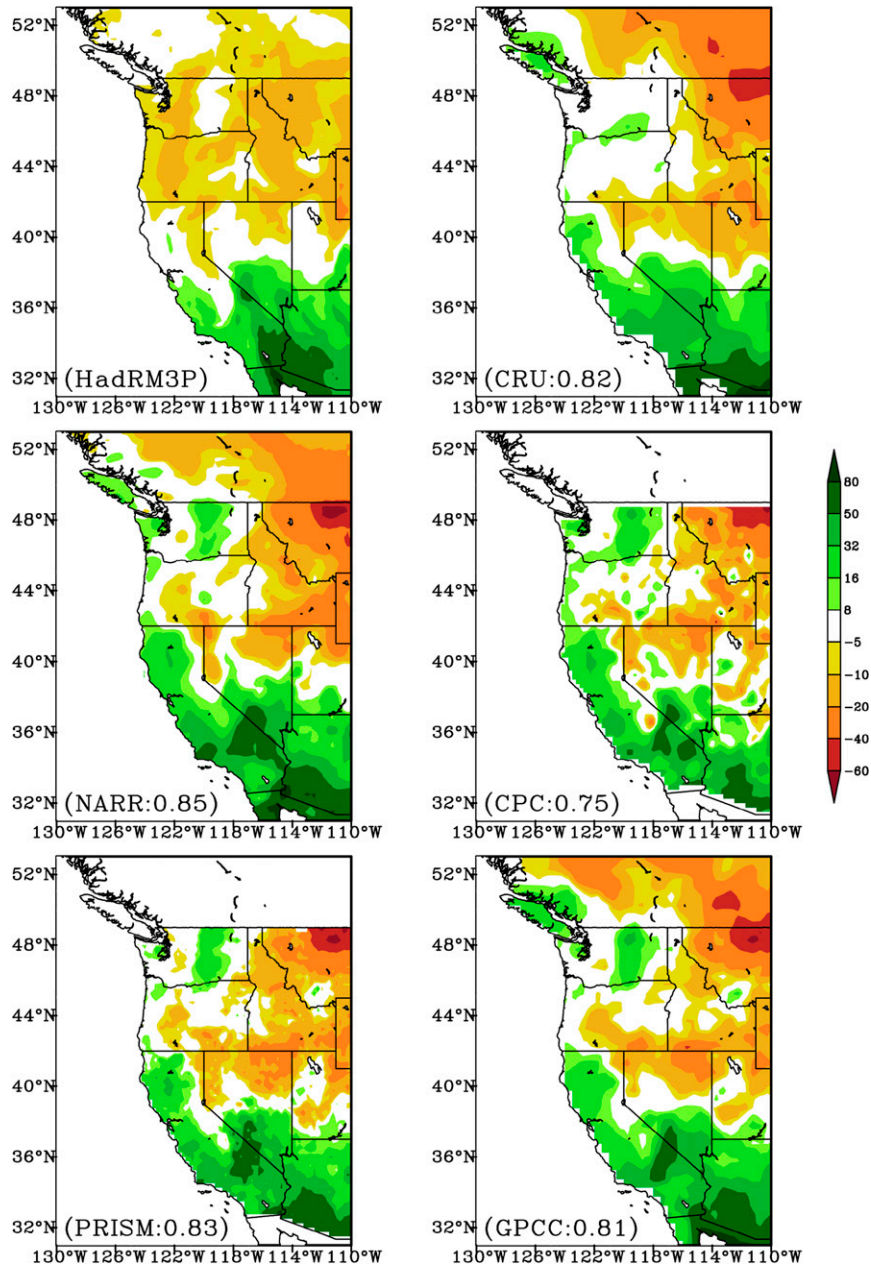
FIG. 9. El Niño years seasonal (DJF) mean precipitation anomaly (%) from HadRM3P and other observational datasets. The number in the bottom-left corner in each panel is the spatial correlation of that dataset with HadRM3P.

