

# Dynamically and statistically downscaled seasonal temperature and precipitation hindcast ensembles for the southeastern USA

J. T. Schoof,<sup>a\*</sup> D. W. Shin,<sup>b</sup> S. Cocke,<sup>b</sup> T. E. LaRow,<sup>b</sup> Y.-K. Lim<sup>b</sup> and J. J. O'Brien<sup>b</sup>

<sup>a</sup> Department of Geography and Environmental Resources, Southern Illinois University, Carbondale, IL, 62901, USA

<sup>b</sup> Center for Ocean-Atmosphere Prediction Studies, Florida State University, Tallahassee, FL, 32306-2840, USA

**ABSTRACT:** We present results from a 15-year 10-member warm season (March–September) hindcast ensemble of maximum and minimum surface air temperatures and precipitation in southeast USA. The hindcasts are derived from the Florida State University/Center for Ocean-Atmospheric Prediction Studies Global Spectral Model (FSU/COAPS GSM) and downscaled using both the FSU/COAPS Nested Regional Spectral Model (NRS) and a statistical downscaling method based on stochastic weather generator techniques. We additionally consider statistical bias correction of the dynamical model output. Basic descriptive statistics indicate that the bias-corrected and statistically downscaled data reduce the FSU/COAPS GSM bias considerably in terms of basic climatology. Statistics describing the daily precipitation process are improved by both downscaling techniques relative to the bias-corrected GSM. Improvement in monthly and seasonal hindcasts relative to FSU/COAPS GSM is spatially and temporally varying. Precipitation hindcasts are generally less skillful than those for temperature, although useful precipitation predictability exists at many locations. Hindcast improvements due to downscaling are greatest over peninsular Florida. The smallest root mean square errors (RMSE) for temperature hindcasts are found in the southern part of the study region during the spring months of March, April and May (MAM) for maximum surface air temperature, and in the summer, June, July and August (JJA), for minimum surface air temperature. Overall, there is no indication that either downscaling method has a direct advantage over the other. Copyright © 2008 Royal Meteorological Society

KEY WORDS downscaling; seasonal hindcast; regional climate model

Received 12 June 2007; Revised 3 April 2008; Accepted 5 April 2008

## 1. Introduction

Seasonal climate forecasts provide useful information for many applications, ranging from agricultural and hydrological modelling (e.g. Shin *et al.*, 2006) to prediction of disease (e.g. Morse *et al.*, 2005). While global climate models (or general circulation models, GCMs) are currently capable of providing qualitative seasonal forecasts, informing the end-user of either below normal, near normal, or above normal conditions, many applications require climate information at much finer spatial and/or temporal scales than those provided by global models. In order to make the GCM output useful for applications, it is necessary to downscale the model output, using either dynamical or statistical methods.

Dynamical downscaling consists of running a high-resolution regional climate model (RCM) using the GCM output for initial and lateral boundary conditions. The RCM is therefore used to account for forcings at spatial scales finer than those offered by the parent GCM. Although more computationally expensive, dynamical

downscaling methods have the potential to outperform statistical techniques, particularly because the resulting downscaled climates are necessarily physically consistent with the GCM output from which they are derived. However, this also means that biases can be passed from the GCM to the RCM resulting in biased regional climate estimates (Giorgi *et al.*, 2001). RCMs can exhibit additional bias resulting from parameterization or inconsistencies with the parent GCM.

Statistical methods of downscaling rely on established relationships between GCM output and a variable of interest at some scale finer than that of the GCM. Statistical downscaling can, therefore, be carried out inexpensively using any number of available methods ranging in complexity from simple interpolation methods and regression models to stochastic weather generators and artificial neural networks. The primary shortcoming of the statistical approach is the assumption that the relationship between the predictors and predictand(s) is stationary, which cannot be proven in advance. However, because statistical downscaling methods are based on observed relationships, they have the potential to incorporate local or regional environmental factors that are not resolved by even the finest RCMs.

\* Correspondence to: J. T. Schoof, Department of Geography and Environmental Resources, Southern Illinois University, Carbondale, IL, 62901, USA. E-mail: jschoof@siu.edu

Studies comparing dynamically and statistically down-scaled GCM (e.g. Hanssen-Bauer *et al.*, 2003; Busuioc *et al.*, 2006) output have generally found similar skill resulting from the two methods. However, these studies have generally focussed on climate change simulations. Here, we extend the comparison to the seasonal timescale and examine the skill associated with the direct model output, as well as dynamically and statistically down-scaled seasonal climate hindcasts, for southeastern USA (Figure 1) as part of a larger project aimed at providing seasonal climate information for applications in agriculture, forestry and water resources management (see <http://secc.coaps.fsu.edu>). It is acknowledged that predictability in this region is greatest during the winter season, primarily via the ENSO teleconnection (Quan *et al.*, 2006; Cocke *et al.*, 2007). However, for applications in the agricultural sector (e.g. Shin *et al.*, 2006), it is imperative to assess the quality of warm season forecasts. This article, therefore, focusses on the evaluation of temperature and precipitation hindcasts from a global climate model and downscaled using dynamical and statistical methods.

In Section 2, we describe the data used in the analysis, which includes the FSU/COAPS GSM, the European Centre for Medium-Range Weather Forecasting (ECMWF) reanalysis data, and the gridded surface climate information provided by the Florida Climate Center.

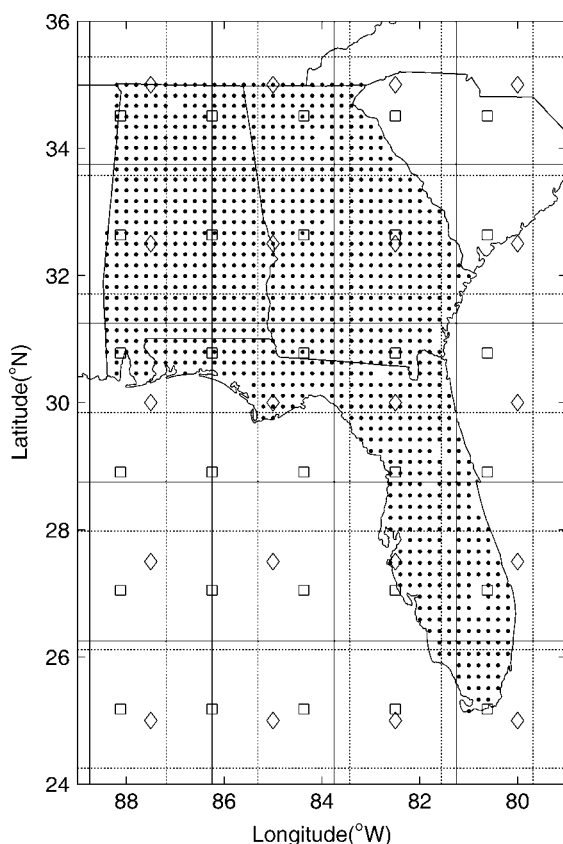


Figure 1. Map of the study area showing various data sources: surface grid points (●), ECMWF grid points (◇), and GSM grid points (□). The vertical and horizontal lines depict the grid boxes associated with the ECMWF (solid) and GSM (dash-dotted) grids.

In Section 3, we describe the downscaling methodology, including the details of the FSU/COAPS Nested Regional Spectral Model (NRSM) and the statistical downscaling methods. The results of the analysis are presented in Section 4 and summarized and discussed in Section 5.

## 2. Data

### 2.1. The Florida State University/Center for Ocean–Atmospheric Prediction Studies Global Spectral Model (FSU/COAPS GSM)

Both the dynamically and statistically downscaled data are driven by the Florida State University/Center for Ocean Atmospheric Studies Global Spectral Model (FSU/COAPS GSM) (Cocke and LaRow, 2000), hereafter referred to as the GSM. The horizontal model resolution is T63, corresponding to a model resolution of approximately  $1.875^\circ \times 1.875^\circ$  in the tropics, with 27 vertical levels. The model has recently been upgraded by including the National Center for Atmospheric Research (NCAR) Community Land Model (CLM2) for the land surface component (see Bonan *et al.*, 2002; Shin *et al.*, 2005). The results described here employ the Simplified Arakawa-Schubert (SAS) convective scheme (Pan and Wu, 1995). In this study, we employ a 10-member ensemble from the GSM obtained by starting the simulations on 10 consecutive days using initial conditions from ECMWF reanalysis data provided by NCAR. The model is then run for the period of interest, until 30 September, using prescribed sea surface temperatures (SSTs), which are updated weekly to provide information regarding the ocean state to the atmospheric model. The simulations here, therefore, provide an upper limit to model-based predictability since, in a predictive mode, SSTs would also be simulated, introducing additional error. The GSM grid points used in this study are shown in Figure 1. Additional details regarding the GSM are provided by Cocke and LaRow (2000) and Shin *et al.*, (2005).

### 2.2. ECMWF reanalysis (ERA-40)

The ECMWF reanalysis dataset (Uppala *et al.*, 2005) is used with the gridded surface data to develop statistical relationships, which are then applied to output from the GSM. The data are gridded atmospheric fields with  $2.5^\circ \times 2.5^\circ$  resolution and were obtained from the ECMWF data server. A large suite of potential predictor variables, including sea-level pressure, specific humidity, temperature, geopotential height, zonal and meridional wind speed, and relative vorticity were evaluated for use in the statistical downscaling (Table I). The ECMWF reanalysis grid points used in this study are shown in Figure 1.

### 2.3. Surface station data

The surface data used in this study consists of gridded data derived from the National Weather Service (NWS) Cooperative Observing Program network (see <http://www.nws.noaa.gov/climate>) and provided by the

Table I. Predictor variables used for statistical downscaling.

