

Statistical downscaling of general circulation model output: A comparison of methods

R. L. Wilby,^{1,2} T. M. L. Wigley,¹ D. Conway,³ P. D. Jones,³
B. C. Hewitson,⁴ J. Main,⁴ and D. S. Wilks⁵

Abstract. A range of different statistical downscaling models was calibrated using both observed and general circulation model (GCM) generated daily precipitation time series and intercompared. The GCM used was the U.K. Meteorological Office, Hadley Centre's coupled ocean/atmosphere model (HadCM2) forced by combined CO₂ and sulfate aerosol changes. Climate model results for 1980–1999 (present) and 2080–2099 (future) were used, for six regions across the United States. The downscaling methods compared were different weather generator techniques (the standard “WGEN” method, and a method based on spell-length durations), two different methods using grid point vorticity data as an atmospheric predictor variable (B-Circ and C-Circ), and two variations of an artificial neural network (ANN) transfer function technique using circulation data and circulation plus temperature data as predictor variables. Comparisons of results were facilitated by using standard sets of observed and GCM-derived predictor variables and by using a standard suite of diagnostic statistics. Significant differences in the level of skill were found among the downscaling methods. The weather generation techniques, which are able to fit a number of daily precipitation statistics exactly, yielded the smallest differences between observed and simulated daily precipitation. The ANN methods performed poorly because of a failure to simulate wet-day occurrence statistics adequately. Changes in precipitation between the present and future scenarios produced by the statistical downscaling methods were generally smaller than those produced directly by the GCM. Changes in daily precipitation produced by the GCM between 1980–1999 and 2080–2099 were therefore judged not to be due primarily to changes in atmospheric circulation. In the light of these results and detailed model comparisons, suggestions for future research and model refinements are presented.

1. Introduction

The present generation of global general circulation models (GCMs) and higher-resolution limited area models (LAMs) of the climate system are restricted in their usefulness for many subgrid scale applications (including those to hydrology) by their coarse spatial resolution and the uncertain reliability of their output on timescales of months or less, especially for variables pertaining directly to the hydrologic cycle [Carter *et al.*, 1994]. As Hostetler [1994] has observed, the parameterizations used in GCMs and in hydrological models are least reliable on the scale(s) at which these models interface. Hydrological models are frequently concerned with small, subcatchment scale processes and must parameterize regional-scale ones, whereas atmospheric models deal most proficiently with fluid dynamics at the planetary scale and parameterize many regional and smaller-scale processes.

Climate model resolution issues have important implications

for studies driven by the outputs of such models because of the potential magnitude of future climate change and the demonstrated sensitivity of resource systems to such change [Gleick, 1987]. This sensitivity, in turn, frequently arises because changes in water availability can have major effects on vegetation and, consequently, on all surface energy fluxes [Henderson-Sellers, 1993]. The scale mismatch between vegetation and land-surface processes, and the large-scale atmospheric circulation leads to problems that are further exacerbated because they often involve the most uncertain components of climate models, water vapor, and cloud feedback effects [Rind *et al.*, 1992].

Indeed, the International Geosphere-Biosphere Programme (IGBP) and the GEWEX Continental Scale International Project (GCIP) were established with the specific mandate to investigate the complex interactions between the physical and biological components of the environment and their responses to anthropogenic change. A major focus of the Biological Aspects of the Hydrologic Cycle (BAHC) component of IGBP has been the development of tools for generating the high-resolution meteorological inputs required for modeling ecohydrological processes [Bass, 1996]. Statistical “downscaling” approaches have subsequently emerged to satisfy the need to interpolate regional-scale atmospheric predictor variables (such as area averages of precipitation or temperature, and circulation characteristics such as mean sea level pressure or vorticity) to station-scale meteorological series [Hay *et al.*, 1991, 1992; Karl *et al.*, 1990; Kim *et al.*, 1984; Wigley *et al.*, 1990].

¹National Center for Atmospheric Research, Boulder, Colorado.

²Division of Geography, University of Derby, Derby, England.

³Climatic Research Unit, University of East Anglia, Norwich, England.

⁴Environmental and Geographical Science, University of Cape Town, Cape Town, South Africa.

⁵Atmospheric Science Group, Cornell University, Ithaca, New York.

Copyright 1998 by the American Geophysical Union.

Paper number 98WR02577.
0043-1397/98/98WR-02577\$09.00

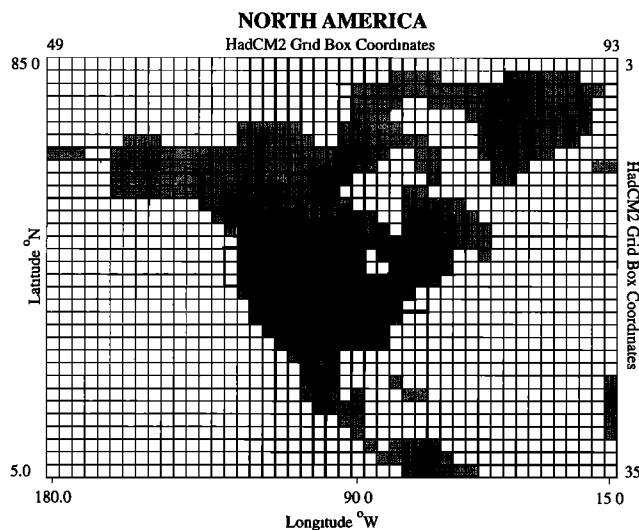


Figure 1. Location of the six North America study regions with respect to the grid of the Hadley Centre coupled ocean/atmosphere model, HadCM2 [Johns et al., 1997; Mitchell and Johns, 1997].

Fundamental to the approach is the assumption that stable empirical relationships can be established between atmospheric processes occurring at disparate temporal and/or spatial scales.

Since the early 1970s, many such relationships have been identified. Mesoscale weather patterns have been used to model numerous meteorological parameters such as precipitation occurrence in Washington State [Pittock, 1977; Hughes and Guttrop, 1994]; space-time daily rainfall patterns in the Ruhr catchment [Bardossy and Plate, 1992] and eastern Nebraska [Matyasovszky et al., 1993]; monthly-mean temperature and precipitation in Oregon state [Wigley et al., 1990]; extreme precipitation events and drought conditions in the Delaware River basin [Hay et al., 1991]; low-frequency precipitation events in the British Isles [Wilby, 1998]; wintertime rainfall in Iberia [von Storch et al., 1993]; and estimates of daily pan evaporation rates in southern Louisiana [McCabe and Muller, 1987]. Secondary relationships between circulation patterns and environmental time series include those encapsulated in studies of flooding in Arizona [Duckstein et al., 1993]; low-flow frequency analyses in the River Coln, United Kingdom [Wilby et al., 1994]; sea level anomalies in the Japan Sea [Maochang et al., 1995] and the Baltic Sea [Heyen et al., 1996]; episodic soil loss from farmland in southern England [Favis-Mortlock et al., 1991]; and surface water acidification in the East Midlands [Wilby, 1993].

The general limitations, theory and practice of downscaling are now well described in the literature [see, e.g., Giorgi and Mearns, 1991; Grotch and MacCracken, 1991; Kattenberg et al., 1996; von Storch et al., 1993; Wilby and Wigley, 1997]. The aims of this paper are more specific, namely, (1) to compare the performance of several statistical downscaling methods for six contrasting regions in the United States using a suite of daily precipitation diagnostics; (2) to compare changes in daily precipitation between a “present-day” (1980–1999) and a “perturbed” (2080–2099) climate scenario produced directly by a general circulation model (GCM), with the corresponding changes produced by the different statistical downscaling methods. The latter set of experiments were undertaken in

order to demonstrate that significant differences in precipitation scenarios can arise when a common set of GCM predictors are applied to different statistical downscaling techniques.

