

Evaluation of present and future North American Regional Climate Change Assessment Program (NARCCAP) regional climate simulations over the southeast United States

S. Sobolowski^{1,2} and T. Pavelsky¹

Received 18 June 2011; revised 3 October 2011; accepted 26 October 2011; published 5 January 2012.

[1] In order to make well-informed decisions in response to future climate change, officials and the public require reliable climate projections at the scale of tens of kilometers, rather than the hundreds of kilometers that the current atmosphere–ocean general circulation models provide. Recent efforts such as the North American Regional Climate Change Assessment Program (NARCCAP) aim to address this need. This study has two principal aims: (1) evaluate the seasonal performance of the NARCCAP simulations over the southeast United States for both present (1971–2000) and future (2041–2070) periods and (2) assess the impact of a performance-based weighting scheme on bias and uncertainty. Application of the weighting scheme results in a substantial reduction in magnitude and percent area exhibiting significant bias in all seasons for both temperature and precipitation. The weighting scheme is then expanded to evaluate future change. Temperature changes are universally positive and outside the bounds of natural variability over the entire region and in all seasons. Application of the weighting scheme tightens confidence intervals by as much as 1.6°C. Future precipitation changes are modest, are of mixed sign, and vary by season and location. Though uncertainty is reduced by as much as 50%, the projected changes are generally not outside the bounds of natural background variability. Thus, under the NARCCAP simulations, stress on water resources is most likely to come from increased temperatures and not changes in mean seasonal precipitation. For energy use, the implication is that the ~3°C temperature increase during the peak use summer season may place additional strain on power grids.

Citation: Sobolowski, S., and T. Pavelsky (2012), Evaluation of present and future North American Regional Climate Change Assessment Program (NARCCAP) regional climate simulations over the southeast United States, *J. Geophys. Res.*, *117*, D01101, doi:10.1029/2011JD016430.

1. Introduction

[2] Currently, large-scale atmosphere–ocean general circulation model (AOGCM) experiments such as those employed by the Intergovernmental Panel on Climate Change (IPCC) do not provide consistent, reliable information, for the present or the future, at scales over which humans tend to act (e.g., cities, states, municipalities) [Oreskes *et al.*, 2010]. In order to make well-informed decisions with respect to climate change adaptation and/or mitigation, elected officials, planners and the public require information at spatial scales on the order of tens, not hundreds, of kilometers. This issue is particularly acute over regions where global models exhibit substantial disagreement with respect to magnitude and even sign of present

conditions and/or projected changes in climate variables. Development of reliable multimodel regional-scale climate model ensembles, improved model performance, and reduction of uncertainty (e.g., due to model error, disagreement, internal variability) have been identified as pressing current research needs following the IPCC's Fourth Assessment Report (AR4), published in 2007 [Doherty *et al.*, 2009]. The present study aims to address some of these needs through an evaluation of a six member ensemble of regional climate models over the southeast United States. There are two goals: one is to evaluate the multimodel performance of the North American Regional Climate Change Program (NARCCAP) relative to observations over the southeast United States and improve multimodel simulation of present-day seasonal temperature and precipitation patterns at the subregional scale. The second goal is to evaluate future changes in temperature and precipitation over this region and reduce their associated uncertainties using an expanded version of the reliability ensemble averaging (REA) technique developed by Giorgi and Mearns [2003]. The resulting weighted projections of future change and estimates of uncertainty provides information on a scale (e.g., 50 km²) that

¹Department Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

²Now at Bjerknes Centre for Climate Research, Bergen, Norway.

is useful for water resource planners, town and state officials, the agricultural sector and the general public.

[3] The climate modeling community has been aware of the need for better regional-scale climate projections for some time, and numerous studies employing regional climate models (RCMs) have been performed in order to evaluate mean temperature and precipitation [e.g., *Mearns et al.*, 2003; *Chen et al.*, 2003] and extreme events [e.g., *Diffenbaugh et al.* 2005] at higher spatial resolutions. These investigations generally employ only a few models and/or a limited number of realizations and are thus not particularly robust. Because of intermodel differences, particularly in handling subgrid-scale parameterizations and land-atmosphere interactions, recent studies suggest that multimodel means almost always provide estimates superior to any one “good” model [*Phillips and Gleckler*, 2006; *Gleckler et al.*, 2008]. Recently, coordinated international efforts have made inroads in providing the research community with multimodel regional climate simulations through efforts such as NARCCAP and the European PRUDENCE and ENSEMBLES programs (<http://prudence.dmi.dk/>). NARCCAP provides multimodel output for present and future climate at a spatial scale of 50 km² over the entire North American continent and is the primary data source for the present study [*Mearns et al.*, 2009].

[4] Even though the higher resolution of RCMs helps to resolve finer-scale aspects of climate they suffer from many of the same systematic biases and errors inherent in their AOGCM counterparts. For example, RCMs must still parameterize many subgrid-scale physical processes (e.g., convection, microphysics, surface-atmosphere interactions, boundary layer processes). Thus, model error due to imperfect parameterization schemes and dynamical specifications increases the uncertainty surrounding simulated climate. Another source of uncertainty is the large influence of internal variability (i.e., weather and climate “noise”) due to the small number of realizations in most modeling experiments. *Deser et al.* [2011] argue that the contribution of internal variability to total uncertainty is, in most cases larger than intermodel variability and that a large number of realizations (e.g., 20+) of each model need to be run in order to obtain robust results for variables other than surface temperature. A final source of uncertainty is scenario uncertainty due to the estimated trajectories of global development and future emissions. This source of uncertainty is one over which the researcher has little control, but it can be mitigated by evaluating a number of more or less possible future scenarios. Recent studies argue that attempts to reduce uncertainty should focus on improving model performance and on adequately accounting for the influence of internal variability [*Hawkins and Sutton*, 2009]. Others acknowledge that some uncertainties can never be significantly reduced (e.g., scenario uncertainty) and suggest a more broad-based approach that allows for decision making under deep uncertainty [*Mearns*, 2010]. While these approaches for mitigating and reducing the various sources of uncertainty are not mutually exclusive in their application (one could conceivably employ all of them) we choose here to focus on reducing model error and future uncertainty given a prescribed scenario. One way to do this is to apply weighting schemes that preferentially weight models on the basis of performance relative to present-day conditions and/or level of future

agreement [e.g., *Giorgi and Mearns*, 2003; *Raisanen et al.*, 2010]. Recent research suggests that bias relative to present-day observations may be an effective predictor of the accuracy of a model’s future climate projection, lending support to approaches that employ multimodel weighting schemes based, at least in part, on model bias [*Tebaldi et al.*, 2005; *Watterson*, 2008; *Matsueda and Palmer*, 2011]. Others offer generally positive evaluations of performance-based multimodel weighting but advocate adding additional measures that also account for intermodel similarity in order to improve the accuracy of future simulations [*Raisanen et al.*, 2010]. However, it should be noted that some studies suggest there is limited advantage to selecting models on the basis of present-day performance and advocate larger ensembles and greater numbers of realizations in order to reduce bias and uncertainty [*Santer et al.*, 2009; *Pierce et al.*, 2009; *Reifen and Toumi*, 2009]. It is not the purpose of this manuscript to enter into a debate on the merits and drawbacks to weighting but to apply one approach to model evaluation, report on the outcome and assess the weighting scheme’s efficacy. Given the small number of models and realizations in the NARCCAP database at the time of evaluation (six), we choose to take the performance-based weighting approach.

[5] The southeast United States is an ideal candidate for the performance-based model weighting approach just described. The observational record over this region is robust, giving us high confidence in our ground truth data. It is an economically important region that has experienced exceptional growth over the past two decades. For example, North Carolina and Georgia report population increases of 42% and 50% since 1990, respectively (data are available from <http://www.census.gov/>). This increase in population has put pressure on water and energy resources and has reduced the region’s capacity to weather multiyear droughts, two of which have occurred in the past 13 years (1998–2002 and 2005–2008). As *Seager et al.* [2009] point out, the recent 2005–2008 drought was typical in amplitude and duration compared to the historical record but its effects were exacerbated because of increased demand. This inability to buffer the effects of the drought led to severe municipal water rationing measures and losses in excess of \$1.3 billion [*Manuel*, 2008]. In addition to stresses on water resources, energy demands have also risen along with population, economic activity and personal income. Over the period of 1997–2006 overall energy consumption in the Southeast increased 13%, over twice that of the national average [*Damassa*, 2009]. Rising temperatures will likely further increase energy demand over the region, particularly during peak use months from late spring to early fall. Because of this rising demand for water and energy the southeast United States is particularly vulnerable to the effects of global climate change on regional temperature and precipitation. However, the southeast United States is a climatically complex region, bordered on the south and east by the Gulf of Mexico and Atlantic Ocean and influenced by the orographic effects of the Appalachian and Blue Ridge mountains. Assessment of potential impacts is complicated by the fact that 20th century trends in seasonal temperature and precipitation are modest [*Karl and Knight*, 1998; *Knutson et al.*, 2006] and that there is a lack of agreement on future changes with respect to precipitation [*Christensen*

et al., 2007; *Seager et al.*, 2009]. In most seasons the line of zero precipitation change typically runs through the Southeast as it lies on the poleward flank of the projected subtropical drying associated with expansion of the Hadley Cell. Thus, the Southeast is a region where elected officials, planners and the public require reliable information at spatial scales on the order of tens, not hundreds, of kilometers in order to make well-informed decisions with respect to climate change adaptation and/or mitigation.

[6] In this study we evaluate seasonal multimodel mean temperature and precipitation output from six RCM-AOGCM pairs. These data are obtained from the publicly available NARCCAP archive [*Mearns et al.*, 2009]. Once multimodel mean bias and average model skill relative to observations are identified, the enhanced REA weighting scheme is applied in order to reduce bias in the present-day simulations and reduce uncertainty in future projections. We expand on previous studies which focused on either a single RCM-AOGCM pair [e.g., *Giorgi et al.*, 1998; *Mearns et al.*, 2003] or applied weighting to regional averages rather than at subregional scales [e.g., *Giorgi and Mearns*, 2003].

[7] In section 2 we describe the NARCCAP modeling experiments and the output used in this study. We also discuss the gridded observation-based data sets used as ground truth for present-day comparisons in section 2. The weighting methodology is introduced in section 3. Results of the analysis for the present-day and future simulations are presented in sections 4 and 5, respectively. Discussion and conclusions are presented in section 6.

2. Models and Data

[8] The modeled precipitation and temperature data used in this study are publicly available from the NARCCAP database and are the result of work performed by participating modeling groups (L. Mearns, The North American Regional Climate Change Assessment Program Dataset, National Center for Atmospheric Research Earth System Grid data portal, 2007, updated 2011). (See the NARCCAP Web site (<http://www.narccap.ucar.edu>) for complete descriptions of the program and its various experiments.) The RCMs are driven with AOGCM boundary conditions and cover a domain that includes the conterminous United States and Canada at a spatial resolution of 50 km. Here we focus on six RCM-AOGCM pairings that each have one realization covering the late 20th century (1971–2000) and the mid 21st century (2041–2070). Each pairing has one realization for each time period resulting in a 6 member multimodel ensemble. The RCM-AOGM pairs are as follows: the Canadian Regional Climate Model version 4 forced by the Community Climate System Model (CRCM-CCSM in the NARCCAP archive), the Canadian Regional Climate Model version 4 forced by the Coupled Global Climate Model version 3 (CRCM-CGCM3 in the archive), the Hadley Centre Regional Model version 3 forced by the Hadley Centre Climate Model version 3 (HRM3-HadCM3 in the archive), the fifth-generation Pennsylvania State University–NCAR Mesoscale Model forced by the Community Climate System Model (MM5I-CCSM in the archive), the International Centre for Theoretical Physics Regional Climate Model version 3 forced by the Coupled Global Climate Model

(RCM3-CGCM in the archive) and the International Centre for Theoretical Physics Regional Climate Model version 3 forced by the NOAA Geophysical Fluid Dynamics Laboratory climate model (RCM3-GFDL in the archive). We do not describe the particular characteristics of individual RCMs in depth here as this study is concerned with an evaluation of overall, multimodel mean performance rather than an intercomparison between individual models.

[9] For the present-day simulation the AOGCMs have been initialized by historical observations while the 21st century runs have been forced according to the SRES A2 emissions scenario. The A2 emissions scenario envisions heterogeneous growth with the global population rising to over 10 billion by 2050 and CO₂ concentrations of 575 (870) ppm by the middle (end) of the 21st century. Though A2 is at the high end of SRES emissions scenarios it is not the highest and was chosen by the NARCCAP group because, from an adaptation and mitigation perspective, a higher-emission scenario potentially provides more information than a low-emission scenario would.

[10] Gridded data sets of monthly mean precipitation rate (P , mm/d) and surface temperature (T , °C), are employed as ground truth for comparison to present-day (historical) RCM-AOGCM output. Terrestrial precipitation and temperature data are obtained from the *Willmott and Matsuura* [1995] (hereinafter WM) $0.5^\circ \times 0.5^\circ$ gridded monthly time series from 1900 to 2008 (available at <http://climate.geog.udel.edu/~climate/>). These data sets are developed from observations and are especially well constrained over regions with high concentrations of point data such as the southeast United States. WM employ a combination of techniques based on observational data and Climatologically Aided Interpolation (CAI) coupled with an enhanced distance weighting method to build their data sets [*Legates and Willmott*, 1990; *Willmott and Matsuura*, 1995; *Willmott and Robeson*, 1995].

