THE NORTH AMERICAN REGIONAL CLIMATE CHANGE ASSESSMENT PROGRAM
Overview of Phase I Results

BY LINDA O. MEARNS, RAY ARRITT, SÉBASTIEN BINER, MELISSA S. BUKOVSKY, SETH MCGINNIS, STEPHAN SAIN, DANIEL CAYA, JAMES CORREIA JR., DAVE FLORY, WILLIAM GUTOWSKI, EUGENE S. TAKLE, RICHARD JONES, RUBY LEUNG, WILFRAN MOUFOUMA-OKIA, LARRY MCDANIEL, ANA M. B. NUNES, YUN QIAN, JOHN ROADS, LISA SLOAN, AND MARK SNYDER

To investigate uncertainties in regional-scale climate projections, NARCCAP evaluated temperature and precipitation results from six regional climate models driven by NCEP–DOE Reanalysis II boundary conditions for 1980–2004.

From a global climate point of view, three main uncertainties have been identified regarding projections of future climate in the twenty-first century: internal (natural) variability of the climate system; the trajectories of future emissions of greenhouse gases and aerosols; and the response of the climate system (represented by various global climate models) to any given set of future emissions/concentrations (Cubasch et al. 2001; Meehl et al. 2007; Hawkins and Sutton 2009). However, as greater interest and concern has been focused on the regional scale of climate change and the desire for greater regional detail continues to grow, the uncertainty due to the application of regional climate models to the climate change problem introduces an additional uncertainty (Giorgi et al. 2001; Christensen et al. 2007a): the uncertainty entailed in dynamical downscaling from coarse to fine resolution. This uncertainty in the regional climate response has been documented in numerous contexts (Giorgi et al. 2001; Christensen et al. 2007a,b; de Elía et al. 2008) and extends to uncertainties in climate impacts (Mearns et al. 2001; Mearns 2003; Wilby et al. 2000; Wood et al. 2004;

Oleson et al. 2007; Morse et al. 2009). European research has moved forward to systematically examine the combined uncertainty in future climate projections from global and regional models (Christensen et al. 2007b, 2009), but no such research program has heretofore been developed over North America (NA). We developed the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al. 2009) to fill this research gap.

The fundamental scientific motivation of NARCCAP is 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 regional climate models (RCMs). An additional and equally important (and related) motivation for this project is to provide the climate impacts and adaptation community with regionally resolved climate change projections that can be used as the basis for studies of the societal impacts of climate change. Because we are using multiple regional and global climate models, impacts researchers will have the ingredients to produce impacts assessments that

characterize multiple uncertainties. Additional goals of the program include the following: to evaluate regional model performance over North America by nesting the RCMs in National Centers for Environmental Prediction–Department of Energy (NCEP–DOE) Reanalysis; to explore some remaining uncertainties in regional climate modeling (e.g., importance of compatibility of physics in nesting and nested models); and to enhance collaboration among the U.S., Canadian, and European climate modeling groups, leveraging the diverse modeling capability across the countries.

As impressive as the two European programs exploring regional future climate uncertainty are, they also have limitations. In Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE), most of the regional models used only one or at most two AOGCMs for boundary conditions and few used more than one emission scenario (A2). Ensemble-Based Predictions of Climate Changes and Their Impacts (ENSEMBLES) improved on the structure of PRUDENCE by using more global models to drive more RCMs (Christensen et al. 2009), but there was no effort to balance which and how many AOGCMs were used by the 15 RCMs involved. In NARCCAP, we have created a smaller but more balanced program that focuses on the uncertainty of the different AOGCMs and RCMs, using only one emissions scenario.

AFFILIATIONS: MEARNS, BUKOVSKY, MCGINNIS, SAIN, AND MCDANIEL—National Center for Atmospheric Research, Boulder, Colorado; ARRITT, FLORY, GUTOWSKY, AND TAKLE—Iowa State University, Ames, Iowa; BINER AND CAYA—Ouranos, Montreal, Quebec, Canada; CORREIA, LEUNG, AND QIAN—Pacific Northwest National Laboratory, Richland, Washington; JONES AND MOUFOUMA-OKIA—Hadley Centre, Exeter, United Kingdom; NUNES—Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, and Scripps Institute of Oceanography, University of California, San Diego, La Jolla, California; ROADS*—Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California; SLOAN AND SNYDER—University of California, Santa Cruz, Santa Cruz, California

* Deceased.

CORRESPONDING AUTHOR: Linda O. Mearns, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO, 80307

E-mail: lindam@ucar.edu

PROGRAM STRUCTURE. The program includes two main phases: phase I, wherein six RCMs use boundary conditions from the NCEP–DOE Reanalysis II (R2) for a 25-yr period (1980–2004), and phase II, wherein the boundary conditions are provided by four AOGCMs for 30 years of current climate (1971–2000) and 30 years of a future climate (2041–70) for the Special Report on Emissions Scenarios (SRES) A2 emissions scenario (Nackicenovic et al. 2000). The simulation domain of the RCMs covers northern Mexico, all of the lower 48 U.S. states, and most of Canada (up to about 60°N) (Fig. 1).

The six RCMs used are as follows: the Canadian RCM (CRCM; Caya and LaPrise 1999), the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5; Grell et al. 1993), the Met Office Hadley Centre's regional climate model version 3 (HadRM3; Jones et al. 2003), the Regional Climate Model version 3 (RegCM3; Giorgi et al. 1993a,b; Pal et al. 2000, 2007), the Scripps Experimental Climate Prediction Center (ECPC) Regional Spectral Model (RSM; Juang et al. 1997), and the Weather Research and Forecasting model (WRF; Skamarock et al. 2005). Five of these models (all but HadRM3) have previously been run over a North American domain using boundary conditions from both reanalyses and AOGCMs.[1] Most have also participated in the Project to Intercompare Regional Climate Simulations (PIRCS; Takle et al. 1999). These particular models were chosen to provide a variety of model physics and/or to use models that have already performed multiyear climate change experiments, preferably in a transient mode. A table displaying the major characteristics of the regional models may be found online (at www.narccap .ucar.edu). Note that some of these model systems (e.g., WRF and MM5) provide multiple options for model parameterizations and submodels; the evalu-

---

[1] MM5 [Leung et al. (2003a,b, 2004) for western United States and Leung and Gustafson (2005) and Gustafson and Leung (2007) for conterminous United States]; various versions of RegCM over the western United States (Giorgi et al. 1998; Bell et al. 2004; Snyder et al. 2002; Diffenbaugh et al. 2004), the southeastern United States (Mearns et al. 2003), and the entire continental United States (Pan et al. 2001; Diffenbaugh et al. 2005); the RSM over the continental United States (Roads 2003; Han and Roads 2004; Nunes and Roads 2007); the Canadian RCM [Laprise et al. (1998, 2003) and Caya and LaPrise (1999) over western and eastern Canada and Plummer et al. (2006) for the whole continent]; and WRF over the western United States (Salathé et al. 2008, 2010).

ations in this paper are valid only for the particular model configurations specified in the table.

Two of the RCMs (RSM and CRCM) use spectral nudging, which provides information from the nesting model not only for the lateral domain boundaries and initial conditions but also throughout the domain and vertical levels. Thus, these models are more directly constrained to follow the nesting model, and in some analyses we compare ensembles of the nudged and non-nudged models with individual models and the full ensemble.

Phase I: Simulations using reanalysis boundary conditions. The RCM simulations for phase I use boundary conditions from the NCEP–DOE R2 for a 25-yr period (1980–2004). Evaluation of such runs is a crucial prerequisite to generating climate scenarios and characterizing their uncertainties (Pan et al. 2001) and was performed in the Modelling European Regional Climate, Understanding and Reducing Errors (MERCURE) program precursor to PRUDENCE (Christensen et al. 1997), its successor ENSEMBLES (Christensen et al. 2009), the Regional Climate Model Intercomparison Project (RMIP) over China (Fu et al. 2005), and the PIRCS program (Takle et al. 1999; Pan et al. 2001; Anderson et al. 2003).

Phase II: Current and future climate simulations. Although this paper concerns phase I, we present a brief overview of phase II to establish the full context of the program and because results of phase I contribute to the characterization of uncertainty in phase II.

In phase II, we use boundary conditions from four AOGCMs: the NCAR Community Climate System Model, version 3 (CCSM3; Collins et al. 2006); the Canadian Climate Centre Coupled General Circulation Model version 3 (CGCM3; Scinocca and McFarlane 2004; Flato 2005); the HadCM3 (Gordon et al. 2000; Pope et al. 2000); and the Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL CM2.1; GFDL Global Atmospheric Model Development Team 2004). Simulations using the SRES A2 emissions scenario have been performed with all of these models, and all have saved output at 6-hr intervals appropriate for driving RCMs. Phase II also includes two high-resolution (50 km) time-slice experiments using the Community Atmosphere
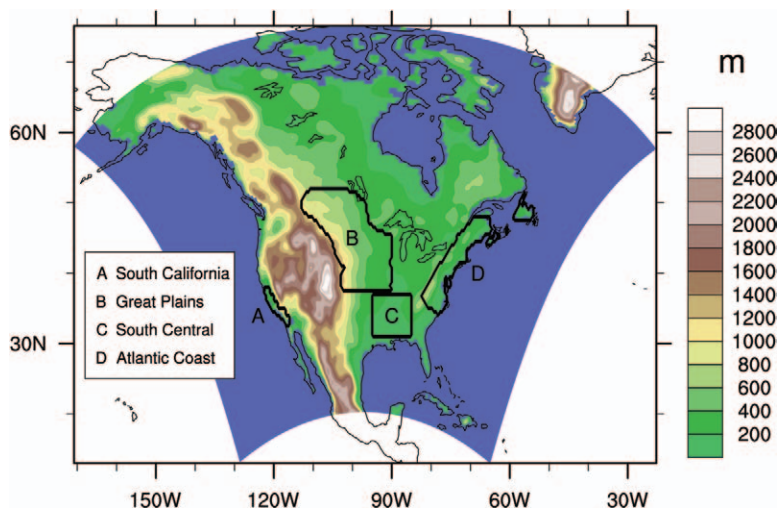

FIG. 1. Domain for NARCCAP as represented by the RegCM3 regional climate model, at 50-km gridpoint spacing over North America showing topography. Colors refer to elevations in meters. Four regions analyzed in detail are delineated by black outlines.

Model (CAM; Govindasamy et al. 2003) and the atmospheric component of the CM2.1 (AM2.1; GFDL Global Atmospheric Model Development Team 2004), the respective atmospheric model components of the CCSM and GFDL AOGCMs. In these experiments, the atmospheric models are run using observed SSTs and sea ice for lower boundary conditions in the current period and the same observations with an offset calculated from current and future runs of the corresponding AOGCM in the future period.

Characterization of uncertainty. Because limited funding precluded the simulation of all 24 nesting combinations, we adopted a balanced fractional factorial design to sample half of the 4 × 6 matrix in a statistically meaningful way that maximizes the amount of information that can be obtained from the experiment. In this type of design as applied here, in the 12 pairings so chosen, each AOGCM provides boundary conditions for three different RCMs and each RCM uses two different AOGCMs for boundary conditions. In addition, each RCM uses one of the AOGCMs that has a corresponding time-slice experiment. Table 1 shows the resulting matrix of AOGCM–RCM pairings.

NARCCAP is paying particular attention to quantifying uncertainty, using a Bayesian probabilistic approach (cf. Tebaldi et al. 2004, 2005) to characterize the joint uncertainty in multimodel ensembles on a regional scale for temperature and precipitation (Tebaldi and Sansó 2009).We will improve on these methods by employing nested Bayesian models that

Table 1. RCM–AOGCM combinations simulated in NARCCAP.