The GCM used was the U.K. Meteorological Office, Hadley Centre’s coupled ocean/atmosphere model (HadCM2) forced by combined CO₂ and albedo (as a proxy for sulphate aerosol) changes [Johns et al., 1997; Mitchell and Johns, 1997]. In this experiment, referred to as the SUL run, the model run begins in 1861, and the model is forced with an estimate of historical forcing to 1990 and a projected future forcing scenario over 1990–2100. The historical forcing is only an approximate representation of the “true” forcing (which itself is quite uncertain) so the GCM results for model years 1980–1999 would not be expected to accurately represent present-day conditions (see Appendix A). Nevertheless, they are as good as one can get with currently available GCM data. For the future, we chose model years 2080–2099 to maximize the changes. We use 20-year periods for the GCM data as a balance between increasing the signal-to-noise ratios for change while minimizing the change occurring through the period, and, at the same time, providing enough data to calibrate the statistical models.

2. Standard Data Sets and Precipitation Diagnostics

Seven data sets were compiled for the purpose of the model calibration and comparison exercise:

1. Observed daily precipitation over 1979–1995 for five sites in each of six regions shown in Figure 1. The “target” regions were 2.5° latitude by 3.75° longitude boxes corresponding to the HadCM2 grid. Sites were either within the target box, or within 1° latitude and longitude of the target box. The five sites were used to calculate unweighted, area-average daily precipitation, and an “optimum” station (or “key site”) was selected from each region for single-site downscaling. The station data were obtained from the U.S. National Climatic Data Center (NCDC), while the chosen data sampling period (1979–1995) corresponds to the availability of climate reanalysis data, described in items 2 and 3 below.
2. Observed daily grid point data spanning 1979–1995 for mean sea level pressure (MSLP) and 500 mbar height for a net of nine points centered on each of the six regions (Table 1; Figure 1). The data were obtained from the National Center for Environmental Prediction (NCEP) reanalysis [Kalnay et al., 1996], regridded to the HadCM2 grid.
3. Daily surface (2 m) temperature data (mean of four 6-hourly values) spanning 1979–1995 for the six central grid points of item 2 from the NCEP reanalysis, also regridded to the HadCM2 grid.
4. Daily gridpoint MSLP and 500 mbar height data for a

Table 1. Downscaling Study Regions: Central Latitudes and Longitudes for the Primary HadCM2 Grid Boxes Together With an Identifying Geographical Location

Location	Abbreviation	Latitude	Longitude
Salem, Oregon	SLM	45°N	123.75°W
Minneapolis, Minn.	MSP	45°N	93.75°W
Philadelphia, Pa.	PHL	40°N	75°W
Yucca Mountain, Nev.	YUC	37.5°N	116.25°W
Oklahoma City, Okla.	OKC	35°N	97.5°W
Jackson, Miss.	JKS	32.5°N	90°W

Table 2. Standard Daily Rainfall Diagnostics Used in Comparisons

Diagnostics	Unit
<i>Daily Rainfall Diagnostics</i>	
Mean wet-day amount	mm
Standard deviation of wet-day amount	mm
Median wet-day amount	mm
95th percentile of wet-day amount	mm
p_{00} (probability of a dry-day conditional on the previous day being dry)	
p_{11} (probability of wet-day conditional on the previous day being wet)	
π_w (unconditional probability of a wet-day)	
<i>Spell Statistics</i>	
L_w (mean wet-spell length in days, which is directly related to p_{11})	days
L_d (mean dry-spell length in days, which is directly related to p_{00})	days
Standard deviation of L_w and L_d	days
90th percentiles of L_w and L_d	days
<i>Low-Frequency Diagnostic</i>	
Standard deviation of monthly precipitation total	mm

A wet day was defined as a day with nonzero precipitation ≥ 0.05 mm.

net of nine points centered on each of the six regions from the HadCM2 SUL run, spanning 1980–1999 and 2080–2099. All GCM data were provided by the Hadley Centre (D. Viner, personal communication, Climate LINK project, Climatic Research Unit, 1996).

5. Daily temperature data for the six central grid points spanning 1980–1999 and 2080–2099 from the HadCM2 SUL run based on the mean of T_{\max} and T_{\min} .

6. Daily precipitation data for the six central grid points from the HadCM2 SUL run for 1980–1999 and 2080–2099.

7. Derived GCM and observed (NCEP) gridpoint vorticity

data based on the MSLP data in items (2) and (4). Vorticity is one of three indices (namely, vorticity, airflow strength and direction) that were originally employed by Jones *et al.* [1993] and Hulme *et al.* [1993] in an objective weather typing system for the British Isles. These indices have since been used as predictor variables for downscaling studies in Europe [Conway *et al.*, 1996], Southeast Asia [Wilby *et al.*, 1998] and North America [Wilby, 1998b].

In addition to using a common data set, the comparisons were integrated by using a standard suite of diagnostic statistics (Table 2). These diagnostics give an overview of the character of daily rainfall variability and enable the various models to be evaluated by comparing statistical descriptors of their daily precipitation with those for observed data. The diagnostics describe the daily variability via first and second moments, wet and dry-day renewal process probabilities and spell lengths, and low-frequency variability, via the standard deviation of monthly totals.

In general, the significance of differences in diagnostic statistics between models or between models and observations was assessed using empirical sampling distributions. In specific cases, idealized estimates of the sampling distribution variances were available, and these allow rough estimates of the significance of differences for such cases. The diagnostics were also used to assess differences between sites, differences between single-site and area-average characteristics, and changes between the present and the future.

3. Comparison of Downscaling Methods

Table 3 summarizes the GCM and downscaling models used in the comparison. More detailed descriptions of the statistical models may be found in Appendix B, in Wilby *et al.* [1996] or in the references provided in Table 3. There are four broad categories of model: the HadCM2 general circulation model;

Table 3. Precipitation Models Used in the Comparison Exercise

Notation	Description	References
HadCM2	Hadley Centre coupled ocean-atmosphere model forced by combined CO ₂ and sulfate aerosol forcing. Two 20 year periods were selected: present climate (1980–1999) and a perturbed, future climate (2080–1999).	Johns <i>et al.</i> [1997], Mitchell and Johns [1997]
ANN1	Artificial Neural Network calibrated against observed single site and area average precipitation, and forced using daily mslp and 500 hpa heights obtained from the two HadCM2 periods.	Hewitson and Crane [1992a, b, 1994, 1996], Crane and Hewitson [1997]
ANN2	Artificial Neural Network calibrated against observed single site and area average precipitation, and forced using daily mslp and 500 hpa heights and near-surface temperatures obtained from the two HadCM2 periods.	Hewitson and Crane [1992a, b, 1994, 1996], Crane and Hewitson [1997]
WGEN	First-order, two-state Markov process of daily rainfall occurrence. Wet-day precipitation amounts are modelled using gamma distributions. Downscaled future precipitation was produced by perturbing WGEN parameters in proportion to the changes in model parameters calibrated using the HadCM2 present and future data.	Richardson [1981], Wilks [1989, 1992], Gregory <i>et al.</i> [1993], Wilby <i>et al.</i> [1996]
SPEL	An alternating negative binomial recurrence process for wet- and dry-spell lengths. Wet-day precipitation amounts are modelled using gamma distributions. Downscaled future precipitation was produced by perturbing SPEL parameters in proportion to the changes in model parameters calibrated using the HadCM2 present and future data.	Wilby <i>et al.</i> [1996]
B-Circ	Binned vorticity method of resampling observed daily rainfall sets. Downscaled precipitation is modeled by sampling rainfall occurrence and amounts from discrete vorticity classes. Only the distribution of daily vorticity values is assumed to change between the present and future climate.	Conway and Jones [1996, 1998], Conway <i>et al.</i> [1996]
C-Circ	Semistochastic precipitation occurrence and intensity driven by vorticity. Downscaled precipitation is modelled using nonlinear empirical relationships between vorticity and wet-day occurrence/persistence/mean amounts. Only the distribution of daily vorticity values is assumed to change between the present and future climate.	Wilby <i>et al.</i> [1996, 1998], Conway <i>et al.</i> [1996]