[11] In order to facilitate comparison all model data are interpolated to a common $0.5^\circ \times 0.5^\circ$ grid on the basis of the observations. Further, monthly and seasonal means are computed from modeled 3 h precipitation and temperature outputs. We define the spatial domain of the southeast United States as the land areas from 92.75°W – 75.25°W and 29.25°N – 37.75°N . The study area encompasses the states of North Carolina, South Carolina, Georgia, Tennessee, Alabama, Mississippi, and parts of six others (Figure 1). The Florida peninsula is left out of the analysis because the driving AOGCMs generally do not allow for clear representation of this feature.

3. Bias-, Skill-, and Distance-Based Weighting Scheme

[12] In this study we adapt the reliability ensemble averaging (REA) techniques developed by *Giorgi and Mearns* [2003] to the NARCCAP ensemble over the southeast United States. Whereas *Giorgi and Mearns* [2003] applied the REA scheme to large-scale regional averages (e.g., all of North America is divided into 5 regions), we apply weights at each 50 km grid cell over the study area. Further, we add an additional weighting criterion that is based on each model's ability to match the probability density function (pdf) of the observations. These weights are applied to the

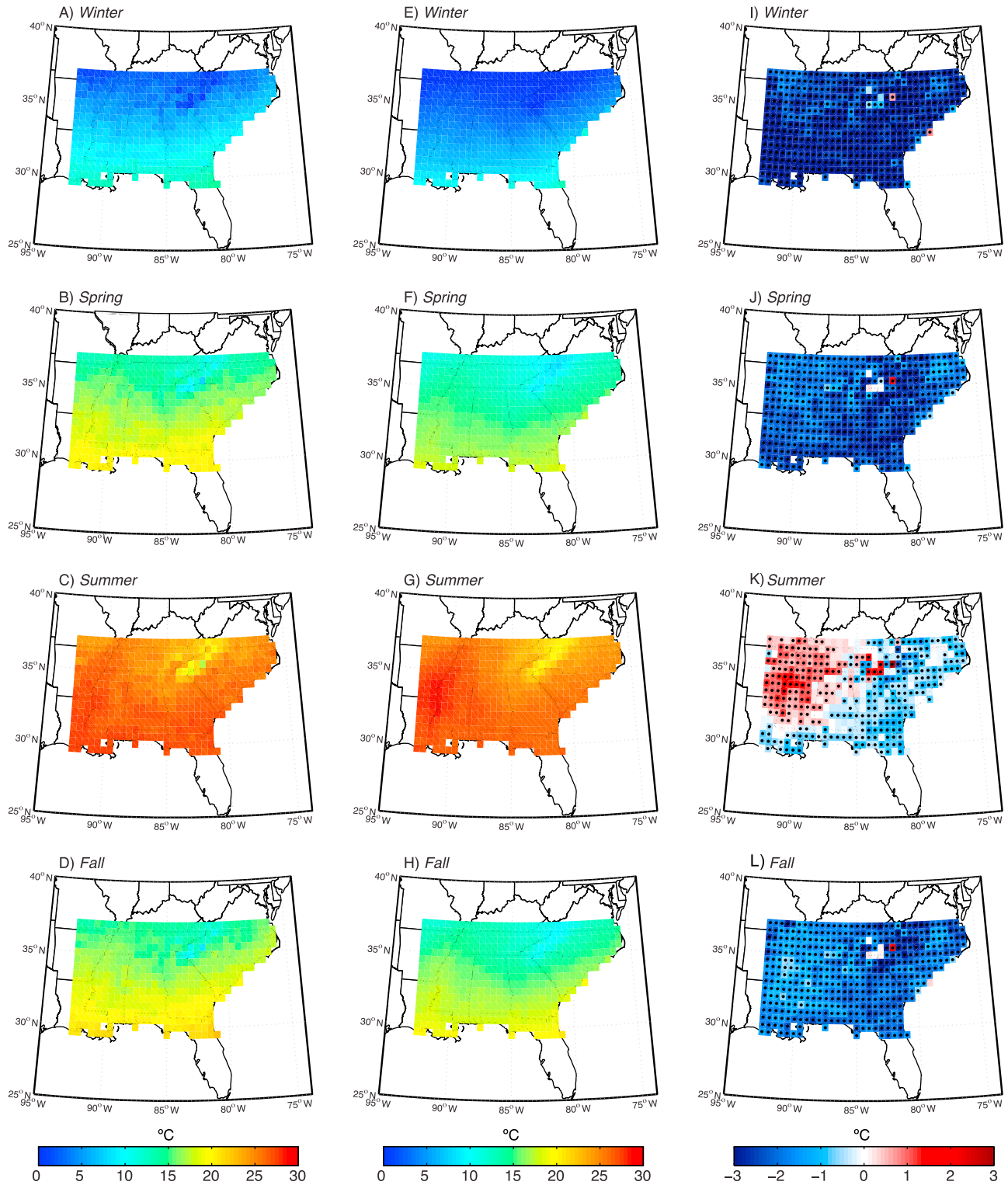


Figure 1. (a–d) Observed mean seasonal surface temperature ($^{\circ}\text{C}$), (e–h) North American Regional Climate Change Assessment Program (NARCCAP) multimodel mean seasonal surface temperature, and (i–l) mean seasonal bias. Stippling indicates grid cells where bias is statistically significant (at the 5% level), as measured by two-sided Student's t tests.

present-day NARCCAP output in order to illustrate present-day bias reduction *and* future change in P and T (ΔP and ΔT , respectively) where the weighting scheme reduces uncertainty on the basis of model reliability and agreement.

We define future change for each model as the difference between the 2041–2070 and 1970–2000 means. In this section we provide a brief overview of the weighting techniques, essential equations and assumptions. For all equations

we follow the naming and numbering conventions of [Giorgi and Mearns, 2003] for consistency.

[13] For the unweighted case the multimodel mean estimated climate change is calculated for T and P (note that ΔP is presented as percent change). The uncertainty surrounding this estimate is measured by calculating the root-mean-square difference (RMSD), $\delta_{\Delta T}$ for temperature. If the pdf of actual future changes is close to Gaussian, $\delta_{\Delta T}$ is simply the standard deviation (describing the 68.3% confidence bounds) and the 95% confidence bounds may be approximately described by $\pm 2\delta_{\Delta T}$. In this case, the confidence bounds are only approximately defined because we do not really know the shape of the future pdf and invoke the central limit theorem to justify the assumption of normality. If the shape of the future change pdf is uniform, then bounds described by the RMSD are about 58% and ± 2 times the RMSD encompasses the entire distribution and is thus not particularly informative.

[14] For the REA weighted case, the multimodel mean change, taking T as an illustrative example, may be described as follows:

$$\widetilde{\Delta T} = \frac{\sum_{i=1}^n R_i \Delta T_i}{\sum_{i=1}^n R_i} \quad (1)$$

where n is the number of RCM-AOGCM pairs and R_i is a weighting factor that is composed of three components. These components account for model bias relative to the observations ($R_{B,i}$), model skill at reproducing the distribution of present-day observations ($R_{S,i}$) and future convergence toward the multimodel mean ($R_{D,i}$). Thus, R_i may be decomposed as follows:

$$R_i = [R_{B,i}^m R_{S,i}^n R_{D,i}^p] \quad (2)$$

$$R_{B,i} = \frac{\epsilon_T}{\text{abs}(T_i - T_{\text{obs}})} \quad (3)$$

$$R_{S,i} = \sum_1^n \min(Z_i, Z_{\text{obs}}) \quad (4)$$

$$R_{D,i} = \frac{\epsilon_T}{\text{abs}(\Delta T_i - \widetilde{\Delta T})}. \quad (5)$$

[15] The bias component ($R_{B,i}$) is defined as the difference between each RCM-AOGCM pair and observations over the 1970–2000 historical period. Epsilon is an estimated measure of background, or natural, variability. In order to calculate ϵ the 1900–2008 temperature and precipitation time series at each grid cell of the WM data sets are linearly detrended and smoothed, using a 30 year moving average to remove long-term trends. Next, ϵ is estimated as the difference between the 97.5th and 2.5th percentiles of the smoothed time series.

[16] The skill score component ($R_{S,i}$) is a metric adapted from Perkins *et al.* [2007] and is not part of the original REA methodology described by Giorgi and Mearns [2003]. Z_i , Z_{obs} are the frequency distributions across all bins (n) of the models and observations, respectively. Bins are set at every 1 mm/d for P , while the bins for T are set at every 2°. Thus, the skill score measures how well the pdf of each model matches the pdf of the observations over the 1971–2000 period, with 1 being a perfect match and 0 being perfect

disagreement. The skill scores are computed here using monthly mean values and therefore measure the models' ability to replicate large-scale average conditions (e.g., effects of large-scale circulation, low-frequency variability, etc.) rather than the synoptic conditions that a skill score computed with daily values might capture.

[17] The distance criterion ($R_{D,i}$) measures how close the projections of individual models are to the multimodel mean. For the first iteration ΔT is used as a “best guess” estimate of the future change. Subsequently we update the weighted average and compute the distance criterion again. Since $\widetilde{\Delta T}$ is updated after each calculation of the distance criterion this represents an iterative procedure and is repeated until convergence occurs (about five iterations). Note also that weights themselves may be weighted on the basis of either subjective or objective criteria (m , n , p in equation (2)). For the present study these values are all held at 1.

[18] Uncertainty in the weighted case is also measured by calculating RMSD, $\tilde{\delta}_{\Delta T}$, in a manner analogous to the unweighted case. However, in this instance $\widetilde{\Delta T}$ and R_i are employed as follows:

$$\tilde{\delta}_{\Delta T} = \left[\frac{\sum_{i=1}^n R_i (\Delta T_i - \widetilde{\Delta T})^2}{\sum_{i=1}^n R_i} \right]^{1/2} \quad (6)$$

[19] As before, if we assume the future change to be normal the 95% uncertainty bounds may be approximated by $\pm 2\tilde{\delta}_{\Delta T}$.

[20] The REA weights are applied at the monthly level and then the weighted seasonal averages are computed. In addition to applying weighting factor, R_i to the multimodel mean future change as described previously, we also analyze the combined effects of $R_{B,i}$ and $R_{S,i}$ on the present-day simulations in order to assess the performance of the NARCCAP RCMs and the ability of the weighting scheme to reduce bias in the present-day multimodel mean. In the latter case the weights are applied to the monthly means of each model in all years so that significance and confidence bounds of the bias may be assessed via a Student's t test. The difference between the bias computed by weighting each model's mean monthly values for all years and weighting each model's monthly climatological mean is negligible.

4. Present-Day Climate and Model Bias

4.1. Seasonal Temperature

[21] The present-day observed seasonal temperature climatology (1971–2000 mean) is shown in Figures 1a–1d. The Southeast is characterized by relatively homogeneous spatial distribution of seasonal temperatures with a distinctive north-south gradient in the winter which weakens somewhat in the spring and fall. The summer season is uniformly hot and humid with the exception of the Appalachian mountains where temperatures are cooler. Figures 1e–1h show the seasonal multimodel mean temperature from the NARCCAP simulations. While the spatial patterns of the modeled temperatures are similar to the observations there is a clear cold bias over much of the Southeast in winter, spring and fall.

Differences between the models and observations in summer are modest and appear to be positive (negative) in the west (east). Seasonal bias is shown explicitly in Figures 1i–1l. Stippling indicates grid cells where the bias is statistically significant at the 5% level, as measured by a Student's t test. Using these measures as guidance, the multimodel mean exhibits significant negative bias over much of the Southeast in winter, spring and fall with regional averages of -2.7°C , -2.0°C , and -1.5°C , respectively. In many areas, the bias exceeds -3°C . In summer the magnitude of the bias is lower (generally $< \pm 1^{\circ}\text{C}$) and fewer regions exhibit statistically significant bias than fall–spring. However, there are regions of significant positive bias along the Mississippi, negative bias over the Appalachian mountains and southeast Georgia.

[22] Another way in which we assess multimodel performance is to examine the skill scores averaged across all models and the differences in variability between the models and observations. The multimodel mean is not appropriate for these comparisons as it reduces variability as an effect of averaging. Rather, these metrics are computed for each model individually and then averaged across all models. Figures 2a–2d show the average model skill in reproducing the seasonal pdf's of the observations. Generally, regions of poor skill correspond to regions of high bias shown in Figure 1. In winter and spring there is a north (south) pattern of high (low) skill scores. During summer, the pattern has a more east (west) orientation with the eastern portion of the Southeast exhibiting high skill scores (> 0.7) while along the Mississippi river skill scores are around 0.5. Generally, high skill scores are exhibited through the central and extreme southern portions of the Southeast during fall with low scores seen around the Blue Ridge mountains. We may infer from these results that model skill varies widely depending on both location and season.

[23] The difference in standard deviation between the models and observations is shown in Figures 2e–2h (we show the standard deviation averaged across all models, not the multimodel mean standard deviation). These differences are generally modest in the winter, spring and fall ($< \pm 0.3^{\circ}\text{C}$), though there is region-wide and systemic underestimation of variability in the winter season. During summer, however, the models overestimate variability over the entire Southeast with some areas along the Mississippi river exhibiting spreads that are larger than the observations by a degree or more.

