

Probabilistic Projections of Climate Change for the Mid-Atlantic Region of the United States: Validation of Precipitation Downscaling during the Historical Era*

LIANG NING AND MICHAEL E. MANN

Department of Meteorology, and Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania

ROBERT CRANE

Department of Geography, The Pennsylvania State University, University Park, Pennsylvania

THORSTEN WAGENER

Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 21 September 2010, in final form 5 July 2011)

ABSTRACT

This study uses a statistical downscaling method based on self-organizing maps (SOMs) to produce high-resolution, downscaled precipitation estimates over the state of Pennsylvania in the mid-Atlantic region of the United States. The SOMs approach derives a transfer function between large-scale mean atmospheric states and local meteorological variables (daily point precipitation values) of interest. First, the SOM was trained using seven coarsely resolved atmospheric variables from the National Centers for Environmental Prediction (NCEP) reanalysis dataset to model observed daily precipitation data from 17 stations across Pennsylvania for the period 1979–2005. Employing the same coarsely resolved variables from nine general circulation model (GCM) simulations taken from the historical analysis of the Coupled Model Intercomparison Project, phase 3 (CMIP3), the trained SOM was subsequently applied to simulate daily precipitation at the same 17 sites for the period 1961–2000. The SOM analysis indicates that the nine model simulations exhibit similar synoptic-scale features to the (NCEP) observations over the 1979–2007 training interval. An analysis of the sea level pressure signatures and the precipitation distribution corresponding to the trained SOM shows that it is effective in differentiating characteristic synoptic circulation patterns and associated precipitation. The downscaling approach provides a faithful reproduction of the observed probability distributions and temporal characteristics of precipitation on both daily and monthly time scales. The downscaled precipitation field shows significant improvement over the raw GCM precipitation fields with regard to observed average monthly precipitation amounts, average monthly number of rainy days, and standard deviations of monthly precipitation amounts, although certain caveats are noted.

1. Introduction

Recent climate research indicates the likelihood of increasing precipitation extremes in a warmer climate (e.g., Meehl et al. 2007). Indeed, such changes are already

apparent as the atmosphere has warmed over the past century (Trenberth et al. 2007; Ning and Qian 2009). The underlying physics of this overall response is relatively well understood, involving basic factors such as those described by the Clausius–Clapeyron equation, which prescribes increased atmospheric water vapor mixing ratios with warming tropospheric temperatures. There is reason to believe that current-generation atmosphere–ocean general circulation models (AOGCMs) provide credible estimates of changes in the hydrological cycle at continental and larger spatial scales (Randall et al. 2007). However, the resolutions of GCMs are usually too coarse to provide detailed regional information about climate change on local scales, and the parameterizations

* Supplemental information related to this paper is available at the Journals Online Web site: <http://dx.doi.org/10.1175/2011JCLI4091.s1>.

Corresponding author address: Liang Ning, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802.
E-mail: lun115@psu.edu

of subgrid-scale processes, such as precipitation, also result in some degree of uncertainty in the grid-scale projections. Understanding and projecting changes in the distribution of precipitation at regional spatial and short temporal scales most relevant for decision making and as input into hydrological models, therefore, requires a more nuanced approach (Wagener et al. 2010).

Downscaling has become a popular technique for exploring the relationship between local-scale climate change and synoptic-scale climate forcing (Hewitson and Crane 1992b,c, 1996, 2002; Wood et al. 2002). For example, Hewitson and Crane (1992a) use a technique based on an artificial neural network to demonstrate that the local precipitation variability in southern Mexico resulted from changes in the near-surface and 500-hPa circulation fields. Wilby and Wigley (1997) describe four categories of downscaling techniques: regression methods; weather-pattern-based approaches; stochastic weather generators, which belong to statistical downscaling; and limited-area modeling, generally referred to as dynamic downscaling. Downscaling techniques have continued to evolve and their use has matured since the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report (Houghton et al. 2001). In the fourth IPCC report, Christensen et al. (2007) evaluate many downscaling methods over different regions of the world, and they conclude that downscaling is an effective way to enhance the regional climate details of the AOGCM-simulated data.

With respect to the mid-Atlantic and Northeast regions of the United States, Crane and Hewitson (1998) apply artificial neural networks and find that anthropogenic greenhouse gas forcing leads to changes in storm-track and humidity fields over eastern North America, which, in an early version of the Goddard Institute for Space Studies (GISS) model, resulted in a substantial increase in spring and summer rainfall. More recent high-resolution projections of future climate change across the northeastern United States, using IPCC emission scenarios combined with both statistical and dynamical downscaling suggest temperature increases, especially at higher latitudes and inland, as well as potential precipitation pattern changes (Hayhoe et al. 2006). According to Christensen et al. (2007), annual mean precipitation is very likely to increase in Canada and the northeastern United States, and it is likely to decrease in the southwestern United States. They also indicate that over the mid-Atlantic region, most GCMs agree on increases of annual and winter mean precipitation, while for summer only about half of the GCMs predict increases. Climate change impacts over the mid-Atlantic region identified in these studies are influenced by several key processes, including midlatitude cyclones, ENSO, and the North Atlantic Oscillation (NAO)/Atlantic Oscillation (AO).

The ridge and valley province of the Appalachian Mountains dominate a large part of Pennsylvania, and statistical downscaling has proven to perform well at downscaling local temperature and precipitation in similar regions of high topographic variability (Benestad 2005; Hanssen-Bauer et al. 2005; Hewitson and Crane 2006).

In many cases it is necessary to propagate climate change projections of meteorological variables, such as precipitation, through hydrological models to yield the variables of interest (e.g., streamflow or soil moisture). Hydrologic models have a much finer resolution than GCMs, and the GCMs' output generally has to be downscaled before becoming useful for these models. Wood et al. (2004) use three different statistical downscaling methods: linear interpolation, spatial disaggregation, and bias correction and spatial disaggregation (BCSD). Each of these methods are applied to both the Parallel Climate Model (PCM) and regional climate model (RCM) to downscale climate model output to drive the Variable Infiltration Capacity (VIC) model at a $1/8^\circ$ spatial resolution (Liang et al. 1996, 1999). They compared the results from the hydrological model driven by the downscaled data from the three other approaches, and they found that the BCSD methods successfully reproduce the main features of the observed hydrometeorology from the retrospective climate. Maurer (2007) also shows that winter streamflow over California will increase, while late spring and summer flow will decrease based on the VIC model driven by downscaled climate change projections from 11 GCMs under both the higher-emission Special Report on Emissions Scenarios (SRES) A2 scenario and lower-emission SRES B1 scenario.

Self-organizing maps (SOMs) represent a nonlinear technique that supports the analysis of variability in large multivariate and multidimensional datasets through the derivation of a spatially organized set of generalized patterns of variability from the data (Reusch et al. 2007). Cavazos (1999) uses SOMs to examine the relationships between large-scale circulation-humidity fields and local daily precipitation events in northeastern Mexico and southeastern Texas. SOMs are applied to combine the precipitation records of individual stations into a regional dataset by Crane and Hewitson (2003). Hewitson and Crane (2006) use SOMs to downscale synoptically controlled daily precipitation over South Africa, while Reusch and Alley (2007) find that SOM-based patterns concisely capture the spatial and temporal variability in monthly Antarctic sea ice edge position data through the examination of area anomalies of Antarctic sea ice coverage.

In this paper, we use the SOM-based downscaling methodology introduced by Hewitson and Crane (2006) to reproduce historical daily precipitation observations

for stations over Pennsylvania (United States). We evaluate how well GCMs reproduce the observed synoptic-scale atmospheric conditions and assess their usefulness in projecting current precipitation variability over the region.

2. Data and methodology

a. Data

In this study, we use three sets of data for the down-scaling procedure: National Centers for Environmental Prediction (NCEP) reanalysis of daily gridded atmospheric data, observed daily station precipitation data, and GCM daily gridded atmospheric data. The daily gridded atmospheric data are constructed from 6-hourly NCEP reanalysis data from 1979 to 2007 with a resolution of $2.5^\circ \times 2.5^\circ$.

The SOM procedure uses seven variables: u and v components of the wind at 10 m and 700 hPa, respectively; relative humidity at 850 hPa; air temperature anomaly at 10 m; and the lapse rate of temperature from 850 to 500 hPa. All seven variables are physically related to the local precipitation. The u and v components of the wind determine low-level convergence and divergence, while

TABLE 1. Locations and elevations of the 17 stations over Pennsylvania. Here and in subsequent tables, ID indicates identifier number.

Station ID	Station name	Lat (°N)	Lon (°W)	Elev (m)
360106	Allentown	40.66	-75.44	118
363054	Chambersburg	39.94	-77.64	195
363028	Franklin	41.39	-79.82	302
363526	Greenville	41.42	-80.37	344
363699	Harrisburg	40.22	-76.85	103
364385	Johnstown	40.34	-78.92	370
365915	Montrose	41.84	-75.87	475
366233	New Castle	41.02	-80.37	251
366689	Palmerton	40.80	-75.62	125
367477	Ridgway	41.42	-78.75	414
368449	State College	40.80	-77.87	357
368596	Stroudsburg	41.01	-75.19	146
368905	Towanda	41.76	-76.42	229
369050	Uniontown	39.92	-79.72	291
369298	Warren	41.86	-79.16	369
369464	West Chester	39.97	-74.64	137
369933	York	39.92	-76.75	119

water vapor content of the lower atmosphere relates to relative humidity and surface temperature. The 850–500-hPa lapse rate determines whether the initial conditions for convection are met.