		AOGCM			
		GFDL	CGCM3	HadCM3	CCSM3
RCM	MM5			X	X
	RegCM3	X	X		
	CRCM		X		X
	HadRM3	X		X	
	RSM	X		X	
	WRF		X		X
	Corresponding time-slice experiment				
		AM2.1			CAM3

because it was developed at the same spatial resolution as the regional model experiments, covers the same spatial and temporal range, and includes an elevation correction for temperature. The station data used include thousands of station records of monthly temperature and precipitation developed in Legates and Willmott (1990a,b), which employs a st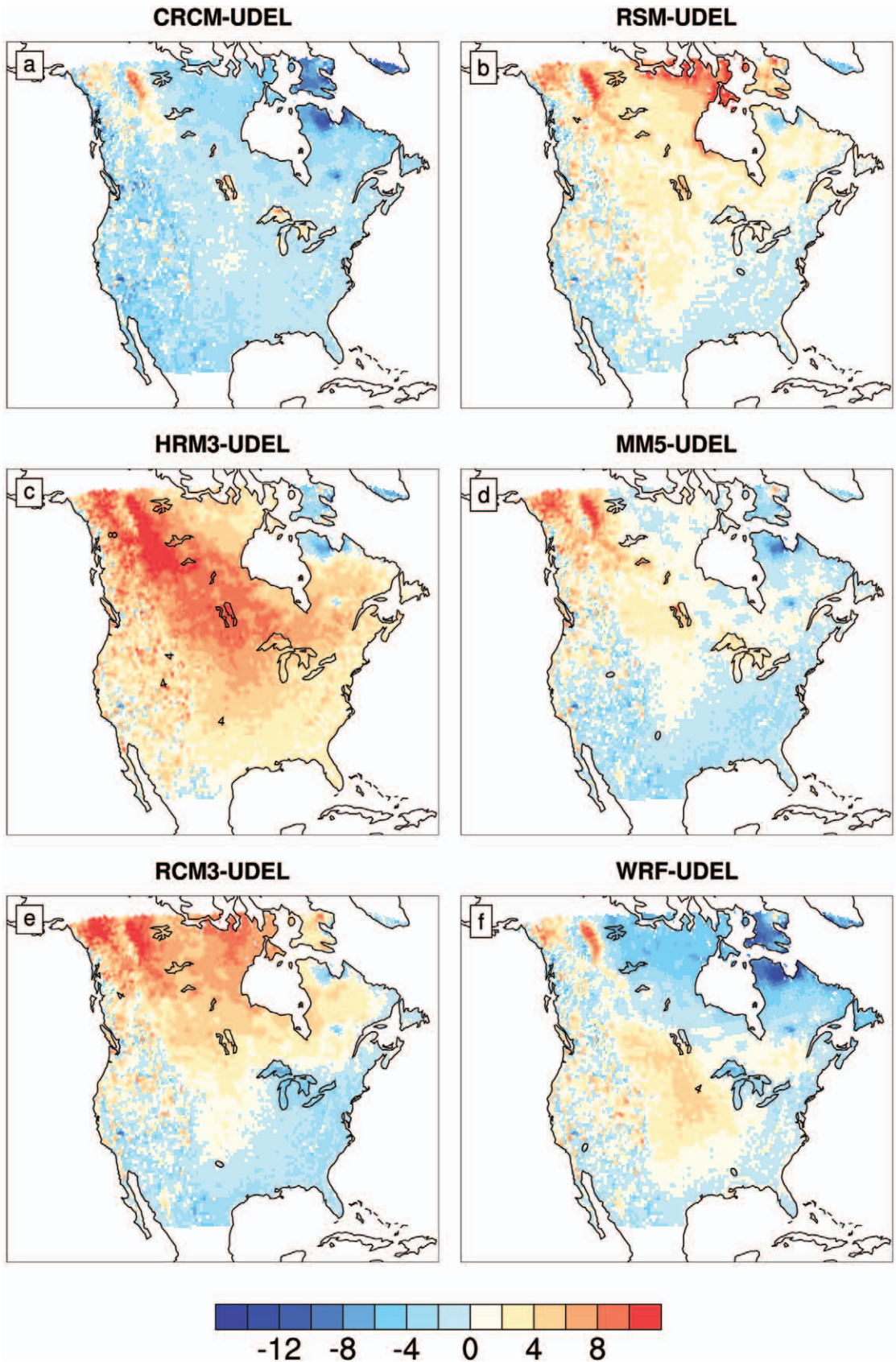andard interpolation method that uses an enhanced distance weighting approach (Willmott et al. 1985) and correction for elevation for temperature based on environmental lapse rates.

We also use two other observed datasets to examine uncertainty in observations that may affect the calculation of model biases: we employ version 2.10 of the Climatic Research Unit (CRU) monthly time series of temperature and precipitation (Mitchell 2008; Mitchell and Jones 2005), which includes elevation correction for temperature and precipitation (New et al. 2000) and the Precipitation–Elevation Regressions on Independent Slopes Model (PRISM) dataset (Daly et al. 1994, 2011). CRU and UDEL use data from thousands of stations around the world interpolated to a half-degree grid over land, whereas PRISM is much higher resolution (4 km), uses a more sophisticated elevation correction scheme (for both temperature and precipitation) ,and includes data from around 8,000 stations, but covers only the conterminous United States. PRISM is therefore likely more accurate for estimates of precipitation in topographically complex regions but could not be used as the primary observed dataset because of its coverage limitations.

Because the models have different map projections and "sponge zones" (edge regions where boundary conditions from the nesting model are blended with RCM calculations), each model's domain is slightly different. For comparison with observations, we interpolated the model data to the UDEL grid over the domain common to all models. The NCEP reanalysis, with a native grid of about 2.5° latitude and longitude (~250-km spatial resolution), was also interpolated to the half-degree UDEL grid.

Domain-wide seasonal temperature. Figures 2 and 3 show bias plots (model minus observed) of mean seasonal temperature (winter and summer, respectively) for each

account for both the global and regional model uncertainties and by adopting a refined approach to weighting the different simulations, a critical improvement over Tebaldi et al. (2005), that will use detailed analysis of climate model results to establish expert judgments of simulation credibility for subregions of interest within the domain and then translate these judgments into a differential weighting scheme used in the Bayesian statistical models. The NCEP-driven RCM results evaluated in the present paper will be used in the development of the weighting scheme and to determine the boundary forcing bias: that is, the difference between the reanalysis-driven RCM runs and the current-period GCM-driven RCM runs (Pan et al. 2001).

Data archive and website. We store RCM output for more than 50 variables at 3-hourly resolution in standards-compliant, GIS-compatible Network Common Data Form (NetCDF) format. The files are organized similarly to the CMIP archive and distributed for free (registration required) via the Earth System Grid data portal. The NARCCAP website (http://narccap.ucar.edu) provides extensive information and guidance in support of three main interests: further dynamical and statistical downscaling; climatological analysis of model results; and exploration of impacts, adaptation strategies, and other policy-related issues.

PHASE I RESULTS FOR TEMPERATURE AND PRECIPITATION. *Observed datasets used for comparisons.* From the various available observed datasets, we selected version 2.01 of the University of Delaware dataset (UDEL) of monthly temperatures and precipitation (Willmott and Matsuura 1995; Matsuura and Willmott 2010) as our primary observed dataset for comparison with model simulations

Fig. 2. Winter [December–February (DJF)] temperature bias (model − UDEL observed; °C) plots for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF.

Fɪɢ. 3. Summer [June–August (JJA)] temperature bias (model − UDEL observed; °C) plots for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF.
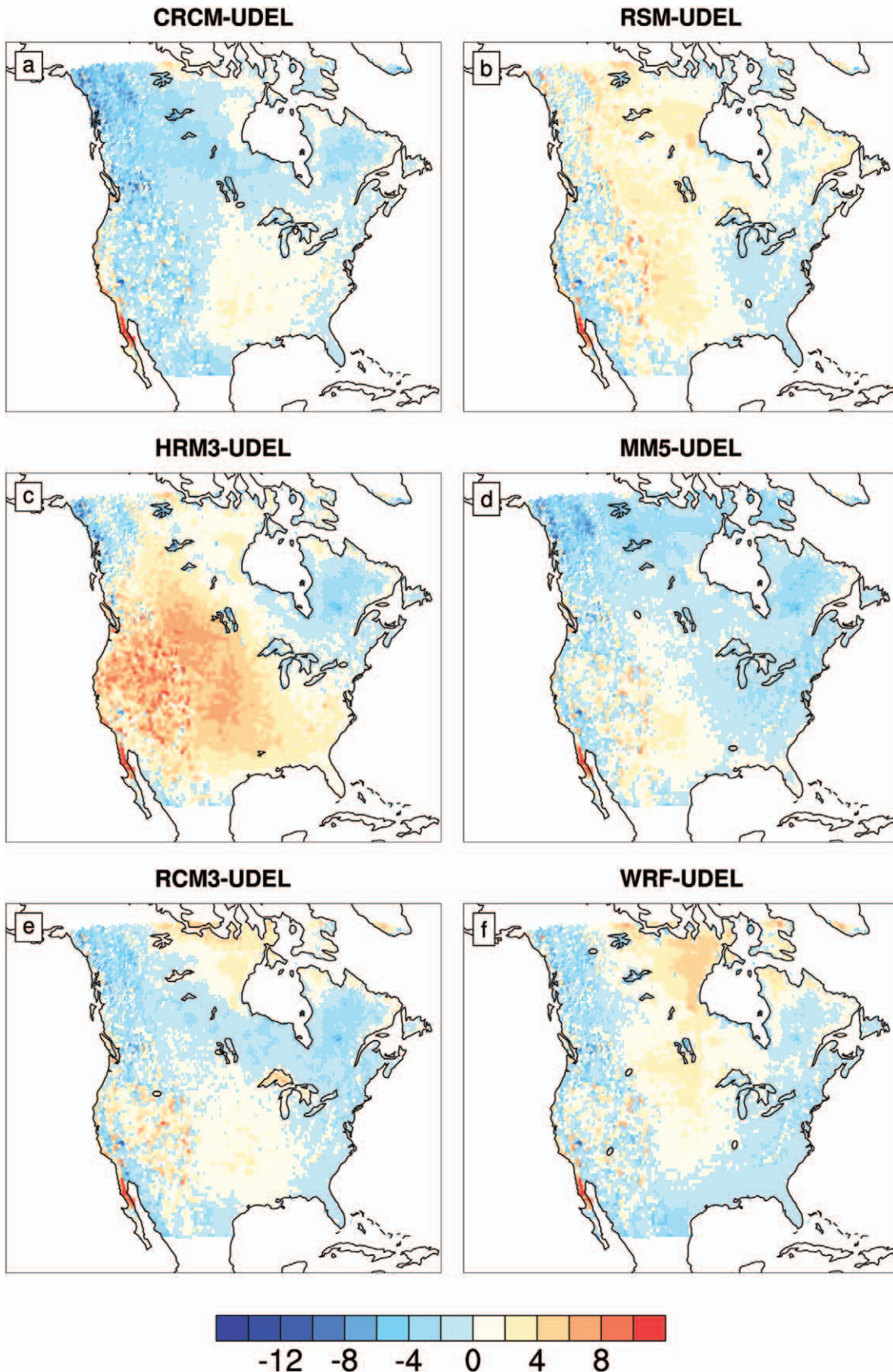
of the six RCMs for the period 1980–2004, and Table 2a lists the root-mean-square errors (RMSEs) for seasonal temperature. The table also lists values for three ensembles: the full ensemble mean averaging all six models, the nudged ensemble mean averaging CRCM and RSM (the models that use spectral nudging), and the nonnudged ensemble mean averaging the other four models. The results for NCEP R2 are also provided as background. The percentage of grid points significantly different from UDEL at the 0.05 level is also included in Table 2. The statistical significance of the bias (at each grid point) was calculated using bootstrapping with bias correction and acceleration to estimate confidence intervals following von Storch and Zwiers (1999) and Efron and Tibshirani (1993). Lower and upper tail critical values (0.05 and 0.95, respectively) were calculated from 1,000 bootstrap samples and then corrected, following the above references. These calculations do not include the effect of the multiplicity of and spatial autocorrelation in the tests, and thus the tests do not determine the field significance of the difference between model simulations and observations but can be used as a relative measure of the differences among the models. This is also the case with other statistical tests presented in this paper.

The largest temperature biases are shown by HadRM3, which exhibits a pronounced warm bias in all seasons but particularly in winter, when temperatures more than 8°C above the observed are found in central to northern Canada (Fig. 2c). CRCM exhibits mainly cold biases in winter, whereas most of the other models tend toward warm biases in the northern part of the domain and cold biases in the south during winter. WRF's warm bias in the central Great Plains extends up into

Canada but is surrounded by cold biases elsewhere. In summer, there is a common warm bias for all models through the central plains of the United States and up into Canada of around 2°C, with small cold biases towards both coasts. The exception is HadRM3, which exhibits a larger warm bias (2°–8°C) and no coastal cold bias. Based on RMSE values (Table 2a), the full ensemble mean outperforms most of the individual models in all seasons, but certain individual models perform as well as the ensemble (MM5 in winter and spring and RSM in spring). However, the nudged ensemble outperforms the full ensemble in all seasons, and the nonnudged ensemble usually performs more poorly than the full ensemble and many of the individual models. In addition, the NCEP reanalysis outperforms many of the regional models in most seasons but not the full or nudged ensemble. [However, it should be noted that it is difficult to establish meaning in comparisons between the NCEP reanalysis itself and the RCM results because the reanalysis assimilates a variety of observed

TABLE 2. Spatial RMSEs for seasonal means of temperature and precipitation over the common domain. The percentage of grid points that statistically differ from the UDEL observations at the 0.05 level is given in parentheses. The lowest value for each season is underscored and the highest is in bold.

a. Temperature (°C)	Winter	Spring	Summer	Fall
CRCM	3.1 (78)	3.0 (81)	2.4 (79)	2.5 (91)
RSM	3.6 (78)	2.4 (63)	2.2 (82)	1.9 (69)
HadRM3	**5.9** (96)	3.9 (93)	**3.6** (88)	**3.1** (84)
MM5	2.8 (62)	2.4 (74)	2.3 (79)	2.9 (81)
RegCM3	4.5 (84)	3.4 (66)	2.1 (75)	1.9 (59)
WRF	3.6 (73)	**4.0** (88)	2.3 (82)	2.9 (75)
ENSEMBLE	2.8 (74)	2.4 (71)	1.8 (66)	1.8 (68)
ENS-Nudge	2.5 (57)	2.3 (64)	1.7 (69)	1.7 (66)
ENS-NonNudge	3.7 (80)	2.8 (79)	2.0 (77)	1.8 (69)
NCEP R2	3.1 (76)	2.4 (70)	2.2 (76)	1.9 (65)

b. Precipitation (mm day⁻¹)	Winter	Spring	Summer	Fall
CRCM	0.86 (73)	0.66 (77)	0.70 (68)	0.82 (63)
RSM	1.01 (88)	**1.12** (91)	0.80 (66)	0.89 (74)
HadRM3	**1.35** (79)	1.01 (81)	0.60 (76)	**1.22** (76)
MM5	1.09 (83)	0.84 (83)	0.60 (60)	0.97 (73)
RegCM3	1.19 (86)	**1.12** (90)	1.10 (80)	1.15 (79)
WRF	0.94 (64)	0.69 (70)	0.77 (67)	0.87 (66)
ENSEMBLE	0.93 (82)	0.82 (87)	0.57 (66)	0.85 (70)
ENS-Nudge	0.87 (84)	0.84 (91)	0.62 (63)	0.81 (71)
ENS-NonNudge	1.00 (81)	0.82 (85)	0.61 (67)	0.92 (70)
NCEP R2	0.84 (66)	0.64 (70)	**1.53** (83)	0.89 (63)

data and is produced at a much coarser resolution (2.5° vs 0.5°)]. CRCM, RSM, and MM5 perform similarly well, with most RMSE values less than 3°C. HadRM3, WRF, and RegCM3 tend to have the larger RMSEs. All models and ensembles perform better in summer than winter based on RMSEs, though not necessarily from the point of view of percent of grid points that differ significantly from observations. In almost every case, more than 60% of the grid points are significantly different from the observations.