examples (depending on availability) from HadRM3P. Therefore, relative to the observed teleconnections, internal variabilities were reduced by averaging in the modeled ensemble, and this could account for some of the discrepancies between HadRM3P simulations and the observed. To verify this, we applied a *t* test at the 5% level to determine the statistical significance of the composites. Areas where the anomalies were statistically significant are shown in Figs. S3 and S4 of the supplemental material. For

each observation record, with a small sample size, the El Niño pattern was significant over a very small part of the domain—a very conservative view would compare models to observations only at those points that are statistically significant, although a more lenient view would consider the whole pattern even though in some places the anomalies cannot be differentiated from random noise. To explore the sensitivity of our comparison to sample size, we lowered the criteria to select ENSO events to at or above
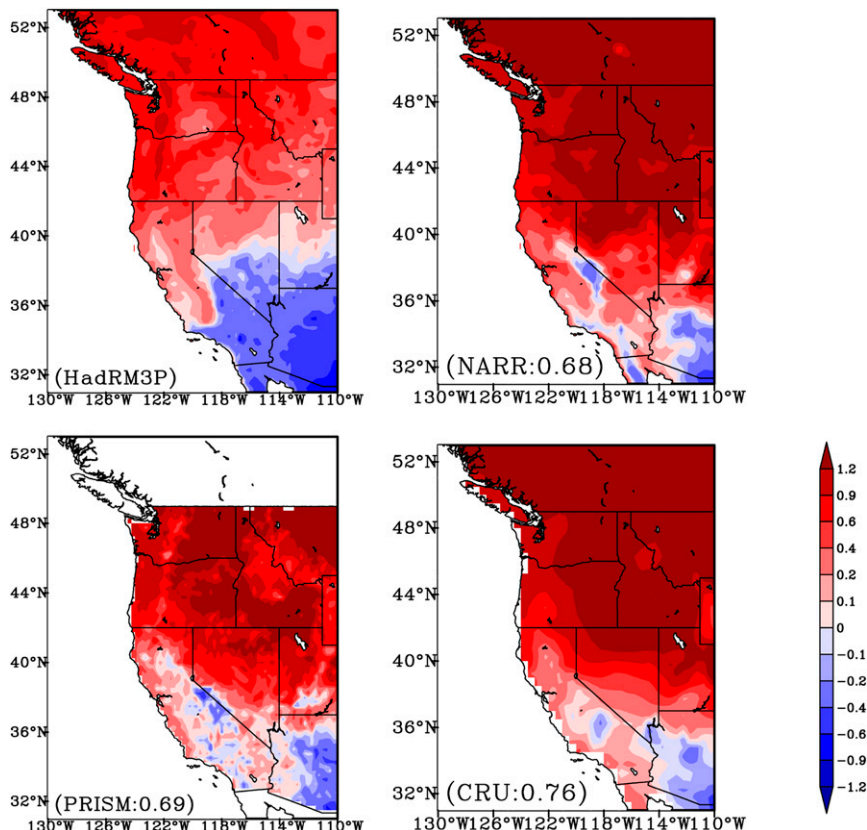
FIG. 10. As in Fig. 9, but for mean surface temperature anomaly (°C).

(at or below) +0.5°C (−0.5°C) for warm (cold) phase and expanded the period of record to 1960–2009, which increased the number of warm ENSO events to 16. NARR begins in 1979, so we excluded NARR in the following analysis. The result of changing the selection criteria was that the areas where the observed anomalies are statistically significant only decreased (Figs. S5 and S6 in the supplemental material). One possible explanation for this is that winter temperature and precipitation in the western United States are affected by many factors other than ENSO; the influence of ENSO on temperature and precipitation should be more detectable during strong ENSO episodes than during weak ones. When averaging strong and weak ENSO events together, the intensity of the impacts is less strong and the random noise is more evident than when looking at strong ENSO events alone. Based on the analysis done here, ENSO teleconnections are a weak metric for evaluating model performance, as noted also by Rupp et al. (2013).

5. Assessment of added value from a superensemble

We compared our results from weather@home to NARCCAP, a multimodel ensemble study covering the