4.2. Seasonal Precipitation

[24] The present-day observed seasonal precipitation climatology (1971–2000 mean) is shown in Figures 3a–3d. The observations illustrate the complex climate dynamics of the Southeast region. From late fall through spring moisture is transported out of the Gulf of Mexico in a roughly southwest to northeast direction as it is entrained in midlatitude storm systems. During summer convective precipitation dominates and the influence of the land-ocean contrast and tropical systems are clear along the coasts. Fall exhibits precipitation minima across the region. The orographic effects of the Appalachian mountains are visible in all seasons to a greater or lesser degree. The multimodel mean seasonal precipitation from the NARCCAP simulations is

shown in Figures 3e–3h. The models tend to overestimate the orographic effects of the Appalachian mountains in all seasons but most markedly in winter and summer. The multimodel mean is also generally unable to capture the spatial patterns of precipitation in any season with possible exception of spring. As Figures 3i–3l illustrate, the models tend to significantly underestimate (overestimate) precipitation in the western (eastern) part of the study area during winter and spring. Precipitation is significantly underestimated along the coasts and along the Mississippi in summer and region-wide in fall. These patterns suggest the RCMs reproduce at least some of the well known, systematic errors of the driving GCMs. In particular, the errors suggest that the models have difficulty reproducing moisture transport and synoptic-scale activity during the cool seasons, convective processes and land-ocean contrasts during warm seasons and orographic effects [e.g., *Martin et al.*, 2010, and references therein]. There are also land surface process biases that likely contribute to these patterns as well.

[25] Precipitation skill scores are shown in Figures 4a–4d. The lowest skill scores are seen along the Gulf coast and the Appalachian mountains during summer, otherwise skill scores are generally above 0.6 across the region in all seasons. Figures 4e–4h show the difference in variability between the models and observations. The models tend to underestimate variability in the western region during winter and spring. During fall variability is underestimated over the entire Southeast, while summer exhibits modest and heterogeneous differences in variability.

4.3. Effects of Weighting on Simulations of Present-Day Climate

[26] Biases in the NARCCAP simulations of temperature and precipitation appear in all seasons, but the patterns and magnitudes are not consistent across seasons. This spatial and temporal variability in the bias is likely linked to subregional differences in climate, the physical processes that dominate the Southeast at different times of the year and the models' difficulty in reproducing moist processes and orographic effects. Weights based on monthly model bias ($R_{B,i}$) and skill score ($R_{S,i}$) are applied to each individual model and the seasonal multimodel means are computed. Figure 5 shows the weighted bias for temperature (compare to Figures 1i–1l). Although statistically significant biases are still evident over much of the Southeast during winter, spring and fall, the magnitude and area affected are both greatly reduced. Temperature bias is all but eliminated in summer. Figure 6 shows the effects of the weighting scheme on the estimated pdf's of temperature at three selected locations across all four seasons. This result illustrates how the effectiveness of the weighting scheme varies depending on location and season.

[27] Figure 7 shows the bias reduction for precipitation (compare to Figures 3i–3l). As with temperature the bias is reduced in both magnitude and area affected in all seasons. Winter still exhibits statistically significant negative biases in the western portion of the study area. However, spring bias is virtually eliminated and summer bias is limited to the coastal areas. The negative bias in fall is still region-wide and the reduction in magnitude is small but nonnegligible. Estimating the pdf's of precipitation under normal assumptions

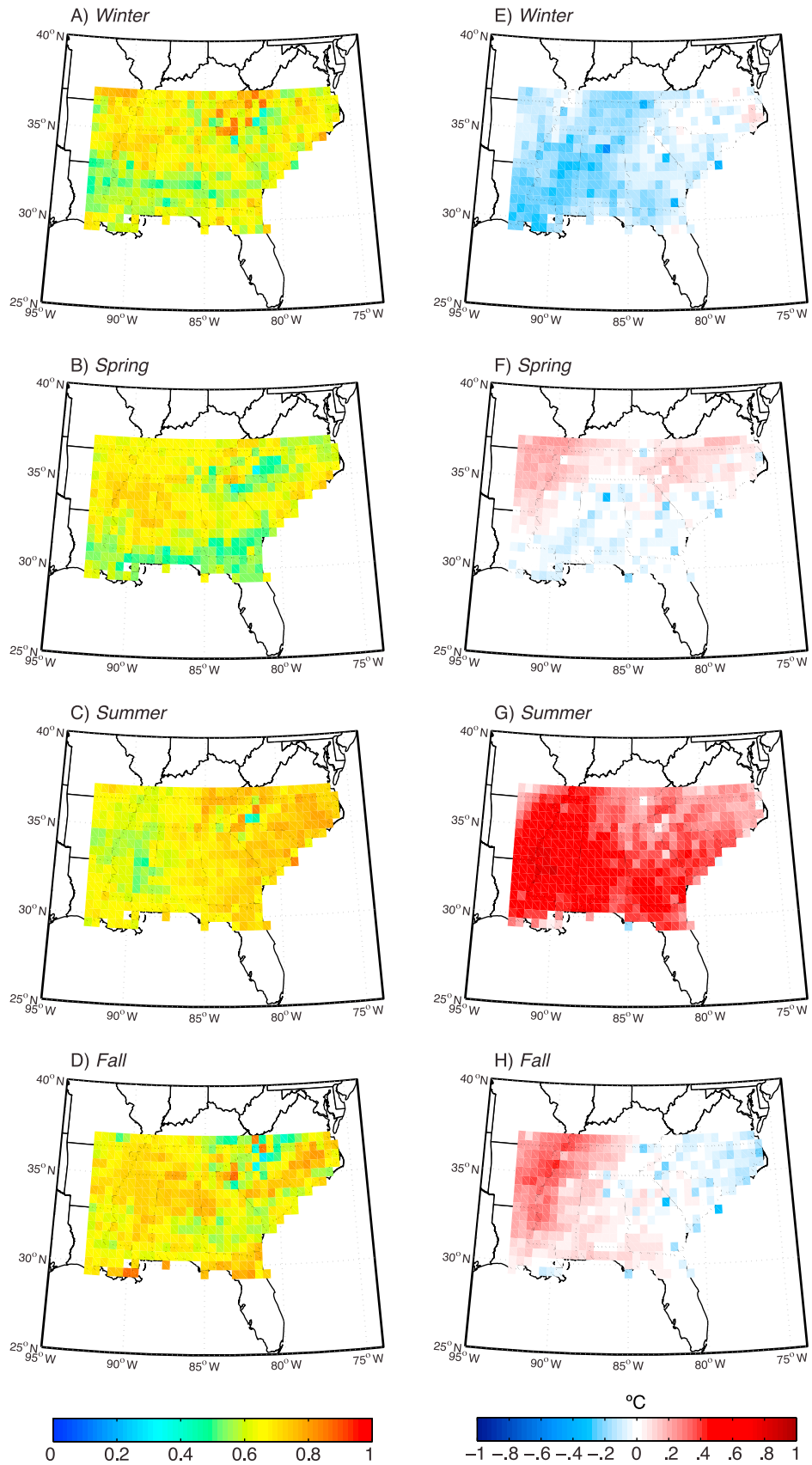


Figure 2. (a–d) Seasonal temperature skill score averaged across all models and (e–h) average difference in standard deviation between models and observations (°C).

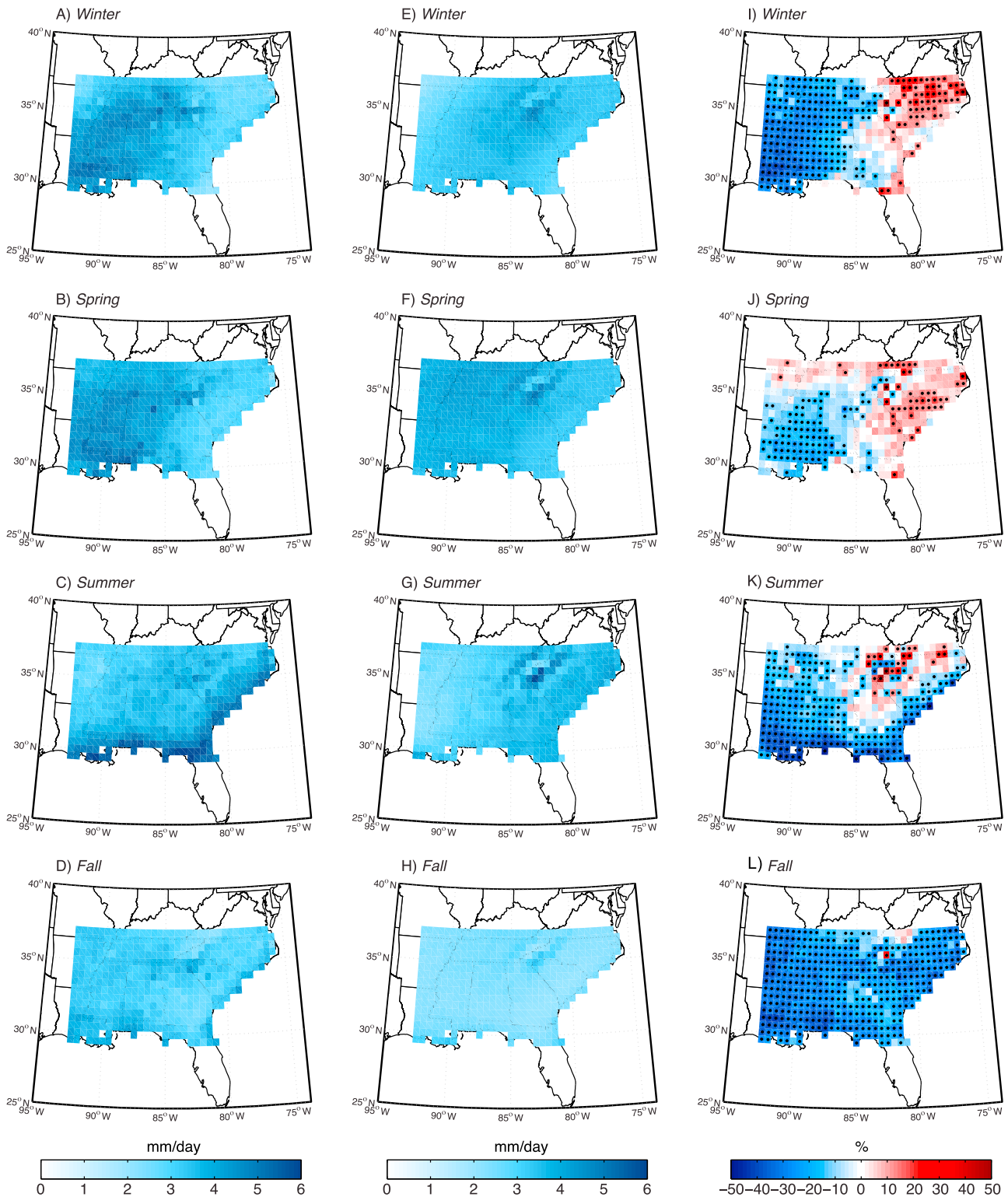


Figure 3. (a–d) Observed mean seasonal precipitation (mm/d), (e–h) NARCCAP multimodel mean seasonal mean precipitation, and (i–l) mean seasonal bias (shown as percentage). Stippling indicates grid cells where bias is statistically significant (at the 5% level), as measured by two-sided Student’s *t* tests.

produces similar, spatially and temporally varying, results as Figure 6, though both the weighted and unweighted multi-model means underestimate the spread of the observations (not shown).

[28] Figure 8 shows the regional average bias for temperature (Figure 8a) and precipitation (Figure 8b) with 95% confidence bounds for both the unweighted and weighted cases. Also shown in Figure 8 is the percent area of the