The biases and RMSEs of these RCM simulations are within the range found in many regional modeling comparisons (see Takle et al. 1999; Pan et al. 2001; Anderson et al. 2003; Leung et al. 2004; Fu et al. 2005; Christensen et al. 2007a; Gutowski et al. 2007; Christensen et al. 2010), several of which have used the same or earlier versions of the RCMs used in NARCCAP. For example, central U.S. temperature errors in continental U.S. simulations are typically around 2°C (e.g., Takle et al. 1999).

There are some uncertainties based on different observational datasets. Figure 4a portrays the maximum difference in temperature among the three observed datasets for winter (1980–2004) for the lower 48 U.S. states at the 0.5° spatial scale. Across most of the domain, differences are quite small, less than 0.6°C, but in the mountainous west differences (in any season)

can be greater than 2°C and thus could affect the calculation of bias. These differences result from differences in the stations involved in the formation of the datasets, the exact nature of the elevation correction used, and the initial resolution of each dataset. In regions of complex terrain, it is difficult to determine if one set is closer to "the truth" in the context of the resolution of the RCMs. For example, in parts of the northern Rockies of the United States, UDEL's temperatures are about 3°C lower than CRU's. However, throughout the Rocky Mountains there is no consistent bias between the two datasets (i.e., one is not consistently higher or lower than the other), and thus performing the bias analysis with CRU instead of UDEL would not have produced consistently different results.

HADRM3 AS AN OUTLIER. It is clear from the above assessment that the HadRM3 performs significantly worse than the others in terms of seasonal temperature biases. However, application of this model over other regions indicates that it does not have a systematic tendency to simulate temperatures significantly warmer than observed [e.g., Xu et al. (2006) over China, Marengo et al. (2009) over South America, and Kamga and Buscalet (2006) over Africa]

In most previous studies, HadRM3 has been driven by the 15-yr European Centre for Medium-
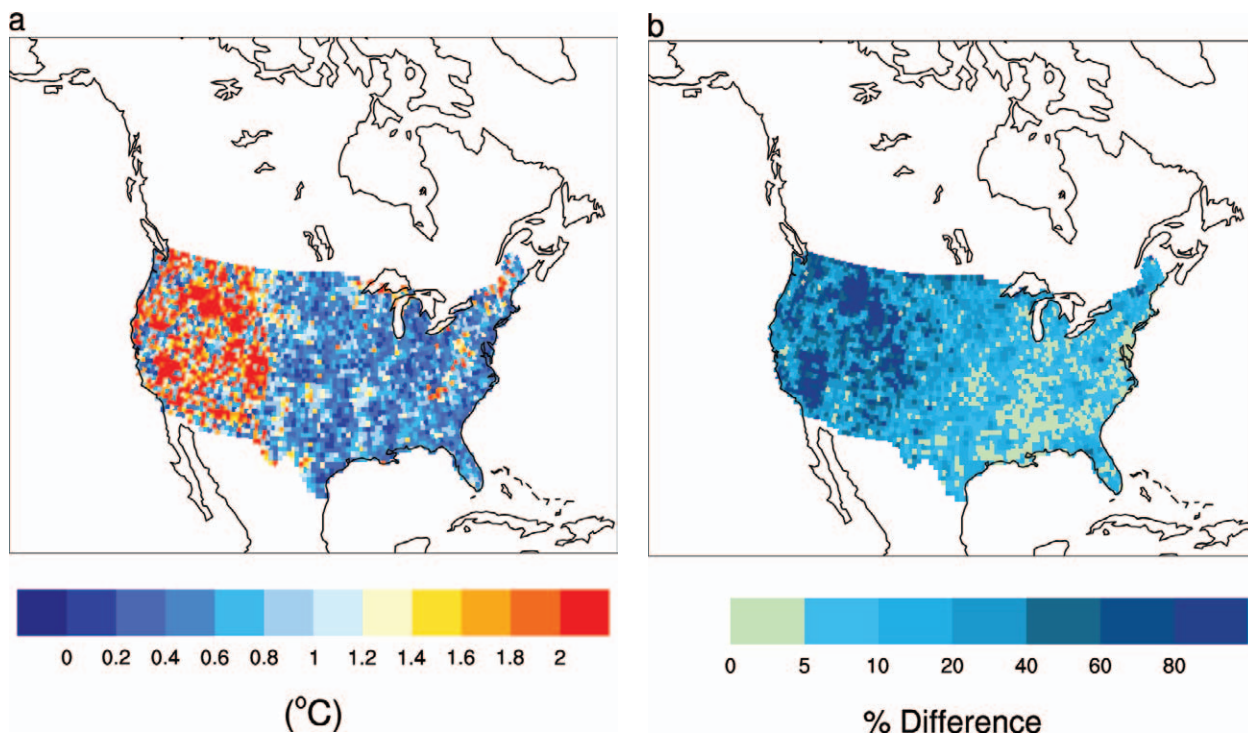

FIG. 4. Comparison of three observed datasets for winter over the lower 48 states for (a) temperature and (b) precipitation. Temperature values displayed are the maximum difference (°C) in each grid box among the three datasets. For precipitation the maximum percent difference across the three datasets is displayed.

Range Weather Forecasts (ECMWF) Re-Analysis (ERA-15), ERA-40, or ERA-Interim. Rerunning HadRM3 using the ERA-Interim data over a subset of years common to that analyzed above (1990–99) produces significantly different temperature biases (Fig. 5). The patterns of the biases are similar, especially in winter, but the magnitudes are significantly less, up to 5°C less, when using the ERA-Interim data for boundary conditions, now of similar magnitude to biases in the other RCMs.

These results indicate clearly that there must be differences between the two reanalyses. Comparing them on the boundary of the NARCCAP domain (not shown) demonstrates that the NCEP data are both warmer and moister both in the lower troposphere and in the upper troposphere/stratosphere. For example, at the 850-mb level at the inflow (western) and outflow (eastern) boundaries, NCEP relative humidity is around 12% and 9% higher, respectively, than that of ERA Interim, whereas temperature along the inflow boundary is about 1.0°–1.5°C warmer.

Both differences would lead the NCEP-driven simulation to be warmer, via the direct increase in temperature in the boundary conditions and via greater downward longwave radiation resulting from higher humidity (possibly further enhanced in winter if the higher humidity led to increased cloud cover). The difference in simulated surface temperatures is greater than in the boundary conditions, which lends support to these additional mechanisms being involved, which may also be enhanced by surface feedbacks.

In addition, in winter the warmer temperatures would lead to reduced snow cover, enhancing solar radiation absorption at the surface. In summer they could lead to drier soils and thus less cloud cover, also enhancing surface solar radiation and consistent with high negative precipitation biases in the NCEP-driven simulation (not shown).

Temperature pattern correlations and variances. In addition to overall bias, we also consider the standard Pearson spatial pattern correlation of mean seasonal temperature (and precipitation). Figure 6a

portrays this metric for the full domain (NA) as well as for three of the four subregions (discussed below). The pattern correlation of temperature for the full domain is very high for all models in both seasons
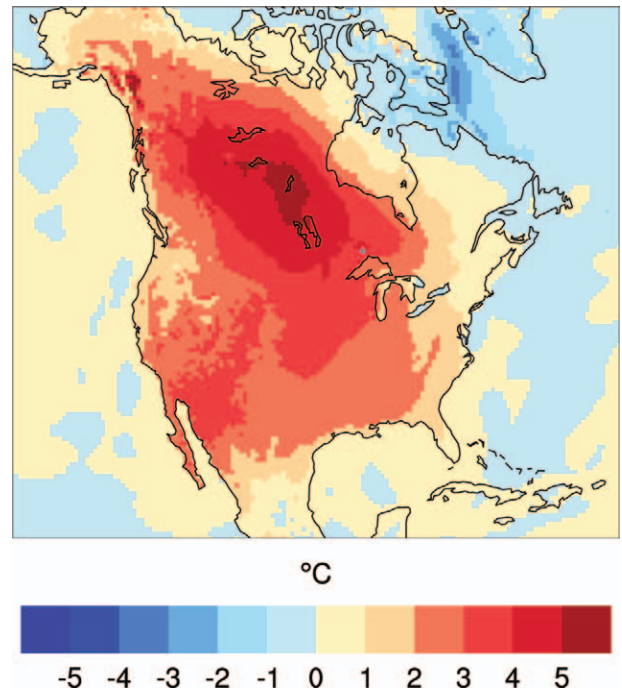

FIG. 5. Difference in winter temperature simulated for 10 years (1990–99) by HadRM3 driven by NCEP vs ERA-Interim boundary conditions.
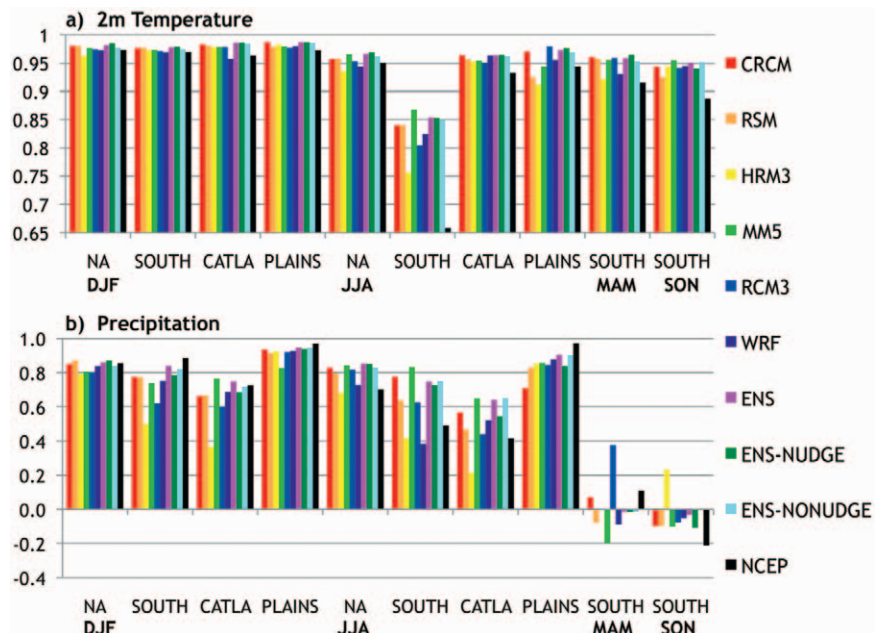

FIG. 6. Spatial pattern correlation for seasonal (a) temperature and (b) precipitation for the full domain (NA) and three of the four subregions, the Atlantic coast (denoted as CATLA), the south-central United States (denoted as SOUTH), and the Great Plains (denoted as PLAINS).
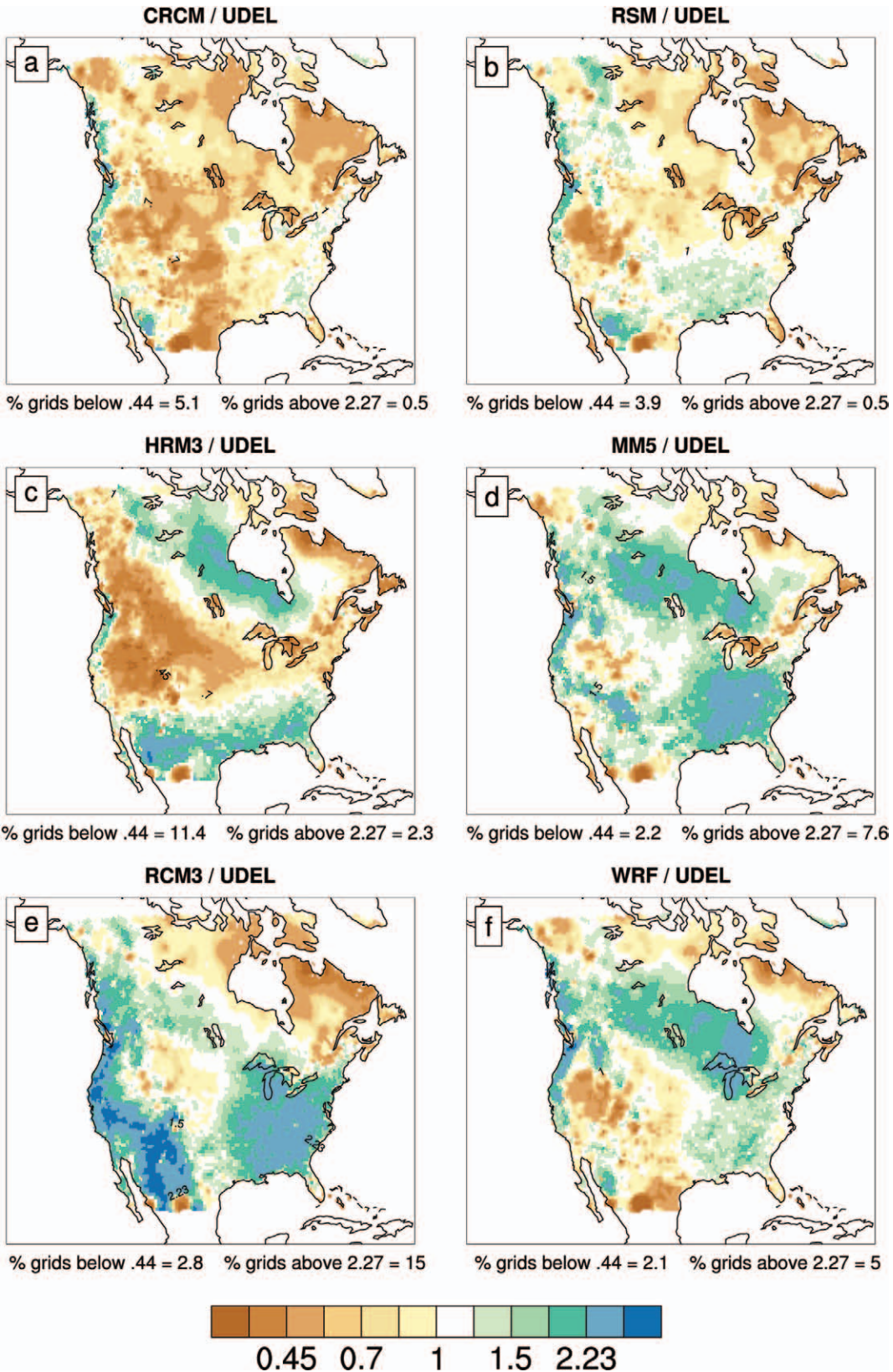
Fig. 7. Ratio of model to observed interannual variance of winter seasonal temperature for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF. The F statistic at the 0.0275 level is 0.44 and at the 0.975 level is 2.27. The percentage of grids below (above) those critical values is indicated below each panel.