Variable name	Pressure levels (hPa)	Symbol
Specific humidity	500, 700, 850	$Q_{500}, Q_{700}, Q_{850}$
Relative vorticity	500, 700, 850	$RV_{500}, RV_{700}, RV_{850}$
Sea-level pressure		SLP
Air temperature	500, 700, 850	$T_{500}, T_{700}, T_{850}$
Zonal wind speed	500, 700, 850	$U_{500}, U_{700}, U_{850}$
Meridional wind speed	500, 700, 850	$V_{500}, V_{700}, V_{850}$
Geopotential height	500, 700, 850	$Z_{500}, Z_{700}, Z_{850}$

Florida Climate Center ([http://www.coaps.fsu.edu/climate\\_center](http://www.coaps.fsu.edu/climate_center)). The data consists of maximum and minimum daily surface air temperatures ( $T_{max}, T_{min}$ ) and total daily precipitation ( $P$ ), which have been objectively interpolated to a  $20 \times 20$  km grid using the Cressman objective analysis scheme (Cressman, 1959). The grid point locations are depicted in Figure 1.

### 3. Methodology

3.1. The Florida State University/Center for Ocean–Atmospheric Prediction Studies Nested Regional Spectral Model (FSU/COAPS NRSM)

The RCM used in this study is the FSU Nested Regional Spectral Model, hereafter referred to as the NRSM. The NRSM can be run at any horizontal resolution using base fields from the global model for boundary conditions. The regional model forecasts perturbations to the global model solution, and the perturbations, which need not be small, are represented spectrally by double Fourier series. The regional model employs a lateral boundary relaxation of the perturbations that has an effective width of about 4 grid points. While the utility of regional models is often tested using base fields from reanalysis products for boundary conditions, problems can result from differing parameterizations (e.g. convective schemes) being used in the reanalysis and RCM. Therefore, in this study, a 6-h nesting interval is used, and the regional model uses the same physics and physical parameterizations as the GSM, including the SAS convection scheme and NCAR CLM2 land model. The GSM and NRSM includes options for six different convection schemes, and tests indicate that the SAS scheme generally performs better than mass flux schemes for the domain and season of this study. By utilizing the same physical parameterizations (save for some resolution-dependent parameters) and dynamical core, we ensure a high degree of compatibility between the global and regional model solution. As a result, the large-scale features of the regional model solution are very consistent with that of the global model solution. A full description of the NRSM, along with an evaluation of its simulations, is given in Cocke (1998) and Cocke and LaRow (2000). For the experiment presented here, the NRSM was run at 20 km resolution. The simulations consist of a 10-member ensemble consistent with the parent GCM

described in Section 2.1. The integrations are performed for 15 years (1987–2001) and for seven months, beginning on 1 March and ending on 30 September of each year. The NRSM has recently been applied to the prediction of cold season precipitation (Cocke *et al.*, 2007) and warm season maximum surface air temperature (Lim *et al.*, 2007). Here, we extend the analysis to include warm season predictions of precipitation and both maximum and minimum surface air temperatures.

### 3.2. Statistical downscaling

Statistical downscaling can be performed using a wide variety of methods, ranging from simple interpolation, regression and analog techniques, to more complex methods, including stochastic weather generators and artificial neural networks. In this study, a hybrid approach was used in which weather generator parameters were downscaled using multiple regression techniques. Most weather generators are variants of the autoregressive approach introduced by Matalas (1967) and refined by Richardson and Wright (1984). Such models have separate components for precipitation and non-precipitation variables. After precipitation has been generated, the remaining variables are generated with dependence on wet/dry status.

#### 3.2.1. Precipitation

The precipitation component of the statistical model has two parts: a binary occurrence process (i.e. it either rains or does not rain) and an intensity process (i.e. the amount of precipitation given on a wet day).

#### 3.2.2. Precipitation occurrence

The occurrence process has historically been addressed using Markov chain models or spell length approaches while wet day amounts have been chosen from a number of statistical probability distributions. Here, precipitation occurrence is downscaled using a logistic regression model of the form:

$$p = 1 - \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (1)$$

where  $p$  is the probability of precipitation,  $\beta_0$  is the regression constant, and  $\beta_{1..k}$  are the slope coefficients for the explanatory variables ( $x_{1..k}$ ). A uniform random number on the interval [0 1],  $u$ , is then generated. If  $p$  is less than  $u$ , then the daily precipitation amount is zero. If  $p$  is greater than or equal to  $u$ , a non-zero precipitation amount,  $P$ , must be determined.

For model construction, predictor variables are extracted from the ECMWF reanalysis (Table I). The variables were chosen *a priori* based on their expected influence on regional surface precipitation and air temperature variations. The optimal model is found by testing each model on the GSM data using a cross-validation framework and by cycling through the potential predictor variables and years. For example, to downscale the

GSM data from 1987, a model like that in Equation (1) would be constructed for each potential predictor variable using ECMWF reanalysis data and station data (for 1988–2001). In this way, the optimal model can be determined by examining overall performance based on this cross-validation. Here, the criterion tested is the wet-day probability. The best model is the one that, when averaged over all years (as constructed above), minimizes the difference in observed and downscaled wet-day probability. This approach also limits the variable selection to those variables that perform well when the model is driven by output from the GSM. The primary variables included in the final models were 850 mb geopotential height and mid-troposphere (700 mb) winds.

### 3.2.3. Precipitation intensity

The non-zero precipitation amounts are chosen randomly from the gamma distribution:

$$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta\Gamma(\alpha)}, \quad x, \alpha, \beta > 0 \quad (2)$$

where  $x$  is the daily precipitation amount,  $\alpha$  and  $\beta$ , are the gamma distribution parameters (the shape and scale parameters, respectively), and  $\Gamma(\alpha)$  is the gamma function evaluated at  $\alpha$ . The gamma distribution has been used by a large number of studies (e.g. Richardson, 1981; Wilks, 1992). Application of this distribution to the downscaling problem requires identifying statistical relationships between the parameters,  $\alpha$  and  $\beta$ , and the large-scale variables. For each year (1987–2001),  $\alpha$  and  $\beta$ , are estimated from observations. Anomalies of  $\alpha$  and  $\beta$  are then regressed (separately) against anomalies of ECMWF variables (Table I) using a cross-validation framework similar to that described in Section 3.2.2. The best model is the model that maximizes the variance explained when GSM anomalies replace the ECMWF data used to construct the model. Additional variables are added if they increase variance explanation by 1.5%. The anomalies that are produced by the model are then added to the mean values of  $\alpha$  and  $\beta$ . The primary predictor chosen for  $\alpha$  was mid-troposphere (500 mb) temperature, with zonal wind and 850 mb and 500 mb chosen as secondary predictors at some stations. The primary predictors for  $\beta$  were 500 mb specific humidity and 500 mb temperature with 500 mb zonal wind and geopotential height as secondary predictors at some stations.

### 3.2.4. Maximum and minimum daily air temperatures ( $T_{\max}$ and $T_{\min}$ )

The autoregressive weather generator models  $T_{\max}$  and  $T_{\min}$  using:

$$X_i = AX_{i-1} + B\varepsilon \quad (3)$$

where  $X_i$  is a vector containing the current day's standardized values of  $T_{\max}$  and  $T_{\min}$ ,  $X_{i-1}$  is a vector containing the previous day's standardized values of  $T_{\max}$

and  $T_{\min}$ ,  $\varepsilon$  is a vector of independent standard Gaussian values, and  $A$  and  $B$  are matrices given by:

$$A = M_1 M_0^{-1} \quad (4)$$

$$BB^T = M_0 - M_1 M_0^{-1} M_1^T \quad (5)$$

where  $M_0$  is the matrix of lag-0 cross correlations and  $M_1$  is the matrix of lag-1 cross correlations.  $A$  can be computed directly, while  $B$  requires the use of an eigenvalue decomposition technique (see Greene, 2000). After generation of the standardized values using Equation (3), properly dimensioned values are obtained by multiplying a standard deviation and adding a mean value. In this study, we employ monthly means and standard deviations for this purpose.

Previous studies (e.g. Schoof and Robeson, 2003; Qian *et al.*, 2005; Schoof *et al.*, 2007) have shown that  $M_0$  and  $M_1$  vary considerable over space and on a seasonal basis, but report only small changes in  $M_0$  and  $M_1$  through time. We therefore use monthly values of  $M_0$  and  $M_1$  estimated from the observed data for the period 1987–2001, leaving out the year that is being hindcast. These data are used to determine values of  $A$  and  $B$ . The temperature downscaling problem is then reduced to determining values for the monthly means and standard deviations of  $T_{\max}$  and  $T_{\min}$ .

For each year in the historical data (1987–2001), the monthly means and standard deviations of  $T_{\max}$  and  $T_{\min}$  are estimated from observations. Anomalies of each of these predictands are then regressed (separately) against anomalies of ECMWF variables (Table I) using a cross-validation framework similar to that described in Sections 3.2.2 and 3.2.3. The best model is the model that maximizes the variance explained when GSM anomalies replace the ECMWF data used to construct the model. Additional variables are added if they increase variance explanation by 1.5%. The anomalies that are produced by the model are then added to the mean values of the means and standard deviations of  $T_{\max}$  and  $T_{\min}$  estimated using the 1987–2001 data (again leaving out the current year).