Table 4. Root-Mean-Square Errors (RMSE) for Single (Key) Site Analyses: Seasonal Models

Diagnostic	Average Observed Value	ANN2 RMSE	B-Circ RMSE	C-Circ RMSE	WGEN RMSE	SPEL RMSE
Mean	8.41	4.53 (5.83) 0.79*	8.40 (1.53) 0.94*	8.36 (0.19) 0.99*
Median	3.84	3.61 (2.12) 0.85*	3.77 (0.97) 0.94*	4.16 (1.06) 0.95*	4.16 (0.57) 0.99*	4.05 (0.60) 0.99*
SD wet	11.84	3.66 (9.71) 0.70*	11.88 (2.47) 0.90*	11.75 (1.65) 0.96*	11.34 (0.92) 0.99*	11.42 (0.90) 0.99*
95% wet	31.02	11.75 (24.66) 0.73*	31.17 (5.61) 0.93*	30.09 (4.07) 0.97*	31.14 (2.23) 0.99*	31.25 (2.35) 0.99*
p_{00}	0.74	0.46 (0.33) 0.70*	0.74 (0.03) 0.96*	0.73 (0.02) 0.99*
p_{11}	0.55	0.83 (0.36) 0.38*	0.55 (0.05) 0.93*	0.55 (0.03) 0.99*
π_w	0.37	0.76 (0.47) 0.61*	0.37 (0.06) 0.93*	0.37 (0.01) 0.99*
Mean L_d	5.01	1.93 (4.33) 0.76*	4.95 (0.89) 0.96*	4.99 (0.40) 0.99*
Mean L_w	2.47	6.86 (5.88) 0.32*	2.47 (0.39) 0.93*	2.51 (0.14) 0.99*
SD L_d	5.04	1.30 (5.52) 0.83*	5.13 (2.04) 0.88*	4.45 (1.35) 0.96*	4.52 (1.41) 0.97*	4.90 (1.37) 0.94*
SD L_w	1.92	6.97 (6.66) 0.36*	1.95 (0.43) 0.93*	1.92 (0.22) 0.98*	1.92 (0.15) 0.99*	1.94 (0.15) 0.99*
90% L_d	11.13	3.46 (13.87) 0.75*	10.58 (2.32) 0.95*	10.84 (1.33) 0.99*	11.00 (1.02) 0.99*	11.33 (1.21) 0.99*
90% L_w	4.88	15.54 (10.43) 0.45*	4.83 (1.37) 0.86*	5.02 (0.50) 0.99*	5.00 (0.46) 0.98*	5.08 (0.61) 0.98*
SD month	54.77	54.55 (16.98) 0.98*	55.51 (16.36) 0.80*	49.10 (11.23) 0.93*	47.83 (11.30) 0.94*	51.15 (7.31) 0.97*

Observed (1979–1995) values are provided in order that the RMSEs may be compared with the mean diagnostics for the sites. RMSE values were calculated over six sites times four seasons ($n = 24$). Linear correlation coefficients for observed versus simulated diagnostics ($n = 24$) are indicated by asterisks. The dots in the WGEN and SPEL columns indicate that the calibration procedure necessarily produces a perfect fit.

two artificial neural network approaches (ANN1 and ANN2); two stochastic rainfall models (WGEN and SPEL); and two methods based on vorticity (airflow) indices (B-Circ and C-Circ). All the downscaling models were calibrated using observed (1979–1995) precipitation and/or circulation data.

The initial aim of the project was to compare the performances of the different statistical downscaling methods in simulating both single-site and area-average observed daily precipitation data. Note that, for some diagnostics, WGEN and SPEL simulate the observations precisely. In these cases, a comparison of results is meaningless in intermodel terms. Of the two ANN approaches, ANN2 (which includes temperature as a predictor variable) was selected for the comparison because we considered the additional predictor would improve its performance when applied to the future climate state. Both the binned vorticity (B-Circ) and the continuous vorticity (C-Circ)

models were included in the comparison, although it should be noted that they have strong conceptual similarities.

Tables 4 and 5 give details of the goodness of fit in model calibration, for single sites and area averages, respectively. Goodness of fit can be judged by comparing the mean values of the various statistics (averaged over the six sites and four seasons, i.e., $n = 24$) but is further quantified using an overall root-mean-square error (RMSE), and linear correlation coefficients. RMSE values and correlation coefficients were calculated for each downscaling model by comparing pairs of downscaled and observed diagnostics for the 1979–1995 calibration period. For example, for diagnostic “X”, each comparison involves 24 pairs of values (six sites by four seasons). The RMSE is determined by summing over the 24 squares of model-observed differences. The importance of these RMSE values can be assessed qualitatively by comparing them with the ob-

Table 5. Root-Mean-Square Errors for Area-Average Analyses: Seasonal Models

Diagnostic	Average Observed Value	ANN2 RMSE	B-Circ RMSE	C-Circ RMSE	WGEN RMSE	SPEL RMSE
Mean	4.26	3.75 (1.17) 0.94*	4.27 (0.15) 0.99*	4.20 (0.09) 0.99*	... (...) (...) ...
Median	1.94	3.01 (0.80) 0.90*	1.98 (0.12) 0.99*	1.98 (0.37) 0.95*	1.93 (0.21) 0.99*	1.84 (0.26) 0.99*
SD wet	5.92	3.01 (3.28) 0.93*	5.92 (0.16) 0.99*	6.33 (0.93) 0.96*	6.04 (0.40) 0.99*	6.14 (0.45) 0.99*
95% wet	16.20	9.67 (7.78) 0.94*	16.16 (0.58) 0.99*	15.56 (1.52) 0.98*	16.38 (1.03) 0.99*	16.62 (1.32) 0.99*
p_{00}	0.60	0.46 (0.19) 0.83*	0.56 (0.05) 0.98*	0.58 (0.03) 0.98*	... (...) (...) ...
p_{11}	0.77	0.86 (0.15) 0.40*	0.75 (0.03) 0.99*	0.76 (0.02) 0.99*	... (...) (...) ...
π_w	0.63	0.79 (0.25) 0.60*	0.63 (0.01) 0.99*	0.63 (0.01) 0.99*	... (...) (...) ...
Mean L_d	2.69	1.88 (1.34) 0.73*	2.43 (0.33) 0.98*	2.58 (0.22) 0.97*	... (...) (...) ...
Mean L_w	4.83	8.27 (6.39) 0.21*	4.36 (0.48) 0.99*	4.64 (0.29) 0.99*	... (...) (...) ...
SD L_d	2.36	1.27 (1.83) 0.69*	1.85 (0.71) 0.95*	2.01 (0.51) 0.95*	2.16 (0.38) 0.97*	2.37 (0.09) 0.99*
SD L_w	4.55	8.25 (6.70) 0.33*	3.93 (0.94) 0.98*	4.08 (0.94) 0.97*	4.36 (0.66) 0.98*	4.52 (0.81) 0.96*
90% L_d	5.63	3.42 (3.59) 0.54*	4.79 (1.18) 0.95*	5.19 (0.67) 0.97*	5.54 (0.65) 0.95*	5.79 (0.58) 0.97*
90% L_w	10.54	18.33 (15.10) 0.21*	9.33 (1.48) 0.98*	10.04 (1.07) 0.97*	10.63 (0.76) 0.99*	10.75 (1.70) 0.96*
SD month	41.36	41.23 (6.96) 0.98*	33.28 (10.67) 0.89*	32.62 (12.11) 0.84*	31.38 (12.44) 0.87*	37.66 (6.87) 0.93*