CRCM / UDEL

(a)

% grids below .44 = 6.3   % grids above 2.27 = 3.7

RSM / UDEL

(b)

% grids below .44 = 2.5   % grids above 2.27 = 12.6

HRM3 / UDEL

(c)

% grids below .44 = 6.3   % grids above 2.27 = 26

MM5 / UDEL

(d)

% grids below .44 = 6.3   % grids above 2.27 = 5.7

RCM3 / UDEL

(e)

% grids below .44 = 4.5   % grids above 2.27 = 31.7

WRF / UDEL

(f)

% grids below .44 = 5.3   % grids above 2.27= 14.6

0.45 0.7   1   1.5 2.23

Fɪɢ. 8. Ratio of model to observed interannual variance of summer seasonal temperature for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF. Values of the F statistic at the 0.275 level is 0.44 and at the 0.975 level is 2.27. The percentage of grids below (above) these critical values is indicated below each panel.
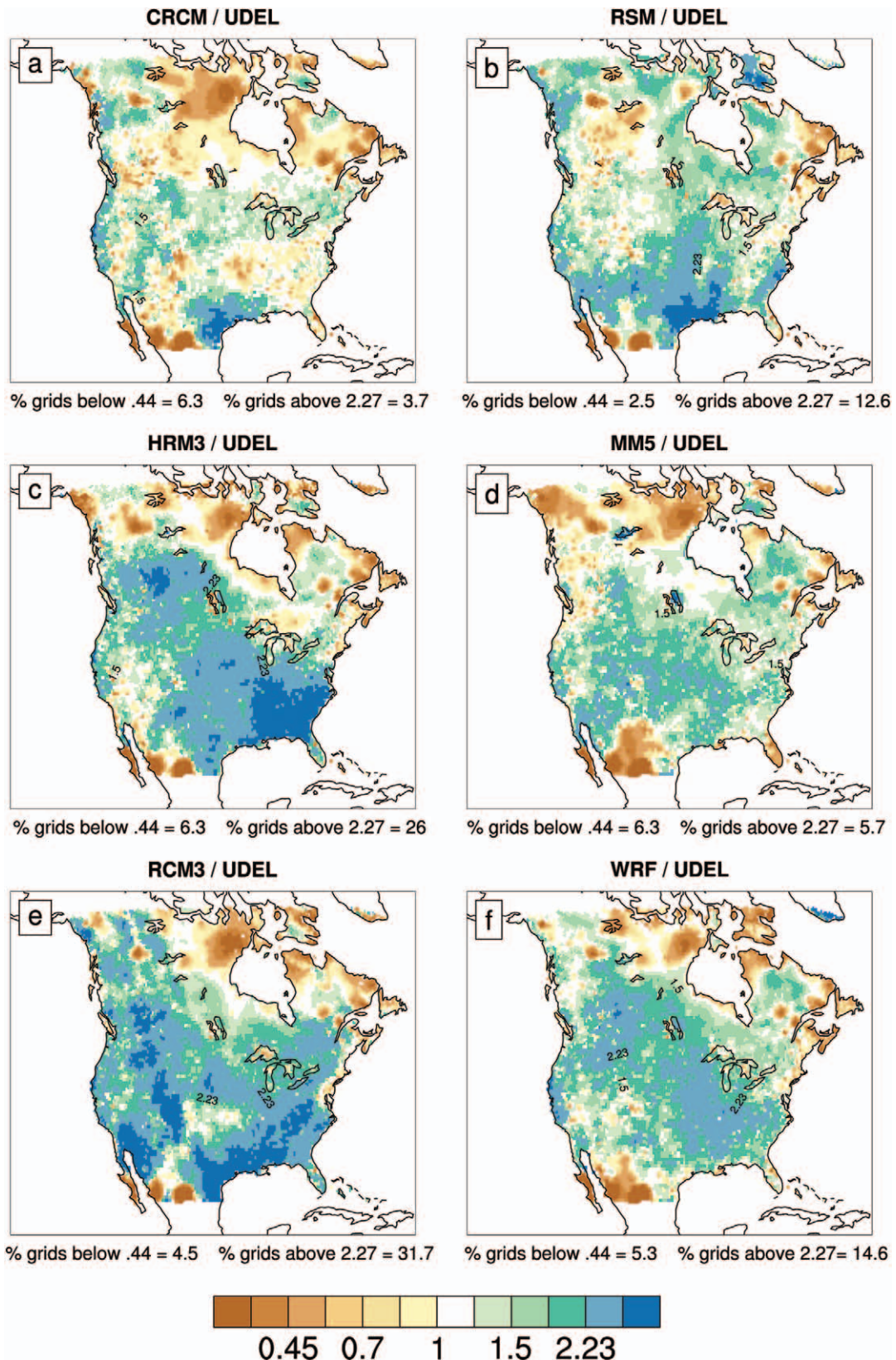
but is slightly higher in winter than in summer. There is little difference across the models, with all winter values above 0.95 and most summer values likewise.

Figures 7 and 8 present interannual winter and summer seasonal temperature variance ratios (model/observations) for the six RCMs, respectively, including the 0.0275 and 0.975 levels of the F test, which indicate ratios that are significantly different from 1.0. In accordance with observations, all models demonstrate higher variance in winter than in summer and higher values in the continental interior (figures not shown), but biases are pronounced for many of the models in both seasons.

In winter, CRCM mainly underestimates variance, particularly in the interior parts of the domain and land areas east of Hudson Bay (Fig. 7a), whereas MM5 and RegCM3 (Figs. 7d,e) mainly overestimate it. RSM reproduces it relatively well (Fig. 7b) over most of the domain, except for an area of underestimation over the American Rocky Mountains. HadRM3 (Fig. 7c) mainly underestimates variance through the center of the domain and extending northward through the Canadian Rocky Mountains. WRF (Fig. 7f) reproduces the observed variance in the center of the domain but overestimates it in large portions of Canada.

In summer, when observed variability is substantially less than in winter, most models overestimate the variance over most of the domain (Fig. 8). CRCM and MM5 produce fewer grid points with significant differences than the other models, and these are a mixture of small underestimations and overestimations. RSM performs relatively well but displays areas of overestimation in the southern parts of the domain and up through the U.S. Midwest. HadRM3 and RegCM3 produce the most substantial overestimations of variance throughout most of the domain. WRF produces a pattern of bias similar to that of HadRM3, and in some areas the overestimations are of the same magnitude. The biases in variance are relatively independent from the biases in mean temperature in all four seasons. The correlations of the two types of biases are generally between +0.3 and −0.3 for most model–season combinations.

Domain-wide seasonal precipitation. Figures 9 and 10 present bias plots (% difference) for precipitation in winter and summer, respectively. Table 2b gives RMSE values and percentages of grid points that are significantly different from observations.

In winter, most models produce predominantly wet biases, with one of the most extreme being RSM, with biases of 80% or more over much of the domain. On the other hand, RSM has the lowest bias in the south-central United States (lower Mississippi River basin), which is a region of high precipitation in the observations. This may be because RSM uses spectral nudging and receives more large-scale information from the NCEP reanalysis. Most of the models underestimate this subregional concentration of precipitation.

In summer, more models exhibit underestimates of precipitation, particularly WRF, MM5, and HadRM3 in the central plains of the United States and Canada. Of the six RCMs, CRCM produces the lowest RMSE in every season but summer, whereas the producer of the highest RMSE varies from season to season. The ensemble average performs best only in summer, whereas the nudged ensemble performs best only in fall, when its performance is little better than CRCM's. The NCEP reanalysis produces the largest RMSE of all models in summer; this is most likely due to the well-known and well-documented assimilation/spindown problem present in the R2 that leads to excessive precipitation, particularly in summer (Kalnay et al. 1996; Bukovsky and Karoly 2007).