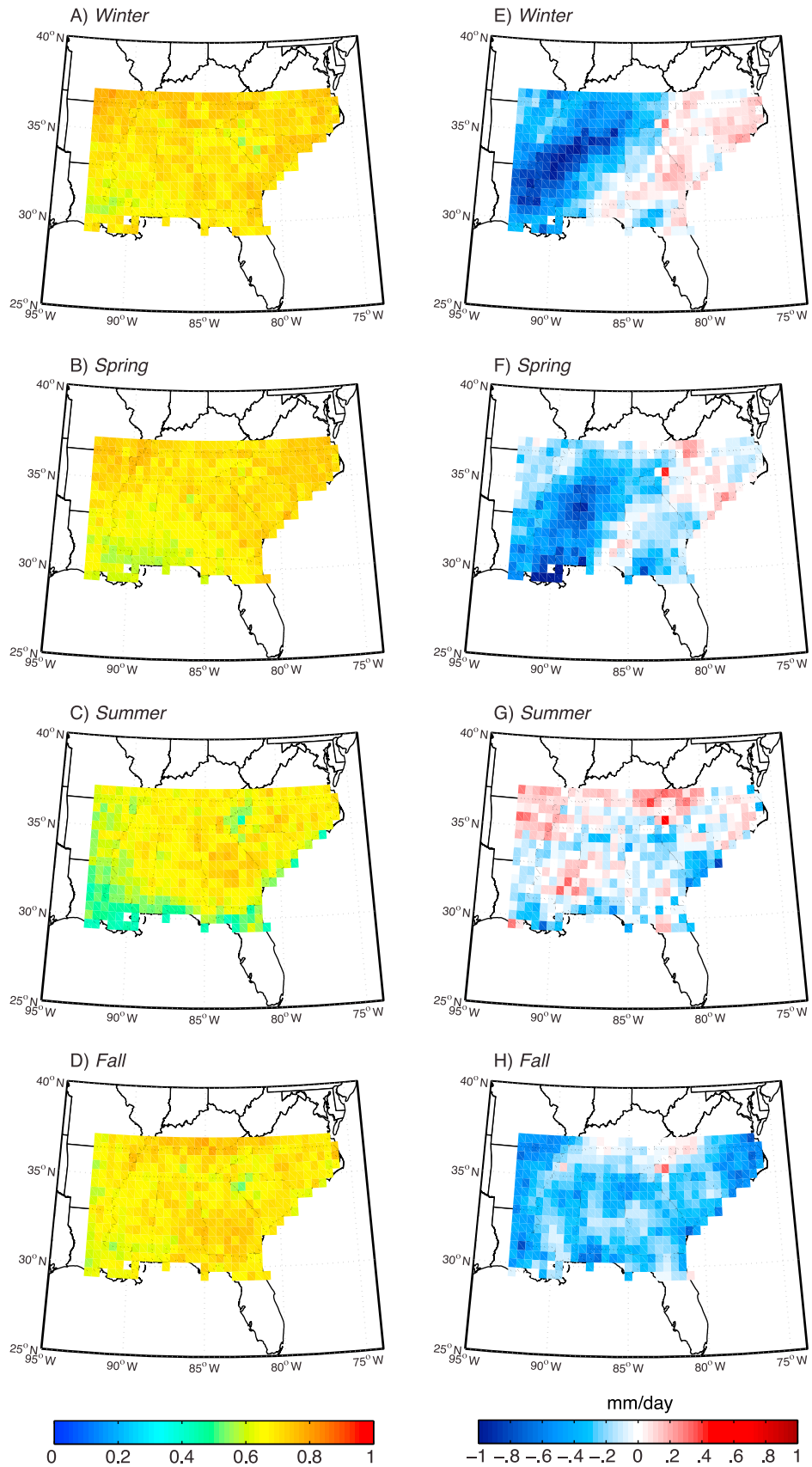


Figure 4. (a–d) Seasonal precipitation skill score averaged across all models and (e–h) average difference in standard deviation between models and observations (mm/d).

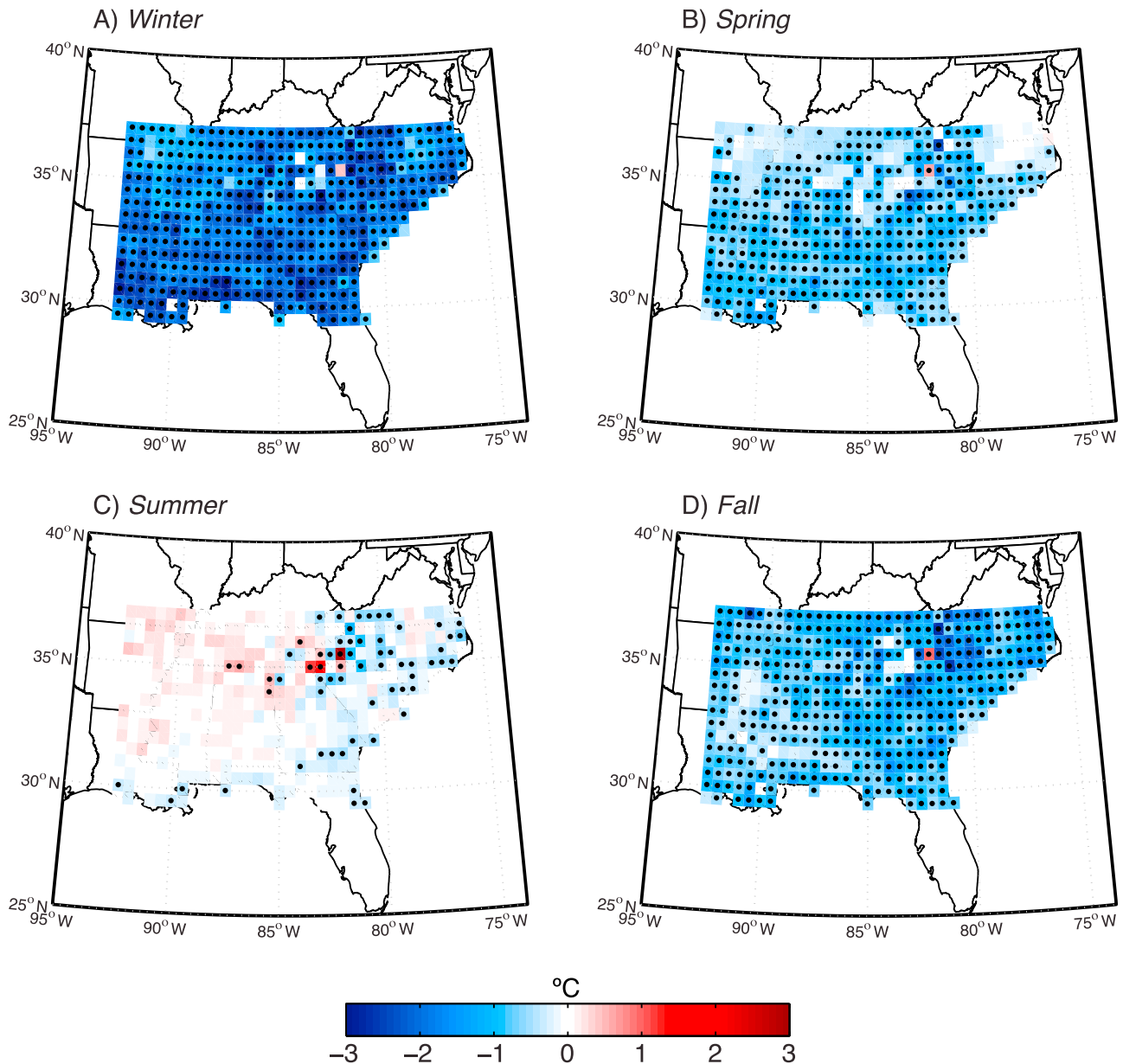


Figure 5. Multimodel mean seasonal temperature bias after application of performance-based weighting ($^{\circ}\text{C}$). Stippling indicates grid cells where bias is statistically significant (at the 5% level), as measured by two-sided Student's t tests.

Southeast that exhibits statistically significant bias for both the unweighted and weighted cases. A decrease in the absolute value of seasonal bias is exhibited in all seasons for both temperature and precipitation (absolute values are used in order mitigate the cancellation between regions of negative and positive bias). Temperature exhibits substantial decreases in the magnitude of bias in all seasons, with spring showing the greatest improvement from -2°C to -0.8°C . The percent area exhibiting statistically significant temperature bias is reduced from 72% to 14% in the summer and 99% to 78% in spring. However, the other season exhibit only modest reductions in percent area of a few percentage points. Conversely, precipitation exhibits more modest reductions in absolute value of seasonal bias relative to

temperature, while exhibiting much more substantial reductions in percent area (Figure 8b). The average regional precipitation bias improves by a few percentage points in all seasons, while percent area decreases by over 20% in spring and summer, 10% in winter and a more modest 7% in fall.

[29] As noted in the introduction there is debate over the efficacy of weighting and model selection schemes. While it is not the purpose of the present study settle, or even enter, this debate, it is important to make some remarks regarding the fidelity of the bias reduction scheme. To evaluate the bias reduction approach the techniques of *Reifen and Toumi* [2009] are adapted to the present case. In short, weights are developed (“trained”) over a variety of 20 year subsets of the 1971–2000 NARCCAP simulation period and are then

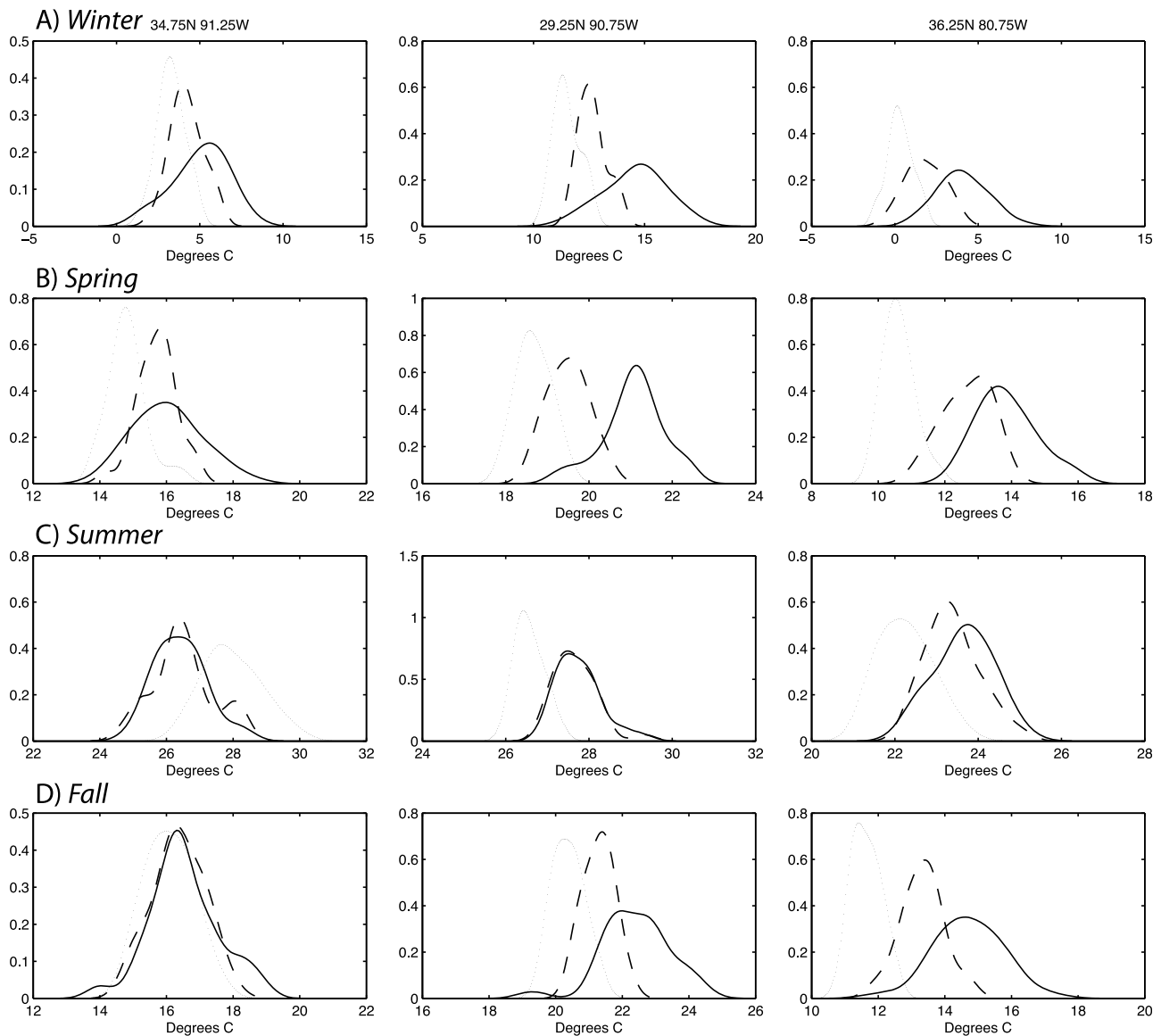


Figure 6. Probability density estimates of temperature for three select points showing the effects of weighting. From left to right the locations correspond to eastern Arkansas, southeast Louisiana, and north central North Carolina. Observations (solid lines), multimodel mean (dotted lines), and reliability ensemble averaging (REA) weighted multimodel mean (dashed lines) are shown for (a) winter, (b) spring, (c) summer, and (d) fall.

applied to the remaining 10 years. These 10 year “test” cases are then compared to the concurrent observations to see if the training weights improve performance during test periods. While the relatively short 30 year simulation period is not ideal for this type of cross-validation test, the results give some indication of the fidelity of the scheme. Figures 9 and 10 show the region-wide results of the test cases for the two periods and should be compared to Figure 8. The reductions in seasonal regional bias and percent area are generally quite similar to those shown in Figure 8. The results are largely insensitive to selection of training and testing periods though there is a reduction in performance as the training period is decreased down to 10 years (not shown). It should also be

noted that the REA scheme has not been optimized in this study. Investigating the optimal magnitudes of the weights (by altering the exponents in equation (2)) and varying bin sizes for the skill score by season or month is the subject of ongoing investigation and has the potential to lead to increased performance.