United States, to demonstrate how our superensemble can augment studies like NARCCAP. For the western United States, normalized standard deviations and correlation coefficients for 21-yr annual average (December 1979 through November 1999) spatial fields of temperature (tas) and precipitation rate (pr) from 50-per-year weather@home ensemble members and for 21-yr annual averages over 1979–99 from 11 NARCCAP RCM–GCM simulations are shown in a Taylor (2001) diagram in Fig. 11. The reference field for the normalization and the correlation was NARR 1979–99. We compared only the time period for which all three datasets overlapped. Normalized standard deviation was calculated as the spatial standard deviation of a simulation divided by the spatial standard deviation of NARR. Note that a perfect simulation would have both a normalized standard deviation and a correlation equal to unity. The skill in simulated temperature was better in the 25-km resolution HadRM3P than in the 50-km resolution NARCCAP models, with correlation for HadRM3P typically >0.98 and normalized standard deviation all clustered close to unity. However, HadRM3P did not show better skill in precipitation simulation than NARCCAP. It is worth pointing out
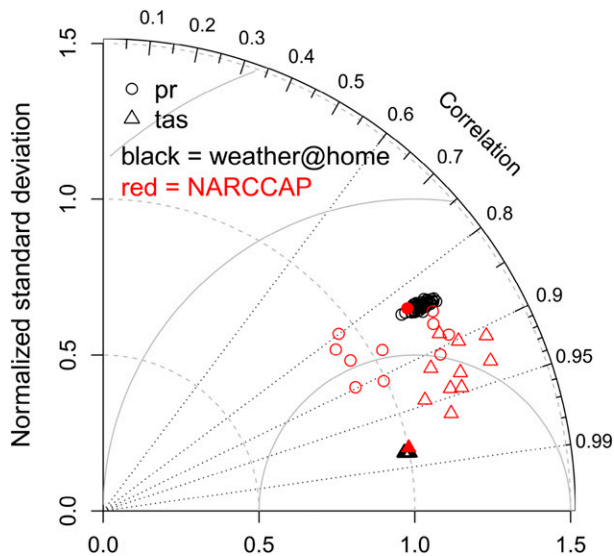
FIG. 11. Taylor (2001) diagram showing normalized standard deviations (radius) and correlation coefficients (angle) for annual tavg (tas; triangle) and precipitation rate (pr; circle) from weather@home (black), with each data point representing one of 50 ensemble members, averaged over 21 years (December 1979 through November 1999). Also shown are the same variables from 11 simulations from NARCCAP for the same time period (red). The solid red symbols correspond to the HadRM3–HadCM3 combination used in NARCCAP. The reference field for the normalization and the correlation is the NCEP NARR.

that the coupled HadRM3–HadCM3 (HRM3+HadCM3 in NARCCAP convention) demonstrated similar skills in simulating temperature (Fig. 11; filled triangle) and precipitation (Fig. 11; filled circle) as HadRM3P–HadAM3P, even though HadAM3P is an atmosphere-only model and SSTs are specified whereas HadCM3 is a coupled ocean–atmosphere model. This similarity between HadRM3P–HadAM3P and HadRM3–HadCM3 suggests that the dynamical coupling between ocean and atmosphere in NARCCAP did not explain most of the difference between HadRM3P–HadAM3P and the various NARCCAP RCM–GCM pairings but that the differences were due mainly to the atmospheric dynamics.

Looking at the 50 ensemble members from weather@home, even though each ensemble member was already averaged over 21 years, there was still notable spread for precipitation. This result highlights the need to run a number of simulations starting from different initial conditions with a given climate model to get to the "true" behavior of the model. Because any single model simulation contains internally generated variability (noise) and externally forced signal, only by averaging ensemble members starting from different initial states, while subjecting the model to the same external forcing,

can the random sequence of internal variability be reduced enough to reveal the true model behavior, as pointed out also by Deser et al. (2014).

The real power of large ensembles lies in the potential to simulate extreme events, as pointed out by Massey et al. (2015). To demonstrate this, we calculated the frequency distribution of errors with increasing ensemble size for the following statistics: JJA mean temperature, DJF mean precipitation, JJA 98th-percentile temperature, and DJF 2nd-percentile precipitation for one year over Los Angeles (Fig. 12). This analysis began with 2106 ensemble members taken from HadRM3P for the year 2008 averaged over a $1° \times 1°$ box using Los Angeles as the center point. From the population of 2106 values, 10 000 random samples of ensemble size N, with replacement, were taken, and the statistic (mean or percentile) was calculated for each ensemble. In Fig. 12, the box (inner quartiles) and whiskers (5th and 95th percentile) summarized the frequency distribution of the statistic for each N. This figure demonstrated the power of creating multiple realizations of 1 year's worth of "weather" to narrow the confidence limits about estimation by the model. Figure 12a showed that to be within $±1°$C of the true simulated summer mean temperature at the given level of confidence (within the 5th and 95th percentile), we needed at least 16 ensemble members; for $±0.5°$C, we needed 64 ensemble members. Figure 12b showed that 16 ensemble members would get us within -60% to $+80\%$ of the true simulated winter precipitation for Los Angeles, while it would take 64 to be within $±50\%$.