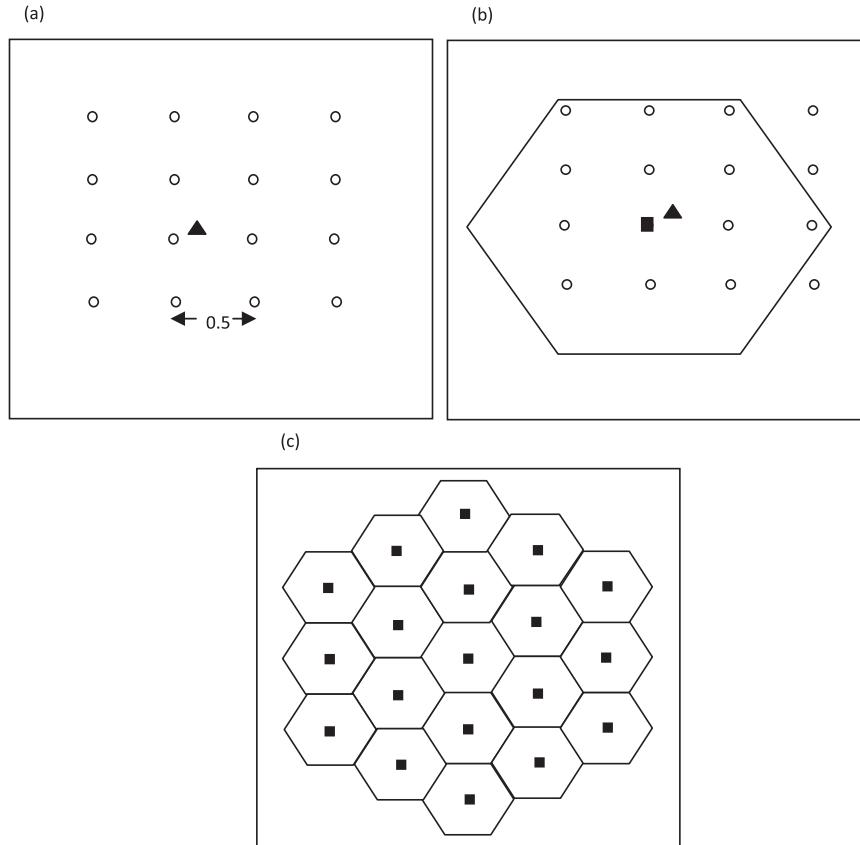


FIG. 1. The SOMs preprocessing (triangle: target location).

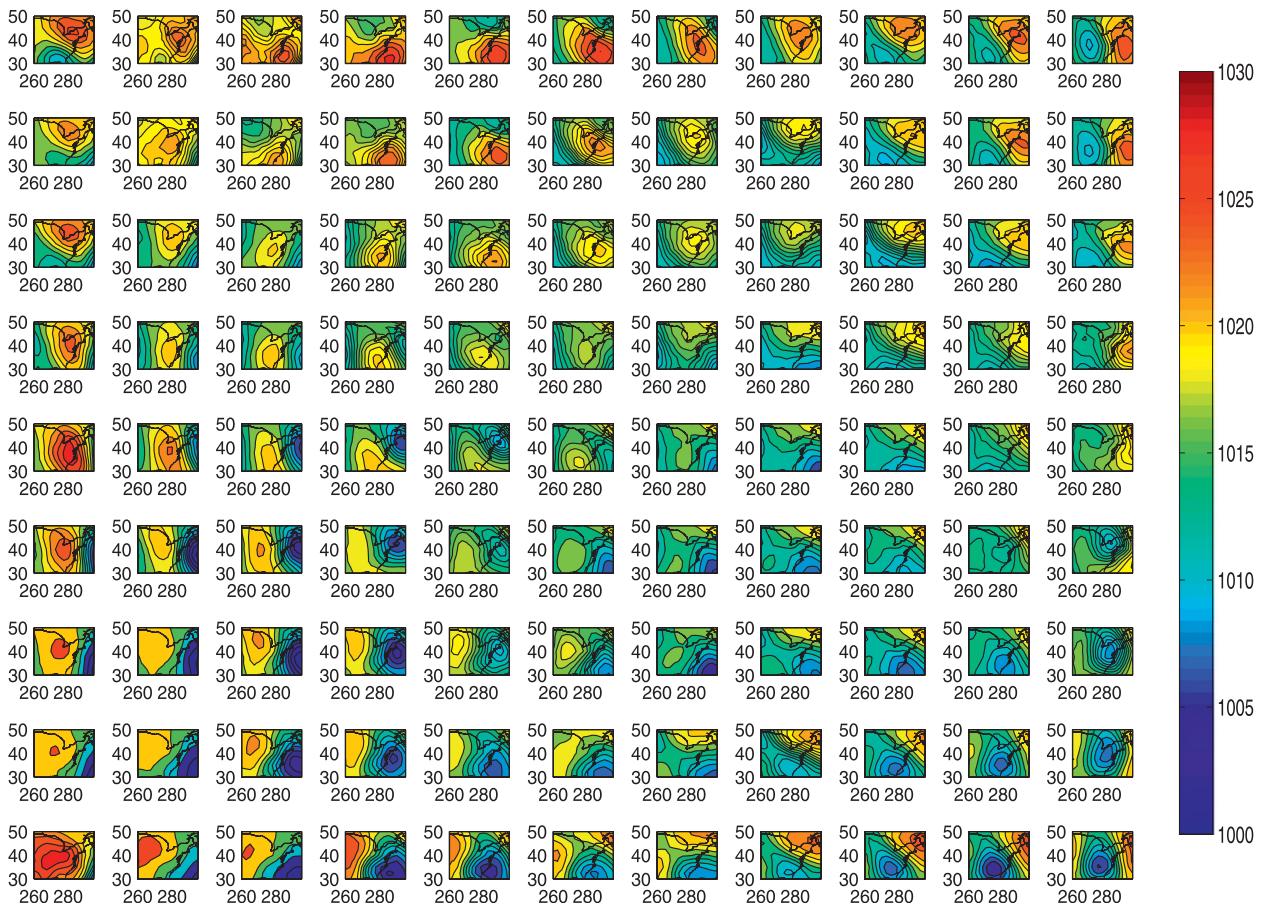


FIG. 2. SLP distribution corresponding with 99 SOM nodes (hPa).

Detailed comparisons (B. C. Hewitson 2009, personal communication) of the present SOM-based statistical downscaling to the application of climate regimes in Africa have shown that the additional use of specific humidity makes little difference to predictions of rainfall during the historical period but leads to predictions of greater increases in rainfall in response to future anthropogenic warming. Comparisons with dynamically downscaled estimates based on regional climate models suggest that these larger precipitation increases represent overestimates. The effects of using the additional humidity parameter will be discussed later in the context of our own downscaling results.

We limit our analysis to the post-1979 NCEP data when satellite observations improve the quality and climatological continuity of the product (Sturaro 2003; Tennant 2004). The precipitation data for Pennsylvania come from 17 stations for the period 1961–2005. The names, locations, and elevations of the stations are given in Table 1. The GCM circulation data from 1961 to 2000 are taken from the twentieth-century simulation using historical greenhouse gas concentrations [Twentieth-Century

Climate in Coupled Model (20C3M) scenario] of the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project, phase 3 (CMIP3), for nine different models: Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (CGCM3.1); Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3 (CNRM-CM3); Commonwealth Scientific and Industrial Research Organisation, Mark 3.0 (CSIRO Mk3.0); Geophysical Fluid Dynamics Laboratory Climate Model, version 2.0 (GFDL CM2.0); Goddard Institute for Space Studies Model E-R (GISS-ER); L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4); Meteorological Institute of the University of Bonn, ECHO-G Model (MIUBECHOG); Max Planck Institute (MPI) ECHAM5; and Meteorological Research Institute Coupled General Circulation Model, version 2.3.2a (MRI CGCM2.3.2a). The data and descriptions of the GCMs can be found at the WCRP CMIP3 Multi-Model Data Web site (<https://esg.llnl.gov:8443/index.jsp>). The variables simulated by the GCM data are the same as those for the NCEP data. In the

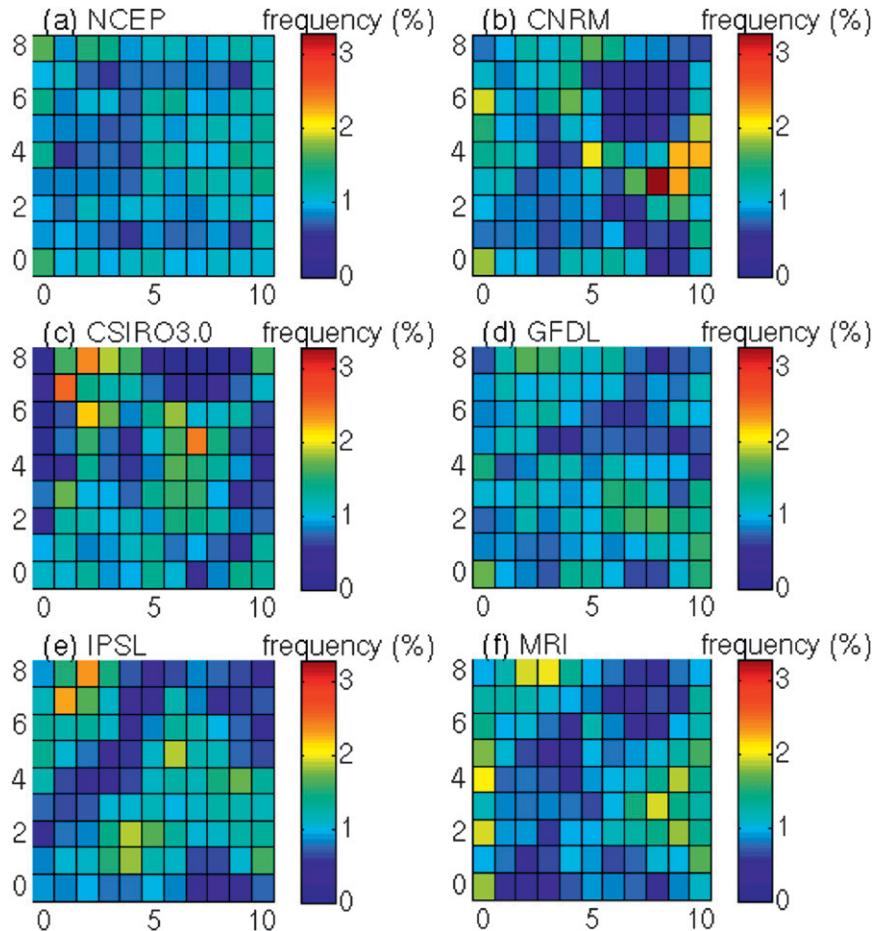


FIG. 3. Frequency distributions across the SOM nodes for atmospheric circulation from (a) NCEP and models (b) CNRM, (c) CSIRO, (d) GFDL, (e) IPSL, and (f) MPI centered on 40.0°N, 76.5°W (%).

assessment of the downscaled product, we also use daily NCEP sea level pressure (SLP) from 1979 to 2007, and the precipitation rate of the nine GCMs from 1961 to 2000.

b. The downscaling procedure

The first step in the downscaling procedure used here involves training the SOMs. SOMs are analogous to a fuzzy-clustering algorithm and are usually used to visualize and characterize multivariate data distributions (Kohonen 1989, 1995). A SOM is typically depicted as a two-dimensional array of nodes, where each node is described by a vector representing the average of the surrounding points in the original data space. For an input dataset that is described by a matrix of n variable data points and m observations, each node in the SOM is described by a reference vector having length n . The initial step in the SOM training involves assigning random values to each node reference vector and then

comparing the data record with each node vector. The reference vector that most closely matches the data vector is defined as the “winning” node. Then, the reference vector of the winning node is updated slightly toward the direction of the input data by a factor termed the “learning rate.” All the surrounding nodes are also updated in the direction of the input data by a smaller learning rate. The entire process is then repeated for multiple iterations until the differences between iterations are smaller than a selected threshold value. This training procedure is described in detail in Crane and Hewitson (2003) and illustrated in Fig. 1 therein.