[30] Thus, the weighting scheme is effective at reducing overall regional bias of both temperature and precipitation. However, we note that substantial bias still exists in some seasons and locations. Generally, the models exhibit negative temperature bias over the entire Southeast in all seasons except summer. Precipitation bias varies more with location, season and sign but, in general, the models

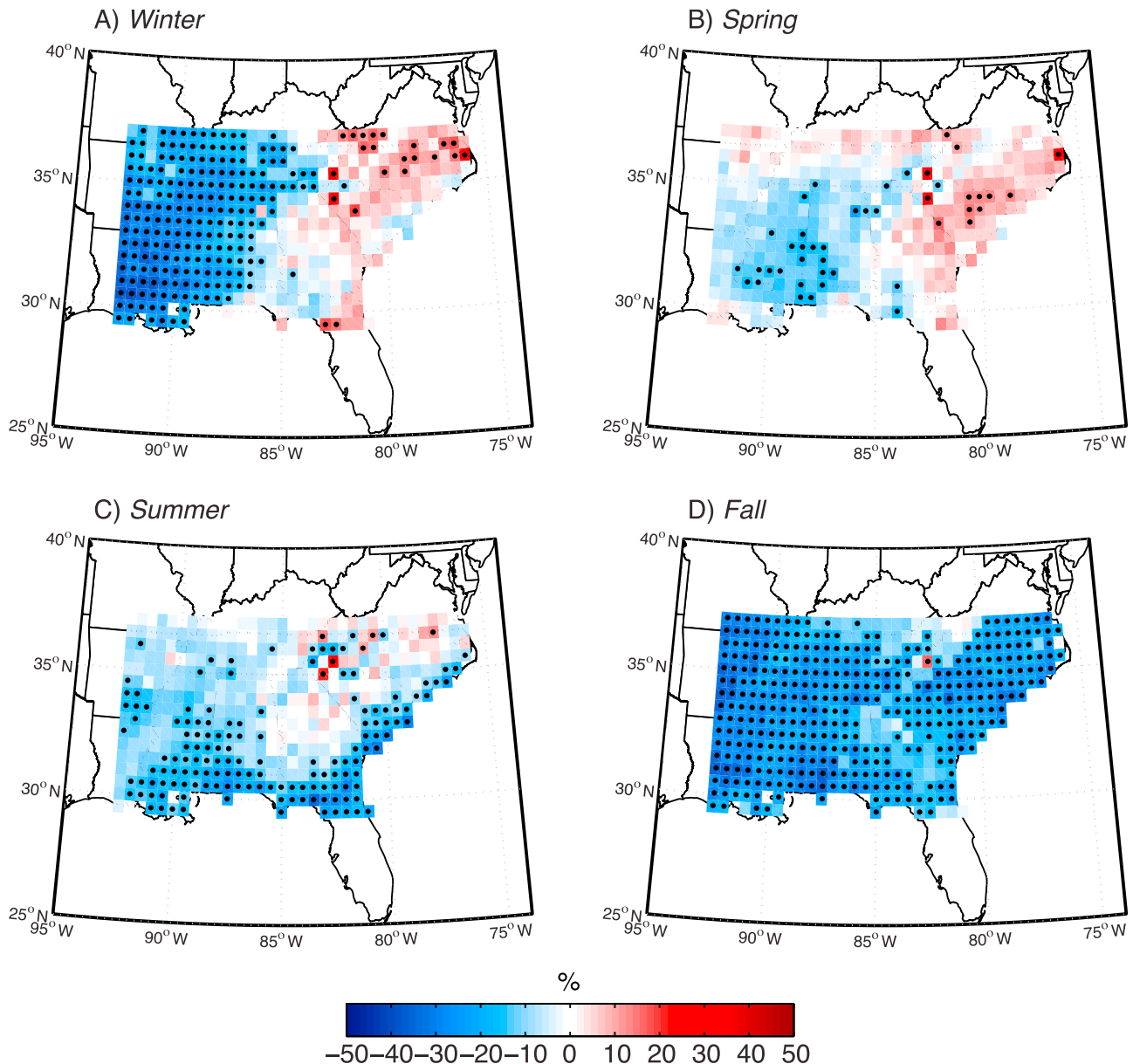


Figure 7. Multimodel mean seasonal precipitation bias after application of performance-based weighting (shown as percentage). Stippling indicates grid cells where bias is statistically significant (at the 5% level), as measured by two-sided Student's t tests.

underestimate precipitation. Weighting was also performed with skill scores and bias separately. However, the combined weighting method was found to result in the largest bias reductions for both variables (not shown).

5. Future Change and Uncertainty

[31] Having established the efficacy of the bias- and skill score-based weighting scheme on the present-day simulations, we add the distance criterion ($R_{D,i}$) and apply the complete REA scheme (R_i) to the 21st century change in temperature and precipitation. It should be noted that the future multimodel mean (ΔT), used in the computation of R_i , is simply a best guess estimate of the future conditions; we

make no claim that it is the true future value given the described scenario.

[32] Figure 11 shows the REA weighted 21st century change in temperature in Figures 11a–11d and the unweighted and weighted uncertainty in Figures 11e–11h and 11i–11k, respectively. The temperature change and uncertainty are in degrees Celsius. As discussed in section 3, if the pdf of future change is nearly Gaussian, then ± 2 times the RMSD approximately defines the 95% confidence bounds. The temperature change in by the middle of the 21st century is positive over the entire southeast United States. The stippling indicates grid cells where the entire 95% confidence bounds of the projected change lie outside the bounds of natural variability as defined by ϵ in section 3. $\Delta \overline{T}$

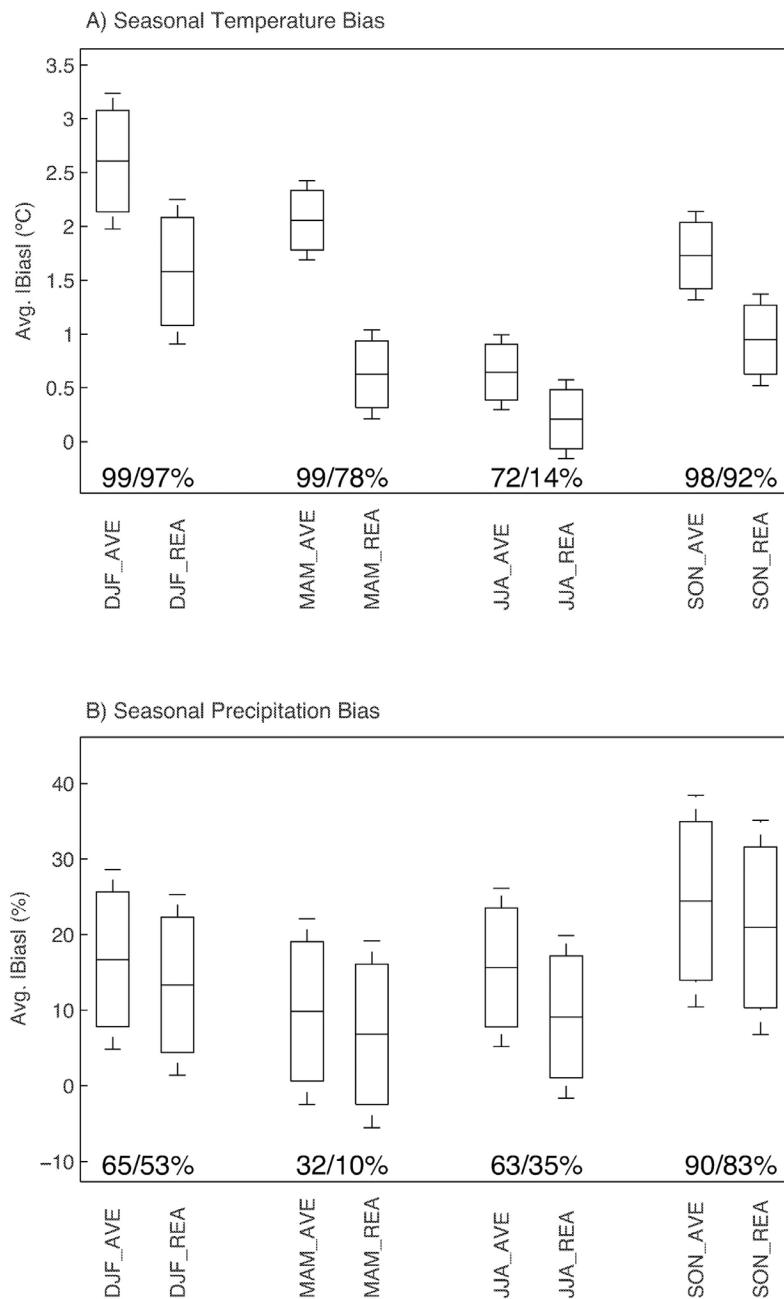


Figure 8. Regionally averaged seasonal (a) temperature and (b) precipitation bias for the unweighted (AVE) and weighted (REA) cases. Absolute values are used to mitigate cancellation due to subregional variations in sign. The 95% confidence intervals calculated from Student's t tests are denoted by whiskers. The percentage of total area exhibiting statistically significant bias for AVE/REA cases is shown on the bottom.

is generally homogeneous over the region in each season, with slight north-south, coast-inland gradients (i.e., more warming north (inland) less south (coasts)). Maximum $\widetilde{\Delta T}$ occurs in the summer with temperatures rising over 3°C over much of the Southeast. This intense warming continues into the fall and decreases slightly in the winter and spring when $\widetilde{\Delta T}$ is between 1.5°C and 2°C . The unweighted 21st century changes are not shown as they are not qualitatively different,

though slightly fewer grid cells exceed natural variability at the 95% level.

[33] Figures 11e–11h show the upper half of the approximate 95% unweighted uncertainty bounds, or $2\delta_{\Delta T}$. The unweighted uncertainties are over 0.5°C across most of the Southeast in all seasons. The Gulf states (spring), North Carolina and Mississippi (summer) and the northern states (fall) exhibit coherent areas with uncertainties as high as 1°C , indicating 95% confidence intervals of around 2° .

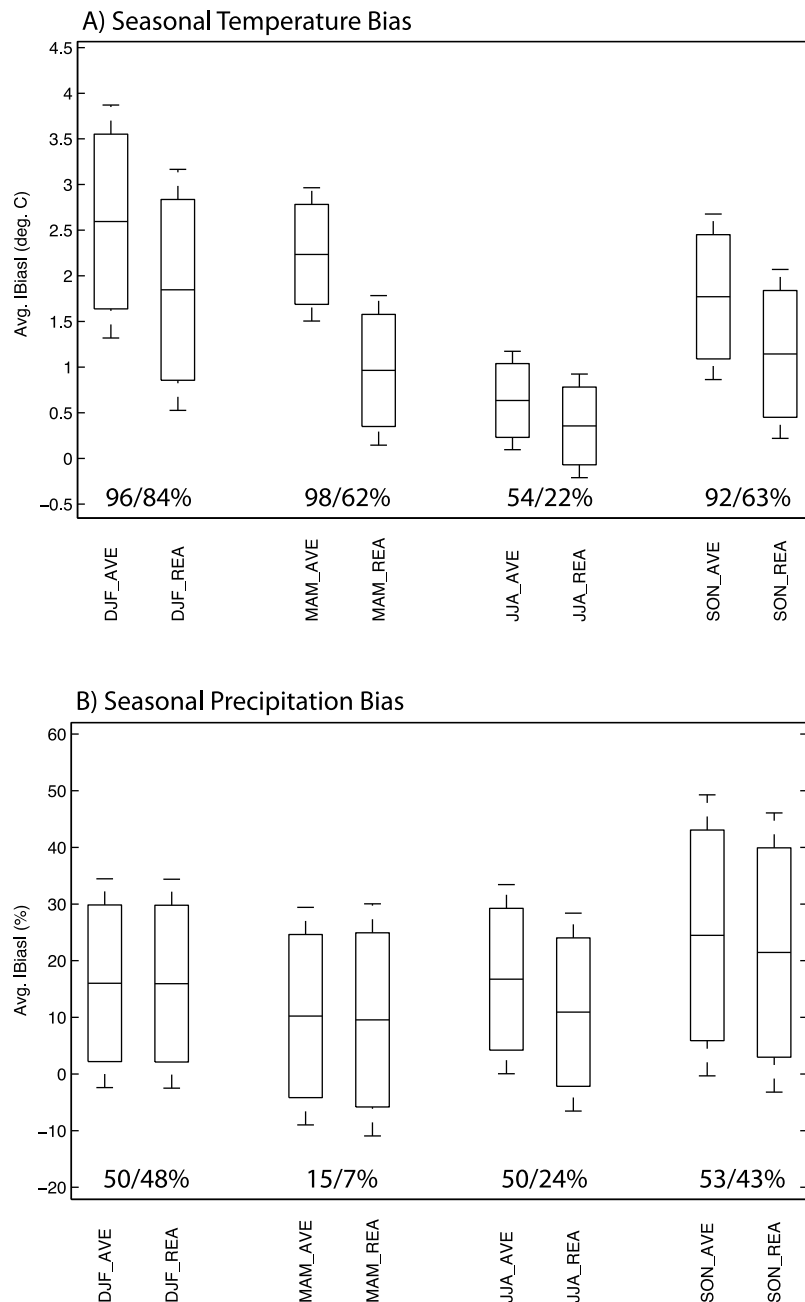


Figure 9. Weighting scheme test results. Training was performed in the first 20 years of the simulation period, and testing was on the last 10 years. Regionally averaged seasonal (a) temperature and (b) precipitation bias for the unweighted (AVE) and weighted (REA) cases. Absolute values are used to mitigate cancellation due to subregional variations in sign. The 95% confidence intervals calculated from Student's t tests are denoted by whiskers. The percentage of total area exhibiting statistically significant bias for AVE/REA cases is shown on the bottom. The similarity to Figure 8 suggests that weighting based on skill in simulating present-day conditions is warranted.

This range of uncertainty sets the lower limits of change as low as 1°C in some areas during winter and spring. Conversely, the upper limits may be as high as 4.5°C – 5°C over some areas during summer. While consideration of the unweighted uncertainties does not change the story of warming over the entire southeast United States by the middle of the 21st century, the range of potential warming

is increased everywhere and specific seasons and locations (e.g., spring along the Gulf coast) exhibit confidence intervals that approach 2° .