The primary predictor chosen for both the monthly mean  $T_{\max}$  and monthly mean  $T_{\min}$  is mid-troposphere (500 mb, 700 mb) temperature. At a few stations, geopotential height at mid-troposphere levels is chosen. For the standard deviation of  $T_{\max}$ , the primary variable chosen is geopotential height, but the optimal vertical level varies among stations. Geopotential height is also chosen as the primary predictor for the standard deviation of  $T_{\min}$  at some stations, while at other stations 850 mb temperature and 500 mb zonal wind are chosen.

### 3.3. Evaluation tools

Because there is no *a priori* expectation that the daily downscaled data should exhibit a one-to-one correspondence with the daily observed data, we focus on the statistical properties of the daily data, which reflect the long-term climatological agreement between observations

and simulations, and the errors associated with the hindcasts at the monthly and seasonal timescales. For the former, we examine precipitation occurrence statistics and means and variances of the daily temperatures and precipitation amounts. For the latter, we evaluate the hindcasts in terms of root mean square errors (RMSEs).

## 4. Results

### 4.1. Climatological statistics

Initial examination of the daily GSM data indicated substantial bias in the temperature and precipitation data for some parts of the region. Through the nesting process, this bias can be passed down to the NRSM. The sources of the bias are currently under investigation. To ensure a fair comparison to the statistically downscaled data, which are derived using observed statistics, a simple bias correction was applied to both the GSM and NRSM output. The correction applied here is described by Wood *et al.*, (2002) and consists of re-mapping the cumulative distribution function (CDF) of the predicted data to that of the observed data. For example, if a minimum temperature value of 285 K corresponds to the 60th percentile in the GSM or NRSM output, then the bias-corrected value would be the 60th percentile value from the observations. Specifically, for each year and month, the simulated CDF is corrected by re-mapping the CDF simulated for the month in the other years to the observed CDF for all the other years. With respect to precipitation, the GSM and NRSM were found to produce too many wet days relative to observations in accordance with results from previous studies (e.g. Mavromatis and Jones (1999)). The bias correction simultaneously addresses overestimation of precipitation occurrence and precipitation amount by the GSM and NRSM by re-mapping trace precipitation amounts to zero. The GSM is bias corrected and evaluated by re-sampling the coarse resolution GSM grid point data at the resolution of the surface data and NRSM output.

#### 4.1.1. Precipitation occurrence

The bias correction described above greatly improves the agreement between the GSM and NRSM and observations with respect to the wet-day probability, and the post-bias-correction GSM and NRSM provide a similar level of agreement with observations as the statistically downscaled data. Averaged over all ensemble members, the coefficients of determination ( $R^2$ ) for the bias-corrected GSM and NRSM are both 0.98 (compared to 0.26 and 0.31 for the raw GSM and NRSM output, respectively). The variance explained by the statistically downscaled results is 0.99. These results indicate that the overall wet-day probability in the GSM and in both the dynamically and statistically downscaled hindcasts exhibits a high level of agreement with observations during the hindcast period.

The overall wet-day probability is useful as a summary statistic, but day-to-day (e.g. high frequency) variability

may be more important for some applications. Transition probabilities were, therefore, computed for each dataset (GSM, NRSM, and statistically downscaled, hereafter referred to as STAT) to assess the realism of the occurrence process in the simulated and downscaled daily precipitation. Here, we limit the assessment to first-order transition probabilities:  $p_{01}$  (the probability of a wet day following a dry day) and  $p_{11}$  (the probability of a wet day following a wet day). These transition probabilities fully describe the first-order precipitation process since  $p_{00} = 1 - p_{01}$  and  $p_{10} = 1 - p_{11}$ .

Prior to bias correction, the GSM and NRSM both produce too many wet days. The bias-corrected data exhibit good agreement with observations (Figure 2). As shown, the persistence of the daily precipitation process is too strong in the GSM data, with the probability of a wet day following a dry-day low relative to observations, and the probability of a wet day following a wet-day high relative to observations. After being downscaled, the level of agreement with observations, as determined by the  $R^2$ , value is much larger ( $p_{01}$ :  $R^2 = 0.81$  for NRSM and 0.82 for STAT;  $p_{11}$ :  $R^2 = 0.85$  for NRSM and 0.88 for STAT relative to values of 0.66 and 0.74 for  $p_{01}$  and  $p_{11}$  from the GSM respectively.). The bias-corrected dynamical results are a substantial improvement over the raw GSM and NRSM output, which explained only 10–30% of the variance in these transition probabilities. Like the bias-corrected GSM, the transition probabilities derived from the statistically downscaled data exhibit systematic error, with the likelihood of a wet day following a dry day too high relative to observations, and the likelihood of a wet day following a wet day too low relative to observations (Figure 2). This result suggests that the statistical model employed underestimates the temporal autocorrelation in the precipitation process. The dynamically downscaled precipitation data, with transition probabilities in the same range as those observed, may be more attractive to end users due to better representation of the temporal autocorrelation, and hence, high frequency information, in the observed precipitation series.

#### 4.1.2. Precipitation intensity

To assess the simulated precipitation intensity, mean wet-day precipitation amounts were computed from the observations, GSM, and data downscaled using the NRSM and the statistical method. For the bias-corrected GSM and NRSM and the statistically downscaled data, there is a high-level agreement between observed and simulated mean wet-day precipitation amounts ( $R^2 > 0.95$ , Figure 3). These results and those presented in Section 4.1.1 suggest that the bias will correct GSM and NRSM, and statistically downscaled precipitation data exhibit good agreement with observations. Hindcast performance at longer timescales will be assessed in Sections 4.2 and 4.3.

#### 4.1.3. Maximum surface air temperature ( $T_{max}$ )

The GSM and NRSM exhibit considerable bias in maximum surface air temperature ( $T_{max}$ ) (Figures 4 and 5). Spatial examination of the results presented in

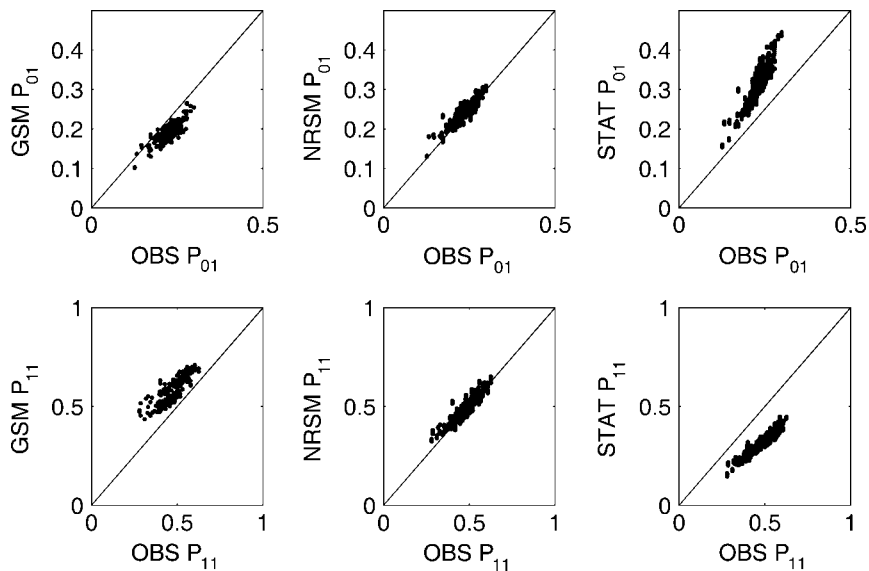


Figure 2. Observed and hindcast values of the transition probabilities for dry day to wet day ( $p_{01}$ , top 3 panels) and wet day to wet day ( $p_{11}$ , bottom 3 panels). The three columns correspond to GSM, NRSM, and statistically downscaled hindcasts. Each point represents an ensemble average for a single grid point.

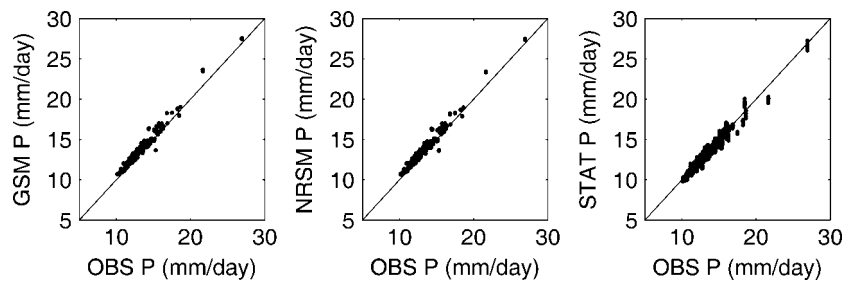


Figure 3. Observed and predicted values of the mean daily precipitation amount for (from left to right) direct (GSM) and dynamically (NRSM) and statistically (STAT) downscaled GSM output. Each point represents an ensemble average for a single grid point.