Observed (1979–1995) values are provided in order that the RMSEs may be compared with the mean diagnostics for the sites. RMSE values were calculated over six sites times four seasons ($n = 24$). Linear correlation coefficients for observed versus simulated diagnostics ($n = 24$) are indicated by asterisks. The dashes in the WGEN and SPEL columns indicate that the calibration procedure necessarily produces a perfect fit.

served mean values of each diagnostic. Note that the calibration diagnostics produced by the two airflow models were the mean values produced from 100 stochastic simulations, whereas for ANN2 only one precipitation realization was used to calculate the diagnostics. In the airflow models, the airflow index (vorticity) defines only the parameters of the precipitation distributions. Individual precipitation time series are then produced stochastically, day by day, by sampling from these distributions. In contrast, in the ANN models, the circulation-precipitation link is deterministic. Only one precipitation time series is produced corresponding to the observed circulation sequence. For WGEN and SPEL, several RMSE results were obtained using analytical methods. Note also that the RMSE and correlation coefficients are used here solely as overall indicators of model performance for the 1979–1995 period, not as a metrics for comparing the spatial patterns of modeled and observed diagnostics.

Both the RMSE and correlation statistics reveal that the WGEN and SPEL methods were superior to all other methods for the majority of the diagnostics. These two approaches have an a priori advantage, however, because they are constrained to reproduce the mean wet-day amount and the renewal process diagnostics (p_{00} , p_{11}) exactly (and hence π_w and mean spell lengths, which are functions of the renewal process probabilities). Since the geometrically-distributed spell lengths required by the Markov-chain occurrence formulation used in WGEN are a special case of the negative binomial spell length distribution used in the SPEL model (see Appendix B) it is not surprising that the results are similar for both models. Differences between these models were most noticeable for longer-duration wet and dry spell simulations. For example, even though the overall spell-length distributions were similar (Table 4), specific items differed markedly: winter wet spells at Philadelphia in the range of 10–12 days were ~2–3 times less

likely under SPEL, while fall dry spells at Salem in the range of 20–25 days were approximately three to five times more likely under SPEL compared with WGEN. These differences may arise from the small sample sizes and reflect “overfitting” of the SPEL model parameters in cases where the geometric spell lengths of the WGEN model are adequate.

The comparison of model performances in Tables 4 and 5 provides useful insights into the limitations of each downscaling technique. A major limitation of both the circulation-based models and of WGEN and SPEL approaches was their marked failure to capture realistically the standard deviation of monthly rainfall totals. In other words, these models do not adequately simulate lower frequency variations in precipitation variability. In most cases, the downscaling methods underestimate the low-frequency variability. This behavior is typical of commonly employed stochastic weather generator models [e.g., Gregory *et al.*, 1993; Mearns *et al.*, 1996], but it has not previously been pointed out in the circulation-based downscaling approach. It is noteworthy that the SPEL method performs substantially better than WGEN in this regard.

By comparing the generically similar B-Circ and C-Circ models it is evident that the B-Circ model simulates the distribution of wet-day amounts better than the C-Circ model at the area-average level, and worse at the single-site level. In simulating occurrence statistics (including spell lengths), the C-Circ model is slightly superior at both spatial scales. These findings are consistent with previous results obtained for single sites in the British Isles [Conway *et al.*, 1996]. The relatively poor simulation of area-average wet-day amounts by the C-Circ model is attributed to a simplifying assumption regarding the distribution of wet-day amounts (see Appendix B). By amending this distribution, it may be possible to produce more realistic daily precipitation amounts using C-Circ.

Tables 4 and 5 show both airflow models consistently perform better during the calibration period than the ANN2 method. In fact, the ANN2 method is superior for only one RMSE (the standard deviation of the area-average monthly precipitation amounts) and two correlation coefficients (the standard deviation of both the single site and area-average monthly precipitation amounts). This relatively poor performance of ANN2 can be attributed to the model’s tendency to produce far too many “wet” days. Because of this limitation, all the conditional wet- and dry-day probabilities and the spell lengths were in serious error and differed markedly from results with the B-Circ, C-Circ, WGEN, and SPEL models. This problem arises because the ANN model, as used here, does not separate the amount and occurrence aspects of the precipitation process. Instead, the model produces an amount for every day, and labels days wet or dry according to some threshold: here, a dry day was defined by amounts (which could be negative) <0.5 mm. A possible solution to this would be to adopt a two-tiered ANN approach in which the first step determines whether or not precipitation has occurred and, if it has, the second step calculates the actual wet-day amount.

The fact that ANN2 generally produced better simulations of the standard deviation of monthly precipitation totals for the area averages and single sites suggests that the upper atmosphere and surface temperature data used to drive the model are perhaps most pertinent to precipitation simulation over time intervals greater than a single day. In contrast, low-frequency variations in precipitation amounts are the most problematic in the WGEN and SPEL methods. Unlike the ANN and the vorticity models, there are no physical predictors

in WGEN or SPEL with which to force longer-term variations that may be related to such processes.

Overall, the results presented in Tables 4 and 5 show that the vorticity based models (i.e. C-Circ and B-Circ) can simulate single-site, daily precipitation amounts to within an average error (RMSE divided by observed value) of $<2.5\%$, unconditional wet-day probabilities (π_w) to within 2%, and mean wet-/dry-spell lengths to within 9%. In contrast, the standard deviation of monthly precipitation totals was captured only to within 10–30%. As noted previously, the WGEN and SPEL methods necessarily have zero error for calibrated daily precipitation renewal process probabilities and mean wet-day amounts. However, in common with other empirically based approaches, there is no certainty that the parameter values in any of the statistical models will remain constant as the climate changes.

4. Changes in Downscaled Precipitation Between 1980–1999 and 2080–2099

In this section we examine changes in precipitation character predicted by the various downscaling methods both at the single (key) site and area-average level. These experiments demonstrate the range of downscaled precipitation scenarios that can arise from different uses of a common set of GCM predictor variables. For the area-average precipitation, we also compare the downscaled changes with those taken directly from the GCM. For WGEN and SPEL, the changes are driven by the HadCM2 precipitation changes directly. For the other downscaling methods, the precipitation changes are driven by HadCM2 circulation changes (as well as temperature in the case of the ANN2 neural network method).

Table 6 summarizes results for individual stations in terms of the average of the absolute percentage changes in downscaled precipitation diagnostics between 1980–1999 and 2080–2099. Overall, the two ANN methods exhibited the greatest proportional changes at the key sites, and the C-Circ vorticity model the least. Excluding the ANN model results, the diagnostic with the smallest change when averaged over all sites and seasons was the conditional dry- to dry-day probability (p_{00}), and the diagnostic with the biggest change was the standard deviation of monthly precipitation amounts. There was no consensus regarding the region with the smallest changes (when averaged over all diagnostics), but 4/6 models suggested greatest changes for Yucca Mountains. Similarly, 4/6 models point to December-January-February (DJF) as the season with the largest changes, and 5/6 models suggested March-April-May (MAM) as the season with smallest changes.