For the mean, multiple years in a time series can, to some degree, substitute for multiple ensemble members for a single year. The advantage of the large ensemble is clearer when analyzing the tails of the distribution. For example, for the summer 98th-percentile temperature (Fig. 12c), 128 ensemble members were required to have an error within a range of $±0.5°$C at the given confidence level. For the winter 2nd-percentile precipitation (Fig. 12d), more than 128 ensemble members were needed for an error range of $±100\%$. Running large numbers of simulations of possible weather under the same external forcing affords us the opportunity to simulate impact-relevant extreme events and to obtain statistics on extreme weather.

So far, we have focused our analysis on the SP ensemble. Our PP experiment affords us the opportunity to explore parameter uncertainties and the effect of different parameter perturbations. As shown in section 4, HadRM3P SP simulations exaggerated the orographic enhancement across the coastal mountains, Cascades, and Sierra Nevada compared with observations. Here we examine whether certain parts of the parameter space lead to better model performance in
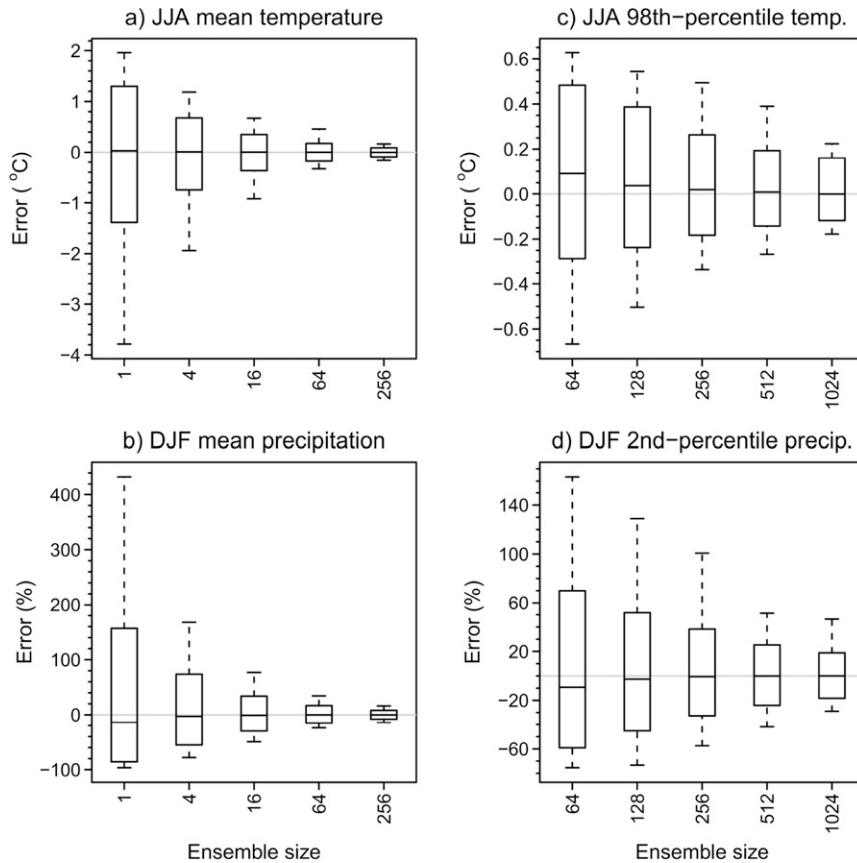
FIG. 12. The distribution of the deviation, or error, in the ensemble statistic from the population statistic with increasing ensemble size for (a) JJA mean temperature, (b) DJF mean precipitation, (c) JJA 98th-percentile temperature, and (d) DJF 2nd-percentile precipitation from the HadRM3P for one year (December 2008 through November 2009) for Los Angeles.

simulating orographic enhancement and whether there are certain physical processes that could be parameterized better to give a more realistic simulation. To explore this possibility, in Fig. 13 we showed the winter season precipitation from parameter sets P1, P2, and P3 along a 47.75°N transect, as in Fig. 6. Each transect in Fig. 13 was the averaged results of 110 different initial condition perturbations, so the internal variability had been minimized, and the differences should be mainly due to different parameter values. There was a clear distinction between the precipitations produced by P1, P2, and P3, especially across the coastal mountains and Cascades; results for all three parameter sets were very similar east of the Cascades. The default P1 lay in between the drier P2 and the wetter P3, as expected. These results underscore the effects of different parameter perturbations on regional precipitation, and the large variation of precipitation values produced by different parameter perturbation combinations suggests that there are certain perturbations that will produce a model variant that simulates the orographic enhancement

more realistically than the standard version. A thorough discussion of parameter combinations and relevant physical processes involved is beyond the scope of this paper and will be explored in a subsequent paper; we have presented these results as an example of the strength of being able to generate supersensembles.