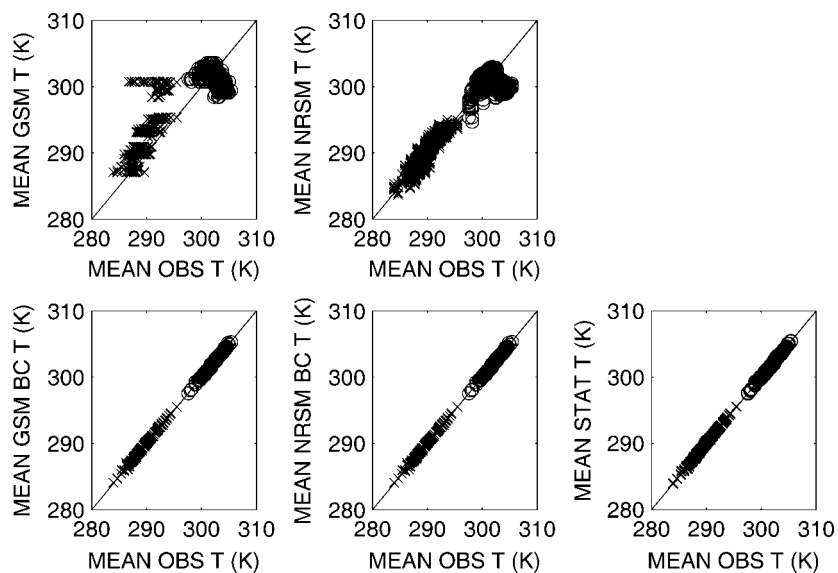


Figure 4. Observed and simulated mean values of daily maximum (O) and minimum (x) surface air temperature. Results are shown for the GSM (left), NRSM (center), and statistically downscaled (STAT, right) data. The second row shows the GSM and NRSM results after statistical bias correction.

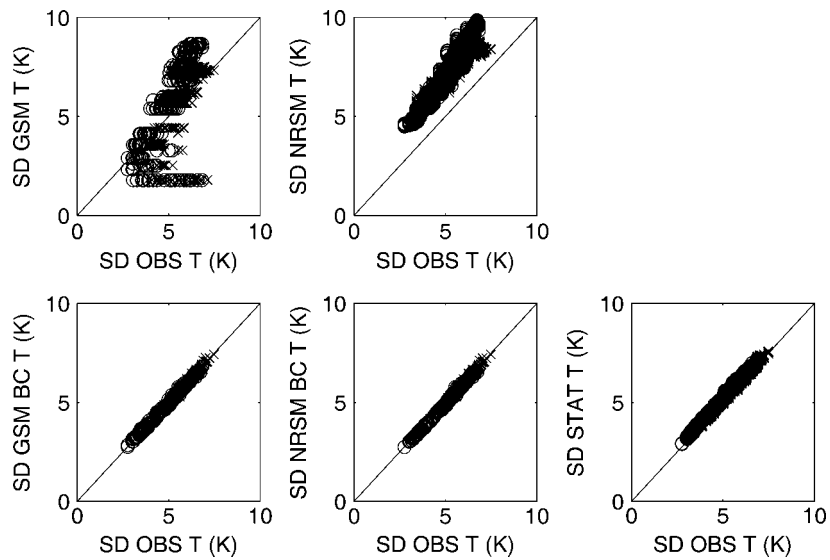


Figure 5. Observed and simulated standard deviations of daily maximum (○) and minimum (×) surface air temperature. Results are shown for the GSM (left), NRSM (center), and statistically downscaled (STAT, right) data. The second row shows the GSM and NRSM results after statistical bias correction.

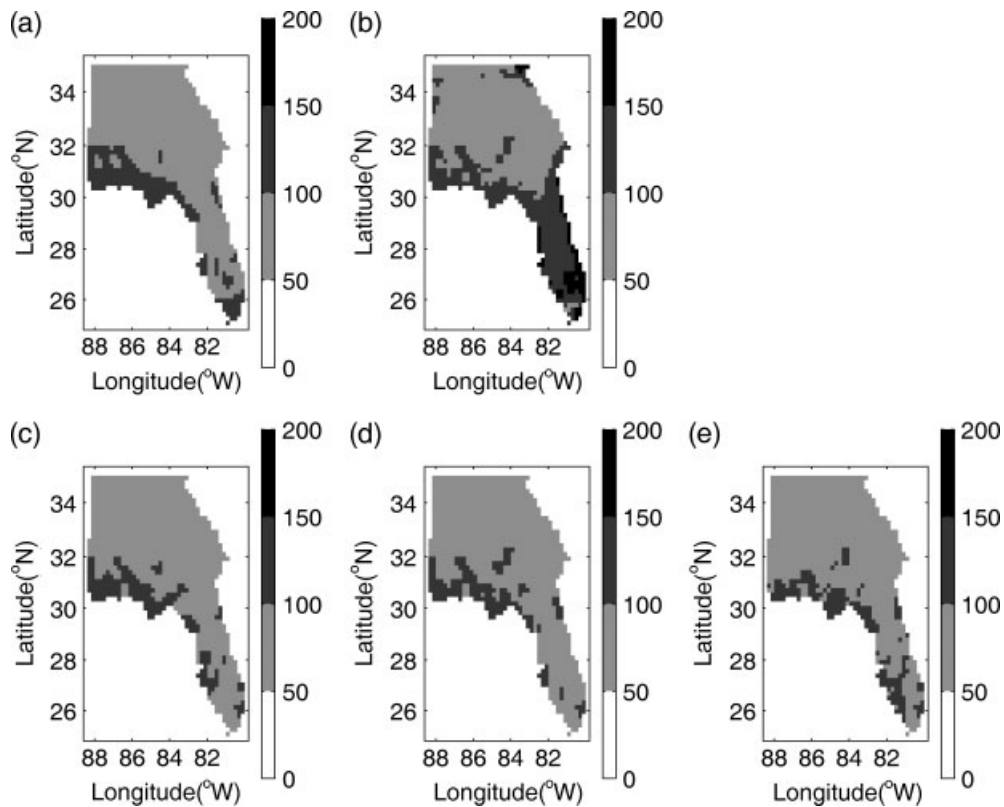


Figure 6. Root mean square errors (RMSE; mm) for monthly total precipitation hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output.

Figure 4 indicates that both the GSM and NRSM exhibit cold bias over the Florida peninsula and warm bias over the remaining parts of the study region. Application of the bias correction to the GSM and NRSM data improves the relationship between the observed and simulated values of  $T_{max}$  (Figures 4 and 5). The statistically downscaled data likewise exhibits a high level of agreement with observations in terms of the overall

mean and standard deviation. The GSM and NRSM also suffer from overestimation of variance (Figure 5), particularly, at northern stations where temporal variability is greater, while the downscaled data exhibit excellent agreement with observations ( $R^2 > 0.96$ ). This shortcoming in the GSM and NRSM data is also improved through application of the bias correction (Figures 4 and 5).

#### 4.1.4. Daily minimum surface air temperature ( $T_{\min}$ )

Although the GSM also exhibits bias with respect to  $T_{\min}$ , the simulations exhibit slightly better agreement than that described for  $T_{\max}$  (Figure 4). As shown in Figure 4, the GSM has a warm  $T_{\min}$  bias at most stations. As with  $T_{\max}$ , the downscaled data result in considerable improvement over those directly simulated by the GSM. Application of the bias correction results in excellent agreement between the GSM, downscaled and observed daily  $T_{\min}$ . The variance characteristics of  $T_{\min}$  are very similar to those described for  $T_{\max}$  in the previous section (Figure 5).

#### 4.2. Monthly and seasonal hindcast assessment

The results presented in Section 4.1 reflect the most basic properties of the simulated data – reproduction of the basic statistical properties of the data. For use in climate-based applications, it must also be demonstrated that the downscaled data exhibit skill on timescales that would allow the results to be used to assist decision makers. The 15 years of hindcasts from 10-ensemble members allow a reasonable assessment of the hindcast skill (including spatial variations) associated with the model simulations. Here, we explore the skill associated with the ensemble mean. The results are presented in three sections corresponding to total precipitation, mean  $T_{\max}$ , and mean  $T_{\min}$ . For each variable, results are presented for monthly, spring (March, April and May;

MAM), and summer (June, July and August; JJA) total precipitation and mean  $T_{\max}$  and  $T_{\min}$ .

#### 4.2.1. Total precipitation

Table II shows the RMSE for monthly and seasonal precipitation totals averaged over all stations and for each state. Figures 6–8 show the spatial distribution of the RMSE for monthly, MAM, and summer JJA precipitation totals. The observed mean monthly total precipitation averages 122.5 mm over the domain, with larger totals over Florida (140.5 mm) than over the northern states (119.6 and 109.8 mm for Alabama and Georgia, respectively). During spring, the mean precipitation amount is 309.7 mm, with the lowest total precipitations amounts in Florida (268.4 mm). During summer, the sea breeze front often triggers afternoon thunderstorms along the extensive Florida coast, leading to larger precipitation amounts (545.8 mm on average for all Florida grid points).