Table 7 summarizes results for area averages in terms of the average of the absolute percentage changes in downscaled precipitation diagnostics between 1980–1999 and 2080–2099. HadCM2 results are shown for comparison. Once again, the two ANN methods exhibited the greatest proportional changes, and the C-Circ vorticity model the least. There was no consensus as to the diagnostic with the largest change, but 7/8 models (including HadCM2) record the conditional wet-wet day probability as changing least. As before, the majority of models point to Yucca Mountains as the grid box with largest changes; 3/8 models agree that Jackson changes the least. Seven out of eight models select DJF as the season with largest changes, and 7/8 select SON as the season with smallest changes.

The magnitudes of change generated by the area-average

statistical models may be compared with those obtained directly from HadCM2. For WGEN and SPEL, changes in all diagnostics are necessarily similar to those for HadCM2 (Table 7) because the WGEN and SPEL parameters were perturbed using HadCM2 precipitation data directly. WGEN and SPEL, nevertheless, still tend to underestimate changes in spell length characteristics relative to the direct GCM changes. Interestingly, WGEN and SPEL generate the largest percentage changes in monthly precipitation variability, even though these models tend to underestimate the magnitude of this statistic.

For the ANN models, changes in precipitation amount statistics (first four rows in Table 7) are considerably smaller than the HadCM2 changes, even for ANN2, which uses temperature as an additional driver for change (as a proxy for changes in atmospheric moisture content under the assumption that relative humidity changes are small). In contrast, renewal process probability and spell length statistics for the ANN models show larger changes than HadCM2. As noted above, however, the ANN models perform very poorly in simulating these statistics, so these results should be treated circumspectly. Monthly-total precipitation variability changes are much smaller for ANN than is observed in HadCM2.

For the B-Circ and C-Circ models, changes in all diagnostics are much less than those generated by HadCM2 (except for the monthly variability changes simulated by B-Circ). For precipitation amount statistics, this is clearly due to the fact that purely circulation-based methods can never capture effects due to all relevant physical processes; particularly, in this case, because of circulation-independent changes in the moisture-holding capacity of the atmosphere. The result is more surprising for the occurrence and spell statistics, where one might expect these changes to be explained more fully by changes in the circulation.

It is clear from the above analyses that in an average sense, changes in diagnostics derived directly from the GCM are generally larger in magnitude than changes derived from the area-average statistical downscaling models. For the GCM, the average results reflect, at the individual-region and seasonally specific level, the many instances where the model produces statistically significant changes in the diagnostics. How, then, do the downscaling model and GCM results compare at this detailed level? The former clearly give smaller changes, but are they at least in the same direction as those produced by the GCM? To answer this question, we correlated the daily precipitation changes, diagnostic by diagnostic, across all regions and seasons ($n = 6 \times 4 = 24$). The results are shown in Table 8, where only correlations significant at the 0.05 level are listed.

Table 8 shows that the change results for WGEN and SPEL correlate significantly with the GCM results for all diagnostics (although, for some renewal statistics such as p_{00} , the correlations are quite low). This is a consequence of the methods used to perturb the WGEN and SPEL parameters, which are based solely on GCM precipitation changes (see Appendix B). For both ANN1 and ANN2, changes in the diagnostics show strong correlations with those from the GCM for all diagnostics except those that refer to dry-spell simulation. In contrast, neither the B-Circ nor C-Circ model diagnostics changes were significantly correlated with the GCM results for the vast majority of diagnostics.

These two results (i.e., the failure of the vorticity models to give precipitation changes anywhere near as large as given by the GCM; and the poor correlations between the vorticity-

Table 6. Single (Key) Site Results for Different Downscaling Models: Averages of Absolute Percentage Changes in Precipitation Diagnostics Between 1980–1999 and 2080–2099

Diagnostic ($n = 24$)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ
Mean wet	26	29	12	11	4	3
Median wet	24	25	15	13	8	3
SD wet	28	37	9	8	3	5
95% wet	28	33	10	10	4	4
p_{00}	5	2	21	22	2	2
p_{11}	11	9	6	5	3	2
π_w	14	18	15	14	4	4
Mean L_d	17	12	34	30	6	5
Mean L_w	12	7	50	44	3	2
SD L_d	16	18	52	38	6	6
SD L_w	16	12	50	47	4	4
90% L_d	20	14	46	37	7	6
90% L_w	16	10	50	50	4	4
SD monthly	28	37	8	8	26	10
Grid box ($n = 56$)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ
SLM	16	9	32	44	7	7
MSP	15	27	17	19	4	4
PHL	14	14	18	14	5	5
YUC	21	35	47	7	13	5
OKC	21	19	29	25	9	3
JKS	14	16	40	20	4	4
Season ($n = 84$)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ
DJF	26	26	30	35	19	6
MAM	15	13	14	15	11	3
JJA	23	22	54	36	9	6
SON	15	24	25	32	3	3
Overall mean ($n = 336$)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ
All data	20	22	32	29	12	5

The top group gives absolute percentage changes in each diagnostic averaged over regions and seasons ($n = 6 \times 4$). The second group averages over diagnostic and seasons ($n = 14 \times 4$) for each region. The third group averages over diagnostics and regions ($n = 14 \times 6$) for each season. The final value gives averages over diagnostics, regions and seasons ($n = 14 \times 6 \times 4$).

driven and direct-GCM changes) have some important implications. First, they imply that, while vorticity is an important determining factor in explaining the interannual variability of present-day precipitation diagnostics, it is less important in controlling future changes on longer timescales. (The likelihood of this was the reason for including temperature in the ANN2 simulations.) This problem is exacerbated by the fact that vorticity changes in the GCM are small at all sites and in all seasons (well below statistical significance). This does not mean circulation is not an important determinant, since the ANN results, which include circulation data via individual grid point values, show much larger changes. However, the poor ANN performance in simulating present-day statistics casts doubt on the credibility of the ANN change results in general. Our conclusion, in spite of the ANN results, is that additional predictor variables beyond vorticity (and beyond circulation more generally) are required before one can confidently downscale future climate with methods such as those used here.

Table 7. Area-Average Results for the Different Downscaling Models and HadCM2: Averages of Absolute Percentage Changes in Precipitation Diagnostics Between 1980–1999 and 2080–2099

Diagnostic (<i>n</i> = 24)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ	HadCM2
Mean wet	26	29	10	10	3	3	27
Median wet	25	27	13	15	5	3	36
SD wet	28	34	8	7	2	5	26
95% wet	28	31	9	9	2	4	29
P_{00}	6	4	15	15	2	3	9
P_{11}	5	5	5	7	1	1	4
π_w	6	7	14	13	2	2	6
Mean L_d	10	8	26	21	2	3	11
Mean L_w	12	8	28	34	3	3	14
SD L_d	17	13	37	25	3	4	22
SD L_w	14	12	32	38	4	34	23
90% L_d	12	10	31	26	3	5	17
90% L_w	15	12	36	32	3	3	15
SD monthly	27	30	8	8	27	13	24
Grid box (<i>n</i> = 56)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ	HadCM2
SLM	13	9	24	34	3	6	18
MSP	13	20	13	13	3	5	16
PHL	14	16	11	11	4	5	17
YUC	31	33	35	27	11	4	34
OKC	19	19	27	16	10	5	20
JKS	11	10	12	17	4	4	15
Season (<i>n</i> = 84)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ	HadCM2
DJF	23	26	26	28	20	8	28
MAM	14	15	13	11	10	4	18
JJA	20	20	32	20	9	4	21
SON	13	15	13	22	3	3	14
Overall mean (<i>n</i> = 336)	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ	HadCM2
All data	18	19	22	21	7	5	21

The downscaling models used here were calibrated using observed data. The top group gives absolute percentage changes in each diagnostic averaged over regions and seasons ($n = 6 \times 4$). The second group averages over diagnostic and seasons ($n = 14 \times 4$) for each region. The third group averages over diagnostics and regions ($n = 14 \times 6$) for each season. The final value gives averages over diagnostics, regions and seasons ($n = 14 \times 6 \times 4$).