In our application, a separate SOM with $9 \times 11 = 99$ nodes is trained for each station, with each node representing a characteristic atmospheric state. The choice of SOM size is ultimately subjective—fewer nodes increases generalization, while an increased number of nodes results in too few days being mapped to each node to derive a representative rainfall distribution function.

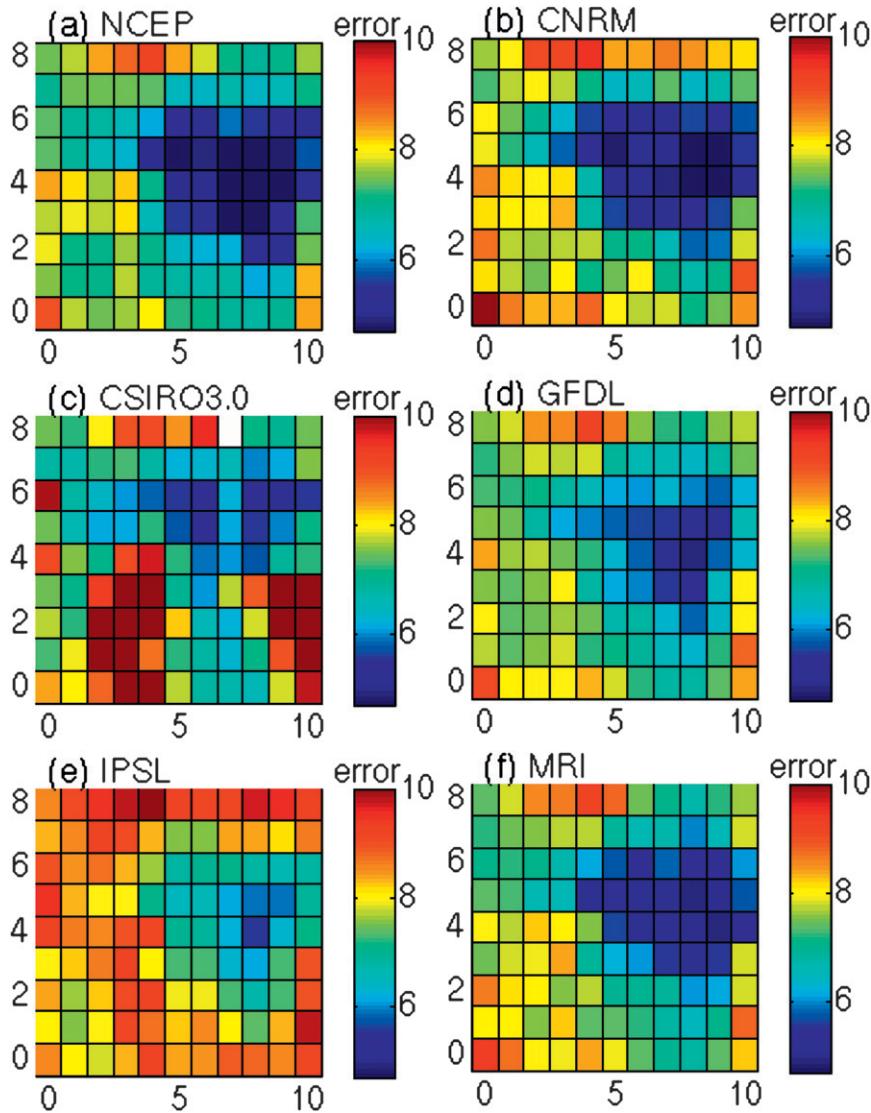


FIG. 4. Average quantization error distributions across the SOM nodes for atmospheric circulation from (a) NCEP and models (b) CNRM, (c) CSIRO3.0, (d) GFDL, (e) IPSL, and (f) MPI centered on 40.0°N , 76.5°W .

However, statistical validation experiments, as described later, can be used to assess the sensitivity of the results to this precise choice.

Prior to the training step, the study area was divided into a grid with a resolution of 0.5 degrees (Fig. 1a), and for any given target station, the nearest cell is identified (Fig. 1b). For example, the grid cell centered on 40.0°N , 76.5°W is the nearest cell for Harrisburg (40.22°N , 76.85°W). For each of the seven variables, 19 hexagonal grids are created, with the grid cell being in the center. For each hexagonal grid, four NCEP data points surrounding each of the six triangular centroids are extracted and regridded to the centroid through a weight inversely proportional to the distance between the

NCEP data point and the hexagonal centroid. The value over each of the 19 hexagonal grids is then calculated through averaging the values of six triangular centroids (Fig. 1c). Finally, all the NCEP data are separately standardized. Thus, for each node, we use seven variables, creating a 19×7 -member vector to describe each day's atmospheric state around the station.

For each station, we compare the observed daily atmospheric data to the SOM nodes and map each day to one particular node. For each SOM node, we take all the days that map to that particular node and then rank the precipitation on those days from low to high. A spline is fit to the ranked precipitation data to define a continuous cumulative distribution function (CDF)

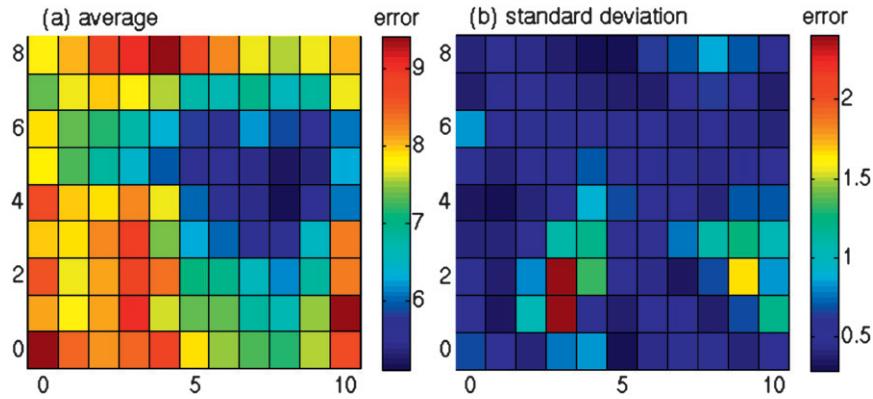


FIG. 5. The (a) average and (b) standard deviation of the averaged quantization errors across the SOM nodes for NCEP and nine GCMs' circulation data centered on 40.0°N, 76.5°W.

of the node's rainfall. This procedure is repeated for all the nodes in the SOM and then for all the stations.

To downscale the precipitation for a given station over a particular period, we first compare the circulation data (either from the GCMs or from the observations) to the SOM, associating each day with a node. For each day, a random number generator is used to select a value of precipitation from the CDF for the node to which the day is mapped. The procedure is repeated many times to produce an ensemble of time series for each station, any one of which can be considered a representative sample of the distribution characterizing the downscaled dataset. We chose to produce ensembles

consisting of 1500 realizations for the purpose of our downscaling applications.

3. Results

There are a large number of empirical downscaling methodologies currently being applied to regional climate datasets. However, many simply describe the application of a technique to a single region using one or maybe two GCMs to drive the downscaling and to compare the results to observed monthly rainfall. Here we attempt to go further by downscaling all available GCMs from the CMIP3 archive that have the required daily

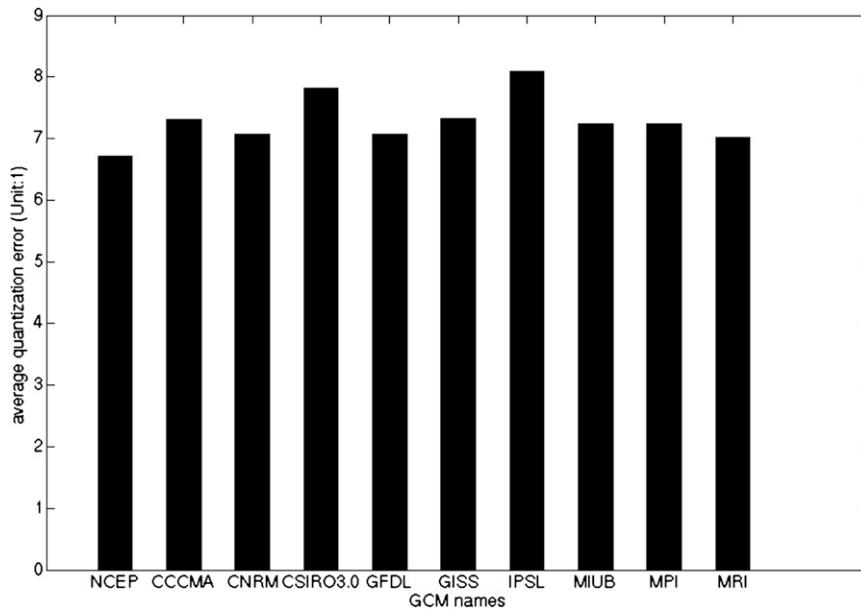


FIG. 6. The average values of the averaged quantization errors over all the SOM nodes for NCEP and nine GCMs' circulation data centered on 40.0°N, 76.5°W.