We examined the uncertainty in bias estimation based on the uncertainty in precipitation observations. Figure 4b shows the largest percent difference in the three observational datasets for the contiguous United States for the winter. Although most of the differences in the eastern two-thirds of the United States are relatively small (within 5–10 percentage points) in the west, in areas of complex terrain some differences are substantial, such as in the northern U.S. Rocky Mountains and in the southwest margin of California. In these areas, differences can be as much as 80%. It is likely that the PRISM dataset, which tends to have the largest value in these regions, is closer to the true values, given the care that was taken to correct its precipitation values for elevation. This means that, in areas like the northern U.S. Rockies, the positive biases indicated by the comparison with UDEL are likely too high. However, given the magnitude of the biases for many of the models, the biases over large subregions would not be substantially altered due to comparison with a different observed dataset. In general, the biases in precipitation fall in the range seen in other modeling efforts. Precipitation errors in continental U.S. simulations are generally between 0.6 and 1.4 mm day^{-1} (Anderson et al. 2003; Gutowski et al. 2007). Leung et al. (2003c) compared three regional simulations for the western United States based on MM5 and RSM driven by different global reanalyses and found that, although RCM simulations were too wet compared to observations by 20%–75%, the same RCM (MM5) driven by two different global reanalyses can produce very different results because
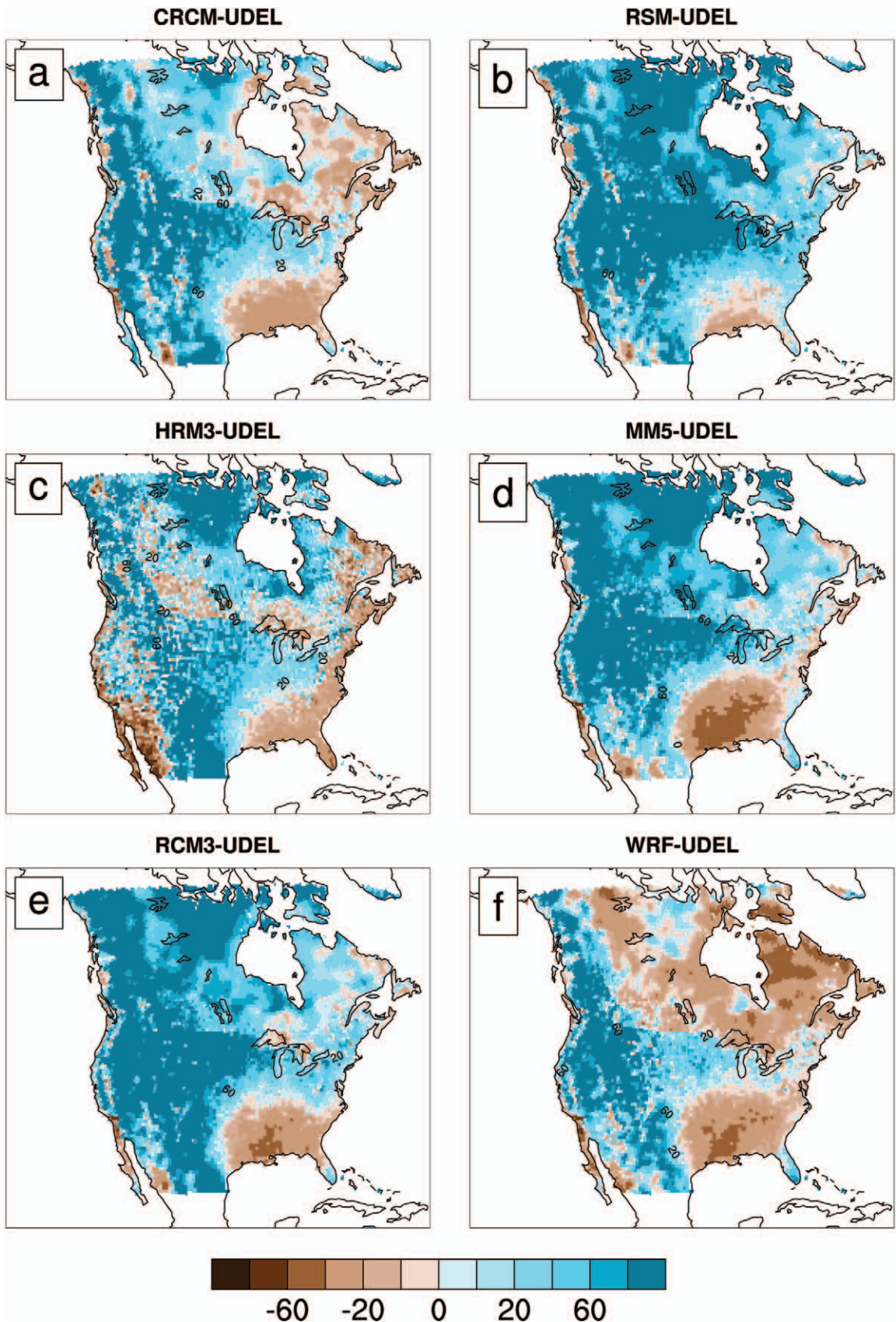
Fig. 9. Winter precipitation bias (% difference; model − UDEL observed) plots for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF.
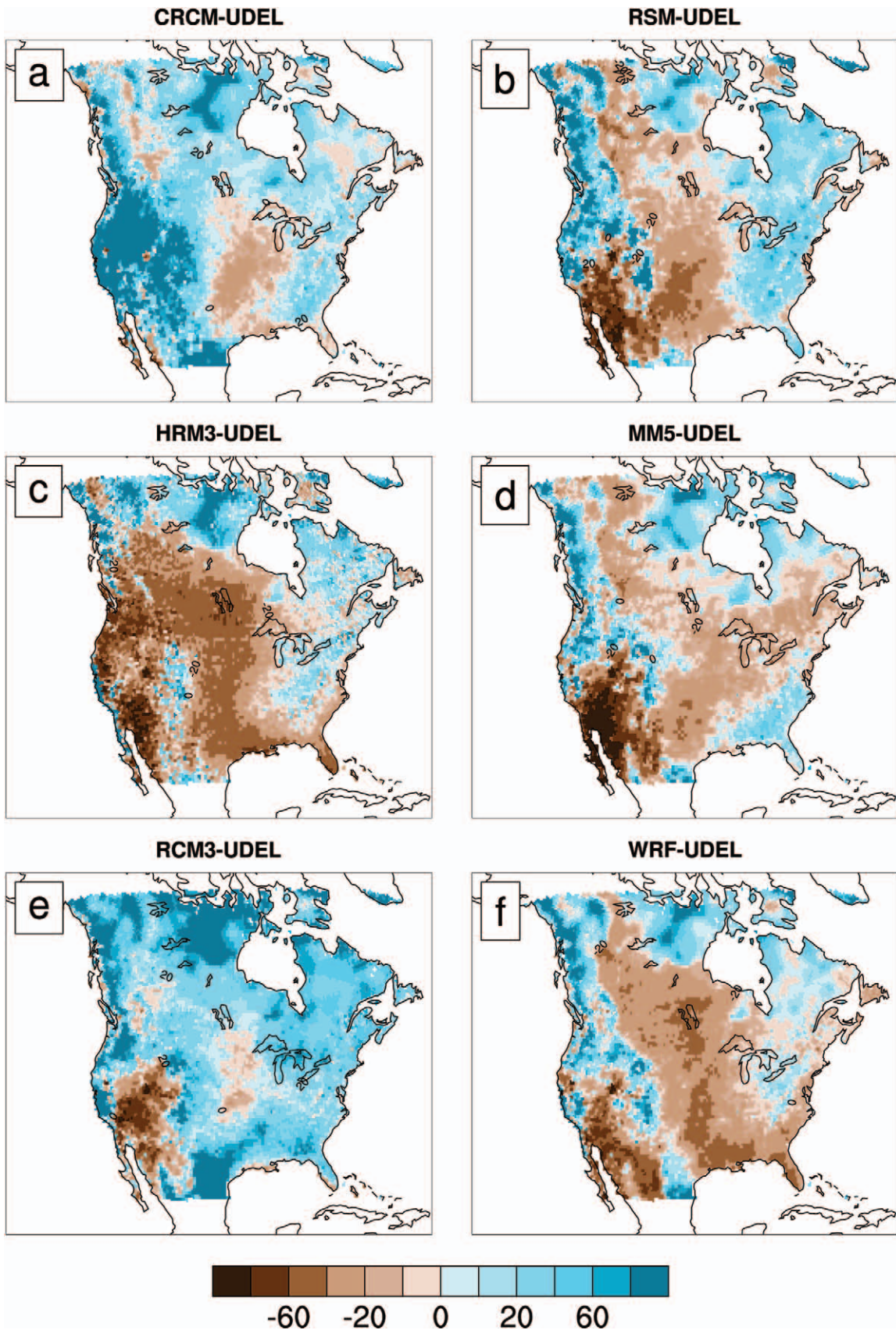
Fɪɢ. 10. Summer precipitation bias (% difference; model − observed) plots for the six regional models: (a) CRCM, (b) RSM, (c) HadRM3, (d) MM5, (e) RegCM3, and (f) WRF.

the moisture fluxes in the reanalyses are not well constrained by observations over the oceans.

The spatial pattern correlations for precipitation (Fig. 6b) are lower than those for temperature but are still relatively high for all models, especially in winter, when they are generally greater than 0.8. In summer most values are greater than 0.8, although HadRM3, WRF, and NCEP are around 0.7.

In winter (Fig. S1; see http://dx.doi.org/10.1175/BAMS-D-11-00223.2), most models show variance ratios in the 3–4 range in the western half of the continent, but some of this is due to the dry bias of UDEL in the mountainous part of the domain. In the southeast, most models exhibit an underestimation of variance, except for the HadRM3, which shows slight overestimations. Areas of underestimation are much more common in summer (Fig. S2) than in winter for all models but RSM and RegCM3, where overestimations prevail.

In contrast to temperature, for precipitation, the correlation between mean biases and biases of variance is considerable. For 22 of the 24 model–season combinations, the correlations are greater than 0.5 and all are positive. Hence, models that have positive biases in mean seasonal precipitation tend to have positive biases in the variance as well. (Note that a log transformation was applied to the data to render the distributions close to normal for this analysis.)

More detailed subregional analyses. We also analyzed four subregions of North America (displayed in Fig. 1) to sample how well the models reproduce climatically distinct subregions of the domain. Southern California has a Mediterranean climate (Köppen classifications Csa and Csb; Köppen 1900) with a pronounced summer dry season and mild, damp winters. There is a strong ENSO signal in this region. The Great Plains has a midlatitude continental climate (mostly Köppen classifications Dfa and Dfb) with a large annual temperature range, having warm summers and cold winters. The region has a precipitation maximum in late spring and early summer that is largely attributable to organized mesoscale convective systems. The south-central region has a humid subtropical climate (Köppen classification Cfa) with a modest wintertime precipitation maximum and temperatures that do not often reach below freezing. The Atlantic coastal region has a moist midlatitude climate affected by proximity to the Atlantic Ocean, with substantial variation from north to south in both temperature and precipitation and Köppen classifications ranging from Cfa in the south to Dfb and Dfc in the north.

Our regional analysis includes pattern correlations for seasonal temperature and precipitation, as well as correlations between temperature and precipitation, which give some indication of how well the models reproduce the factors that govern the overall climatology of the subregion. We also examine the monthly biases averaged over each region, the observed versus modeled temporal correlations, and a comparative quartile analysis of the seasonal biases. We also examine the uncertainty of the observed datasets in some of these analyses to determine under which conditions uncertainty in observations has an important effect on model evaluation.

COMPARATIVE QUARTILE ANALYSIS. Any climatological variable of interest (e.g., seasonal temperature or precipitation) will have a spatial distribution over a given region. To perform comparative quartile analysis, we determine the median, lowest, and highest quartiles (q50, q25, and q75) of the distribution for both model and observations on a seasonal basis. We then construct a bias index by comparing them as follows:

$$q50(model) > q75(obs) \rightarrow Index +1,$$

$$q50(model) < q25(obs) \rightarrow Index -1,$$

$$q25(model) > q50(obs) \rightarrow Index +1, \text{ and}$$

$$q75(model) < q50(obs) \rightarrow Index -1.$$

This gives index values between −2 and +2. Each increase (decrease) of 1 means there is a positive (negative) bias of the model compared to the observations. A nonzero index value indicates that the model median is outside the central half of the observed distribution, that the upper or lower quartile of the model distribution is shifted past the median of the observations, or both.

Mean seasonal cycle of temperature and precipitation. In this analysis, we include all three observational datasets for the two subregions (Southern California and south-central United States) fully within the continental United States and two datasets for the other two subregions (Great Plains and the Atlantic coast) that extend outside the PRISM coverage area. To further quantify the bias ranges based on the different datasets, we provide ranges of seasonal bias based on different observational datasets for temperature and precipitation for the first two subregions (Table 3).

SOUTHERN CALIFORNIA. All models reproduce the observed climatology of warm, dry summers and cool, moist winters (Figs. 11a,e), but nonetheless

there are notable variations. Observational uncertainty is significant in the winter for precipitation and in the summer for temperature. HadRM3 underestimates winter precipitation by between about 1.5–2 mm day⁻¹, a 60%–70% underestimation (Table 3) of the observed rates, whereas the RSM

bias ranges from −2% to +29%, depending on the observed dataset used for comparison (Fig. 12e and Table 3). The ensemble mean corresponds to the observed annual cycle of precipitation from the UDEL observations especially closely but underestimates winter precipitation based on the other two observed

TABLE 3. Range of seasonal biases for (a) temperature (°C), (b) precipitation (percentage difference), and (c) precipitation (difference in mm day⁻¹) for two subregions based on the three different observed datasets for DJF, March–May (MAM), JJA, and September–November (SON).

a.

	Range in temperature bias (°C)							
	Southern California				South-central United States			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
CRCM	−3.0 : −2.6	−1.2 : 0.0	2.3 : 4.7	−1.5 : −0.5	−1.2 : −0.9	−0.8 : −0.5	1.0 : 1.4	−0.9 : −0.6
RSM	−0.8 : −0.4	−0.2 : 1.0	0.6 : 2.9	0.1 : 1.2	−0.3 : 0.0	0.1 : 0.4	0.4 : 0.8	0.5 : 0.9
HadRM3	1.7 : 2.1	2.0 : 3.3	3.7 : 6.1	2.9 : 4.0	3.0 : 3.3	1.7 : 2.1	3.5 : 3.9	2.2 : 2.5
MM5I	−0.1 : 0.3	−0.3 : 0.9	−0.1 : 2.2	−0.7 : 0.4	−1.7 : −1.3	−1.2 : −0.8	−0.2 : 0.1	−2.3 : −2.0
RegCM3	−0.1 : 0.3	0.3 : 1.5	0.5 : 2.9	0.3 : 1.4	−1.4 : −1.1	−0.7 : −0.3	0.8 : 1.2	−2.2 : −1.8
WRFG	0.2 : 0.6	−0.6 : 0.6	0.5 : 2.9	0.1 : 1.1	0.4 : 0.8	0.1 : 0.5	−0.9 : −0.5	−1.4 : −1.0
ENS	−0.3 : 0.1	0.0 : 1.2	1.3 : 3.6	0.2 : 1.3	−0.2 : 0.2	−0.1 : 0.2	0.8 : 1.2	−0.7 : −0.3
NCEP	−0.6 : −0.2	−1.7 : −0.5	−2.3 : 0.1	−1.0 : 0.0	−0.9 : −0.6	0.1 : 0.5	0.0 : 0.4	−0.4 : 0.0

b.

	Range in precipitation bias (%)							
	Southern California				South-central United States			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
CRCM	−16 : 11	−2 : 24	92 : 210	−3 : 38	−25 : −23	−5 : −3	1 : 3	−32 : −28
RSM	−2 : 29	4 : 32	−39 : −2	−6 : 34	−8 : −6	4 : 5	1 : 3	−32 : −27
HadRM3	−71 : −62	−16 : 7	34 : 117	−81 : −72	−16 : −14	−8 : −7	−8 : −6	−41 : −37
MM5I	−23 : 2	−16 : 7	−60 : −35	−24 : 8	−43 : −42	−8 : −7	4 : 6	−38 : −34
RegCM3	−17 : 11	−15 : 9	2 : 64	−19 : 15	−34 : −33	1 : 3	42 : 45	−36 : −31
WRFG	−33 : −11	−40 : −23	−37 : 1	−33 : −5	−37 : −36	−19 : −17	−35 : −34	−33 : −28
ENS	−27 : −3	−14 : 9	−2 : 59	−27 : 3	−27 : −26	−6 : −5	1 : 3	−35 : −31
NCEP	−47 : −29	−47 : −33	−55 : −27	−44 : −21	−39 : −38	−15 : −13	80 : 83	−16 : −10

c.