[34] The effect of the REA weighting scheme can be seen by comparing Figures 11e–11h and 11i–11l. Figures 11i–11l show 2 times the REA RMSD, or $2\delta_{\Delta T}^{\text{REA}}$. The average change shown in Figures 11a–11d plus or minus this uncertainty

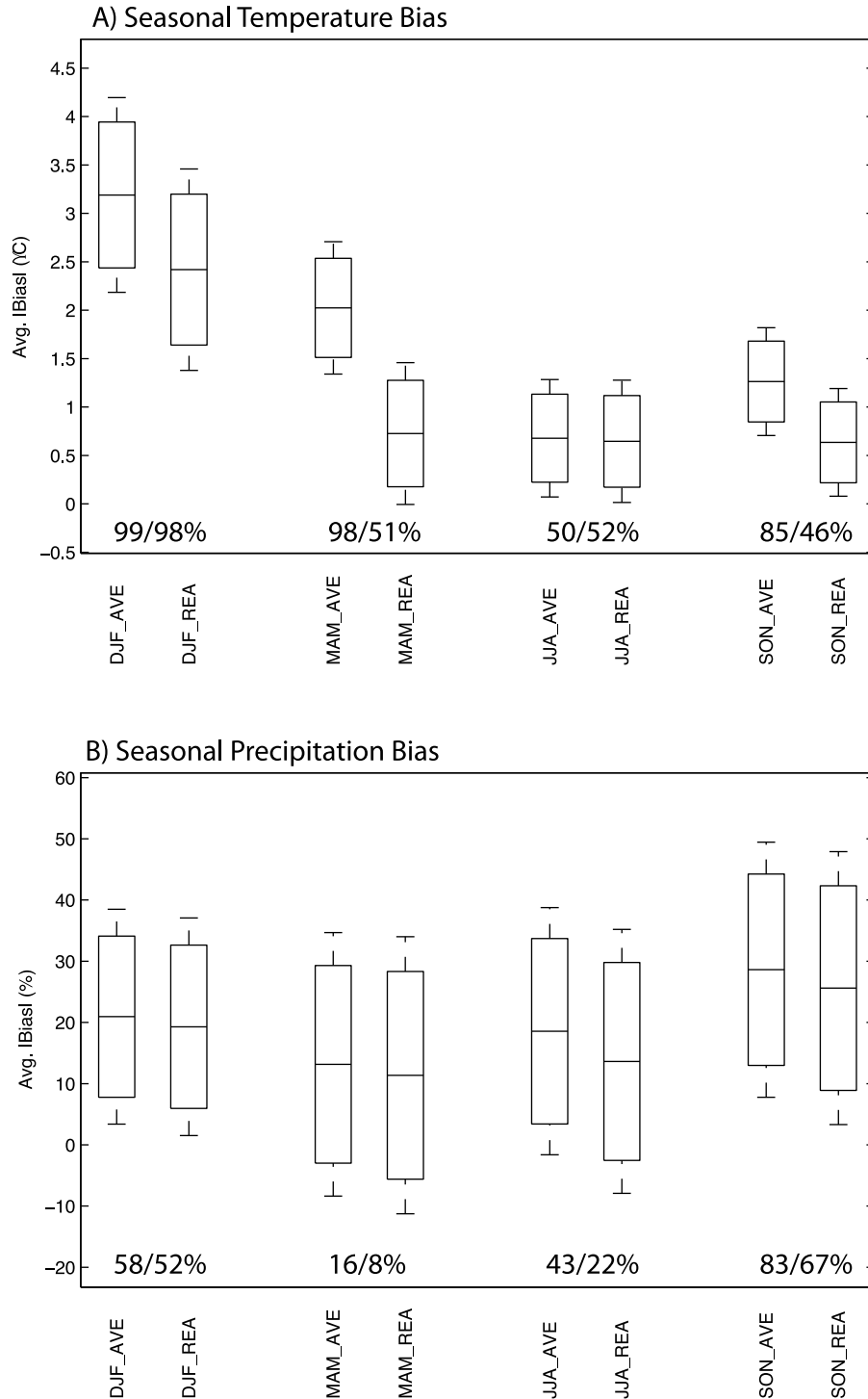


Figure 10. Weighting scheme test results. Training was performed in the first 20 years of the simulation period, and testing was on the last 10 years. Regionally averaged seasonal (a) temperature and (b) precipitation bias for the unweighted (AVE) and weighted (REA) cases. Absolute values are used to mitigate cancellation due to subregional variations in sign. The 95% confidence intervals calculated from Student’s *t* tests are denoted by whiskers. The percentage of total area exhibiting statistically significant bias for AVE/REA cases is shown on the bottom. The similarity to Figure 8 suggests that weighting based on skill in simulating present-day conditions is warranted.

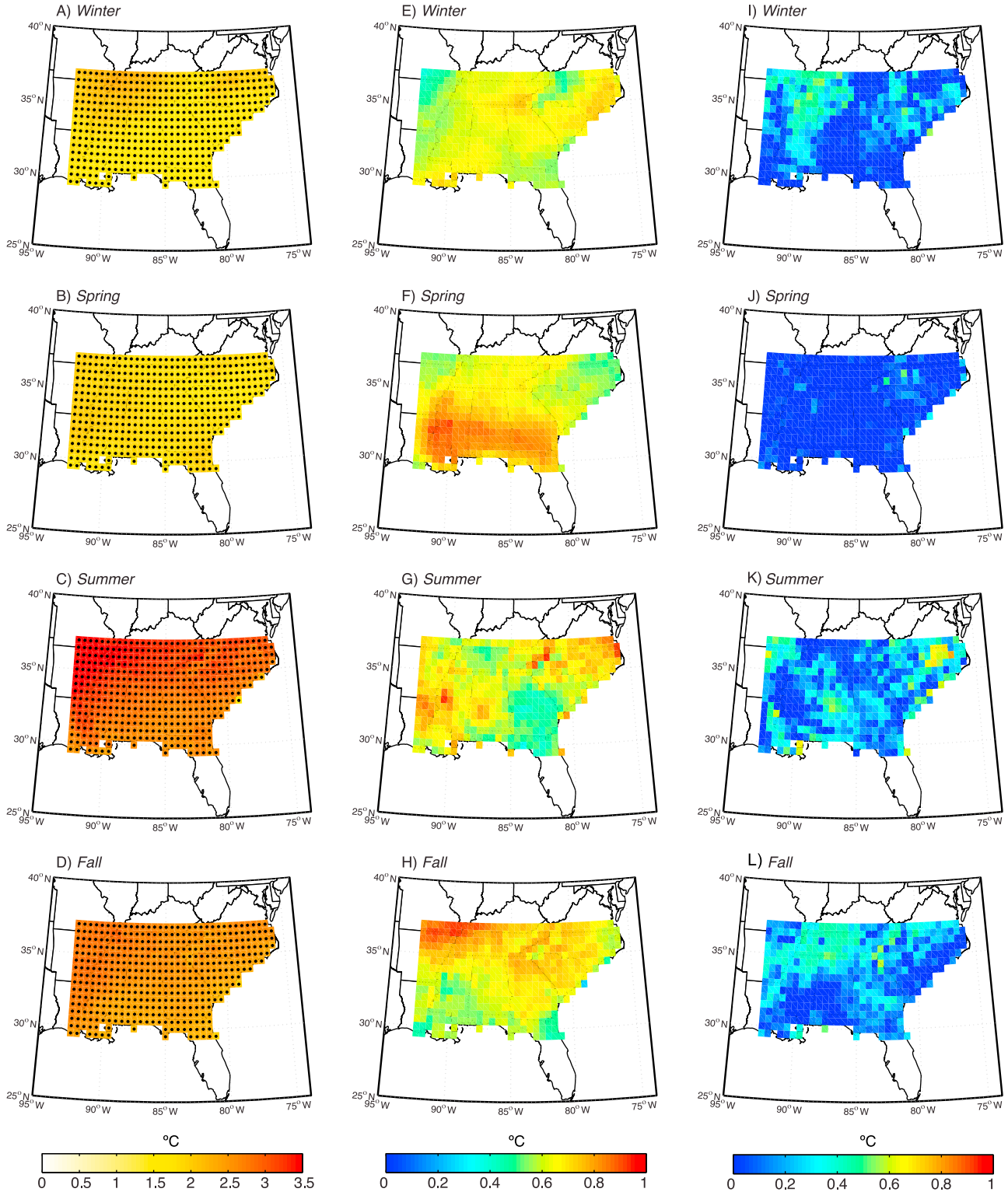


Figure 11. (a–d) Weighted 21st century multimodel mean seasonal temperature change (2041–2070 minus 1971–2000), where stippling indicates areas where the weighted 95% confidence intervals lie outside the range of natural variability as measured by ϵ . (e–h) Unweighted uncertainty shown as 2 times the unweighted root-mean-square difference (RMSD) of the projected change. (i–l) REA weighted uncertainty shown as 2 times the weighted RMSD of the projected change (equation (6)). All units are in °C.

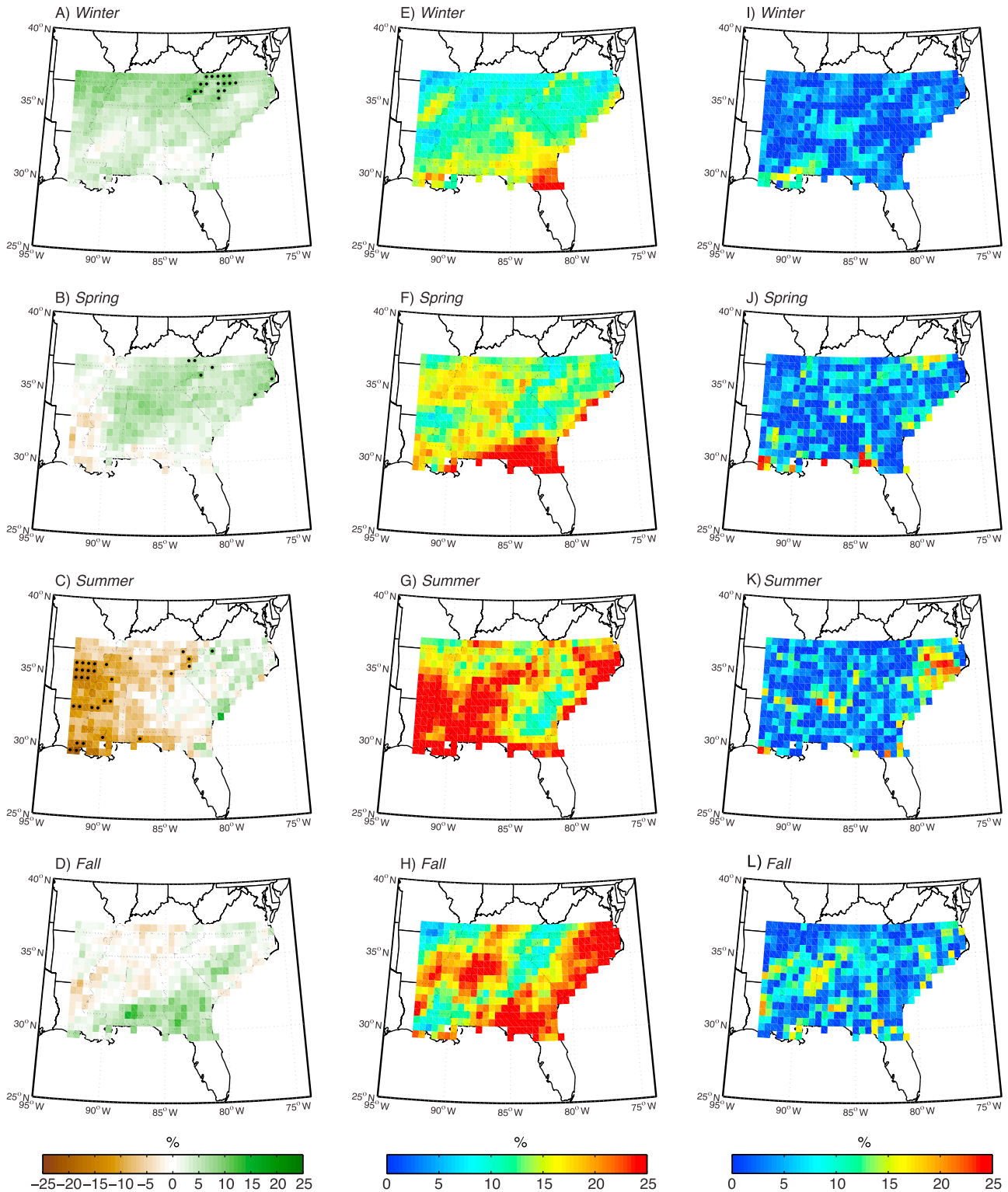


Figure 12. (a–d) Weighted 21st century multimodel mean seasonal precipitation change (2041–2070 minus 1971–2000), where stippling indicates areas where the weighted mean lies outside the range of natural variability as measured by ϵ . (e–h) Unweighted uncertainty shown as 2 times the unweighted RMSD of the projected change. (i–l) REA weighted uncertainty shown as 2 times the weighted RMSD of the projected change (equation (6)). All results are shown as percentages.