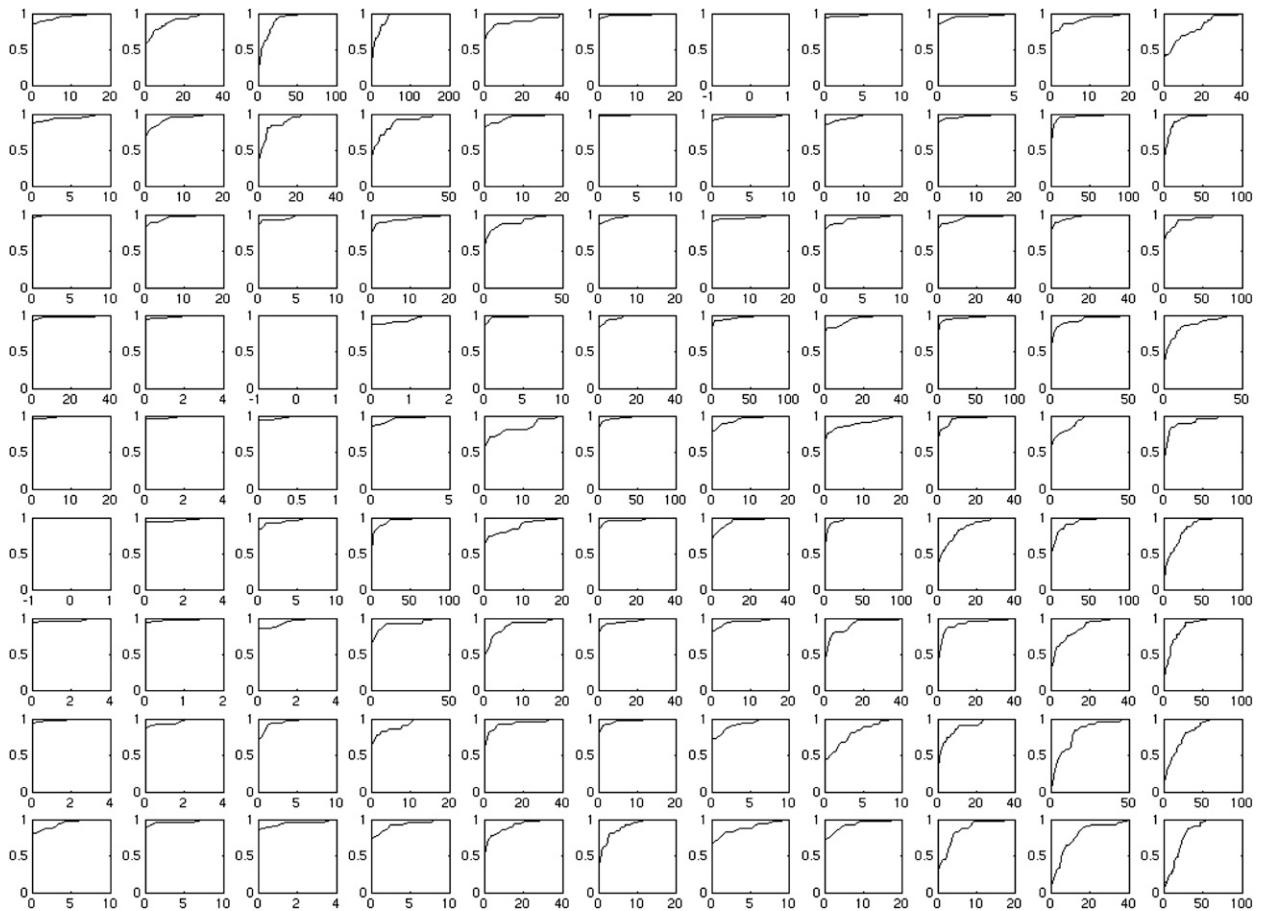


FIG. 7. The CDFs of daily precipitation values corresponding to 99 SOM nodes (x -axis; mm).

parameters. Furthermore, we focus on validation using a variety of measures to achieve a more robust downscaling that also reflects the uncertainty resulting from the variations in the original GCM simulations.

a. Synoptic controls on precipitation

The following basic assumptions are behind our downscaling approach: (i) precipitation at each of the stations will vary as a function of the atmospheric state, (ii) the NCEP variables can adequately describe that state, and (iii) they are to some degree a function of the larger-scale atmospheric state. One appropriate diagnostic of that state is the SLP field. This motivates as an internal consistency check on the validity of our underlying assumptions, assessing how well the nodes of the trained SOM reflect differences in the synoptic state of the atmosphere as indicated by their projection onto the SLP field—a field that was not used to define (train) the SOM.

The projection of the 99 SOM nodes onto the SLP field for the Harrisburg (40.0°N, 76.5°W) site is shown in Fig. 2. This site is chosen because of its central location, but results

are similar for all sites. Each projection of a given node is defined by the average of the SLP field over all days mapped to that node. The figure demonstrates that, although the SLP data are not directly used in training the SOM, the SLP distributions are clearly well differentiated by the SOM nodes: similar patterns locate close to each other in the SOM space, while different patterns locate farther apart. High pressure dominates the nodes to the top and left of the SOM space, while low pressure dominates in the bottom right, with transitional nodes in between.

b. GCM validation

A basic method for evaluating a particular GCM's usefulness in assessing climate change is to test the model's ability to simulate present climate (including variability and extremes). The differences between simulations and observations should be considered insignificant if they are within unpredictable internal variability, have expected differences in forcing, or have uncertainties in the observed fields (Randall et al. 2007). In the present case, we can compare the simulations to

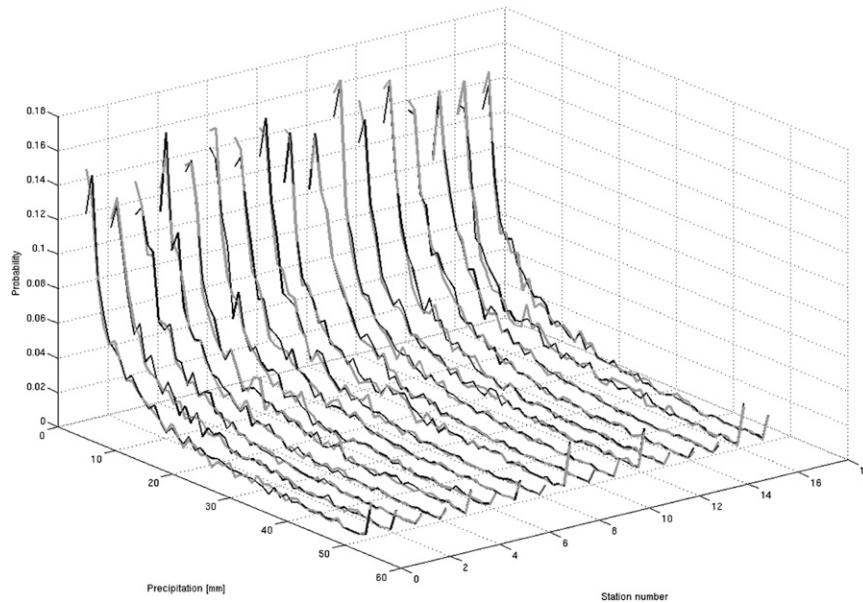


FIG. 8. The probability distributions of observed (black) and downscaled (gray) daily precipitation over 17 stations in Pennsylvania during the period 1979–2005.

observations (NCEP) by mapping the GCM fields to the trained SOM and comparing the results with the NCEP mapping, providing the means for assessing how well the various GCMs reproduce the atmospheric states used to differentiate characteristic rainfall distributions.

As an example, we still consider the Harrisburg (40.0°N, 76.5°W) site (Fig. 3). In Fig. 3, each square represents one node in the SOM. The frequency of a node is equal to the number of the days mapped to this node as a percentage of the number of all the days used in the SOM training. Figure 3a shows the frequency of days mapped to each of the 99 SOM nodes for the NCEP data for the period 1979–2007, while Figs. 3b–f shows the mapping of five GCMs for the period 1961–2000: CNRM, CSIRO, GFDL, IPSL, and MRI. These five models span the range of quantization errors (discussed below) encountered among the full set of GCMs analyzed. Mapping the GCM data onto the SOM trained by the NCEP data shows how well the GCMs reproduce the atmospheric states used to differentiate characteristic rainfall distributions.

The NCEP frequency distribution (Fig. 3a) shows that the frequencies are fairly uniformly distributed across all the nodes, with slightly larger frequencies located at the edges and corners. The GCMs also show a fairly uniform distribution across all nodes, although several of the models show centers of higher frequencies that are not present in the NCEP distributions. CSIRO Mk3.0 and IPSL CM 4 in particular, the two models with the largest quantization errors, show a concentration of variance in fewer nodes. These centers of higher frequencies in

some of the GCM mappings suggest slightly reduced variance in the model fields, but the distribution across the nodes shows that the models do recreate the atmospheric states revealed in the NCEP data, indicating that the GCMs produce realistic synoptic-scale patterns and variability across the region.

The average quantization errors (Fig. 4) measure how well the NCEP or five GCMs circulation fields map onto the available nodes. The quantization error is the mean difference between the node vector and each of the days that map to the node. It shows how closely the days are clustered in the data space or, alternatively, how much of the data space is represented by that node. By analogy with cluster analysis, the quantization error represents the within-group variability. The quantization error of one day is defined as the smallest Euclidean distance between the input vector and its best-matching node when mapping that day to one of the SOM nodes, and it is the measure of how the node reference vector represents the mapped atmospheric vector. Figure 4a shows that the quantization errors over the left side of the SOM are larger than those over the center and right side, indicating that the variances are larger for the synoptic states dominated by high pressure systems (Fig. 2).

The average quantization error distributions of the GCMs (Figs. 4b–f) are very similar to the corresponding distribution for the NCEP fields, although with slightly higher error values in some cases. The blank node (7, 8) in the distribution of model CSIRO Mk3.0 (Fig. 4c) indicates that no single day maps to that node. This

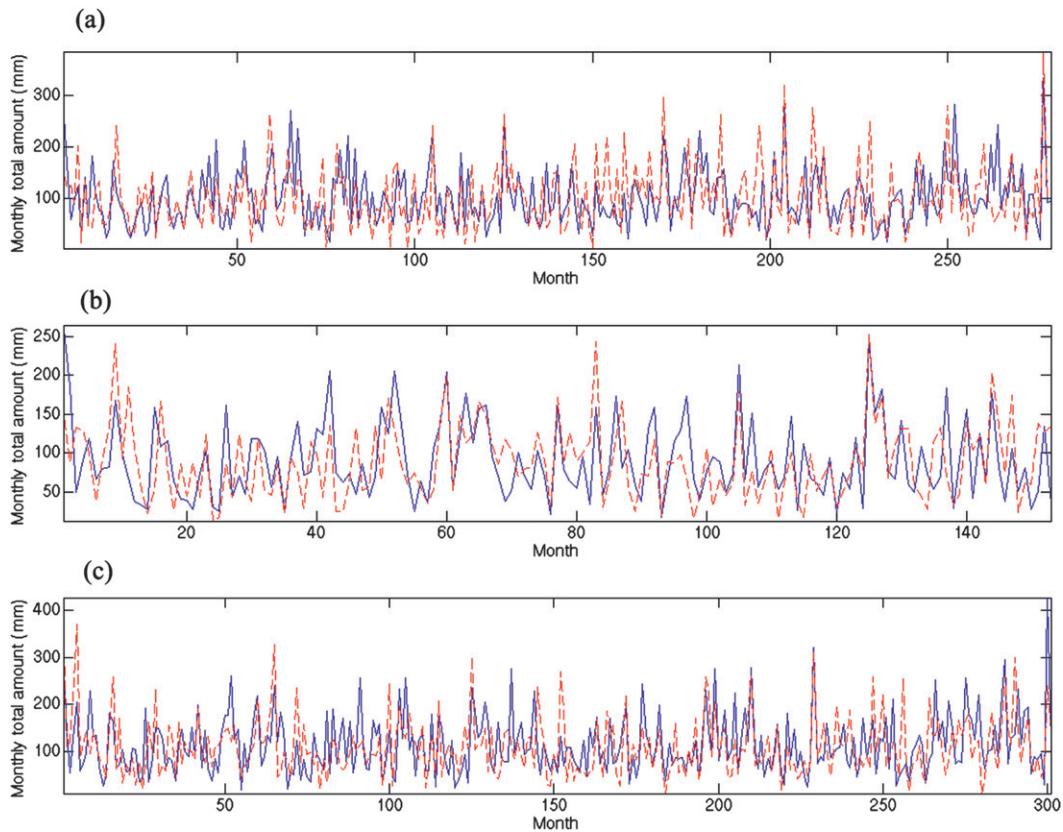


FIG. 9. Observed (blue) and downscaled (red) monthly precipitation amount time series for period 1979–2005 over stations (a) Allentown, (b) Harrisburg, and (c) Towanda (mm).

observation suggests that although the GCMs may have reduced dimensionality compared to the observed data, there is, in some cases, greater variability within the synoptic states mapped to each node.