	Range in precipitation bias (mm day⁻¹)							
	Southern California				South-central United States			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
CRCM	−0.50 : 0.26	−0.03 : 0.27	0.12 : 0.17	−0.02 : 0.24	−0.99 : −0.92	−0.19 : −0.14	0.05 : 0.11	−1.15 : −0.94
RSM	−0.08 : 0.68	0.05 : 0.36	−0.05 : 0.00	−0.05 : 0.21	−0.30 : −0.23	0.15 : 0.20	0.03 : 0.09	−1.14 : −0.92
HadRM3	−2.19 : −1.43	−0.22 : 0.08	0.04 : 0.09	−0.72 : −0.46	−0.64 : −0.57	−0.34 : −0.28	−0.26 : −0.19	−1.47 : −1.25
MM5I	−0.71 : 0.05	−0.23 : 0.07	−0.08 : −0.03	−0.21 : 0.05	−1.72 : −1.65	−0.34 : −0.28	0.15 : 0.21	−1.37 : −1.15
RegCM3	−0.51 : 0.24	−0.21 : 0.10	0.00 : 0.05	−0.17 : 0.10	−1.35 : −1.29	0.05 : 0.11	1.43 : 1.49	−1.28 : −1.06
WRFG	−1.02 : −0.26	−0.56 : −0.26	−0.05 : 0.00	−0.29 : −0.03	−1.50 : −1.43	−0.77 : −0.71	−1.18 : −1.12	−1.17 : −0.95
ENSA	−0.83 : −0.08	−0.20 : 0.10	0.00 : 0.05	−0.25 : 0.02	−1.08 : −1.01	−0.24 : −0.18	0.04 : 0.10	−1.26 : −1.05
NCEP	−1.44 : −0.68	−0.67 : −0.36	−0.07 : −0.02	−0.40 : −0.13	−1.58 : −1.51	−0.60 : −0.55	2.69 : 2.76	−0.57 : −0.35

datasets (e.g., –27%; Table 3b). The large ranges of precipitation biases in summer (Table 3b) result from the very small amounts of precipitation occurring in that season (Table 3c). Temperature biases for the models (Fig. 12a and Table 3) range from a strong cold bias in winter and warm bias in summer (CRCM) to a warm

bias throughout the year (HadRM3). The ensemble mean reproduces the seasonal cycle of temperature quite well except for an overestimation in summer, which is slight based on the PRISM observations but substantial based on UDEL and CRU. Comparative quartile analysis for seasonal temperature (Fig. S3)
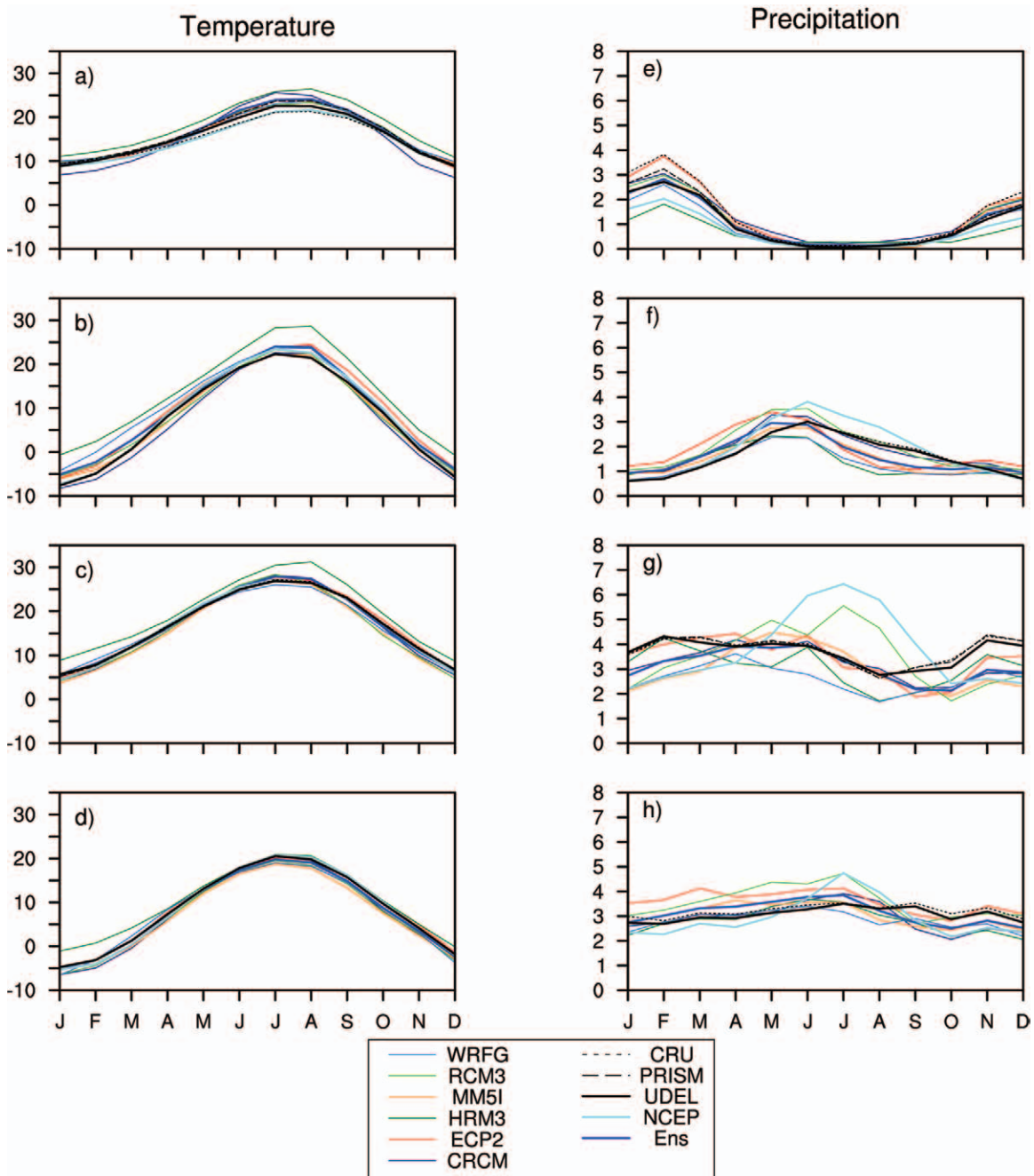

FIG. 11. Seasonal cycle of (left) temperature (°C) and (right) precipitation (mm day^{-1}) for the six models in the four subregions: (a),(e) Southern California; (b),(f) Great Plains; (c),(g) south-central United States; and (d),(h) the Atlantic coast. Results for the ensemble of the six models (denoted as Ens), NCEP, and three observed datasets are also displayed. ECP2 refers to RSM.

shows that most models reproduce all four seasons well, except for CRCM in summer and winter and HadRM3 year-round. This can also be seen in Fig. 12. Most models perform best in the shoulder seasons. Based on quartile analysis for precipitation (Fig. S4), most models perform well in most seasons but summer, when observed precipitation amounts are very

small. Overall, HadRM3 produces the largest total bias index score.

GREAT PLAINS. All models reproduce the general trend of cold winters and warm summers (Fig. 11b), although all but CRCM have a warm bias in winter (Fig. 12b). The observed values for both UDEL and
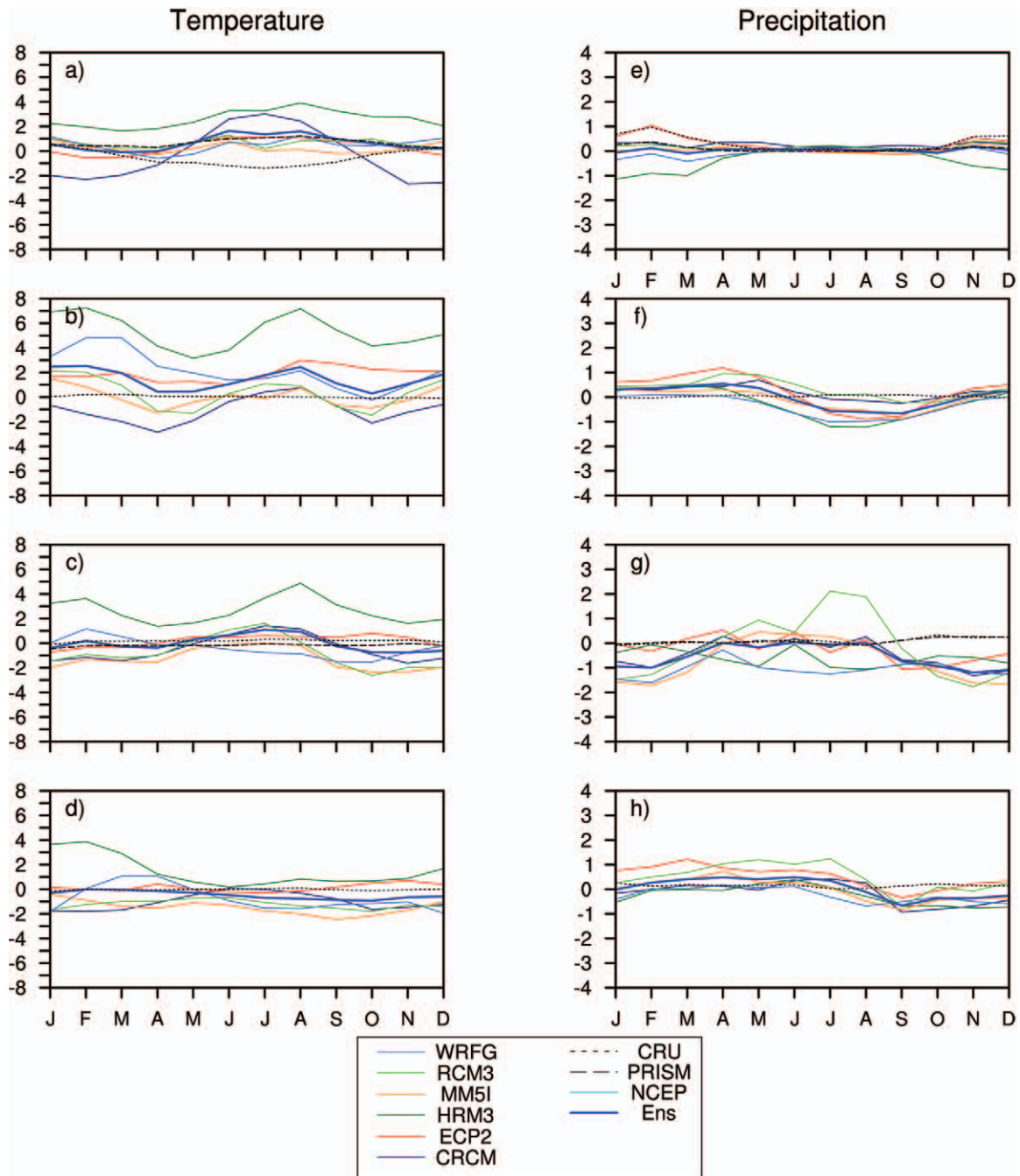

FIG. 12. Monthly biases of (left) temperature (°C) and (right) precipitation (mm day⁻¹) for the six models in the four subregions: (a),(e) Southern California; (b),(f) Great Plains; (c),(g) south-central United States; and (d),(h) the Atlantic coast. Results for the ensemble of the six models (denoted as Ens), NCEP, and two other observed datasets compared to UDEL observations are also displayed. ECP2 refers to the RSM.

CRU for temperature and precipitation are extremely close and do not contribute appreciable uncertainty to the model evaluations. HadRM3 has a pronounced warm bias ranging from 3.2°C in June to over 7°C in March and September. This tendency for the warm bias to peak in spring and fall appears in a more muted form in several other models. The annual cycle of precipitation shows a tendency for wet biases in winter and a dry bias from midsummer through fall (Figs. 11f, 12f). The summer dry bias may reflect the fact that much of the warm-season precipitation in this region is produced by mesoscale systems that are poorly resolved at the 50-km grid spacing used in NARCCAP (Anderson et al. 2007). The warm temperature bias and wet precipitation bias in winter are consistent to the extent that a warmer atmosphere implies greater water vapor content and thus greater potential for precipitation. Quartile analysis shows that HadRM3 has a consistent warm bias in all seasons (Fig. S3) and WRF has a warm bias in winter and spring. Other models perform well in all seasons. All models overestimate precipitation in the winter (Fig. S4), and half the models (particularly HadRM3 and WRF) show dry biases in the summer and/or fall.

SOUTH-CENTRAL UNITED STATES. All three observed datasets for this region are very close in value for both temperature and precipitation and thus do not reflect any significant range in biases for any of the models (Table 3).

As in the Great Plains region, HadRM3 has a warm bias throughout the year, peaking in March and September at 3.6° and 4.9°C, respectively (Figs. 11c, 12c). There is no consistent pattern to the other models, though most have a cool bias in winter and spring. Precipitation shows large intermodel spread in each month, amounting to a variation of about 50% of observed monthly precipitation in most months (Figs. 11g, 12g). The ensemble mean underestimates precipitation in fall and winter but is very close to observations from May through September. Interestingly, over this part of the year, NCEP is the worst performer for both temperature and precipitation but particularly the latter. It exhibits a pronounced and unrealistic maximum in August that overestimates precipitation by about 125%. The RegCM3 most closely follows the NCEP seasonal cycle. This seasonal difference in bias may reflect seasonal differences in the physical character of precipitation. The quartile analysis for temperature (Fig. S3) also shows that HadRM3 has a consistent warm bias and that other models have difficulty reproducing fall and summer temperatures, with only RSM performing well across

all seasons. Precipitation biases, mainly negative, are pronounced in most seasons in most models (Fig. S4). Note that perceived performance of the models varies based on whether one views the seasonal (Fig. S4) or monthly (Figs. 11g and 12g) results.

ATLANTIC COAST. For this region, the differences in the UDEL and CRU observed datasets do not result in an appreciable range in biases for any given model, although the CRU dataset is about 0.2 mm day^{-1} wetter in most months. Most models have a slight cool bias or are close to the observed mean in most months (Figs. 11d, 12d), with the notable exception being HadRM3, with a warm bias throughout the year peaking at 3.9°C in March. The bias of the ensemble mean is generally small and is most extreme in October and November at −0.9°C. It is interesting to note that the skill of the ensemble mean comes about in part because the warm bias in HadRM3 acts as a counterweight to the cool bias in most of the other models. Thus, in this case, a model that could be viewed as having relatively low skill (based on its individual bias) adds value to the ensemble. Quartile analysis of temperature (Fig. S3) shows that all models perform well in all seasons, except for a slight cool bias for MM5 in summer. Note that the warm bias discussed above for HadRM3 is not evident in the seasonal quartile analysis (Fig. S3). This illustrates the role selection of metrics can have in the final perception of model quality. For precipitation (Fig. S4), models exhibit positive and negative biases in more than half of the model–season combinations. Most of the models have a wet bias from March through July and a dry bias from October through December (Figs. 11h, 12h), whereas RegCM3 and RSM show high positive biases in all seasons but fall (Fig. S4).