6. Conclusions and further discussion

This paper evaluated properties of the climate of the western United States as simulated by a regional climate model in the weather@home system. The regional simulations reproduced many climate features that are important in the western United States and added more detail at the regional scales. The spatial patterns of temperature and precipitation in the western United States were much more accurately represented in the regional simulations than in the global simulations.

Overall, temperature was well represented in the simulations and the influences of the major geographical features were reproduced; 95% of grid points were
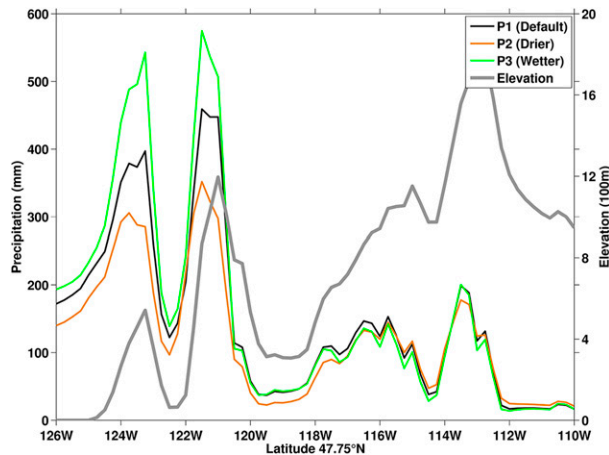
FIG. 13. Simulated DJF precipitation along the same west–east transect as in Fig. 6, from P1 (default), P2 (drier), and P3 (wetter).

within 2.7°, 2.4°, and 3.6°C of NARR for the means in winter, spring, and summer, respectively. There was a pervasive summer warm bias over the western United States except central California and southern Nevada. Though we ruled out bias in solar radiation as a possible cause, further investigation of model physics, particularly cloud-related parameterizations, is warranted. For extreme temperature, in most places, winter extremes of Tmin1 were somewhat lower than observed and summer Tmax1 values were higher than observed. This should be taken into consideration if and when this model is used to make projections of future changes in extreme weather events at the local scale described here. Future work could explore whether parameter perturbations could improve the simulation of extremes.

The overall magnitude of precipitation and its geographical features were reasonably well characterized by the regional simulations for all seasons, though simulated precipitation exaggerated the orographic enhancement across the coastal mountains, Cascades, and Sierra Nevada compared with observations. The importance of topographic control on regional climate conditions was illustrated through the examination of temperature/precipitation–topography relationships. Analyses of temperature/precipitation and topography along an east–west transect showed significant impacts of surface terrain on the spatial pattern of precipitation. Rain shadow effects were captured well along the coast, Cascades, and Sierra Nevada; the transects of precipitation showed many features of the observed transect, including the location of peak precipitation west of the crest and a rapid drop by a factor of 5 or more in the lee.

The HadRM3P simulations produced warm/dry northwest and cool/wet southwest U.S. patterns associated with warm ENSO. However, there were also

notable differences, including the locations of the transition of the response from warm (dry) to cool (wet) in the anomaly fields, with the transition of anomaly signs extending farther south in California and Nevada than what was seen in the observations. However, the sample size of the observational records was too limited to conclude that the placement of the zero line in HadRM3P was significantly different from observed.