Recall that in the training of the SOM, each daily synoptic circulation state is treated as a location in the original multidimensional state space, and that similar circulation states locate close to each other, while those states with large differences locate farther apart. Combining Figs. 3 and 4, it can be concluded that the total volume in state space of the NCEP data is larger than the volume of the GCMs' state spaces; however, around separate characteristic synoptic atmospheric states, the distance among the daily synoptic states in each of the GCMs is usually larger than the distance in NCEP state space.

The average values and standard deviations of the averaged quantization errors through NCEP and nine GCMs are given in Fig. 5. The pattern of average values of the averaged quantization errors is similar to the pattern of averaged quantization errors seen with the NCEP data, with larger errors over the left-hand side of the SOM and smaller errors over the right-hand side

(Fig. 5a). The additional four GCMs not shown in Fig. 4 thus have similar quantization error distributions to the five GCMs that are shown. It can be concluded that the distribution of variance of the synoptic circulation patterns in the GCMs are broadly similar to those for the NCEP data. The standard deviations of the mean quantization errors (Fig. 5b) show that the variability is almost uniform over all the nodes. The average of all GCMs gives a close match to the observations (NCEP). This was also demonstrated for the Pennsylvania region by Shortle et al. (2009), who show that the present-day mean temperature and rainfall simulations for all GCMs in the CMIP3 archive provide a closer match to the observations than any individual model.

As a simple measure of GCM performance in reproducing the observed atmospheric states, we average all the mean quantization errors over the 99 nodes individually for the NCEP data and for each of the GCMs (Fig. 6). The average error for the NCEP data is about 7, and the average error from most of the GCMs is also close to 7, with slightly larger errors from models CSIRO Mk3.0, and IPSL CM4. This suggests that overall these two models are slightly less accurate in

TABLE 2. Ratio of mean-square errors (MSEs) (relative to the observed daily precipitation values) using mean downscaled estimates and climatological mean values.

Station	Spring	Summer	Autumn	Winter	Annual
Allentown	0.72	0.90	0.77	0.70	0.78
Chambersburg	0.77	0.87	0.78	0.72	0.78
Franklin	0.82	0.88	0.83	0.79	0.84
Greenville	0.82	0.85	0.82	0.81	0.83
Harrisburg	0.75	0.91	0.76	0.71	0.78
Johnstown	0.79	0.88	0.81	0.84	0.84
Montrose	0.75	0.86	0.78	0.81	0.79
New Castle	0.83	0.88	0.86	0.80	0.85
Palmerton	0.78	0.89	0.78	0.70	0.80
Ridgway	0.81	0.86	0.81	0.79	0.83
State College	0.85	0.86	0.81	0.80	0.83
Stroudsburg	0.70	0.86	0.73	0.70	0.74
Towanda	0.81	0.87	0.81	0.81	0.82
Uniontown	0.83	0.88	0.78	0.84	0.84
Warren	0.80	0.85	0.81	0.80	0.82
West Chester	0.77	0.85	0.80	0.78	0.80
York	0.71	0.87	0.74	0.72	0.76
Avg	0.78	0.87	0.79	0.77	0.81
95% CI	(0.780, 0.788)	(0.870, 0.873)	(0.791, 0.796)	(0.769, 0.776)	(0.806, 0.811)

reproducing the observed distribution of synoptic-scale atmospheric states compared to the other seven models. Although it does not necessarily follow that the accuracy with which a particular GCM is able to simulate the present climate translates directly into its ability to project the future, using the inverse difference between the GCM and the NCEP quantization errors would be one possible approach to weighting the GCM output when looking at projected changes for future climates over the region based on a multimodel ensemble (e.g., the CMIP3 future climate change projections).

c. The downscaled precipitation

Figure 7 shows the calculated CDFs of daily precipitation values corresponding to the 99 SOM nodes. For the nodes located in regions of the SOM dominated by high surface pressure, most of the CDFs show low or zero precipitation amounts. As would be expected, for nodes located in regions of the SOM dominated by low surface pressure and certain transitional surface pressure patterns, precipitation amounts are much higher. The differences in the CDFs across the SOM indicate

TABLE 3. Ratio of widths (as defined by interquartile range) of downscaled NCEP vs observed climatological precipitation distribution (days with nonzero precipitation only).

Station	Spring	Summer	Autumn	Winter	Annual
Allentown	1.35	0.46	0.89	0.80	0.91
Chambersburg	1.00	0.38	0.79	0.58	0.73
Franklin	0.77	0.41	0.66	0.67	0.68
Greenville	1.08	0.44	0.76	0.76	0.76
Harrisburg	1.07	0.29	0.79	0.79	0.82
Johnstown	0.91	0.51	0.87	0.94	0.80
Montrose	1.02	0.57	0.82	0.82	0.87
New Castle	0.89	0.37	0.75	0.67	0.66
Palmerton	1.12	0.31	0.74	0.61	0.71
Ridgway	0.85	0.53	0.69	0.73	0.75
State College	0.82	0.40	0.76	0.63	0.70
Stroudsburg	1.50	0.59	0.90	0.68	0.92
Towanda	0.97	0.42	0.67	0.53	0.67
Uniontown	1.10	0.47	0.77	0.72	0.82
Warren	0.92	0.50	0.75	0.95	0.79
West Chester	1.14	0.60	0.71	0.53	0.80
York	1.02	0.58	0.99	0.63	0.84
Avg	1.03	0.46	0.78	0.71	0.78
95% CI	(1.012, 1.045)	(0.452, 0.470)	(0.776, 0.791)	(0.692, 0.718)	(0.768, 0.783)

TABLE 4. Comparisons between observed and NCEP downscaled average monthly precipitation amounts, average monthly number of rainy days, and standard deviations of monthly precipitation over the 17 stations during the period 1979–2005.

Station ID	OA ^a	DA ^b	ON ^c	DN ^d	OSD ^e	DSD ^f
Allentown	101.49	96.73	9.61	9.69	55.96	57.75
Chambersburg	93.78	87.08	9.37	9.13	49.47	46.27
Franklin	103.76	101.82	11.87	12.23	46.93	48.75
Greenville	98.03	94.34	12.38	12.31	44.27	42.31
Harrisburg	90.63	87.51	9.77	9.67	50.57	46.81
Johnstown	105.91	103.28	12.71	12.71	50.92	48.01
Montrose	110.56	106.65	12.36	12.67	53.58	52.62
New Castle	86.09	84.72	10.88	11.02	44.16	45.62
Palmerton	91.99	87.79	8.02	8.19	55.10	56.36
Ridgway	101.91	98.97	12.40	12.23	44.81	44.51
State College	94.72	91.59	10.99	11.03	49.73	47.02
Stroudsburg	113.49	108.23	10.36	10.29	62.49	58.95
Towanda	80.26	79.11	9.78	9.69	43.58	43.78
Uniontown	97.52	95.38	11.80	11.91	44.97	43.74
Warren	107.59	105.34	13.37	13.26	44.35	45.75
West Chester	106.37	100.52	9.50	9.41	59.03	54.68
York	95.57	92.59	9.74	9.69	51.79	49.86
Avg	98.80	95.39	10.88	10.89	50.10	48.99
95% CI	(98.07, 99.60)	(94.69, 96.06)	(10.72, 11.06)	(10.78, 11.06)	(49.33, 50.47)	(48.44, 49.40)

^a OA is observed average monthly precipitation amounts (mm).

^b DA is downscaled average monthly precipitation amounts (mm).

^c ON is observed average monthly number of rainy days (day).

^d DN is downscaled average monthly number of rainy days (day).

^e OSD is observed standard deviation of monthly precipitation amounts (mm).

^f DSD is downscaled standard deviation of monthly precipitation amounts (mm).

that the SOM categorization of atmospheric conditions does allow for substantial differentiation between different precipitation states and that these differences make physical sense in the context of the synoptic-scale circulation.

The previous discussion clearly demonstrates that the SOM characterization of atmospheric states captures the synoptic variability, that the different atmospheric states (represented by the SOM nodes) have different precipitation characteristics, and that the GCMs exhibit the same atmospheric states and synoptic variability. The final step in the validation procedure is to examine whether the trained SOM can be used to generate realistic precipitation time series that have the same magnitude and frequency characteristics as the observed data.

The SOM approach to downscaling precipitation acknowledges that similar atmospheric conditions can result in different observed precipitation amounts. By randomly selecting from the rainfall CDF for each node, the approach captures some of this stochastic variability. Because the downscaling is a simplification of reality, and because of its stochastic element, any individual recreation of the precipitation represents only one possible realization of the precipitation regime, which should match in its fundamental statistical attributes, the observed precipitation, while not necessarily

matching the observed time series on a day-to-day basis. To be a valid and useful representation of actual precipitation, the downscaling needs to match the characteristics needed for, for example, hydrologic modeling; that is, the downscaled precipitation should exhibit the same monthly and seasonal precipitation amounts, the same day-to-day variability, and the same number of rain days per month.