At the monthly timescale, the lowest errors in the precipitation hindcasts occur in the northern part of the domain (Figure 6, Table II), where RMSE values range from 50–75 mm/month. The bias-correction procedure results in only minor improvement to the raw GSM results. Additionally, the NRSM appears to introduce additional error to the GSM results, which is reduced through application of the bias correction. The bias-corrected dynamical hindcasts (both GSM and NRSM)

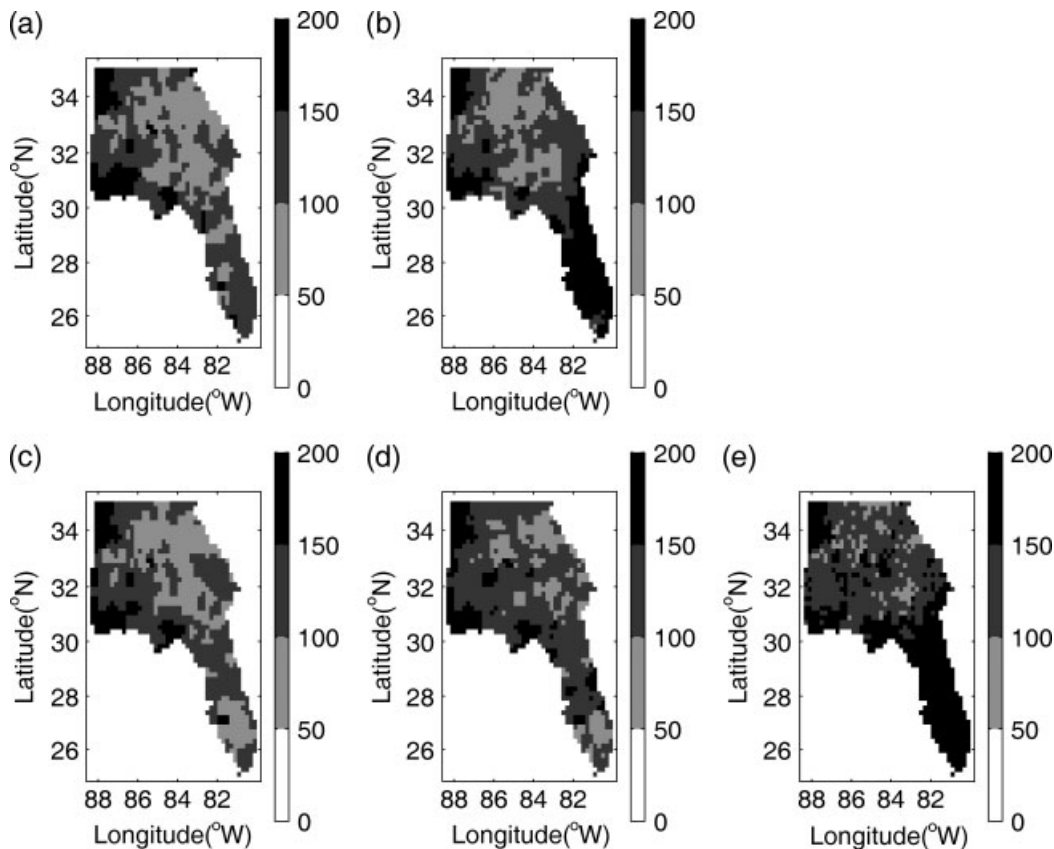


Figure 7. Root mean square errors (RMSE; mm) for spring (MAM) total precipitation hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output.



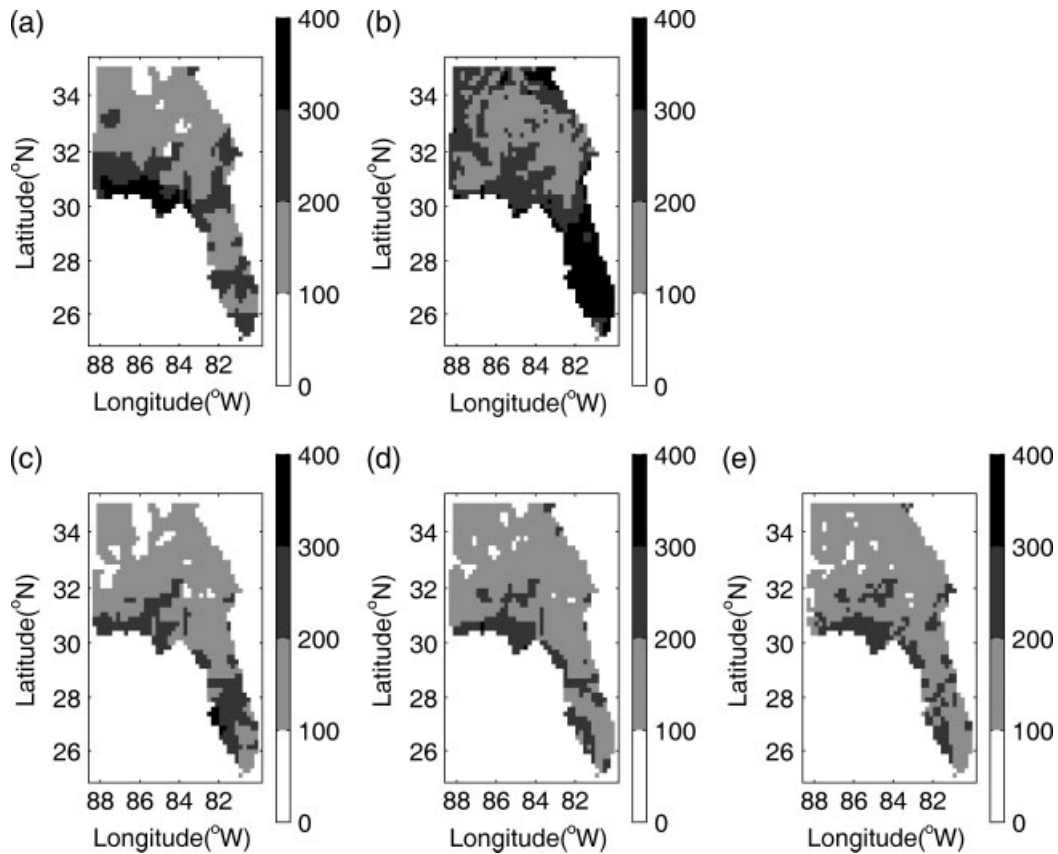


Figure 8. Root mean square errors (RMSE; mm) for summer (JJA) total precipitation hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output.

Table II. Root mean square errors (mm) for monthly and seasonal total precipitation hindcasts. Results are shown for all stations (Figure 1) and for the individual states, and for the direct and bias-corrected GSM output, the direct and bias-corrected NRSM output, and the statistically (STAT) downscaled data.

	GSM	GSM <sub>BC</sub>	NRSM	NRSM <sub>BC</sub>	STAT
Monthly (all stations)	84.6	81.9	100.2	82.3	82.0
Monthly AL	84.6	81.5	87.9	81.6	77.4
Monthly GA	72.9	70.9	86.1	73.6	72.9
Monthly FL	98.4	95.4	129.9	93.4	97.7
MAM (all stations)	117.9	114.1	135.6	124.4	154.5
MAM AL	131.1	129.4	126.2	138.9	138.5
MAM GA	96.9	96.9	110.3	107.3	126.9
MAM FL	128.6	118.2	175.8	129.0	204.5
JJA (all stations)	190.7	162.9	250.4	162.2	164.3
JJA AL	184.6	140.7	201.5	141.6	141.5
JJA GA	160.4	143.3	209.8	149.1	158.0
JJA FL	233.3	210.0	350.9	199.7	196.0

and the statistically downscaled hindcasts generally provide a similar level of skill, although minor differences exist. When averaged over all stations, the RMSEs differ by less than 0.5 mm. For northeast grid points (Georgia),

the bias-corrected GSM has a slight advantage. For northwest grid points (Alabama), the statistically downscaled data exhibits smaller errors. For southern grid points (Florida), the bias-corrected NRSM exhibits the lowest errors.

The GSM hindcasts for total MAM precipitation exhibit much smaller relative errors than their monthly counterparts, although large errors remain for some parts of the region (Figure 7, Table II). Spatially, the best hindcasts have similar error structure, with the smallest errors in the northern states and slightly larger errors for southern grid points. The bias-corrected GSM generally performs better than the downscaled data although the NRSM results in a slight improvement for grid points in Alabama.

The application of both downscaling techniques to JJA total precipitation results in considerable improvement for grid points in Florida relative to the bias-corrected GSM (Figure 8, Table II). When averaged over all stations, the bias-corrected dynamical results and statistical results are very similar, while the bias-corrected GSM bears a slight advantage over the downscaled results in Georgia.

Simulation of precipitation is strongly dependent on issues related to scale. Individual thunderstorms are not resolved by any of the methods employed. However, it is conceivable that the dynamically and statistically downscaled GSM output may be more reflective of

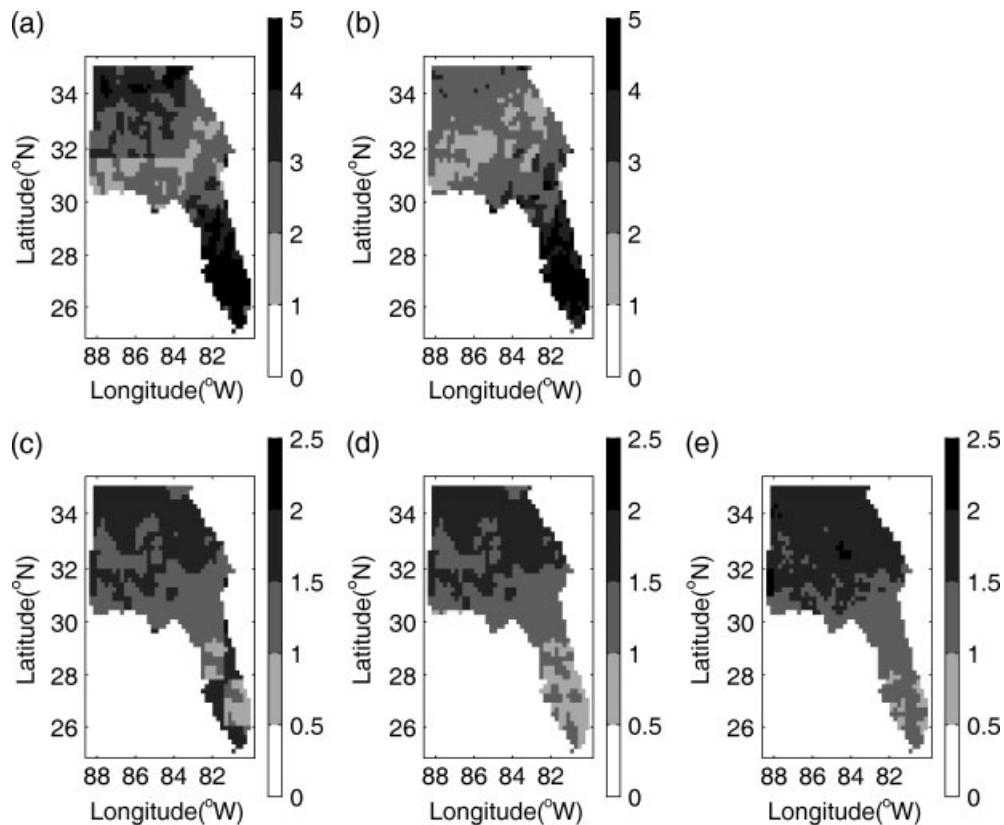


Figure 9. Root mean square errors (RMSE; K) for monthly mean maximum surface air temperature ( $T_{\max}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