A second interpretation of these results is that they may also reflect deficiencies in the way circulation and precipitation are related in the GCM. A full analysis of these issues is beyond the scope of the present paper, but we note in passing that the relationships between precipitation diagnostics and vorticity in the 1980–1999 GCM data were both weaker and of quite different character compared with those in observed data.

5. Conclusions and Future Research Opportunities

In the comparison using observed data, there were marked differences in the circulation-based statistical downscaling models: the B-Circ and C-Circ models performed similarly and were much better than the ANN (neural network) models. The B-Circ model was superior for area-average wet-day amounts, while the C-Circ model was superior for area-average renewal-process statistics, and for all statistics at the single-site level. The C-Circ model's performance on wet-day amounts could probably be improved by fitting a better distribution to the

calibration data. The ANN model's deficiencies were traced to poor simulations of wet-day occurrence (too many wet days), a deficiency that could perhaps be overcome by adopting a two-tier approach separating the occurrence and amount aspects. The WGEN and SPEL models performed similarly, except in their simulations of low-frequency variability where SPEL is superior (albeit, still deficient). For spell-length simulations there is clearly scope to improve the higher moments obtained with SPEL by fitting more sophisticated (and possibly autocorrelation) length distributions.

The area-average precipitation changes simulated by the downscaling models (driven by GCM vorticity changes for B-Circ and C-Circ, by mslp and 500 mbar heights for ANN1 as well as temperature changes for ANN2, and GCM precipitation changes for WGEN and SPEL) showed numerous important differences from those based directly on GCM output. The B-Circ and C-Circ models produced much smaller changes than the GCM, with the vorticity-based changes sometimes being in the opposite direction to those for the GCM

Table 8. Significant ($p \leq 0.05$) Correlations Between Changes in Downscaled Area Average and HadCM2 Precipitation Diagnostics Between 1980–1999 and 2080–2099

Diagnostic	WGEN	SPEL	ANN1	ANN2	B-Circ	C-Circ
Mean wet	+0.98	+0.92	+0.47	+0.27		
Median wet	+0.91	+0.86				
SD wet	+0.89	+0.70	+0.57	+0.46		
95% wet	+0.91	+0.87	+0.47	+0.51		
p_{00}	+0.70	+0.56	+0.40			
p_{11}	+0.83	+0.78	+0.67	+0.71	+0.44	
π_w	+0.84	+0.72	+0.65	+0.63		+0.49
Mean L_d	+0.84	+0.67				
Mean L_w	+0.89	+0.52	+0.79	+0.76		
SD L_d	+0.70	+0.66				
SD L_w	+0.73	+0.43	+0.61	+0.54		
90% L_d	+0.57	+0.45				
90% L_w	+0.80	+0.52	+0.56	+0.48		

Correlations were calculated over all regions and seasons ($n = 24$). The results of the statistical models are those based on observed data.

(since these correlations were not statistically significant they are not shown in Table 8). The small changes produced by C-Circ and B-Circ clearly result from the limited explanatory power of vorticity in the climate-change context, in contrast to the stronger role this circulation variable has in explaining observed interannual variability [see Conway *et al.*, 1996]. In determining changes in the character of precipitation variability, vorticity change is only one of a number of driver mechanisms. Better simulations will require the incorporation of other predictor variables into this type of downscaling model. For the ANN models, much larger changes were simulated (more comparable to those generated directly by the GCM), but the realism of these changes is uncertain because of the poor validation performance of these models. For WGEN and SPEL, an interesting feature is that these models simulate relatively large changes in the interannual variability of monthly precipitation totals, even though this variability is not explicitly forced (e.g., by circulation changes).

Thus the present study has demonstrated significant variations in the behavior of the selected statistical downscaling methods using daily precipitation data drawn from climatologically different regions in the United States. As is often the case, the research has raised many questions concerning the refinement and future application of the downscaling techniques. These issues are distilled here into a number of potentially fruitful research areas.

Although the present study considered six contrasting regions in the US, the total area investigated still amounted to <1% of the North American land area. There is considerable scope therefore for the extension of the study to additional sites within the United States and elsewhere (e.g., to date there has been a real paucity of studies in the southern hemisphere).

Increasing spatial coverage raises the issue of spatial autocorrelation, and the need to investigate spatial variations in downscaling model parameters. By establishing either physical or empirical relationships between model parameters and factors such as topography, aspect, or proximity to large water bodies, it may be possible to extend and improve conventional downscaling methods [e.g., Daly *et al.*, 1994]. Similarly, our analyses have confirmed the problem that weather generators have in simulating low-frequency variability. Both the WGEN and SPEL models could be improved by the inclusion of time-varying parameters that are better able to reflect monthly and

interannual variations in precipitation. This problem was also evident in the circulation-based methods. By relating temporal variations in parameter values to variations in atmospheric and oceanic phenomena such as the El Niño Southern Oscillation (ENSO) phenomenon or the North Atlantic Oscillation (NAO), it may be possible to improve the low-frequency performance of the downscaling methods and so increase confidence in extrapolated details of future climate scenarios [e.g., Wilby, 1997].

The present study used only a limited suite of predictor variables. By increasing the number of driving variables (without going so far as to compromise model parsimony) it may be possible to develop improved and more universally applicable downscaling models. For example, both the binned (B-Circ) and continuous (C-Circ) vorticity methods would benefit from the inclusion of other airflow indices (e.g., airflow strength and direction, and divergence, rather than just vorticity). We noted, in addition, the possible importance of including better measures of atmospheric moisture holding capacity in downscaling models of precipitation.

An important application for downscaling analyses is in GCM validation and diagnosis. The comparison of the B-Circ and C-Circ precipitation change results with the corresponding HadCM2 changes revealed possible inconsistencies in the GCM's relationships between precipitation and circulation that could be investigated more directly. The empirical relationships between NCEP vorticity and observed precipitation (occurrence and wet-day amounts) were not as evident in the equivalent GCM vorticity and GCM precipitation comparisons. In some instances, the GCM produced increases in vorticity accompanied by lower seasonal rainfall, opposite to what is observed in the current real-world climates of the study regions [Wilby, 1998b].

Finally, conventional downscaling methods take, as their initial input, changes in predictor variables between GCM integrations of present and future climates. This presupposes the veracity of the GCM with respect to the observed climatology of the target. An alternative strategy might be to stochastically generate the predictor variables (such as vorticity) based on changes in their statistics. By using downscaling models driven by stochastic atmospheric sequences, it may be possible to explore the range of future possibilities more comprehensively.

Appendix A: O/AGCM Data

The O/AGCM has been run both in an unforced mode, commonly referred to as the “control run,” and forced by time-dependent variations in CO₂ and CO₂-plus-albedo changes. The latter case, used here, is referred to as the “SUL” experiment. In the O/AGCM SUL experiment, the initial (1861) state comes from an initial section of the unforced control run; and the control run in turn uses a CO₂ level of 323 ppmv and begins with *Levitus* [1982] “present-day” ocean temperatures. The experiment then follows a forcing history that roughly parallels observations to 1990 and forcing that corresponds to the central IS92 emissions scenario [Leggett *et al.*, 1992] subsequently. Thus, the initial state does not correspond to 1861 conditions, but the changes in forcing from 1861 onwards do correspond approximately to real-world changes.