In some respects, validation is simply a matter of examining these characteristic statistics and assessing whether the results are good enough for a particular application. “Good enough,” of course, is a subjective decision that depends on the particular application. In this case, and to demonstrate that the downscaling may have broad application, we also ask, does the downscaling give a better result than just using climatology? And, is the downscaling a significant improvement over using the nearest GCM gridcell data? We seek to demonstrate that the downscaled data are a close match to the observations, and that the downscaling gives an appreciable improvement over using the GCM precipitation field directly.

The first step in validating statistical downscaling is to compare the statistical properties of the downscaled time series generated by the reanalysis fields with those of the corresponding observations. Figure 8 compares

TABLE 5. Comparisons of average monthly precipitation amount, average monthly number of rainy days, and standard deviations of monthly precipitation amount for observed, downscaled GCM, and raw GCM precipitation for all nine GCMs, averaged over stations (period 1961–2000, all months). The 95% confidence intervals are given in parentheses.

GCM names	DA ^a	DN ^b	DSD ^c	SA ^d	SN ^e	SSD ^f
Obs ^g	96.87 (96.09, 97.68)	10.89 (10.77, 11.00)	48.58 (48.15, 49.14)	96.87 (96.09, 97.68)	10.89 (10.77, 11.00)	48.58 (48.15, 49.14)
CGCM3.1 ^g	98.07 (97.08, 98.93)	11.12 (10.98, 11.22)	50.95 (50.14, 51.56)	90.89 (90.28, 91.55)	21.64 (21.59, 21.72)	36.08 (35.80, 36.48)
CNRM-CM3 ^g	106.06 (104.87, 107.16)	11.52 (11.39, 11.63)	55.93 (54.97, 56.71)	105.17 (104.88, 106.74)	21.22 (21.12, 21.29)	55.84 (55.68, 57.04)
CSIRO Mk3.0 3.0 ^g	99.06 (98.05, 99.94)	11.50 (11.36, 11.64)	52.94 (51.51, 53.48)	78.71 (78.30, 79.19)	18.09 (17.78, 18.44)	32.29 (31.97, 32.63)
GFDL CM2.0 ^g	103.66 (102.70, 104.57)	11.50 (11.37, 11.61)	52.31 (51.04, 52.72)	99.92 (99.51, 100.18)	19.15 (19.01, 19.34)	41.07 (40.84, 41.18)
GISS-ER ^g	115.69 (114.18, 116.93)	12.17 (12.05, 12.28)	60.22 (58.94, 61.19)	52.86 (52.35, 53.17)	15.75 (15.73, 15.77)	40.33 (39.82, 40.64)
IPSL CM4 ^g	101.30 (100.13, 102.24)	11.28 (11.16, 11.40)	55.01 (53.55, 55.52)	120.35 (119.45, 121.18)	19.23 (19.18, 19.32)	50.17 (49.80, 50.62)
MIUBECHOG ^g	104.40 (103.68, 105.56)	11.19 (11.06, 11.29)	53.55 (52.76, 54.13)	95.68 (95.22, 96.09)	22.21 (22.17, 22.27)	37.68 (37.37, 37.99)
MPI ECHAM5 ^g	100.11 (99.30, 100.99)	11.21 (11.09, 11.32)	52.14 (51.45, 52.59)	106.12 (105.40, 106.63)	15.43 (15.28, 15.61)	51.93 (51.81, 52.11)
MRI CGCM2.3.2a ^g	102.89 (102.16, 104.22)	11.26 (11.13, 11.35)	51.63 (50.76, 52.29)	72.32 (72.15, 72.52)	15.16 (15.14, 15.18)	34.01 (33.95, 34.08)
Avg ^h	103.47 (101.94, 104.15)	11.42 (11.32, 11.45)	53.85 (53.06, 54.22)	91.34 (87.71, 96.15)	18.65 (18.21, 19.09)	42.16 (40.45, 43.39)

^a DA is downscaled average monthly precipitation amounts (mm).

^b DN is downscaled average monthly number of rainy days (day).

^c DSD is downscaled standard deviation of monthly precipitation amounts (mm).

^d SA is raw simulated average monthly precipitation amounts (mm).

^e SN is raw simulated average monthly number of rainy days (day).

^f SSD is raw simulated standard deviation of monthly precipitation amounts (mm).

^g CIs are based on jackknife analysis with respect to the stations.

^h CIs are based on jackknife analysis with respect to the GCMs.

the probability distributions of the observed daily precipitation and the downscaled daily precipitation from one random iteration generated by the NCEP data over 17 stations for the period 1979–2005. In the calculation, only the days on which both observed and downscaled precipitation data are available are counted [note that only precipitation events larger than 0.25 mm (0.01 in.) are considered, consistent with the threshold for defining a “rain day” used in past work; e.g., Fitzpatrick and Krishnan 1967; Hershfield 1971; Gallus and Segal 2004]. The precipitation interval used in the calculation is 1 mm, and the probabilities of extreme precipitation larger than 50 mm are considered together. From Fig. 8 it can be concluded that, although there are some differences between the downscaled and observed probabilities of daily precipitation in the range from 0.25 to 1 mm, the downscaling reproduces the probability distributions extremely well. Moreover, the observed and downscaled probabilities of the largest precipitation events—with daily precipitation greater than 50 mm—are very close, which means that the downscaling, importantly, is also effective in capturing the extreme precipitation events for each station.

We can see that the downscaling captures the temporal variability of the actual observations (Fig. 9). While we generate many (1500) realizations of the downscaled precipitation to construct our ensemble, it is extremely unlikely that we will happen to reproduce the unique sequence of daily precipitation events that characterizes the actual observations. Nonetheless, it is instructive to see if we can find members of our ensemble that not only capture the overall statistical character of the observations but also happen to approximate well the observed sequence of monthly precipitation anomalies. Figure 9 shows the observed and downscaled monthly precipitation amount time series for three stations, with the largest correlation coefficients between the observed and downscaled monthly precipitation: Allentown (Fig. 9a), Harrisburg (Fig. 9b), and Towanda (Fig. 9c).

As one important measure of statistical skill, we evaluate whether the mean downscaled precipitation estimates using the NCEP data (rain days only) are on average closer to the observations than the climatological mean for that season (Table 2; results using median rather than mean of the 1500 downscaling realizations are provided in supplementary Table S1). Ratios less than unity indicate nominally better skill than the null “no skill” prediction of climatological mean values. We used a remove-one-sample-at-a-time jackknife procedure (Efron 1982) to provide nonparametric confidence intervals (CIs) in the mean skill over all 17 stations. If the associated upper 95% confidence limit remains below unity, we conclude that the downscaling procedure

TABLE 6. As in Table 5, but for all 17 stations averaged over the GCMs (period 1961–2000, all months). The 95% confidence intervals are given in parentheses.

Station ID	OA ^a	ON ^b	OSD ^c	DA ^d
Allentown ^j	99.48	9.52	52.90	107.03 (105.40, 107.80)
Chambersburg ^j	91.411	9.23	49.06	97.84 (96.18, 98.90)
Franklin ^j	01.06	11.79	45.12	106.56 (105.53, 107.26)
Greenville ^j	92.19	12.19	40.73	96.81 (95.51, 97.41)
Harrisburg ^j	92.981	9.72	54.95	96.20 (94.80, 96.96)
Johnstown ^j	06.521	12.43	53.13	110.99 (109.42, 111.78)
Montrose ^j	10.25	12.37	52.68	117.54 (116.35, 118.10)
New Castle ^j	83.99	11.18	39.68	86.94 (85.82, 87.60)
Palmerton ^j	92.201	8.52	50.87	98.43 (96.43, 99.54)
Ridgway ^j	00.71	12.81	43.02	105.11 (103.86, 105.80)
State College ^j	90.931	10.84	47.97	98.93 (97.53, 99.63)
Stroudsburg ^j	09.41	10.13	55.43	123.58 (121.33, 124.69)
Towanda ^j	79.81	9.99	42.32	85.13 (84.00, 85.92)
Uniontown ^j	94.611	11.68	44.60	102.15 (101.15, 102.71)
Warren ^j	05.921	13.63	42.42	110.37 (109.39, 111.03)
West Chester ^j	01.97	9.47	56.36	113.53 (110.24, 115.07)
York ^j	93.45	9.67	54.59	101.89 (100.09, 102.73)
Avg ^{g,k}	96.87 (96.04, 97.94)	10.89 (10.77, 11.00)	48.58 (48.09, 49.07)	103.47 (102.59, 104.62)

^a OA is observed average monthly precipitation amounts (mm).

^b ON is observed average monthly number of rainy days (day).

^c OSD is observed standard deviation of monthly precipitation amounts (mm).

^d DA is downscaled average monthly precipitation amounts (mm).

^e DN is downscaled average monthly number of rainy days (day).

^f DSD is downscaled standard deviation of monthly precipitation amounts (mm).

^g SA is raw GCMs simulated average monthly precipitation amounts (mm).

^h SN is raw GCMs simulated average monthly number of rainy days (day).

ⁱ SSD is raw GCMs simulated standard deviation of monthly precipitation amounts (mm).

^j CIs are based on jackknife analysis with respect to the GCMs.

^k CIs are based on jackknife analysis with respect to the stations.

yields a statistically significant improvement above the climatological no-skill baseline. The ratios for each season (and annual mean) are in fact observed to be significantly below unity, indicating, as we would hope, that the downscaling does perform better than simply invoking climatology. Among the four seasons, the least improvement occurs in summer, reflecting the fact that much of the summer precipitation is convective. In this case there is more of a stochastic element and less dependence on the synoptic circulation.

Continuing the process of skill evaluation, we would expect that if the downscaling is providing large-scale information that usefully informs the local distribution of precipitation, drawing from the precipitation CDFs for the appropriate SOM node should yield a narrower distribution than drawing simply from the climatological daily rainfall distribution for the appropriate season. In other words, accounting for the synoptic atmospheric state ought to provide some additional discrimination beyond a random draw from the climatological seasonal distribution. We define the width of the respective PDFs by the interquartile range (i.e., the difference between the 75th and 25th percentiles of the distributions). Table 3

tabulates a skill metric defined as the ratio of the mean squared width of the PDFs constructed from the ensemble of 1500 downscaled precipitation values and the observed climatological distribution. In this analysis, to get the more precise PDFs with larger amounts of samples, we use a different rainy day definition, one that is larger than 0 mm. A ratio below unity indicates that the downscaled values are drawn from a narrower PDF than would be the case for the corresponding climatological distributions, and it is therefore suggestive that conditioning on the large-scale atmospheric state via the downscaling procedure provides some additional predictive skill beyond climatology. We once again use a jackknife procedure to estimate confidence intervals and evaluate statistical significance. Apart from a number of the stations in spring (February–April), the ratios remain significantly below unity. The possible reason for the exception with the spring season results is that a large number of spring rain days are governed by synoptic circulation patterns with CDFs that have wide daily precipitation ranges.