Spatial pattern correlations for all regions. As illustrated in Fig. 6, the pattern correlation in winter temperature is excellent, at greater than 0.95 in all instances. Summer temperature pattern correlations are also good, at greater than 0.90 for all regions except the south-central United States, where the position of the summertime southern U.S. maximum has a strong influence on the temperature gradient's orientation. As expected, model performance on this metric is much more variable for precipitation than temperature; the greatest consistency in performance is in the Great Plains region, where most models perform well (above 0.80).

Precipitation pattern correlations for spring and fall are included in Fig. 6b for the south-central United States, because the models' performance

on this metric in this subregion stood out as particularly poor. This is due to poor positioning of the precipitation maximum in this region. In the spring, for instance, instead of positioning a maximum adjacent to the coast, most models place it towards the northern edge of this region (nearer the Appalachians), switching the direction of the gradient and yielding poor correlations.

Precipitation/temperature correlations. Correlations of precipitation and temperature for winter and summer in each subregion are shown in Fig. 13. In the Great Plains and the south-central United States, where warmer summers are often drier summers, all of the RCMs capture the interannual interplay between precipitation and temperature. Performance in the coastal regions and in winter, where the average correlations are weaker, is less straightforward, but most of the RCMs capture the direction of correlation indicated by UDEL and CRU. One exception is HadRM3 in winter; although these regions are comparatively small, the pattern of the precipitation/temperature correlations is not homogeneous and in most regions there is an area where HadRM3 produces an exaggerated correlation that overwhelms the signal from the rest of the region. For example, in Southern California there is a strong gradient to the correlation from weakly negative in the south to moderately positive in the north (not shown). All of the RCMs capture the direction of this gradient, but HadRM3 has a stronger negative correlation in the south and a weaker positive correlation in the north, resulting in an overall correlation that is weakly negative. The same is true in the Great Plains region, where from north to south the correlation pattern switches sign twice, with the central negative correlation dominating the region (not shown). Some models capture this better than others, but HadRM3 produces a strong positive correlation at the south end that results in an overall positive correlation.

Time series correlations. Because the annual cycle dominates the time series, obscuring interannual and intermodel variations, before computing correlations we subtracted the mean annual cycle from each time series; we refer to the result as "deviation temperature" or "deviation precipitation." Computing deviations also factors out systematic biases in both the model mean and the mean annual cycle. Even models with large biases can be useful if those biases are systematic and can be easily removed.

Correlations of deviation temperatures with UDEL (Table 4a) are around 0.7–0.8 for most models and regions. CRCM has the highest correlation with observations in each region, which may be because it is one of the two models to use spectral nudging. RSM has the second highest correlation and is the other model using spectral nudging. Of the remaining four models, RegCM3 has the lowest regional correlations but is not markedly lower than the other models. Note that, despite HadRM3's large biases, its correlations with the observed deviation time series are comparable to those of the other models, suggesting that the model depicts interannual variability well once systematic biases are removed.

For deviation precipitation, correlations with UDEL are largest for Southern California and smallest for the Great Plains (Table 4b). This may reflect the difference in precipitation seasonality between the two regions; Southern California has a pronounced cool-season precipitation maximum resulting primarily from synoptic-scale processes that are
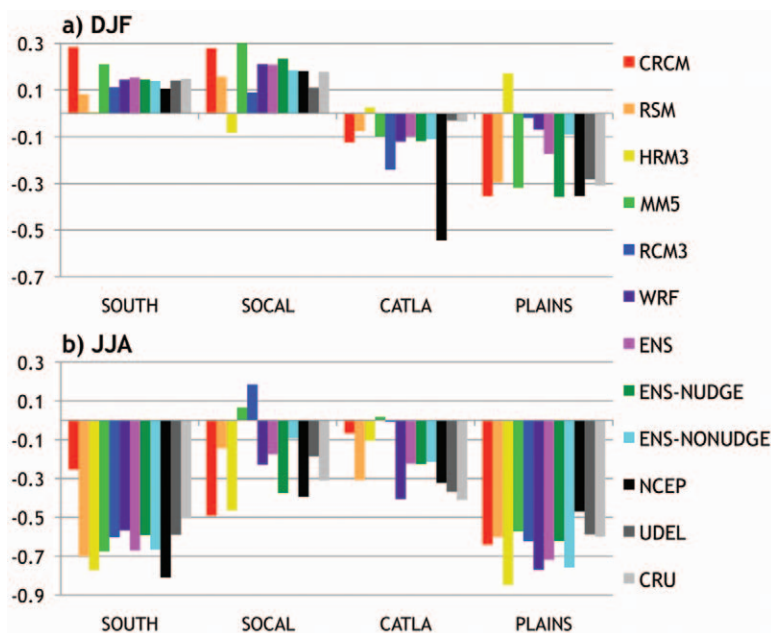

Fig. 13. Correlations of temperature and precipitation for the four subregions for (a) winter and (b) summer. Results for all six models, the full ensemble (denoted as ENS), the subensemble of the two models that used nudging (denoted as ENS-NUDGE), that of the remaining four models (denoted as ENS-NONUDGE), NCEP, and two different observation datasets are displayed.

large enough to be resolved on the 50-km NARCCAP grid, whereas the Great Plains has a warm-season precipitation maximum resulting primarily from mesoscale systems that are poorly resolved on a 50-km grid. The features of observed interannual precipitation variation in Southern California, such as large positive anomalies during the El Niño events of 1982/83, 1994/95, and 1997/98 and the multiyear drought of the late 1980s, are well represented, particularly by the ensemble (Fig. S5a). Models have most difficulty reproducing the variability in the Great Plains, where correlations are lowest. However, the ensemble mean generally reflects observed interannual precipitation variability, with the exception of the extreme wet anomaly of 1993 (Fig. S5b).

For regions other than Southern California, the highest correlations for deviation precipitation are shown by CRCM and RSM, sometimes markedly so (Table 4b). We hypothesize that this is again attributable at least in part to the use of large-scale information in the domain interior via nudging and therefore have computed correlations for a mini-ensemble containing only these two models. This nudged mini ensemble has correlations consistently greater than both the best single model and the full ensemble mean.

DISCUSSION AND CONCLUSIONS. Given the variety of metrics examined, the variability of model performances across them, and the spatial and temporal variability and complexity of the climate, it is difficult to simply state whether the models successfully reproduce the climate over North America for 1980–2004. Our results are within the range of what has been found in other multiple model comparisons: seasonal temperature is relatively well produced by most models but seasonal precipitation is less so. All

| TABLE 4. Correlation of RCM simulated monthly time series with the UDEL observational data. Highest values of the six RCMs and of the two ensembles are underscored. |||||

a. Temperature: Deviation from model's annual cycle

	Southern California	Great Plains	South-central United States	Atlantic coast
CRCM	<u>0.86</u>	<u>0.92</u>	<u>0.85</u>	<u>0.90</u>
RSM	0.83	0.91	0.80	0.86
HadRM3	0.74	0.71	0.55	0.68
MM5	0.80	0.74	0.68	0.66
RegCM3	0.62	0.72	0.51	0.57
WRF	0.74	0.75	0.63	0.73
Ensemble	0.83	0.87	0.80	0.83
Nudged	<u>0.87</u>	<u>0.94</u>	<u>0.87</u>	<u>0.91</u>

b. Precipitation: Deviation from model's annual cycle

	Southern California	Great Plains	South-central United States	Atlantic coast
CRCM	0.68	<u>0.45</u>	<u>0.67</u>	<u>0.64</u>
RSM	<u>0.69</u>	0.39	0.59	0.60
HadRM3	0.63	0.28	0.39	0.20
MM5	0.67	0.33	0.41	0.33
RegCM3	0.66	0.33	0.30	0.34
WRF	0.68	0.30	0.48	0.33
Ensemble	0.70	0.43	0.63	0.56
Nudged	<u>0.69</u>	<u>0.45</u>	<u>0.68</u>	<u>0.70</u>

models have more difficulty reproducing the variability of precipitation than of temperature. They do correctly reproduce the seasonal cycle of temperature variability. Performance also varies substantially from one subregion to another.

There is no single model that performs "best" in terms of simulation of both temperature and precipitation based on the specific metrics we apply. Failure to discern a best model has been documented in recent papers on model evaluation (e.g., Gleckler et al. 2008; Reichler and Kim 2008; Walsh et al. 2008). We maintain that all NARCCAP simulations can provide useful information about current and future climate and that the spread across the simulations underscores the uncertainty in understanding and modeling climate processes.

However, differential model performance is discernible. With regard to seasonal average temperature biases over the whole domain RSM and MM5 had the lowest total RMSEs, whereas HadRM3 had the greatest overall temperature bias. RSM, RegCM3, and

MM5 had the lowest (absolute value) total comparative quartile index scores, giving RSM and MM5 the lowest temperature bias by both metrics combined. Aside from HadRM3, the models reproduced well the seasonal temperatures. We did identify some uncertainty for bias calculation based on choice of observed dataset for small regions in complex terrain.

Regarding precipitation bias, generalizations are harder to make. In addition, we found some important uncertainties in calculation of bias based on choice of observed dataset for regions with complex terrain.

Based on the sum of the four seasonal RMSEs, CRCM has the lowest total RMSE for the full domain, with WRF following closely, whereas HadRM3 and RegCM3 have the highest total RMSEs. Summing the absolute values of the comparative quartile index over all regions and seasons produces a slightly different picture, with MM5, CRCM, and RSM showing the lowest bias (scores of 13, 15, and 15, respectively). However, these models do not produce the lowest scores for each region taken individually. The full ensemble scored 18 as did RegCM3 and WRF, and HadRM3 scored highest at 23. Considering both metrics, the models with lowest precipitation bias are CRCM, MM5, and WRF. Although MM5 ranks among the top-tier performers for both mean temperature and precipitation, it is not a top performer with regard to variance of temperature and precipitation, as noted below.

The ensemble average of all six RCMs performs well on many metrics for mean temperature and precipitation, but it does not perform best in all cases. For example, CRCM outperforms it (as well as the nudged mini-ensemble) with regard to RMSE of mean seasonal precipitation in several seasons. The nudged mini-ensemble often performs better than the full ensemble, clearly reflecting the advantage of having additional information from the driving model within the domain interior.

Regarding temperature variability, RSM, CRCM, and WRF have relatively low total number of grid points showing statistically significant departures from observations in this category in winter, whereas CRCM and MM5 are the models with the lowest numbers in summer. The models in general perform better on this metric in winter, which contrasts with the seasonal tendencies for mean temperature performance. Regarding precipitation variability, all models perform relatively poorly, and there are no clear indications of differential model performance. Performance is slightly better in summer than in winter, which is consistent with the results for mean

precipitation. With regard to regional correlation metrics, results vary across the regions and seasons. The correlations of monthly time series for both temperature and precipitation do show some clear tendencies, with the nudged mini-ensemble generally having the highest correlation in the four regions for both temperature and precipitation. Of the six models, the CRCM has the highest temperature correlation in the four regions, with values very close to those of the nudged ensemble, as well as the highest correlation of the six models in all of the regions for precipitation but one, being surpassed by RSM in Southern California. The fact that both CRCM and RSM score relatively high on reproduction of variability for both temperature and precipitation suggests that the large-scale circulation, which is more faithfully captured in CRCM and RSM due to nudging, has important control over temperature and precipitation variability.

The relatively poorer performance of HadRM3, compared to the other models, for many of the metrics we used may well reflect the "home-court advantage" in that this is the one model that has not previously been used to perform simulations over North America (Takle et al. 2007). The performance of HadRM3 is generally substantially better when driven with ERA-Interim. We make several observations based on the comparison of HadRM3 driven by the different reanalyses: 1) it may be important to recognize the limitations of evaluating a model's performance based on simulations driven by a single set of reanalyses; 2) it may be appropriate to assess the quality of a reanalysis before using it to validate an RCM; and 3) HadRM3 is the only non–North American model being used in NARCCAP and it is unlikely that its current formulation would have been deemed acceptable if it had been tested using NCEP boundary conditions during its development. Thus, we establish here a potentially important uncertainty regarding which reanalysis acts as driver.

Our goal was to provide an overview of the relative performances of the six models both individually and as an ensemble with regard to temperature and precipitation. We have shown that all the models can simulate aspects of climate well, implying that they all can provide useful information about climate change. In particular, the results from phase I of NARCCAP will be used to establish uncertainty due to boundary conditions as well as final weighting of the models for the development of regional probabilities of climate change.

Some researchers have tried to create a single composite metric for weighting of models; for example,

Reichler and Kim (2008) successfully distinguished performance across different generations of models (e.g., CMIP1 vs CMIP3), although they had more difficulty distinguishing between individual models. Walsh et al. (2008) were able to rank global climate models over several nested domains of North America using a composite metric (combined RMSEs for three variables). We prefer to follow the approach of Gleckler et al. (2008) and look at the various metrics individually. Christensen et al. (2010), in working with the ENSEMBLES suite of RCMs, used six metrics that mainly consider temperature and precipitation and found differences among the models. Although a model that performed best was identified when the metrics were combined, there was still variety in which model performed best for each metric. The authors used the metrics and the combined metrics to establish different weighting schemes, not to eliminate poor or only use best models. They note that we still know little about how to comparatively evaluate regional models to the end of differentially ranking or weighting them. A similar conclusion was reached in an Intergovernmental Panel on Climate Change (IPCC) document developed to provide guidance for use of multimodel ensembles (Knutti et al. 2010) for the IPCC Fifth Assessment Report.