Because of this, in none of the experiments performed with this model is the present-day climate simulated in a fully consistent way. A strict validation of the model against observations is therefore not possible. The unforced run corresponds approximately to present-day (1970s) conditions for ocean temperatures and greenhouse-gas concentrations, but not for aerosols. (The 323 ppmv CO₂ level used was the actual level in 1968). In terms of total global-mean forcing (CO₂, other greenhouse gases and aerosols), 323 ppmv was the equivalent CO₂ level in 1979 (based on the IPCC forcing history from *Kattenberg et al.* [1996]). For the sulfate aerosol forcing pattern, however, the unforced run corresponds to preindustrial conditions. Thus the unforced run simulates a climate state that is a mixture of preindustrial and present-day conditions. The same inconsistency problem applies to the SUL perturbation run. If one takes the “present” (e.g., 1980–1999) from the SUL run, this partially simulates aerosol forcing effects but has a CO₂ level that corresponds to conditions around 2020–2040 (depending on how one considers the forcing of non-CO₂ gases). We choose to use the 1980–1999 SUL results for our present-day period but note that this is not strictly a fair test of the model.

Appendix B: Statistical Downscaling Models

B1. Weather Generators (WGEN and SPEL)

B1.1. The models. Two stochastic “weather generators” are considered. The first, referred to as “WGEN” is the conventional chain-dependent-process stochastic model [Katz, 1977a, b] for daily precipitation occurrence and amount. This process constitutes the precipitation part of the commonly used WGEN “weather generator” [Richardson, 1981]. In WGEN, daily precipitation occurrence is represented as following a two-state, first-order Markov chain characterized by the two conditional probabilities

$$p_{01} = Pr\{\text{precipitation on day } t | \text{no precipitation on day } t - 1\} \quad (1a)$$

$$p_{11} = Pr\{\text{precipitation on day } t | \text{precipitation on day } t - 1\} \quad (1b)$$

Equivalently, and more conveniently in the present context, the Markov chain can be characterized by the unconditional wet-day probabilities (π_w) and the lag-1 autocorrelation (r)

$$\pi_w = \frac{p_{01}}{1 + p_{01} - p_{11}} \quad (2)$$

$$r = p_{11} - p_{01} \quad (3)$$

In WGEN as implemented here, precipitation amounts are modeled as independent gamma variates, with probability density

$$f(x) = (x/\beta)^{a-1} \exp[-x/\beta]/\beta\Gamma(a) \quad x, a, \beta > 0 \quad (4)$$

The second stochastic model, which we refer to as “SPEL,” is similar but somewhat more elaborate in structure. Precipitation occurrences are simulated using a renewal process, in which the lengths of alternating wet and dry spells (i.e., runs of one or more consecutive wet or dry days) are represented as being drawn from two-parameter (p, k) negative binomial distributions [Johnson *et al.*, 1992, Chapter 5]. That is, the frequency distributions of lengths (L) of wet and dry spells (in days) are represented by separate probability distribution functions of the form

$$\begin{aligned} Pr\{L = x\} &= \binom{k+x-2}{k-1} p^k (1-p)^{x-1} \\ &= p^k (1-p)^{x-1} \prod_{i=1}^{x-1} \frac{k+i-1}{i}; \quad x = 1, 2, 3, \end{aligned} \quad (5)$$

The first-order Markov model (1) is effectively a special case of (5), since first-order Markov dependence implies a geometric distribution for the spell lengths [e.g., Foufoula-Georgiou and Lettenmaier, 1987]. The geometric distribution is a special case of (5), which occurs when the parameter $k = 1$.

Precipitation amounts on wet days are also modeled using gamma distributions in the SPEL model, but separate distributions are used for single wet days, for the first wet day of a multiday wet spell, and for the subsequent wet days of multiday spells. These three gamma distributions are constrained to have the same shape parameter α but have separate scale parameters β_1, β_2 , and β_3 . Maximum likelihood methods were used to fit all the model parameters for both WGEN and SPEL.

B1.2. Downscaling procedure. Construction of changed-climate scenarios using GCM results and the above stochastic models proceeds in two steps, referred to here as “extrapolation” and “downscaling.” In the extrapolation step, significant differences (at the 0.05 level, as determined using likelihood ratio tests) between the parameters characterizing the present-day (1980–1999) and future (2080–2099) series are applied in a consistent manner to the corresponding parameters characterizing the respective series of area-averaged observations. The second step is to then downscale these changes in the area-average model parameters to the local station level.

Extrapolation adjustments are made separately to the occurrence and intensity parameters. For the WGEN model, climatological wet-dry probabilities (2) are adjusted linearly on a log-odds scale,

$$L(\pi) = \ln[\pi/(1-\pi)], \quad (6)$$

so that

$$\pi_{\text{adj}} = L^{-1}\{L(\pi_0) + [L(\pi_P) - L(\pi_C)]\} \quad (7)$$

The subscripts 0, P , and C here denote observed data, and “perturbed” and “control” GCM series, respectively. The lag-1

correlations (3) are adjusted linearly on the scale of the Fisher Z transform,

$$Z(r) = 0.5 \ln [(1 + r)/(1 - r)] \quad (8)$$

so that

$$r_{adj} = Z^{-1}\{Z(r_0) + [Z(r_P) - Z(r_C)]\} \quad (9)$$

The two-parameter gamma distributions are extrapolated through proportional adjustments to their shape and scale parameters:

$$\alpha_{adj} = \alpha_0 \alpha_P / \alpha_C \quad (10a)$$

$$\beta_{adj} = \beta_0 \beta_P / \beta_C \quad (10b)$$

The adjustments (10) also yield proportional adjustments to the means and variances of the gamma distributions.

For the SPEL model, extrapolation adjustments for the negative binomial spell length parameters are made proportionally to the means and variances of the spell-length distributions, in a way that ensures $\mu > 1$ and $\sigma^2 > 0$:

$$(\mu_{adj} - 1) = (\mu_0 - 1)(\mu_P - 1)/(\mu_C - 1) = \frac{k_0 k_P (1 - p_0)(1 - p_P)}{k_C (1 - p_C)} \frac{p_C}{p_0 p_P} \quad (11)$$

$$\sigma_{adj}^2 = \sigma_0^2 \sigma_P^2 / \sigma_C^2 = (\mu_{adj} - 1) p_C / p_0 p_P \quad (12)$$

Solving (11) and (12) yields the extrapolation expressions

$$p_{adj} = p_0 p_P / p_C \quad (13a)$$

$$k_{adj} = \frac{k_0 k_P (1 - p_0)(1 - p_P)}{k_C (1 - p_C)} \frac{p_C}{p_C - p_0 p_P} \quad (13b)$$

for the adjusted negative binomial parameters. Extrapolated values of the gamma distribution shape and scale parameters are computed using (10a) and (10b), respectively.