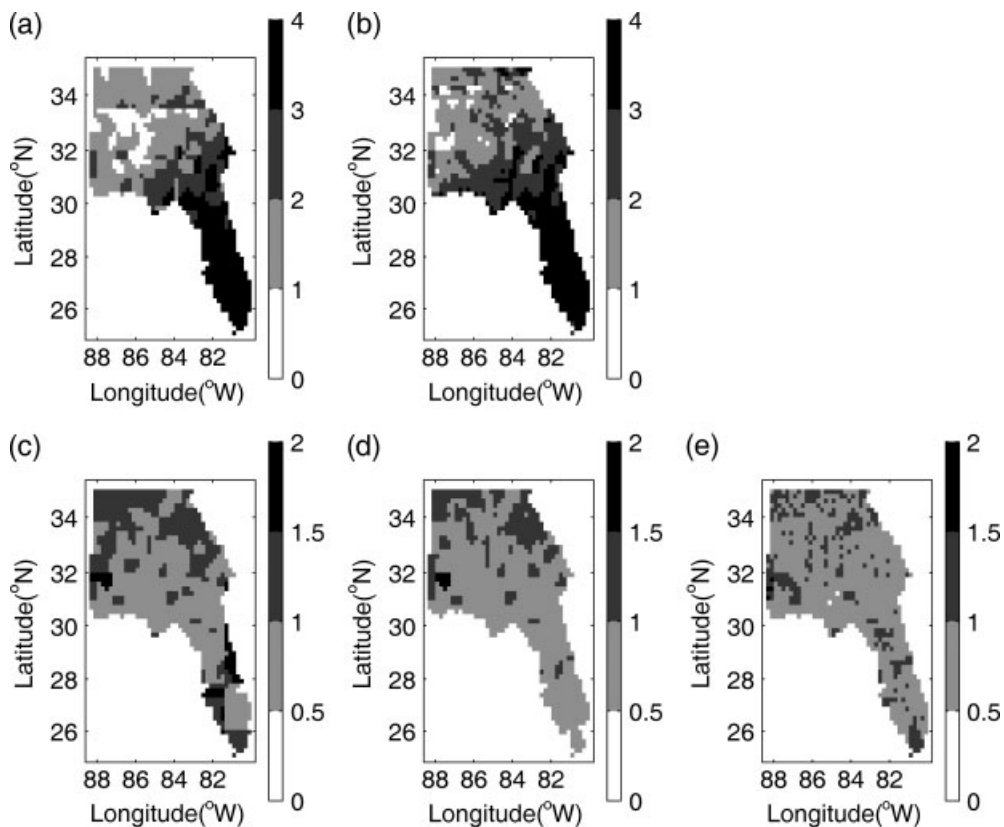


Figure 10. Root mean square errors (RMSE; K) for spring (MAM) mean maximum surface air temperature ( $T_{\max}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

mesoscale events such as squall lines triggered by thermal boundaries associated with sea breezes. While it is not possible to prove this link without more rigorous analyses, previous studies by Pielke *et al.*, (1999) and Marshall *et al.*, (2004) have simulated the Florida sea breeze using a RCM. The seasonally and spatially varying nature of the precipitation hindcast skill reported here is also consistent with the results of previous precipitation downscaling studies, as is the similar level of dynamical and statistical hindcast skill (e.g. Diez *et al.*, 2005).

4.2.2. Maximum surface air temperature ( $T_{max}$ )

Maximum temperature hindcasts exhibit lower relative errors than those described for precipitation in the previous section. Like their daily counterparts, monthly mean maximum surface air temperature hindcasts from the raw GSM and NRSM exhibit considerable bias (Table III, Figure 9). Application of the bias-correction procedure results in very similar hindcast skill for the GSM, NRSM, and statistical downscaling method (Table III, Figure 9). In regions where model resolution issues exist (e.g. with grid boxes that cover land and water, as in the Florida peninsula), the downscaled results exhibit slightly lower RMSEs than the GSM.

MAM and JJA mean maximum air temperature errors are smaller than those for the monthly data (Table III, Figures 10, 11). The bias correction again improves the GSM and NRSM simulations, and the downscaling methods result in slight improvement over the course resolution GSM, particular for the NRSM hindcasts for Florida. Averaged over all stations and over the three individual states, the spring RMSEs from the bias-corrected dynamical methods and statistical method are all less than or equal to 1 K, while those for summer

are less than or equal to 1.3 K. For select grid points, spring and summer mean errors are less than 0.5 K (Figures 10, 11).

4.2.3. Minimum surface air temperature ( $T_{min}$ )

Errors in the monthly mean minimum surface air temperature are of similar magnitude to those reported in Section 4.2.2 for  $T_{max}$ . While the GSM exhibits bias in  $T_{min}$ , the spatial variation of the bias is lower than that for  $T_{max}$ , in accordance with the daily data described in Section 4.1.4. It is particularly noteworthy that the NRSM does not inherit this bias from GSM and the bias correction, therefore, results in less improvement for the NRSM than it does for the GSM (Table IV, Figure 12). This improvement in  $T_{min}$  by the NRSM is somewhat expected since, as shown by Shin *et al.*, (2006),  $T_{min}$  is more sensitive to land surface model features than  $T_{max}$ . As with the  $T_{max}$  hindcasts, the lowest errors are produced by the bias-corrected NRSM results, followed closely by the bias-corrected GSM and statistically downscaled GSM hindcasts (Table IV, Figure 12). Improvement from downscaling is primarily found in Florida.

MAM mean minimum temperature hindcast errors are of similar magnitude to those for the monthly mean, although those for the statistically downscaled data are smaller for northern grid points (Table IV, Figure 13). The JJA mean minimum temperatures exhibit the smallest errors among all the temperature hindcasts, particularly for the bias-corrected and statistically downscaled data (Table IV, Figure 14). Errors for many stations are less than 0.5 K, and domain-wide averages are far less than 1 K for the bias-corrected GSM and NRSM, as well as the statistically downscaled data.

Table III. Root mean square errors (K) for monthly and seasonal mean maximum surface air temperature hindcasts. Results are shown for all stations (Figure 1) and for the individual states, and for the direct and bias-corrected GSM output, the direct and bias-corrected NRSM output, and the statistically (STAT) downscaled data.

	GSM	GSM <sub>BC</sub>	NRSM	NRSM <sub>BC</sub>	STAT
Monthly (all stations)	3.0	1.4	2.7	1.4	1.5
Monthly AL	2.9	1.5	2.2	1.5	1.6
Monthly GA	2.6	1.5	2.4	1.5	1.6
Monthly FL	3.6	1.3	3.7	1.1	1.2
MAM (all stations)	2.4	1.0	2.6	0.9	0.9
MAM AL	1.3	1.0	1.6	0.9	0.9
MAM GA	2.0	1.0	2.3	0.9	0.9
MAM FL	4.0	1.0	4.1	0.8	0.9
JJA (all stations)	3.0	1.1	2.3	1.1	1.1
JJA AL	3.4	1.2	2.0	1.2	1.3
JJA GA	2.5	1.2	1.8	1.2	1.3
JJA FL	3.2	0.9	3.2	0.8	0.8

Table IV. Root mean square errors (K) for monthly and seasonal mean minimum surface air temperature hindcasts. Results are shown for all stations (Figure 1) and for the individual states, and for the direct and bias-corrected GSM output, the direct and bias-corrected NRSM output, and the statistically (STAT) downscaled data.