We next compared the observed and NCEP downscaled precipitation fields with respect to several key measures: mean monthly precipitation, average monthly

TABLE 6. (*Extended*)

DN ^c	DSD ^f	SA ^g	SN ^h	SSD ⁱ
9.96 (9.91, 10.00)	62.61 (61.29, 63.08)	92.81 (87.70, 98.40)	17.45 (16.87, 18.05)	44.31 (42.54, 45.50)
9.68 (9.57, 9.74)	53.64 (52.68, 54.09)	94.02 (90.69, 98.22)	18.37 (17.91, 18.99)	44.86 (43.24, 46.35)
12.66 (12.58, 12.71)	49.37 (48.95, 49.82)	88.79 (86.27, 93.32)	19.82 (19.43, 20.45)	39.53 (37.21, 41.07)
12.35 (12.27, 12.41)	44.62 (44.03, 45.05)	87.79 (85.04, 92.87)	20.11 (17.73, 20.76)	38.35 (36.64, 39.74)
10.16 (10.11, 10.20)	51.48 (50.45, 51.92)	91.42 (87.48, 96.69)	17.85 (17.31, 18.39)	43.42 (41.55, 44.73)
13.47 (13.38, 13.51)	51.09 (49.91, 51.72)	91.62 (88.21, 96.85)	18.35 (17.91, 18.79)	43.16 (41.38, 44.40)
13.33 (13.22, 13.39)	58.11 (57.36, 58.74)	93.64 (88.50, 98.72)	18.95 (18.49, 19.40)	41.94 (39.89, 43.59)
11.46 (11.38, 11.52)	44.72 (43.84, 45.20)	88.24 (84.69, 93.28)	19.40 (18.98, 19.98)	39.82 (38.04, 41.04)
8.61 (8.53, 8.67)	63.48 (61.95, 64.28)	91.70 (86.63, 96.78)	17.20 (16.66, 17.79)	44.10 (42.38, 45.27)
12.60 (12.52, 12.66)	49.51 (49.00, 49.87)	89.06 (86.61, 93.59)	19.93 (19.57, 20.55)	39.06 (36.71, 40.60)
11.58 (11.47, 11.65)	52.67 (52.15, 52.93)	92.29 (90.14, 97.26)	19.08 (18.71, 19.59)	42.32 (41.33, 43.26)
10.94 (10.83, 10.98)	69.52 (68.13, 70.23)	91.56 (86.18, 96.39)	17.76 (17.12, 18.40)	42.95 (41.05, 44.23)
10.51 (10.40, 10.55)	46.73 (46.35, 47.09)	92.68 (87.35, 97.69)	19.35 (18.91, 19.85)	41.17 (39.04, 43.09)
12.53 (12.42, 12.58)	47.07 (46.68, 47.40)	97.28 (94.33, 101.85)	19.07 (18.67, 19.56)	45.45 (43.89, 47.01)
13.64 (13.57, 13.69)	47.99 (47.43, 48.39)	85.46 (82.74, 89.61)	19.83 (19.48, 20.33)	37.02 (35.52, 38.25)
10.14 (9.99, 10.21)	66.56 (65.08, 67.22)	91.30 (87.79, 96.27)	16.99 (16.46, 17.54)	44.00 (42.27, 45.24)
10.43 (10.30, 10.46)	56.33 (55.45, 56.67)	93.05 (89.10, 97.04)	17.58 (17.03, 18.11)	45.19 (43.54, 46.66)
11.42 (11.29, 11.52)	53.85 (53.06, 54.43)	91.34 (91.17, 91.70)	18.65 (18.57, 18.74)	42.16 (41.97, 42.48)

number of rain days, and standard deviations of monthly-mean precipitation (Table 4). The comparison was performed over the 17 stations for the period 1979–2005. For the forgoing discussion, we make use of a randomly selected, representative realization from the ensemble of 1500 downscaling surrogates, though similar results are obtained for any realization. The observed average monthly precipitation amounts vary from roughly 80 to 110 mm among the different sites, with an average of 98.8 mm. The downscaling slightly underestimates the mean precipitation at 95.4 mm. This underestimation bias is statistically significant according to the jackknife error estimates, and it appears to result from the fact that the downscaling underestimates the magnitude of the most intense daily precipitation events (larger than 50 mm). The observed and downscaled average monthly rain days are very close for all stations, with an identical average over all stations of 10.9 days for both downscaled and observed. The observed and downscaled standard deviations of monthly precipitation totals are also similar, with the 95% confidence intervals overlapping (albeit only just so), indicating that the downscaling procedure is able to reproduce the observed variability.

Having validated the downscaling procedure for the late-twentieth-century observations, we then turned to the GCM simulations. We compared the same characteristics of the precipitation field for the downscaled and raw GCM precipitation data over the full data record 1961–2000. Results were analyzed by both GCM (Table 5), averaging of the 17 stations, and station (Table 6), averaging over GCMs. Errors relative to observations for both downscaled and raw GCM precipitation values are compared in Tables 7 and 8, respectively,

while results for individual seasons are provided in the supplementary information (supplementary Tables S2–S9). Uncertainties in averages across the models are determined based on jackknifing with respect to the model, while uncertainties in averages across stations are determined by jackknifing with respect to the station.

These comparisons yield a number of important insights. First of all, the downscaled results are clearly closer to observations than the raw GCM results for nearly all the models and all the stations. This finding is also true for averages over the stations and averages over the models. Extremely large errors are found with the raw precipitation field for individual GCMs with respect to, for example, mean monthly precipitation (see Tables 5 and 7). However, because these errors are often of opposite sign in different models—that is, either considerably below or above the observed value—they tend to cancel. Averages over GCMs are consequently considerably closer to observations than individual GCMs. This finding is consistent with the widely reported finding (e.g., Meehl et al. 2007) that averages over multimodel ensembles often provide more faithful estimates than individual models, presumably because of the cancellation of errors specific to models. In this particular case, the errors in question likely involve the differing convective parameterization schemes used to estimate precipitation in the various models.

While the downscaled model precipitation field is closer in nearly all characteristics to observations than the models' raw precipitation field, it is worth noting that statistically significant biases nonetheless remain in the downscaled model estimates. Mean monthly rainfall totals are, as they were for the late-twentieth-century

TABLE 7. Absolute errors with respect to observations (expressed as percent) for the average monthly precipitation amounts, average monthly number of rainy days, and standard deviations of monthly precipitation amounts across the 17 stations for all the months during the period 1961–2000.

GCM names	DPA ^a	DPN ^b	DPSD ^c	SPA ^d	SPN ^e	SPSD ^f
CGCM3.1	2.12	3.30	8.08	10.48	101.52	25.46
CNRM-CM3	9.42	5.89	15.81	13.11	97.43	17.37
CSIRO Mk3.0	2.65	5.60	10.74	18.26	65.76	33.27
GFDL CM2.0	6.96	5.57	10.57	9.14	77.58	15.33
GISS-ER	19.38	12.03	23.55	45.02	47.07	17.68
IPSL CM4	5.19	4.18	14.83	25.13	79.39	6.35
MIUBECHOG	7.72	3.79	11.30	9.80	107.26	21.94
MPI ECHAM5	3.44	3.69	9.81	11.50	41.99	11.29
MRI CGCM2.3.2a	6.73	5.08	7.62	24.80	41.65	29.11
Avg across nine GCMs	7.07	5.46	12.48	18.58	73.29	19.75

^a DPA is downscaled percentages of absolute errors of average monthly precipitation amounts.

^b DPN is downscaled percentages of absolute errors of average monthly number of rainy days.

^c DPSD is downscaled percentages of absolute errors of standard deviations of monthly precipitation amounts.

^d SPA is raw simulated percentages of absolute errors of average monthly precipitation amounts.

^e SPN is raw simulated percentages of absolute errors of average monthly number of rainy days.

^f SPSD is raw simulated percentages of absolute errors of standard deviations of monthly precipitation amounts.

^g CIs are based on jackknife analysis with respect to the stations.

^h CIs are based on jackknife analysis with respect to the GCMs.

observations (Table 5), biased slightly high (the mean monthly downscaled precipitation over all models and stations is 103.47 mm, while the observed value is 96.87 mm). Rain day numbers are also biased slightly at 11.42 days per month averaged over all models and stations versus the observed value of 10.89 days per month—the difference is small but statistically significant. The monthly standard deviation of the downscaled precipitation field is 53.85 mm averaged over stations and GCMs, while it is 48.58 for the observations—a difference that is once again statistically significant and suggests that the downscaled GCM precipitation is slightly more variable in time than the observations.

Some of the differences between the downscaled GCM values are the result of model bias and differences in the simulated atmospheric states. However, the similarity of the downscaled data suggests that a large portion of the GCM differences is likely due to differences in precipitation parameterization schemes.

That such bias remains in the downscaled estimates is hardly surprising, as there are clearly systematic biases in the various fields of the model (temperature, winds, lapse rates, etc.) from which the downscaled precipitation estimates are derived, and not even downscaling methods

TABLE 8. Absolute errors with respect to observations (expressed as percent) for the average monthly precipitation amounts, average monthly number of rainy days, and standard deviations of monthly precipitation amounts across the nine GCMs for all the months during the period 1961–2000.

Station ID	DPA ^a	DPN ^b	DPSD ^c	SPA ^d	SPN ^e	SPSD ^f
Allentown	7.59	4.62	18.36	17.45	83.24	19.38
Chambersburg	7.53	4.92	9.33	19.21	99.02	15.03
Franklin	5.46	7.38	9.40	17.29	68.15	22.10
Greenville	5.06	2.53	9.54	17.48	64.99	18.91
Harrisburg	4.67	4.58	8.23	16.84	83.72	22.49
Johnstown	4.95	8.35	7.99	16.75	47.58	20.56
Montrose	6.61	7.81	10.30	21.00	53.25	23.74
New Castle	4.13	3.01	12.70	19.69	73.52	17.91
Palmerton	7.96	2.77	24.79	18.87	101.87	19.17
Ridgway	4.62	2.42	15.08	16.66	55.57	19.99
State College	8.79	6.84	9.79	18.19	75.96	17.41
Stroudsburg	12.95	8.02	25.43	21.44	75.30	23.60
Towanda	6.94	5.25	10.41	25.16	93.71	20.95
Uniontown	7.96	7.29	5.61	18.39	63.27	16.11
Warren	4.37	2.05	13.13	19.60	45.46	16.56
West Chester	11.50	7.08	18.09	16.36	79.50	22.51
York	9.03	7.87	3.93	15.51	81.89	19.42
Avg across 17 stations	7.07	5.46	12.48	18.58	73.29	19.75

^a DPA is downscaled percentages of absolute errors of average monthly precipitation amounts.