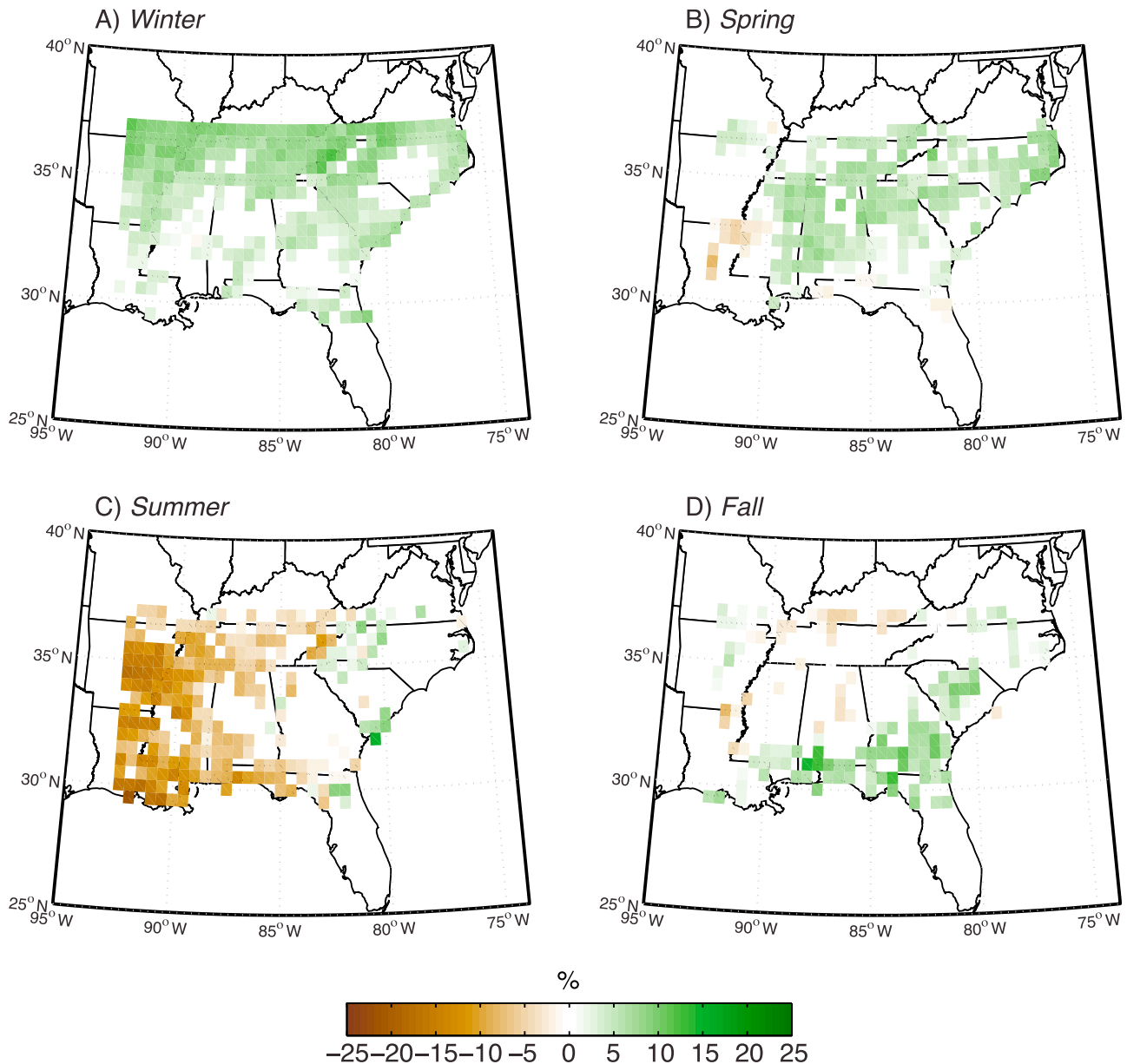


Figure 13. Weighted 21st century multimodel mean seasonal precipitation change (2041–2070 minus 1971–2000) showing only grid cells where the weighted 95% confidence intervals lie completely above or below zero. All results are shown as percentages.

gives the approximate 95% uncertainty bounds. These plots indicate that the REA uncertainty is below 0.5°C over much of the study area in all seasons with spring exhibiting region-wide uncertainty lower than 0.2°C . The highest uncertainties occur in summer over the eastern Carolinas and the Gulf coast states. The Northwest portion of the study area in fall and winter also exhibit uncertainties of 0.5°C . Overall, the pattern of REA uncertainties suggest confidence bounds of much less than one degree, or about half that of the unweighted case, over most of the Southeast.

[35] Figure 12 shows the REA weighted 21st century change in precipitation in Figures 12a–12d and the unweighted and weighted uncertainty in Figures 12e–12h and 12i–12l, respectively. All changes and uncertainties are shown as percentages. Unlike temperature, the projected

changes in precipitation are modest and vary in both sign and magnitude by location and season. Winter and spring exhibit increases of $\sim 10\%$ across the much of the Southeast while decreases of $\sim 15\%$ over the western portion of the study area are indicated for summer. During fall, increases of $\sim 10\%$ are projected for the coastal regions. The stippling in this case represents grid cells where the average weighted change is greater than the background variability (ϵ). This is in contrast to the case of temperature where the stippling represents grid cells where the entire 95% confidence bound lies outside background variability. This less restrictive criteria is chosen because the modest magnitudes of the precipitation changes. Thus, the average precipitation change only exceeds background variability along the Mississippi river in summer and the Appalachian mountains in winter.

[36] The uncertainty in the unweighted case is shown in Figures 12e–12h. Over most of the Southeast the uncertainties are larger than the projected change itself during all seasons. Over specific locations the unweighted 95% confidence intervals are up to a factor four larger than the projected changes (e.g., east coast in fall, Gulf coast in summer and spring). Uncertainties greater than 25% are exhibited over the coasts and some inland areas during spring, summer and fall. This suggests confidence intervals of over 50% in some areas. Winter exhibits generally lower uncertainties inland but uncertainties of $\sim 20\%$ around the Gulf states. The reductions in uncertainty due to the REA weighting scheme are evident from a comparison of Figures 12e–12h and 12i–12l. These patterns suggest that, in all seasons, over most of the Southeast, the REA 95% confidence intervals are now smaller than the projected changes. Over most of the Southeast the uncertainties are around 5% though there are scattered areas which exhibit uncertainties over 15% (e.g., the east coast during summer), suggesting confidence intervals of $\sim 10\%$ – 30% .

[37] While the uncertainty reductions due to weighting in the temperature projections are large they do not change the interpretation of the future change; they merely increase confidence in the robustness of the NARCCAP simulations. Conversely, the uncertainty reductions in the precipitation projections due to weighting tighten the confidence intervals substantially and pull the bounds to either side of zero change over much of the study region. Figure 13 shows only grid cells where the approximate 95% confidence intervals of the REA future precipitation change lie completely to either side of zero change. Using this measure, the increases in winter (over the northern states) and the decreases in summer (along the Mississippi) appear coherent and robust, although the magnitudes of the projected changes are modest. This is in contrast to the unweighted case where the 95% intervals at all grid cells for all seasons straddle the line of zero change (not shown).

6. Discussion and Conclusions

[38] One of the many goals of the multimodel NARCCAP simulations is to provide information and evaluate uncertainty at scales on which humans structure their societies. At the 50 km resolution the NARCCAP simulations should provide information that officials and stakeholders at municipal to state levels will find useful in developing local strategies to respond to climate change. This program implicitly addresses the issues raised by recent claims that climate output at typical GCM scales does not provide actors with sufficient information to make informed decisions about climate change adaptation and/or mitigation [Oreskes *et al.*, 2010]. The present study aims to improve the representation of present-day climate by the NARCCAP simulations and reduce the uncertainty associated with projected future changes in temperature and precipitation. After comparing the multimodel mean of the NARCCAP simulations to observations a weighting scheme is applied on the basis of the REA approach devised by Giorgi and Mearns [2003] that incorporates an additional measure on the basis of the model skill score [Perkins *et al.*, 2007]. It is important to note that the results presented here should be viewed within the context of the set of NARCCAP simulations and the

SRES A2B scenario. Any reduction in uncertainty pertains to the ensemble of models under investigation, not uncertainty related to future climate change directly. In this section we discuss the sources and magnitude of model error and bias, implications of the weighting scheme and issues related to number of realizations and ensemble size. We make recommendations for future research directions that may address outstanding modeling issues. We then discuss the implications of the future projections for temperature and precipitation over the southeast United States.

6.1. Evaluation of NARCCAP Multimodel Performance

[39] As Figures 1 and 3 illustrate, simply increasing model resolution leads to improved representation of the spatial patterns of temperature and precipitation (see Mearns *et al.* [2003] and the NARCCAP Web site for examples of T and P at AOGCM scales). However, as Figures 1i–1l and 3i–3l demonstrate, substantial biases persist. While some of these biases may be explained by propagation of errors from the AOGCM down to the RCMs, the southeast United States is far enough away from the RCM domain boundaries that these types of errors should be minimized. Visual inspection of the NARCCAP simulations forced with NCEP reanalysis data suggests that biases still exist even when the RCMs are forced with observation-based reanalysis data (<http://www.narccap.ucar.edu/results/ncep-results.html>). Thus, it is likely that the RCMs have systematic errors in heat transport, land-atmosphere processes and moist processes (i.e., subgrid-scale parameterizations). Especially troubling are the systematic and significant underestimation of temperatures across the region in winter, spring and fall. The pattern of bias in precipitation fields suggests that the RCMs have trouble with subgrid-scale processes. The results here suggest they are not able to reproduce the cool season moisture transport out of the Gulf of Mexico and subsequent entrainment into synoptic systems and they have difficulty with moist processes in the presence of orography and along the coasts. Despite recent improvements in both RCMs and their AOGCM counterparts, improving these models' ability to reproduce fundamental aspects of the climate systems is clearly a pressing need [Doherty *et al.*, 2009]. There are a few ways to address this issue. For example, one suggestion in the recent literature is to initialize simulations with observations in order to reduce the influence of model error and internal variability on near-term projections [Hawkins and Sutton, 2009]. While this approach is certainly viable for near-term simulations of future climate (e.g., 10 years) it is currently not feasible for experiments with more distant time horizons. Another approach is to implement a performance-based weighting scheme which may improve confidence in multimodel projections. For the present study we chose to implement a weighting scheme that rewards models which show an ability to reproduce the mean and pdf of the observations and leads to improved reproduction of the present-day climate by the NARCCAP simulations and greater confidence in the future projections.

[40] Recent research suggests that performance-based weighting of multimodel results has the potential to improve confidence in estimated future projections and supports efforts to reduce bias in present-day simulations [Raisanen *et al.*, 2010; Matsueda and Palmer, 2011]. One of the

difficulties in applying any weighting scheme is the tacit assumption that the location, sign and magnitude of present-day bias is time invariant. Certainly the present scheme is unable to address this issue although the results of [Matsueda and Palmer, 2011] provide some support that bias may be a useful predictor of the accuracy of the future signal for specific regions and variables. Given the results shown in Figures 1i–1l and 3i–3l, it is likely that optimal weighting scheme over the southeast United States will depend nonlinearly with time of year and magnitude. The weighting scheme as applied here makes two important assumptions that need to be acknowledged. First of these is that the future multimodel mean is the “best guess” estimate of the projected change given the chosen future scenario. The other is that the future changes will be nearly Gaussian and that ± 2 times the RMSD approximately represents the 95% confidence bounds of the future projections. A third caveat involves the skill scores. Computing the skill scores using monthly means results in a measure of large-scale rather than synoptic-scale performance. Thus, we ignore the ability of models simulate the distribution of weather events, which likely results in higher scores than those computed with the more variable information contained in daily data. For example, a recent study over this region using NARCCAP output finds considerably lower skill scores when using daily values than those reported here [Kabela and Carbone, 2011]. This is as one would expect given the greater variability introduced when one considers daily values rather than monthly means. Despite these potential shortcomings the results shown in Figures 5 and 6 indicate that substantial improvements and reductions in both the magnitude and the area affected by model bias. Further, the effect of weighting on the future projections results in significant reductions in uncertainty, particularly in the case of future precipitation change (Figures 8 and 9). As Figure 13 shows this reduction in uncertainty leads to more robust estimates of future precipitation change over the northern states in winter and the Mississippi in summer.

[41] Another potential complication concerns sample size. Recent research suggest that current small ensemble sizes underestimate the contribution of internal variability (i.e., climate or weather noise) to total uncertainty and that this type of variability is, in some cases more important than intermodel variations [Deser et al., 2011]. Further, Deser et al. [2011] find that while only one to three realizations are required to produce detectable, significant forced responses for temperature many more (as many as 20 in some instances) are required for precipitation. The six RCM-AOGCM pairs provided by the NARCCAP simulations have one realization each. Thus, given the modest precipitation response it is likely that many more realizations are needed in order to increase the signal-to-noise ratio. We also note that it is possible that many areas of the southeast United States will experience negligible to modest changes in mean seasonal precipitation and that these changes will never be significant no matter how many ensemble realizations are run.

[42] The small number of realizations also render rigorous probabilistic assessment of the future changes impossible, as the pdf’s of the future changes cannot be known because of the small sample size. What we provide here is a qualitative assessment and evaluation that marks a good first step to

evaluating future changes at the subregional scale. A number of techniques exist to increase the relative sample size and enable construction of pdf’s of the future climate change, though few have been employed at this scale. One approach is to construct Bayesian statistical models, using RCM output, to fully account for uncertainty [e.g., Tebaldi et al., 2004, 2005]. Another is to apply resampling techniques to artificially increase the number of realizations [e.g., Raisanen and Ruokolainen, 2006]. Currently both approaches are being investigated over the southeast United States as an extension of the present study.