	GSM	GSM <sub>BC</sub>	NRSM	NRSM <sub>BC</sub>	STAT
Monthly (all stations)	3.2	1.5	1.9	1.4	1.5
Monthly AL	2.7	1.5	1.9	1.5	1.6
Monthly GA	2.9	1.4	1.9	1.4	1.5
Monthly FL	4.3	1.5	1.8	1.2	1.3
MAM (all stations)	3.1	1.4	1.6	1.3	1.3
MAM AL	2.3	1.5	1.8	1.4	1.3
MAM GA	2.6	1.4	1.7	1.4	1.2
MAM FL	4.4	1.5	1.4	1.2	1.2
JJA (all stations)	2.9	0.6	1.2	0.6	0.6
JJA AL	2.5	0.7	1.1	0.7	0.7
JJA GA	2.4	0.6	1.1	0.6	0.6
JJA FL	3.9	0.6	1.3	0.5	0.6

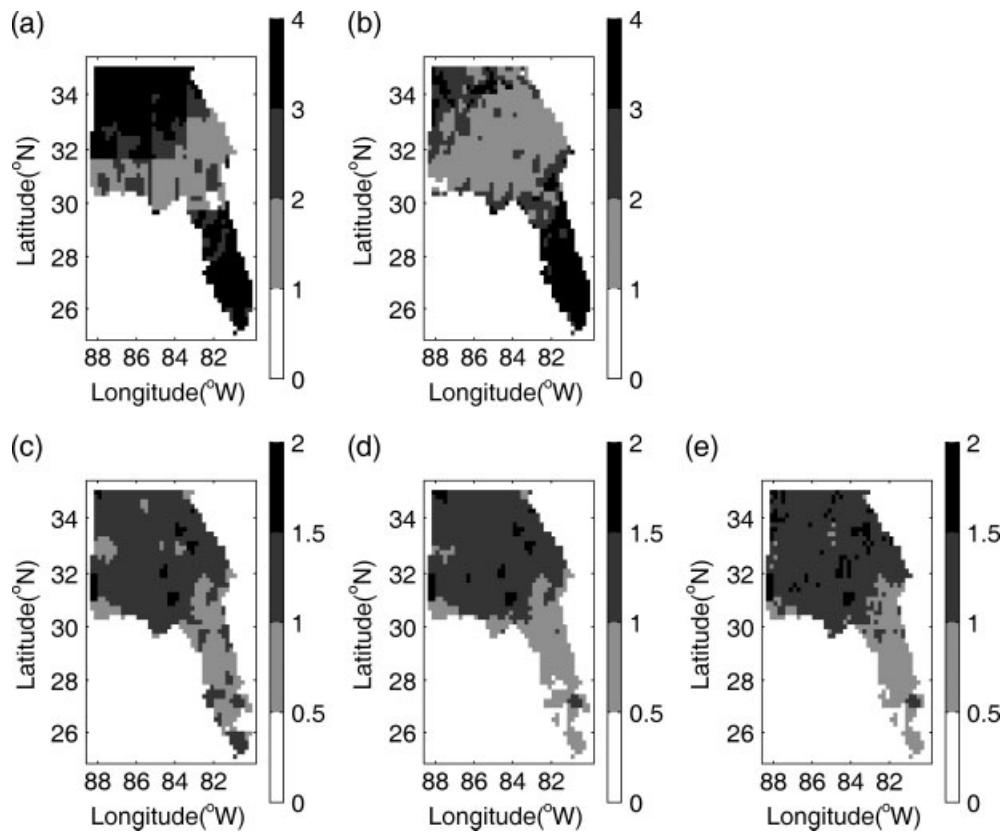


Figure 11. Root mean square errors (RMSE; K) for summer (JJA) mean maximum surface air temperature ( $T_{\max}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

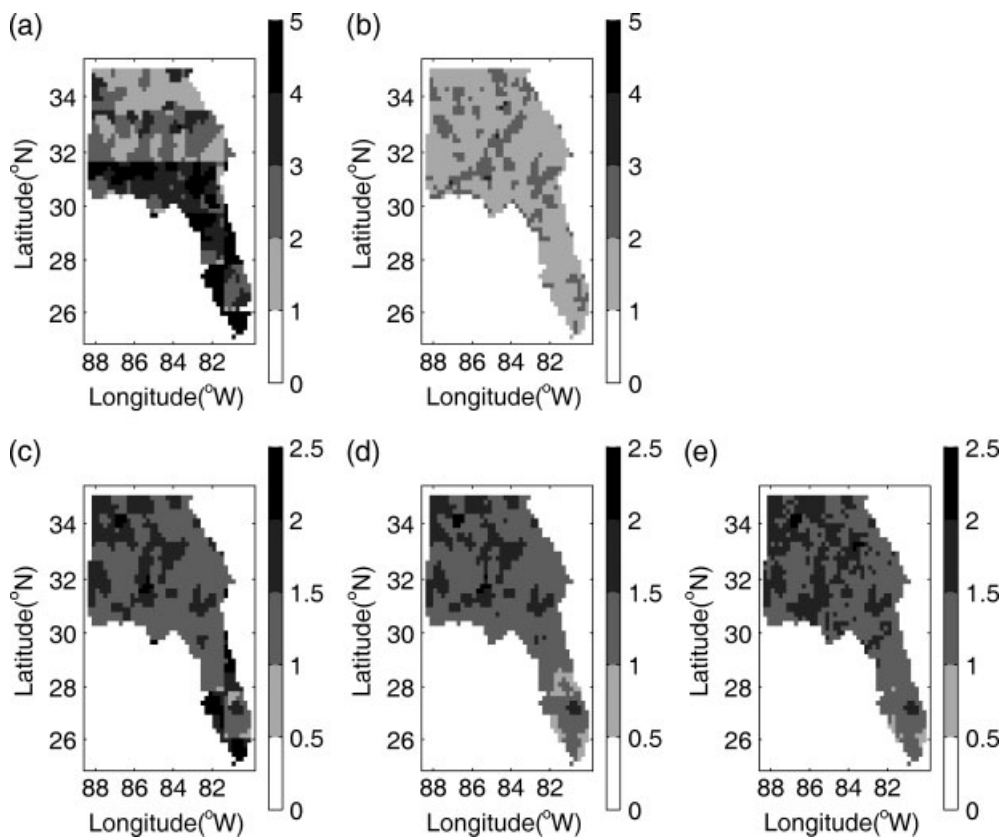


Figure 12. Root mean square errors (RMSE; K) for monthly mean minimum surface air temperature ( $T_{\min}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

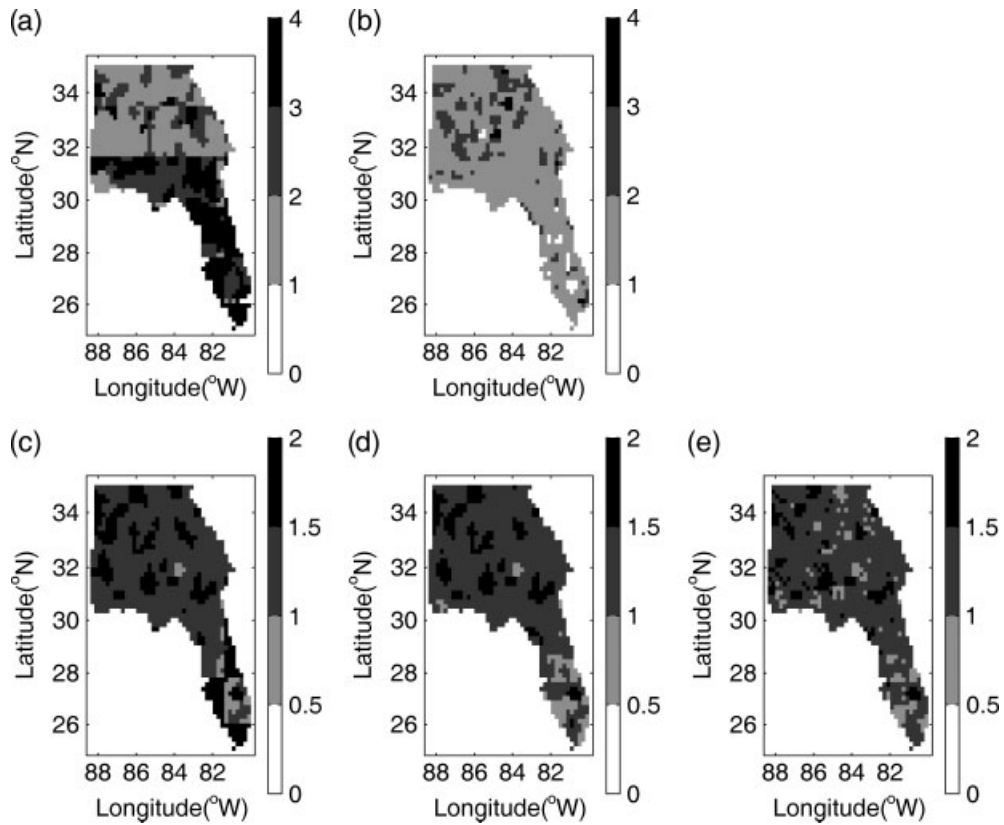


Figure 13. Root mean square errors (RMSE; K) for spring (MAM) mean minimum surface air temperature ( $T_{min}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