In this paper we showed how our superensemble augments other regional ensemble modeling studies through a comparison with NARCCAP. We also demonstrated the power of a superensemble by showing how as more ensemble members were included in the analysis, the signal-to-noise ratio improved sufficiently to estimate with high precision not only the means but more importantly the extremes. Historically, the strength of regional model simulations (viz., a spatial resolution high enough to resolve key features like rain shadows) was offset by their key weakness (viz., only one or a few runs—too few to determine whether differences between runs were meaningful or just statistical noise) (see O'Brien et al. 2011). This experiment allows us to run simulations at a high enough resolution to resolve key regional features and to run multiple ensemble members to provide robust assessments of physically meaningful forced signals as opposed to internally generated variability. Our experimental results can supplement studies like NARCCAP by providing superensemble results for western North America, allowing more complete characterization of natural variability, exploration of the effect of different parameterizations, and better analysis of extreme events. Discussions about uncertainty due to internally generated variability have come to the fore in the past few years, and global models (e.g., CCSM3) have been used in many studies (Deser et al. 2012a,b; Oshima et al. 2012; Kang et al. 2013; Hu and Deser 2013; Wettstein and Deser 2014; Wallace et al. 2015) to investigate uncertainty due to internal variability on a large scale. Our experiment provides the opportunity to investigate uncertainty due to internal variability on a finescale regional level. We also demonstrated that by exploring different parameter combinations, we could produce model variants that do better than the standard version, which could lead to optimal regional performance. Through the control of initial condition perturbations and the sampling of model variants in the parameter space (though not in model structure space), we can utilize this large ensemble to better understand two of the major sources of uncertainty: initial condition and model response, within the models used here, specifically for this region.

REFERENCES

Allen, M. R., 1999: Do-it-yourself climate prediction. *Nature*, **401**, 642, doi:10.1038/44266.

Cayan, D. R., K. T. Redmond, and L. G. Riddle, 1999: ENSO and hydrologic extremes in the western United States. *J. Climate*, **12**, 2881–2893, doi:10.1175/1520-0442(1999)012<2881: EAHEIT>2.0.CO;2.

Christensen, J. H., and O. B. Christensen, 2007: A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change*, **81**, 7–30, doi:10.1007/s10584-006-9210-7.

Daly, C., M. D. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. A. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, doi:10.1002/joc.1688.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, doi:10.1038/nclimate1562.

——, A. S. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:10.1007/s00382-010-0977-x.

——, ——, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, doi:10.1175/ JCLI-D-13-00451.1.

Dickinson, R. E., R. M. Errico, F. Giorgi, and G. T. Bates, 1989: A regional climate model for the western United States. *Climatic Change*, **15**, 383–422, doi:10.1007/BF00240465.

Duffy, P. B., and Coauthors, 2006: Simulations of present and future climates in the western United States with four nested regional climate models. *J. Climate*, **19**, 873–895, doi:10.1175/JCLI3669.1.

Dulière, V., Y. Zhang, and E. P. Salathé Jr., 2011: Extreme precipitation and temperature over the U.S. Pacific Northwest: A comparison between observations, reanalysis data, and regional models. *J. Climate*, **24**, 1950–1964, doi:10.1175/ 2010JCLI3224.1.

Gershunov, A., 1998: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Implications for long-range predictability. *J. Climate*, **11**, 3192–3203, doi:10.1175/1520-0442(1998)011<3192: EIOIER>2.0.CO;2.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168, doi:10.1007/s003820050010.

Grabowski, W. W., 2000: Cloud microphysics and the tropical climate: Cloud-resolving model perspective. *J. Climate*, **13**, 2306–2322, doi:10.1175/1520-0442(2000)013<2306:CMATTC>2.0.CO;2.

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

Hu, A., and C. Deser, 2013: Uncertainty in future regional sea level rise due to internal climate variability. *Geophys. Res. Lett.*, **40**, 2768–2772, doi:10.1002/grl.50531.

Jones, R. G., M. Noguer, D. Hassell, D. Hudson, S. Wilson, G. Jenkins, and J. Mitchell, 2004: Generating high resolution climate change scenarios using PRECIS. Met Office Hadley Centre Rep., 40 pp. [Available online at http://www. metoffice.gov.uk/media/pdf/6/5/PRECIS_Handbook.pdf.]

Kang, S. M., C. Deser, and L. M. Polvani, 2013: Uncertainty in climate change projections of the Hadley circulation: The role of internal variability. *J. Climate*, **26**, 7541–7554, doi:10.1175/ JCLI-D-12-00788.1.

Knight, C. G., and Coauthors, 2007: Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *Proc. Natl. Acad. Sci. USA*, **104**, 12 259–12 264, doi:10.1073/pnas.0608144104.

Leung, L. R., Y. Qian, and X. Bian, 2003a: Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part I: Seasonal statistics. *J. Climate*, **16**, 1892–1911, doi:10.1175/1520-0442(2003)016<1892: HOTWUS>2.0.CO;2.