^b DPN is downscaled percentages of absolute errors of average monthly number of rainy days.

^c DPSD is downscaled percentages of absolute errors of standard deviations of monthly precipitation amounts.

^d SPA is raw simulated percentages of absolute errors of average monthly precipitation amounts.

^e SPN is raw simulated percentages of absolute errors of average monthly number of rainy days.

^f SPSD is raw simulated percentages of absolute errors of standard deviations of monthly precipitation amounts.

can cure these ills. However, there is far greater consistency among the downscaled model estimates with respect to all diagnostics (monthly-mean precipitation, rain days, and monthly standard deviation) than in the raw model precipitation values. The downscaling procedure appears to provide a more reliable determination of whether precipitation is likely, and when it is likely, how much an event will produce. These observations reinforce previous findings (e.g., Hewitson and Crane 2006) that bypassing the convective parameterization schemes in the models yields precipitation estimates that are in all key attributes more consistent among models, closer to observations, and likely more robust with respect to *future projections*—an issue we will discuss further below.

Looking more closely at the computed relative errors (Tables 7 and 8), we obtain some additional important insights. We see, for example, that the ability of the downscaled precipitation fields to reproduced observed characteristics at our sites is highly model dependent

(Table 7). Several models—CGCM3.1, CSIRO Mk3.0, and MPI ECHAM5 to be specific—are able to reproduce observed monthly-mean rainfall totals with less than 5% error relative to observations. These same models reproduce the observed frequency of rain days with less than 6% relative error and the monthly standard deviation with less than 11% relative error. By contrast, relative errors remain high even for the downscaled estimates (though considerably less so than for the raw model precipitation) for certain models, specifically GISS (error is nearly 20% for mean precipitation, roughly 12% for frequency of rain days, and roughly 24% for the standard deviation). Looking at the breakdown of error by site (Table 8), we see that certain sites (e.g., Stroudsburg and West Chester) show particularly large errors in mean monthly precipitation (11%–13%) and monthly standard deviation (18%–26%) (errors remain small for the number of rain days). These are among the few wettest sites in our network, and they also produce among the largest discrepancies for both mean precipitation and monthly standard deviation in the downscaling of modern (NCEP) observations (Table 4). It is reasonable to speculate that the larger biases we see in the downscaled simulations in these locations may relate as much to the intrinsic biases in the application of downscaling at these sites to any features specific to the model simulations themselves.

Finally, we return to the issue of the potential sensitivity of the downscaling procedure to the precise variables used in training the SOM. As we discussed earlier, future projections of precipitation based on statistical downscaling methods show some sensitivity to which humidity variables are used. Previous work in Africa, for example, indicates that use of specific humidity can lead to projections of large future increases in mean precipitation. Comparisons with parallel dynamical downscaling results suggest that these larger projections are unrealistic (B. C. Hewitson 2009, personal communication). A plausible explanation is that the exponential dependence of specific humidity on temperature leads to a large potential extrapolation error when projecting future precipitation changes in a warmer atmosphere based on training over an historical interval for which there are no analog states for characterizing future atmospheric temperature. Arguably, the use of relative humidity as a training variable avoids this problem.

Nonetheless, one might view this sensitivity to the humidity variables used as a reasonable measure of a key structural uncertainty in projecting future precipitation changes. In this spirit, we have performed the same analyses as described earlier but where specific humidity has been added as an additional large-scale predictive variable in training the SOM. This alternative

procedure yields remarkably similar results, including very similar relative errors (see supplementary Tables S10–S16). Thus, it is not possible from the skill assessments presented in this study to objectively favor one choice of humidity variables over the other. In additional work involving the downscaling of future climate change projections, we intend to use the sensitivity of projections to this choice as a measure of structural uncertainty in projecting regional changes in precipitation characteristics (Ning et al. 2011, manuscript submitted to *J. Climate*).

4. Conclusions

Using a specific application—reproducing historical precipitation characteristics in Pennsylvania—we demonstrated how statistical downscaling using self-organization maps to condition local precipitation estimates on large-scale atmospheric states can yield improved representations of precipitation characteristics. Using a variety of skill metrics and internal consistency tests, we showed that the downscaling procedure applied to modern (NCEP) atmospheric observations realistically reproduces observed daily precipitation characteristics at a network of sites throughout Pennsylvania.

Next, we demonstrated that application of the same SOM procedure to a suite of nine simulations from the CMIP3 multimodel historical simulation archive yields local precipitation estimates that form the model simulations that agree better with both each other and with historical observations than the raw GCM precipitation field. While the downscaling procedure does not entirely eliminate biases in the modeled local precipitation statistics, it does reduce these biases considerably, and it does provide greater consistency between the models, suggesting that bypassing the individual model's own varying precipitation parameterization schemes can yield more robust estimates of the distribution of local precipitation from the models.

Finally, we found that similarly skillful precipitation statistics could be obtained using either of two alternative representations of humidity information in training the SOM, one in which only relative humidity is used, another in which both relative and specific humidity are used. Given that one cannot objectively distinguish, on the basis of validation against historical data alone, which of these two schemes is preferable, it is advisable to consider the sensitivity to this choice as one measure of structural error in examining downscaling results applied to future climate change projections, where the two schemes may give somewhat different results.

Acknowledgments. This work was supported by the U.S. Department of Energy (DOE). Bruce Hewitson

(University of Cape Town) kindly provided the downscaling code used, and Lisa Coop (University of Cape Town) provided assistance with the downscaling implementation. The NCEP grid data and station daily precipitation data over Pennsylvania were obtained from the National Centers for Environmental Prediction (NCEP) and the DOE's Carbon Dioxide Information Analysis Center (CDIAC) at Oak Ridge National Laboratory, respectively. The WCRP CMIP3 multimodel GCM dataset is made available by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modeling (WGCM). Support of this dataset is provided by the DOE's Office of Science.

REFERENCES

- Benestad, R. E., 2005: Climate change scenarios for northern Europe from multi-model IPCC AR4 climate simulations. *Geophys. Res. Lett.*, **32**, L17704, doi:10.1029/2005GL023401.
- Cavazos, T., 1999: Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. *J. Climate*, **12**, 1506–1523.
- Christensen, J. H., and Coauthors, 2007: Regional climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 847–940.
- Crane, R. G., and B. C. Hewitson, 1998: Double CO₂ precipitation changes for the Susquehanna basin: Down-scaling from the GENESIS general circulation model. *Int. J. Climatol.*, **18**, 65–76.
- , and —, 2003: Clustering and upscaling of station precipitation records to regional patterns using self-organizing maps (SOMs). *Climate Res.*, **25**, 95–107.
- Efron, B., 1982: *The Jackknife, the Bootstrap, and other Resampling Plans*. Society for Industrial and Applied Mathematics, 92 pp.
- Fitzpatrick, E. A., and A. Krishnan, 1967: A first-order Markov model for assessing rainfall discontinuity in central Australia. *Theor. Appl. Climatol.*, **15**, 242–259.
- Gallus, W. A., and M. Segal, 2004: Does increased predicted warm-season rainfall indicate enhanced likelihood of rain occurrence? *Wea. Forecasting*, **19**, 1127–1135.
- Hanssen-Bauer, I., C. Achberger, R. E. Benestad, D. Chen, and E. J. Førland, 2005: Statistical downscaling of climate scenarios over Scandinavia. *Climate Res.*, **29**, 255–268.
- Hayhoe, K., and Coauthors, 2006: Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dyn.*, **28**, 381–407, doi:10.1007/s00382-006-0187-8.
- Hershfield, D. M., 1971: The frequency of dry periods in Maryland. *Chesapeake Sci.*, **12**, 72–84.
- Hewitson, B. C., and R. G. Crane, 1992a: Large-scale atmospheric control on local precipitation in tropical Mexico. *Geophys. Res. Lett.*, **19**, 1835–1838.
- , and —, 1992b: Regional climates in the GISS global circulation model: Synoptic-scale circulation. *J. Climate*, **5**, 1002–1011.
- , and —, 1992c: Regional-scale climate prediction from the GISS GCM. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **97**, 249–267.
- , and —, 1996: Climate downscaling: Techniques and application. *Climate Res.*, **7**, 85–95.
- , and —, 2002: Self-organizing maps: Applications to synoptic climatology. *Climate Res.*, **22**, 13–26.
- , and —, 2006: Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *Int. J. Climatol.*, **26**, 1315–1337.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Kohonen, T., 1989: *Self-Organization and Associative Memory*. 3rd ed. Springer-Verlag, 312 pp.
- , 1995: *Self-Organizing Maps*. Springer, 362 pp.
- Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global Planet. Change*, **13**, 195–206.
- , —, and —, 1999: Modeling ground heat flux in land surface parameterization schemes. *J. Geophys. Res.*, **104**, 9581–9600.
- Maurer, E. P., 2007: Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emission scenarios. *Climatic Change*, **82**, 309–325.
- Meehl, G. A., and Coauthors, 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 747–845.
- Ning, L., and Y. Qian, 2009: Interdecadal change in extreme precipitation over south China and its mechanism. *Adv. Atmos. Sci.*, **26**, 109–118.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.
- Reusch, D. B., and R. B. Alley, 2007: Antarctic sea ice: A self-organizing map-based perspective. *Ann. Glaciol.*, **46**, 391–396.
- , —, and B. C. Hewitson, 2007: North Atlantic climate variability from a self-organizing map perspective. *J. Geophys. Res.*, **112**, D02104, doi:10.1029/2006JD007460.
- Shortle, J., and Coauthors, 2009: Pennsylvania climate impact assessment report to the Department of Environmental Protection. Environment and Natural Resources Institute, The Pennsylvania State University, 7000-BK-DEP4252, 350 pp.
- Sturaro, G., 2003: A close look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data. *Climate Dyn.*, **21**, 309–316.
- Tennant, W., 2004: Considerations when using pre-1979 NCEP/NCAR reanalyses in the Southern Hemisphere. *Geophys. Res. Lett.*, **31**, L11112, doi:10.1029/2004GL019751.
- Trenberth, K. E., and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 235–336.
- Wagener, T., and Coauthors, 2010: The future of hydrology: An evolving science for a changing world. *Water Resour. Res.*, **46**, W05301, doi:10.1029/2009WR008906.
- Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.*, **21**, 530–548.
- Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, doi:10.1029/2001JD000659.
- , L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216.