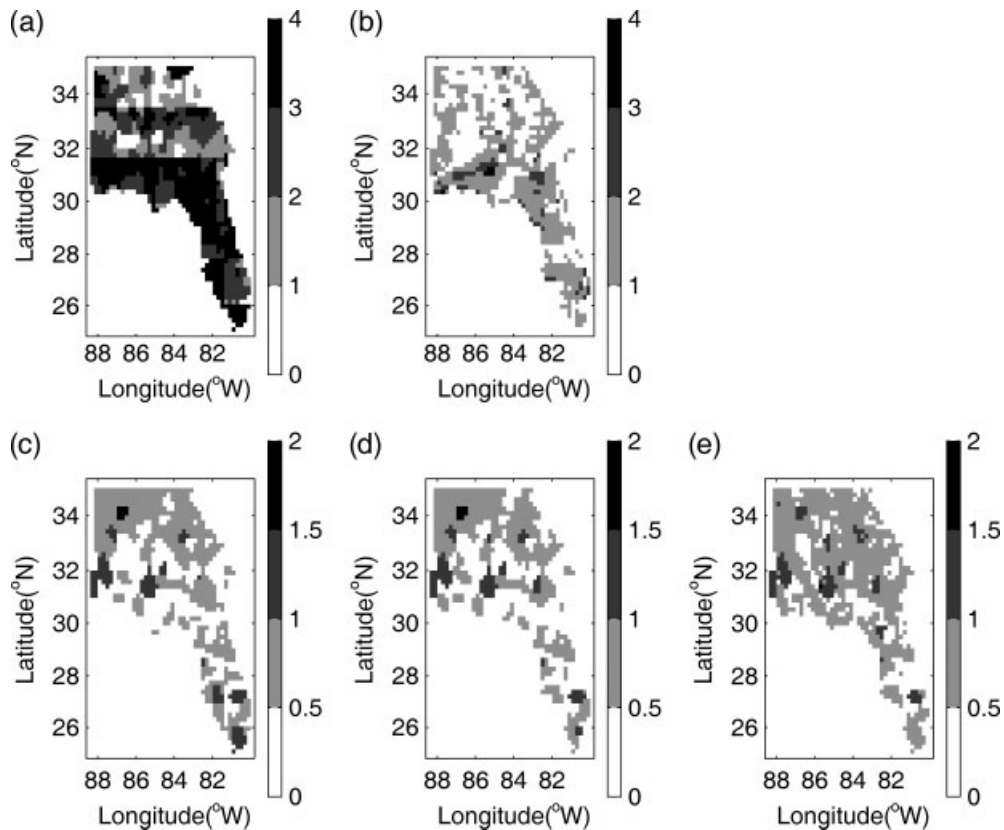


Figure 14. Root mean square errors (RMSEs; K) for summer (JJA) mean minimum surface air temperature ( $T_{min}$ ) hindcasts derived from (a) direct GSM output, (b) direct NRSM output, (c) bias-corrected GSM output, (d) bias-corrected NRSM output, and (e) statistically downscaled GSM output. Note that different scales are used for (a) and (b) compared to (c), (d), and (e).

## 5. Summary and conclusion

In this study, we have presented the results of an ensemble temperature and precipitation hindcast experiment for the southeastern USA. A 10-member ensemble of 15-year warm season (March–September) FSU/COAPS GSM experiments has been downscaled using the FSU/COAPS NRSRM and a statistical downscaling method based on a stochastic weather generator.

The results indicate that the precipitation errors are generally large, although there are regions and times of year in which precipitation is simulated with skill that may be useful for some applications (e.g. northern locations and select southern locations during spring and summer). The GSM and NRSRM produce too many wet days with small precipitation amounts on each day. Application of a bias-correction technique improves the daily data over the entire study region and also improves the monthly and seasonal precipitation total hindcasts of the NRSRM over Florida. Improvement from downscaling of the GSM output is generally limited to peninsular Florida during summer and may be due to improved simulation of mesoscale phenomena, such as sea breezes, which link local precipitation to large-scale processes that may be resolved by the NRSRM. Both the NRSRM and statistical downscaling result in better characterization of the statistics describing the daily precipitation occurrence process.

The maximum and minimum surface air temperatures from the GSM and NRSRM also exhibit bias, which is corrected using a bias-correction procedure. The bias-corrected GSM, NRSRM, and statistically downscaled daily data exhibit excellent agreement with observations in terms of climatological means and standard deviations of the daily data. Improvements based on downscaling for temperature are primarily found over Florida, in accord with the  $T_{\max}$  results reported by Lim *et al.*, (2007), but also in other regions for specific cases (e.g. the slight improvement of spring  $T_{\max}$  in Georgia by statistical downscaling in Figure 9).

The results of the analyses presented here are consistent with the results of recent studies (e.g. Castro *et al.*, 2005), which suggest that RCMs do not improve GCMs in terms of large-scale spatial structure, but could add additional skill by improving the spatial representation of weather systems. The errors reported here are similar to other seasonal climate hindcast studies (e.g. Diez *et al.*, 2005), and low errors for each variable within some part of the study area suggest that there is potential for seasonal predictability in the region. A greater understanding of situations in which downscaling adds skill to coarse model output should be a focus of further work in this arena. Given the lack of a strong ENSO signal in the study region during the warm season, further understanding of the source of forecast skill and its spatial and temporal variation will also be an avenue for further investigation.

## Acknowledgements

This research was partially supported by a subcontract from FSU to the lead author at Southern Illinois University Carbondale. COAPS is a NOAA Applied Research Center funded by the NOAA Climate Program Office. This research was partially supported via NOAA Grant NA06OAR4310070. Support from the National Science Foundation Geography and Regional Science Program (#0648025) is also gratefully acknowledged. This manuscript was improved by constructive comments from two anonymous reviewers.

## References

- Bonan GB, Oleson KW, Vertenstein M, Levis S, Zeng X, Dai Y, Dickinson RE, Yang Z-L. 2002. The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. *Journal of Climate* **15**: 3123–3149.
- Busuioc A, Giorgi F, Bi X, Ionita M. 2006. Comparison of regional climate model and statistical downscaling simulations of different winter precipitation change scenarios over Romania. *Theoretical and Applied Climatology* **86**: 101–123.
- Castro CL, Pielke RA Sr, Leoncini G. 2005. Dynamical downscaling: assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *Journal of Geophysical Research-Atmospheres* **110**(D5): D05108, DOI:10.1029/2004/JD004721.
- Cocke S. 1998. Case study of erin using the FSU nested regional spectral model. *Monthly Weather Review* **126**: 1337–1346.
- Cocke S, LaRow TE. 2000. Seasonal predictions using a regional spectral model embedded within a coupled ocean-atmosphere model. *Monthly Weather Review* **128**: 689–708.
- Cocke S, LaRow TE, Shin DW. 2007. Seasonal rainfall predictions over the southeast United States using the Florida State University nested regional spectral model. *Journal of Geophysical Research* **112**: D04106, Doi:10.1029/2006JD007535.
- Cressman GP. 1959. An operational objective analysis system. *Monthly Weather Review* **87**: 367–374.
- Diez E, Primo C, Garcia-Moya JA, Gutierrez JM, Orfila B. 2005. Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus Series A-Dynamic Meteorology and Oceanography* **57**: 409–423.
- Giorgi F, Hewitson BC, Christensen J, Hulme M, von Storch H, Whetton P, Jones R, Mearns L, Fu C. 2001. Regional climate information – evaluation and projections. In *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ (eds). Cambridge University Press: Cambridge, New York; 881.
- Greene WH. 2000. *Econometric Analysis*. Prentice Hall: Upper Saddle River, NJ.
- Hanssen-Bauer I, Forland EJ, Haugen JE, Tveito OE. 2003. Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Climate Research* **25**: 15–27.
- Lim Y-K, Shin DW, Cocke S, LaRow TE, Schoof JT, O'Brien JJ, Chassignet EP. 2007. Dynamically and statistically downscaled seasonal simulations of maximum surface air temperature over the southeastern United States. *Journal of Geophysical Research* **112**: D24102, DOI: 10.1029/2007JD008764.
- Marshall CH, Pielke RA Sr, Steyaert LT, Willard DA. 2004. The impact of anthropogenic land-cover change on the Florida peninsula sea breezes and warm season sensible weather. *Monthly Weather Review* **132**: 28–52.
- Matalas NC. 1967. Mathematical assessment of synthetic hydrology. *Water Resources Research* **3**: 937–945.
- Mavromatis T, Jones PD. 1999. Evaluation of HadCM2 and direct use of daily GCM data in impact assessment studies. *Climatic Change* **41**: 583–614.
- Morse AP, Doblas-Reyes FJ, Hoshen MB, Hagedorn R, Palmer TN. 2005. A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model. *Tellus* **57A**: 464–475.

- Pan H-L, Wu W-S. 1995. Implementing a mass flux convective parameterization package for the NMC medium-range forecast model. NMC office note 409, 40 (available from NCEP/EMC, 5200 Auth Road, Camp Springs, MD 20746).
- Pielke RA Sr, Walko RL, Steyaert LT, Vidale PL, Liston GE, Lyons WA, Chase TN. 1999. The influence of anthropogenic landscape changes on weather in South Florida. *Monthly Weather Review* **127**: 1663–1673.
- Qian B, Hayhoe H, Gameda S. 2005. Evaluation of the stochastic weather generators LARS-WG and AAFC-WF for climate change impact studies. *Climate Research* **29**: 3–21.
- Quan X, Hoerling M, Whitaker J, Bates G, Xu T. 2006. Diagnosing sources of U.S. seasonal forecast skill. *Journal of Climate* **19**: 3279–3293.
- Richardson CW. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research* **17**: 182–190.
- Richardson CW, Wright DA. 1984. *WGEN: A Model for Generating Daily Weather Variables*. US Department of Agriculture, Agricultural Research Service: Washington, DC, USA, Pub. No. ARS-8;83.
- Schoof JT, Robeson SM. 2003. Seasonal and spatial variations of cross-correlation matrices used by stochastic weather generators. *Climate Research* **24**: 95–102.
- Schoof JT, Pryor SC, Robeson SM. 2007. Downscaling daily maximum and minimum temperatures in the midwestern USA: a hybrid empirical approach. *International Journal of Climatology* **27**: 439–454.
- Shin DW, Cocke S, LaRow TE, O'Brien JJ. 2005. Seasonal surface air temperature and precipitation in the FSU climate model coupled to the CLM2. *Journal of Climate* **18**: 3217–3228.
- Shin DW, Bellow JG, LaRow TE, Cocke S, O'Brien JJ. 2006. The role of an advanced land model in seasonal dynamical downscaling for crop model application. *Journal of Applied Meteorology and Climatology* **45**: 686–701.
- Uppala SM, Kallberg PW, Simmons AJ, Andrae U, da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* **131**: 2961–3012, DOI:10.1256/qj.04.176.
- Wilks DS. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change* **22**: 67–84.
- Wood AW, Maurer EP, Kumar A, Lettenmaier DP. 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research-Atmospheres* **107**: D20.