Although comparing regional models across different metrics provides a useful baseline evaluation, understanding why models perform the way they do is a more critical endeavor. As more of the 3D fields (outputs from the vertical levels) from the simulations are made available for all NARCCAP model runs, we plan to provide process-level analysis for why the models do well in certain regions and less well in others as well as variability in individual model performance. These in-depth analyses will improve our understanding of sources of errors in the regional climate models and, when combined with similar analyses of the RCM simulations driven by GCM current climate simulations, will help establish the ability of the GCM–RCM combinations to reliably simulate the climate of the region. This is a crucial first step in providing reasons for giving greater credibility to certain model results over others and thus providing higher quality information about future climate change and its uncertainties across North America.

REFERENCES

Anderson, C. J., and Coauthors, 2003: Hydrological processes in regional climate model simulations of the central United States flood of June–July 1993. *J. Hydrometeor.,* **4,** 584–598.

——, R. W. Arritt, and J. S. Kain, 2007: An alternative mass flux profile in the Kain–Fritsch convective parameterization and its effects in seasonal precipitation. *J. Hydrometeor.,* **8,** 1128–1140.

Bell, J., L. Sloan, and M. Snyder, 2004: Extreme climate events: A future scenario. *J. Climate,* **17,** 81–87.

Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeor,* **8,** 837–846.

Caya, D., and R. Laprise, 1999: A semi-Lagrangian semi-implicit regional climate model: The Canadian RCM. *Mon. Wea. Rev.,* **127,** 341–362.

Christensen, J. H., B. Machenhauer, R. G. Jones, C. Schär, P. M. Ruti, M. Castro, and G. Visconti, 1997: Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. *Climate Dyn.,* **13,** 489–506.

——, and Coauthors, 2007a: Regional climate projections. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 847–940.

——, T. R. Carter, M. Rummukainen, and G. Amanatidis, 2007b: Evaluating the performance and utility of regional climate models: The PRUDENCE project. *Climatic Change,* **81** (Suppl.), 1–6.

——, M. Rammukainen, and G. Lenderink, 2009: Formulation of very-high-resolution regional climate model ensembles for Europe. *ENSEMBLES: Climate Change and Its Impacts: Summary of Research and Results from the ENSEMBLES Project,* P. Van der Linden, and J. F. B. Mitchell, Eds., Met Office Hadley Center, 47–58.

——, E. Kjellström, F. Giorgi, G. Lenderink, and M. Rummukainen, 2010: Assigning relative weights to regional climate models: Exploring the concept. *Climate Res.,* **44,** 179–194, doi:10.3354/cr00916.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate,* **19,** 2122–2143.

Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 525–582.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.,* **33,** 140–158.

——, W. Gibson, and G. Taylor, cited 2011: 103-year high-resolution precipitation climate data set for the conterminous United States. PRISM Climate Group. [Available online at ftp://ftp.ncdc.noaa.gov /pub/data/prism100.]

de Elía, R., and Coauthors, 2008: Evaluation of uncertainties in the CRCM-simulated North American climate. *Climate Dyn.,* **30,** 113–132.

Diffenbaugh, N. S., M. A. Snyder, and L. C. Sloan, 2004: Could CO_2-induced land cover feedbacks alter near-shore upwelling regimes? *Proc. Nat. Acad. Sci. USA,* **101,** 27–32.

——, J. S. Pal, R. J. Trapp, and F. Giorgi, 2005: Fine-scale processes regulate the response of extreme events to global climate change. *Proc. Nat. Acad. Sci. USA,* **102,** 15 774–15 778.

Efron, B., and R. Tibshirani, 1993: *An Introduction to the Bootstrap.* Chapman & Hall/CRC, 450 pp.

Flato, G. M., cited 2012: The third generation Coupled Global Climate Model (CGCM3). Environment Canada Canadian Centre for Climate Modelling and Analysis. [Available online at www.ec.gc.ca /ccmac-cccma/default.asp?lang=En&n=1299529F-1.]

Fu, C., and Coauthors, 2005: Regional Climate Model Intercomparison Project for Asia. *Bull. Amer. Meteor. Soc.,* **86,** 257–266.

GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmospheric and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate,* **17,** 4641–4673.

Giorgi, F., M. R. Marinucci, and G. T. Bates, 1993a: Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. *Mon. Wea. Rev.,* **121,** 2794–2813.

——, ——, 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. *Mon. Wea. Rev.,* **121,** 2814–2832.

——, L. O. Mearns, C. Shields, and L. McDaniel, 1998: Regional nested model simulations of present day and $2xCO_2$ climate over the central plains of the US. *Climatic Change,* **40,** 457–493.

——, and Coauthors, 2001: Regional climate information: Evaluation and projections. *Climate Change 2001: The Scientific Basis.* J. T. Houghton et al., Eds., Cambridge University Press, 739–768.

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

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.

Govindasamy, B., P. B. Duffy, and J. Coquard, 2003: High-resolution simulations of global climate, Part 2: Effects of increased greenhouse gases. *Climate Dyn.,* **21,** 391–404, doi:10.1007/s00382-003-0340-6.

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

Gustafson, W. I., and L. R. Leung, 2007: Regional downscaling for air quality assessment: A reasonable proposition? *Bull. Amer. Meteor.. Soc.,* **88,** 1215–1227.

Gutowski, W. J., K. A. Kozak, R. W. Arritt, J. H. Christensen, J. Patton, and E. S. Takle, 2007: A possible constraint on regional precipitation intensity changes under global warming. *J. Hydrometeor.,* **8,** 1382–1396.

Han, J., and J. Roads, 2004: U.S. climate sensitivity simulated with the NCEP Regional Spectral Model. *Climatic Change,* **62,** 115–154.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.,* **90,** 1095–1107.

Jones, R., D. Hassell, D. Hudson, S. Wilson, G. Jenkins, and J. Mitchell, 2003: Generating high resolution climate change scenarios using PRECIS. UNDP National Communications Unit Workbook, 34 pp.

Juang, H., S. Hong, and M. Kanamitsu, 1997: The NMC nested regional spectral model: An update. *Bull. Amer. Meteor. Soc.,* **78,** 2125–2143.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kamga, A., and E. Buscalet , 2009: Simulation du climat de l'Afrique de l'Ouest à l'aide d'un modèle climatique régional. *La Meteor.,* **52,** 28–37.

Knutti, R., G. Abromowitz, M. Collins, V. Eyering, P. Gleckler, B. Hewitson, and L. Mearns, 2010: Good practice guidance paper on assessing and combining multi model climate projections. Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections, IPCC, 15 pp.

Köppen, W., 1900: Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt [Attempted climate classification in relation to plant distributions]. *Geogr. Z.,* **6,** 657–679.

Laprise, R., D. Caya, M. Giguère, G. Bergeron, H. Côté, J.-P. Blanchet, G. J. Boer, and N. A. McFarlane, 1998: Climate and climate change in western Canada as simulated by the Canadian Regional Climate Model. *Atmos.–Ocean,* **36,** 119–167.

——, ——, A. Frigon, and D. Paquin, 2003: Current and perturbed climate as simulated by the Canadian Regional Climate Model (CRCM-II) over northwestern North America. *Climate Dyn.,* **21,** 405–421.

Legates, D. R., and C. J. Willmott, 1990a: Mean seasonal and spatial variability global surface air temperature. *Theor. Appl. Climatol.,* **41,** 11–21.

——, and ——, 1990b: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.,* **10,** 111–127.

Leung, L. R., and W. I. Gustafson Jr., 2005: Potential regional climate change and implications to U.S. air quality. *Geophys. Res. Lett.,* **32,** L16711, doi:10.1029/2005GL022911.

——, 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.

——, ——, ——, 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.

——, ——, J. Han, and J. O. Roads, 2003c: Intercomparison of global reanalyses and regional simulations of cold season water budgets in the western United States. *J. Hydrometeor.,* **4,** 1067–1087.

——, ——, 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.

Marengo, J. A., R. G. Jones, L. M. Alves, and M. C. Valverde, 2009: Future change of temperature and precipitation extremes in South America as derived from the PRECIS regional climate modeling system. *Int. J. Climatol.,* **29,** 2241–2255, doi:10.1002/joc.1863.

Matsuura, K., and C. Willmott, cited 2010: Terrestrial air temperature and precipitation: 900-2008 gridded monthly time series (V 2.01). University of Delaware Department of Geography Center for Climatic Research. [Available online at http://climate.geog.udel.edu/~climate/html_pages/archive.html.]

Mearns, L. O., 2003: Issues in the impacts of climate variability and change on agriculture. *Climatic Change,* **60,** 1–6.

——, M. Hulme, T. R. Carter, R. Leemans, M. Lal, and P. Whetton, 2001: Climate scenario development. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 583–638.

——, F. Giorgi, C. Shields, and L. McDaniel, 2003: Climate scenarios for the southeast US based on GCM and regional modeling simulations. *Climatic Change,* **60,** 7–35.

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

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

Mitchell, T. D., cited 2008: CRU TS 2.10. University of East Anglia Climatic Research Unit. [Available online at www.cru.uea.ac.uk/cru/data/hrg/cru_ts_2.10.]

——, and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.,* **25,** 693–712.

Morse, A., C. Prentice, and T. Carter, 2009: Assessments of climate change impacts. ENSEMBLES: Climate change impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Center Rep., 107–129.

Nakicenovic, N., and Coauthors, 2000: *Special Report on Emissions Scenarios.* Cambridge University Press, 599 pp.

New, M., M. Hulme, and P. D. Jones, 2000: Representing twentieth century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate,* **13,** 2217–2238.

Nunes, A. M. B., and J. Roads, 2007: Influence of precipitation assimilation on a regional climate model's surface water and energy budgets. *J. Hydrometeor.,* **8,** 642–664.

Oleson, J. E., and Coauthors, 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.

Pal, J. S., E. E. Small, and E. A. B. Eltahir, 2000: Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM. *J. Geophys. Res.,* **105** (D24), 29 579–29 594.

——, and Coauthors, 2007: Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. *Bull. Amer. Meteor. Soc.,* **88,** 1395–1409.

Pan, Z., J. H. Christensen, R. W. Arritt, W. J. Gutowski Jr., E. S. Takle, and F. Otenio, 2001: Evaluation of uncertainties in regional climate change simulations. *J. Geophys. Res.,* **106,** 17 735–17 752.

Plummer, D. A., and Coauthors, 2006: Climate and climate change over North America as simulated by the Canadian RCM. *J. Climate,* **19,** 3112–3132.

Pope, V. D. and Coauthors, 2000: The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Climate Dyn.,* **16,** 123–146.

Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Amer. Meteor. Soc.,* **89,** 303–311.

Roads, J. O., 2004: Experimental weekly to seasonal, global to regional US precipitation forecasts. *J. Hydrol.,* **288,** 153–169.

Salathé, E. P., R. Steed, C. F. Maas, and P. Zahn, 2008: A high-resolution climate model for the Pacific Northwest: Mesoscale feedbacks and local responses to climate change. *J. Climate,* **21,** 5708–5726.

——, L. R. Leung, Y. Qian, and Y. Zhang, 2010: Regional climate model projections for the state of Washington. *Climatic Change,* **102,** 51–75.

Scinocca, J. F., and N. A. McFarlane, 2004: The variability of modeled tropical precipitation. *J. Atmos. Sci.,* **61,** 1993–2015.

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 Tech. Note NCAR/TN-468+STR, 88 pp.

Snyder, M. A., J. L. Bell, L. C. Sloan, P. B. Duffy, and B. Govindasamy, 2002: Climate responses to a doubling of atmospheric carbon dioxide for a climatically vulnerable region. *Geophys. Res. Lett.,* **29,** 1514, doi:10.1029/2001GL014431.

Takle, E. S., and Coauthors, 1999: Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results. *J. Geophys. Res.,* **104,** 19 443–19 462.

——, and Coauthors, 2007: Transferability intercomparison: An opportunity for new insight on the global water cycle and energy budget. *Bull. Amer. Meteor. Soc.,* **88,** 375–384.

Tebaldi, C., and B. Sansó, 2009: Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *J. Roy. Stat. Soc.,* **172A,** 83–106, doi:10.1111/j.1467-985X.2008.00545.x.

——, L. O. Mearns , D. Nychka, and R. L. Smith, 2004: Regional probabilities of precipitation change: A Bayesian approach. *Geophys. Res. Lett.,* **31,** L24213, doi:10.1029/2004GL021276.

——, R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *J. Climate,* **18,** 1524–1540.

von Storch, H., and F. W. Zwiers, 1999: Statistical Analysis in Climate Research. Cambridge University Press, 484 pp.

Walsh, J. E., W. L. Chapman, C. Romanovsky, J. H. Christensen, and M. Stendal, 2008: Global climate model performance over Alaska and Greenland. *J. Climate,* **21,** 6156–6174.

Wilby, R. L., L. E. Hay, W. J. Gutowski Jr., R. W. Arritt, E. S. Takle, Z. Pan, G. H. Leavesley, and M. Clark, 2000: Hydrological responses to dynamically and statistically downscaled general circulation model output. *Geophys. Res. Lett.,* **27,** 1199–1202.

Willmott, C. J., and K. Matsuura, 1995: Smart interpolation of annually averaged air temperature in the United States. *J. Appl. Meteor.,* **34,** 2577–2586.

——, C. M. Rowe, and W. D. Philpot, 1985: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Amer. Cartogr.,* **12,** 5–16.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical downscaling approaches to downscaling climate model outputs. *Climatic Change,* **62,** 189–216.

Xu, Y. L., Y. Zhang, E. D. Lin, W. T. Lin, W. J. Dong, R. Jones, D. Hassell, and S. Wilson, 2006: Analyses on the climate change responses over China under SRES B2 scenario using PRECIS. *Chin. Sci. Bull.,* **51,** 2260–2267.