6.2. Implications for Water Resources and Energy Use in the Southeast United States

[43] When the issues discussed in the preceding paragraphs are considered, the projected changes in temperature and precipitation present a mixed message for city planners, state officials and water resource managers. The temperature changes shown in Figure 7 are consistent with previous GCM and RCM studies and are likely robust given the A2 SRES scenario. The confidence bounds of these changes lie well outside the range of natural variability and it is not likely that increasing ensemble size, resolution or number of models will significantly alter this finding. We note however, that considerable bias exists between the multimodel mean and observations that is only partially corrected by the weighting scheme, indicating a systematic underestimation of temperature by the models in all seasons except summer, over most of the southeast United States. Whether these biases will be of similar sign, magnitude and location in the future is an open question.

[44] If the A2 SRES assumptions prove correct, the results suggest that average temperatures across the Southeast will increase by over 2.5° in summer and fall with slightly smaller increases of about 1.8° in winter and spring. There is a slight northwest to Southeast gradient in all seasons with largest changes in the Northwest and smallest in the Southeast. With respect to water resources the largest changes are expected for inland regions during times of year which typically experience peak demand (summer) and minimum precipitation (fall). The increased temperatures will likely lead to greater evapotranspiration and plant water stress in the absence of increased precipitation. Further, the largest increases in temperatures are projected for the time of the year (summer) that currently experiences peak energy demand because of widespread air conditioner use in both home and industrial settings. Since we have evaluated seasonal means we do not speculate on the potential increase in heat waves or droughts, but previous research indicates that both extremes may be expected to increase over the southeast United States over the 21st century [Diffenbaugh et al., 2005]. So, in addition to the pressures placed on energy and water resources by population and economic growth, increasing temperature will also likely add to these pressures in the future. Research is needed to quantify the relative contributions of population, economic activity and climate change.

[45] With respect to precipitation the picture is not so clear. None of the changes shown in Figure 12 are significantly greater than background variability, and given the large intermodel spread, it is possible that a new set of NARCCAP simulations, even under the same forcing

scenario, would yield different results. However, Figures 9 and 10 suggest that with the reduction in uncertainty due to the weighting scheme and with more realizations, the patterns of change in winter and summer may be robust though modest. The winter pattern suggests modest increases in precipitation, on the order of 10%–15%, along the northern states of the region (i.e., Arkansas, Tennessee, North Carolina, Virginia, and Kentucky). Conversely, the summer change suggests decreases as high as 20% in the western portion of the Southeast region. This decrease is mostly along the Mississippi but also extends along the Gulf coast. As with temperature we make no claim with respect to future extreme events, but previous research indicates that extreme wet events are likely to increase over the region because of enhanced moisture convergence along the coasts [Diffenbaugh *et al.*, 2005]. There is some promise that the NARCCAP simulations are capable of reproducing historical precipitation extremes over some parts of North America so there is hope that this resource may be useful for investigations of future extremes [Gutowski *et al.*, 2010]. Additional contributions to changes in extreme precipitation may come from changes in intensity and/or frequency of tropical cyclone activity [Knutson *et al.*, 2008a, 2008b]. Contributions to precipitation changes on synoptic scales will also depend on the magnitude and dynamic implications of future changes that are still poorly understood, such as widening of the Hadley circulation [Lu *et al.*, 2007a, 2007b; Frierson *et al.*, 2007] and northward shifting of the North Atlantic storm tracks [Yin, 2005]. The line of zero 21st century precipitation (or $P - E$) change bisects the Southeast latitudinally [Christensen *et al.*, 2007; Seager *et al.*, 2010] and a shift of this line north or south may have profound implications for the numerous large, expanding municipalities across the southeast United States. What the present study reveals however, is that given the small number of realizations and comparatively large uncertainties, the projected precipitation changes for the Southeast are likely to be modest with the possible exception of the northern portion of the region in winter and the western portion in summer. From a water resources perspective, stress is more likely to come from increased temperatures with little contribution (either mitigation or exacerbation) from the seasonal mean precipitation changes. There is potential for a positive contribution from more frequent extreme events, but runoff from extreme precipitation events tends not to contribute to storage as infiltration capacity is greatly exceeded and excess water runs off, eventually escaping to the ocean.

[46] In conclusion, the application of a performance-based weighting scheme allows us to reduce bias in the present and uncertainty in the future NARCCAP multimodel simulations. Further, the analysis of high-resolution model output allows for subregional patterns to emerge in the future projections that are unresolvable at typical GCM scales. While more realizations are currently needed to increase robustness of the projected precipitation response, the temperature response is likely quite robust. Improvements to model physics are also needed to address shortcomings in moist processes, orographic effects and land-ocean moisture transport. Despite these issues, the present study enables officials and the public across the southeast United States to consider a warmer future in all seasons, particularly away from the coasts and during the peak water and energy

demand seasons. At the seasonal scale, precipitation changes are uncertain but likely modest, with increases along the northern boundary in winter and decreases in the west during summer. Future work will build on the analysis presented here in an effort to explain the physical mechanisms driving future changes to the hydrological cycle over the southeast United States.

[47] **Acknowledgments.** The authors would like to acknowledge the three anonymous reviewers, whose comments and suggestions greatly improved the quality of this manuscript.

References

- Chen, M., D. Pollard, and E. Barron (2003), Comparison of future climate change over North America simulated by two regional models, *J. Geophys. Res.*, *108*(D12), 4348, doi:10.1029/2002JD002738.
- Christensen, J., et al. (2007), Regional climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. 11, pp. 849–940, Cambridge Univ. Press, Cambridge, U. K.
- Damassa, T. (2009), Energy by the numbers (data supplement), technical report, World Resour. Inst., Washington, D. C.
- Deser, C., A. Phillips, V. Bourdette, and H. Teng (2011), Uncertainty in climate change projections: The role of internal variability, *Clim. Dyn.*, doi:10.1007/s00382-010-0977-x, in press.
- Diffenbaugh, N., J. Pal, R. Trapp, and F. Giorgi (2005), Fine-scale processes regulate the response of extreme events to global climate change, *Proc. Natl. Acad. Sci. U. S. A.*, *102*(44), 15,774–15,778, doi:10.1073/pnas.0506042102.
- Doherty, S. J., et al. (2009), Lessons learned from IPCC AR4 scientific developments needed to understand, predict, and respond to climate change, *Bull. Am. Meteorol. Soc.*, *90*(4), 497–513, 497, doi:10.1175/2008BAMS2643.1.
- Frierson, D. M. W., J. Lu, and G. Chen (2007), Width of the Hadley Cell in simple and comprehensive general circulation models, *Geophys. Res. Lett.*, *34*, L18804, doi:10.1029/2007GL031115.
- Giorgi, F., and L. Mearns (2003), Probability of regional climate change based on the reliability ensemble averaging (REA) method, *Geophys. Res. Lett.*, *30*(12), 1629, doi:10.1029/2003GL017130.
- Giorgi, F., L. Mearns, C. Shields, and L. McDaniel (1998), Regional nested model simulations of present day and $2 \times \text{CO}_2$ climate over the central plains of the US, *Clim. Change*, *40*(3–4), 457–493.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.
- Gutowski, W. J., et al. (2010), Regional extreme monthly precipitation simulated by NARCCAP RCMs, *J. Hydrometeorol.*, *11*(6), 1373–1379, doi:10.1175/2010JHM1297.1.
- Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteorol. Soc.*, *90*(8), 1095–1107, doi:10.1175/2009BAMS2607.1.
- Kabela, E., and G. Carbone (2011), Downscaling climate change information for water resource and agricultural interests in the southeast United States, paper presented at the 2011 NARCCAP User's Meeting, Pac. Clim. Impact Consortium, Boulder, Colo.
- Karl, T., and R. Knight (1998), Secular trends of precipitation amount, frequency, and intensity in the united states, *Bull. Am. Meteorol. Soc.*, *79*(2), 231–241.
- Knutson, T. R., T. L. Delworth, K. W. Dixon, I. M. Held, J. Lu, V. Ramaswamy, M. D. Schwarzkopf, G. Stenchikov, and R. J. Stouffer (2006), Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models, *J. Clim.*, *19*(9), 1624–1651.
- Knutson, T. R., J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held (2008a), Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions, *Nat. Geosci.*, *1*, 359–364, doi:10.1038/ngeo202.
- Knutson, T. R., J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held (2008b), Erratum: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions, *Nat. Geosci.*, *1*, 479, doi:10.1038/ngeo229.
- Legates, D., and C. Willmott (1990), Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatol.*, *10*(2), 111–127.
- Lu, J., G. A. Vecchi, and T. Reichler (2007a), Expansion of the Hadley cell under global warming, *Geophys. Res. Lett.*, *34*, L06805, doi:10.1029/2006GL028443.

- Lu, J., G. A. Vecchi, and T. Reichler (2007b), Correction to “Expansion of the Hadley cell under global warming,” *Geophys. Res. Lett.*, *34*, L14808, doi:10.1029/2007GL030931.
- Manuel, J. (2008), Drought in the Southeast: Lessons for water management, *Environ. Health Perspect.*, *116*(4), A168–A171.
- Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim (2010), Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate, *J. Clim.*, *23*(22), 5933–5957, doi:10.1175/2010JCLI3541.1.
- Matsueda, M., and T. N. Palmer (2011), Accuracy of climate change predictions using high resolution simulations as surrogates of truth, *Geophys. Res. Lett.*, *38*, L05803, doi:10.1029/2010GL046618.
- Mearns, L. O. (2010), The drama of uncertainty, *Clim. Change*, *100*(1), 77–85, doi:10.1007/s10584-010-9841-6.
- Mearns, L., F. Giorgi, L. McDaniel, and C. Shields (2003), Climate scenarios for the southeastern US based on GCM and regional model simulations, *Clim. Change*, *60*(1-2), 7–35.
- Mearns, L. O., 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. AGU*, *90*(36), 311–312.
- Oreskes, N., D. A. Stainforth, and L. A. Smith (2010), Adaptation to global warming: Do climate models tell us what we need to know?, *Philos. Sci.*, *77*(5), 1012–1028.
- Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney (2007), Evaluation of the AR4 climate models’ simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, *J. Clim.*, *20*(17), 4356–4376, doi:10.1175/JCLI4253.1.
- Phillips, T., and P. Gleckler (2006), Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics, *Water Resour. Res.*, *42*, W03202, doi:10.1029/2005WR004313.
- Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler (2009), Selecting global climate models for regional climate change studies, *Proc. Natl. Acad. Sci. U. S. A.*, *106*(21), 8441–8446, doi:10.1073/pnas.0900094106.
- Raisanen, J., and L. Ruokolainen (2006), Probabilistic forecasts of near-term climate change based on a resampling ensemble technique, *Tellus, Ser. A*, *58*(4), 461–472, doi:10.1111/j.1600-0870.2006.00189.x.
- Raisanen, J., L. Ruokolainen, and J. Ylhaisi (2010), Weighting of model results for improving best estimates of climate change, *Clim. Dyn.*, *35*(2-3), 407–422, doi:10.1007/s00382-009-0659-8.
- Reifen, C., and R. Toumi (2009), Climate projections: Past performance no guarantee of future skill?, *Geophys. Res. Lett.*, *36*, L13704, doi:10.1029/2009GL038082.
- Santer, B. D., et al. (2009), Incorporating model quality information in climate change detection and attribution studies, *Proc. Natl. Acad. Sci. U. S. A.*, *106*(35), 14,778–14,783, doi:10.1073/pnas.0901736106.
- Seager, R., A. Tzanova, and J. Nakamura (2009), Drought in the southeastern United States: Causes, variability over the last millennium, and the potential for future hydroclimate change, *J. Clim.*, *22*(19), 5021–5045, doi:10.1175/2009JCLI2683.1.
- Seager, R., N. Naik, and G. A. Vecchi (2010), Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming, *J. Clim.*, *23*(17), 4651–4668, doi:10.1175/2010JCLI3655.1.
- Tebaldi, C., L. Mearns, D. Nychka, and R. Smith (2004), Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations, *Geophys. Res. Lett.*, *31*, L24213, doi:10.1029/2004GL021276.
- Tebaldi, C., R. Smith, D. Nychka, and L. Mearns (2005), Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, *J. Clim.*, *18*(10), 1524–1540.
- Watterson, I. G. (2008), Calculation of probability density functions for temperature and precipitation change under global warming, *J. Geophys. Res.*, *113*(D12), D12106, doi:10.1029/2007JD009254.
- Willmott, C., and K. Matsuura (1995), Smart interpolation of annually averaged air-temperatures in the United States, *J. Appl. Meteorol.*, *34*(12), 2577–2586.
- Willmott, C., and S. Robeson (1995), Climatologically aided interpolation (CAI) of terrestrial air temperature, *Int. J. Climatol.*, *15*(2), 221–229.
- Yin, J. H. (2005), A consistent poleward shift of the storm tracks in simulations of 21st century climate, *Geophys. Res. Lett.*, *32*, L18701, doi:10.1029/2005GL023684.

T. Pavelsky, Department of Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA. (pavelsky@email.unc.edu)

S. Sobolowski, Bjerknæs Centre for Climate Research, Postboks 7810, NO-5020 Bergen, Norway. (stefan.sobolowski@uni.no)