——, ——, ——, and A. Hunt, 2003b: Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part II: Mesoscale ENSO anomalies. *J. Climate*, **16**, 1912–1928, doi:10.1175/1520-0442(2003)016<1912: HOTWUS>2.0.CO;2.

Massey, N., and Coauthors, 2015: Weather@home: Development and validation of a very large ensemble modelling system for probabilistic event attribution. *Quart. J. Roy. Meteor. Soc.*, **141**, 1528–1545, doi:10.1002/qj.2455.

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

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

——, and Coauthors, 2014: The North American Regional Climate Change Assessment Program dataset. National Center for Atmospheric Research Earth System Grid data portal, accessed 23 April 2015, doi:10.5065/D6RN35ST.

Menne, M. J., C. N. Williams Jr., and R. S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1007, doi:10.1175/ 2008BAMS2613.1.

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

Mote, P. W., M. R. Allen, R. G. Jones, S. Li, R. Mera, D. E. Rupp, A. Salahuddin, and D. Vickers, 2015: Superensemble regional climate modeling for the western United States. *Bull. Amer. Meteor. Soc.*, in press.

O'Brien, T. A., L. C. Sloan, and M. A. Snyder, 2011: Can ensembles of regional climate model simulations improve results from sensitivity studies? *Climate Dyn.*, **37**, 1111–1118, doi:10.1007/s00382-010-0900-5.

Oshima, K., Y. Tanimoto, and S.-P. Xie, 2012: Regional patterns of wintertime SLP change over the North Pacific and their uncertainy in CMIP3 multi-model projections. *J. Meteor. Soc. Japan*, **90**, 385–396, doi:10.2151/jmsj.2012-A23.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, 2013: Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *J. Geophys. Res. Atmos.*, **118**, 10 884–10 906, doi:10.1002/jgrd.50843.

Salathé, E. P., Jr., R. Steed, C. F. Mass, and P. H. Zahn, 2008: A high-resolution climate model for the U.S. Pacific Northwest: Mesoscale feedbacks and local responses to climate change. *J. Climate*, **21**, 5708–5726, doi:10.1175/2008JCLI2090.1.

——, L. R. Leung, Y. Qian, and Y. Zhang, 2010: Regional climate model projections for the state of Washington. *Climatic Change*, **102**, 51–75, doi:10.1007/s10584-010-9849-y.

Sanderson, B. M., C. Piani, W. J. Ingram, D. A. Stone, and M. R. Allen, 2007: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Climate Dyn.*, **30**, 175–190, doi:10.1007/s00382-007-0280-7.

——, and Coauthors, 2008a: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *J. Climate*, **21**, 2384–2400, doi:10.1175/2008JCLI1869.1.

——, C. Piani, W. Ingram, D. Stone, and M. Allen, 2008b: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Climate Dyn.*, **30**, 175–190, doi:10.1007/s00382-007-0280-7.

——, K. Shell, and W. Ingram, 2010: Climate feedbacks determined using radiative kernels in a multi-thousand member ensemble of AOGCMs. *Climate Dyn.*, **35**, 1219–1236, doi:10.1007/s00382-009-0661-1.

Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406, doi:10.1038/nature03301.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, doi:10.1029/2000JD900719.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate*, **5**, 561–576, doi:10.1175/1520-0442(1992)005<0561:SVDOWS>2.0.CO;2.

——, C. Deser, B. V. Smoliak, and A. S. Philips, 2015: Attribution of climate change in the presence of internal variability. *Climate Change: Multidecadal and Beyond*, C. P. Chang et al., Eds., World Scientific Series on Asia-Pacific Weather and Climate, Vol. 6, World Scientific, in press.

Wettstein, J. J., and C. Deser, 2014: Internal variability in projections of twenty-first-century Arctic sea ice loss: Role of the large-scale atmospheric circulation. *J. Climate*, **27**, 527–550, doi:10.1175/JCLI-D-12-00839.1.

Wu, X., 2002: Effects of ice microphysics on tropical radiative–convective–oceanic quasi-equilibrium states. *J. Atmos. Sci.*, **59**, 1885–1897, doi:10.1175/1520-0469(2002)059<1885:EOIMOT>2.0.CO;2.

Zhang, Y., V. Dulière, P. Mote, and E. P. Salathé Jr., 2009: Evaluation of WRF and HadRM mesoscale climate simulations over the U.S. Pacific Northwest. *J. Climate*, **22**, 5511–5526, doi:10.1175/2009JCLI2875.1.