For the second step, downscaling GCM-derived climate changes from the gridbox, or area-average scale to the individual station scale, requires that the relevant relationships between the stochastic model parameters at the two scales be known or estimated. The approach taken here was to estimate these relationships using regression equations linking observed station-level and area-average stochastic model parameters, pooling all six grid boxes and all seasons. For the WGEN models, these equations are

$$\ln [\pi_s / (1 - \pi_s)] = -1.41 + 1.37 \ln [\pi_A / (1 - \pi_A)] \quad (14a)$$

$$R^2 = 79\%,$$

$$Z(r_s) = -.112 + 1.03 Z(r_A) \quad R^2 = 70\%, \quad (14b)$$

$$\mu_s = 2.00 \mu_A \quad R^2 = 73\%, \quad (14c)$$

$$\sigma_s^2 = 3.72 \sigma_A^2 \quad R^2 = 74\%. \quad (14d)$$

Here the subscript “s” indicates station-level parameters, and the subscript “A” indicates the extrapolated (adjusted) area-average parameters defined by (7), (9), and (10). For the SPEL occurrence-process models, the corresponding downscaling regression equations are

$$\ln (\mu_{D,s} - 1) = 0.572 + 1.32 \ln (\mu_{D,A} - 1) \quad (15a)$$

$$R^2 = 79\%,$$

$$\ln (\mu_{W,s} - 1) = -1.40 + 1.27 \ln (\mu_{W,A} - 1) \quad (15b)$$

$$R^2 = 69\%,$$

$$\ln (\sigma_{D,s}^2) = 0.477 + 1.44 \ln (\sigma_{D,A}^2) \quad R^2 = 79\%, \quad (15c)$$

$$\ln (\sigma_{W,s}^2) = -1.87 + 1.04 \ln (\sigma_{W,A}^2) \quad R^2 = 76\%, \quad (15d)$$

where the subscripts “D” and “W” denote dry and wet spells, respectively. Parameters for the three-distribution gamma distribution model for precipitation were downscaled using

$$\mu_{1,s} = 4.44 \mu_{1,A} \quad R^2 = 56\%, \quad (16a)$$

$$\mu_{2,s} = 2.80 \mu_{2,A} \quad R^2 = 31\%, \quad (16b)$$

$$\mu_{3,s} = 1.94 \mu_{3,A} \quad R^2 = 71\%, \quad (16c)$$

$$\sigma_{3,s}^2 = 3.52 \sigma_{3,A}^2 \quad R^2 = 68\%. \quad (16d)$$

One additional step was employed to arrive at the final downscaled station-level parameters to account for the fact that the regression relationships in (14)–(16) are not perfect downscaling algorithms. As an example, in the case of the WGEN parameter π , the “extrapolation” equation (7) is applied again, but now with the subscript “0” pertaining to the original station-series parameter, the subscript “P” indicating downscaled station-level parameters obtained using (14a) with the extrapolated/adjusted area-average parameter, and the subscript “C” indicating downscaled values using the regressions in (14a) with the original fitted area-average π . This final step is intended as an adjustment for the lack of fit of the downscaling regressions, and is imposed so that the final downscaled values reflect primarily the differences between the two GCM integrations, rather than any lack of fit of the regression equations (14)–(16). To the extent that parameter differences between the two GCM series are judged to be statistically significant, this procedure applies those changes to the station-level parameters in a way that cancels most or all of the lack of fit in the downscaling regressions.

B2. Artificial Neural Networks (ANN1 and ANN2)

In ANN downscaling a quantitative function is derived directly relating two data sets (e.g., daily precipitation as a predictand and atmospheric circulation variables as predictors). In this regard, ANNs are analogous to multiple regression, although mathematically dissimilar. In the same way that multiple regression develops a quantitative relationship between a predictand and a set of predictors, so does an ANN, although with no supposition regarding the form of the function or the degree of nonlinearity. In theory, an ANN is capable of representing any arbitrary nonlinear relationship. An overview of ANNs may be found in the work of *Hewitson and Crane* [1994], and a more detailed outline of the procedures as applied in this paper may be found in the work of *Crane and Hewitson* [1997].

An ANN is composed of set of simple processing nodes, each of which receives inputs from other nodes, and outputs values to further nodes. The resultant net of nodes will have some nodes dedicated to receiving primary inputs, and some to providing final outputs from the overall net. Each node is connected to others via weighted links, and calculates a sum of the weighted inputs. This sum is then transformed by some function that may be either linear or nonlinear such as a step function, or a bounded differentiable nonlinear function (like a sigmoid), which becomes the node output value. Nodes are generally arranged in layers, with an input layer connected to

a hidden layer which is in turn connected to an output layer. Given this structure, a useful analogy of how the ANN represents a complex function is that of inverse Fourier transforms. In the same way that inverse Fourier transforms combines simple sine and cosine waves to create a more complex wave, so ANNs combine simple nonlinear functions to create a more complex function.

In the present downscaling application the ANN is configured with only one output node to provide the downscaled information for the target location. All nodes were established with a sigmoid transfer function and were arranged with an input layer of nodes, one hidden layer, and a single node output layer.

Development of the actual function represented by the ANN is accomplished through a training procedure. There are numerous methods that may be used for training, but all, in some form or other, represent a minimization procedure. Initially, the weights in the ANN are set to small random values. The ANN is then presented with input data for which known target output values are available. The full set of input samples is presented sequentially to the ANN, and the error between the ANN output and desired output values noted. The training procedure then adjusts the weights in an attempt to minimize the error, and the training data are again presented to the ANN. In this manner the training algorithm carries out a minimization search procedure over the error surface iteratively to find the global minimum; or, at least, a local minimum close to the global minimum.

In the present downscaling application, the training algorithm used was "back-propagation," whereby the error after each pass of the data is "back-propagated" through the net from the output node to the input nodes, with proportional errors assigned to each node. The weights connecting each node are then adjusted in a direction to minimize the error. In its simplest form this is equivalent to a gradient descent algorithm. The form utilized here, however, makes some additional assumptions about the shape of the error surface in the vicinity of local to each node in order to speed up the convergence, and avoid becoming trapped in a local minimum. Full details of the algorithm are available with the shareware software package NevProp, available at <ftp://ftp.scs.unr.edu/pub/cbmr/nevprop-dir>

The ANN training procedure has some potential pitfalls. First, is the possibility of overtraining (equivalent to overfitting in regression analysis). Because of the ability of the ANN to represent highly complex relationships, it is possible for the ANN to learn to relate the noise (i.e., physically or unrelated variance) in the input to the target output data. To avoid this, a random portion (25%) of the training data is removed from the training process and retained for testing purposes. During the training procedure, the ANN is then repeatedly evaluated against the independent test data until the performance of the ANN on the test data no longer improves. A second pitfall arises from the inherent nonlinearity of the method. As with all nonlinear statistically derived relationships, one must be careful in applying the ANN outside the calibration data domain.

Once the training of the ANN is completed, the ANN function may be applied to further input data with the assumption that the new data falls within the span of the training data. In the downscaling this is carried out using the HadCM2 GCM atmospheric data for daily mean sea level pressure, mean temperature and 500 mbar geopotential heights to generate sub-grid-scale precipitation.

B3. Continuous Vorticity Method (C-Circ)

In this model, second-order polynomial regression equations of the form

$$p_{11} = a + bZ + cZ^2 \quad (17)$$

$$p_{01} = d + eZ + fZ^2 \quad (18)$$

$$\hat{R} = g + hZ + iZ^2 \quad (19)$$

were used to derive empirical relationships between the observed wet-wet (p_{11}) or dry-wet day (p_{01}) conditional probabilities and the vorticity (Z). The observed mean wet-day amounts (\hat{R}) at each location were also modelled using second-order polynomial regression against the Z index.

Using the polynomial equations for the daily precipitation occurrence and intensity processes, daily precipitation amounts are generated stochastically using the following procedure. For each day, the vorticity value determines (through (17) if the previous day was wet, or (18) if it was dry) the probability (p) that the day is a wet day. Then, a uniformly distributed random number r_1 ($0 < r_1 \leq 1$) is chosen. If the value of r_1 is less than or equal to p the day is selected as wet, otherwise dry. The expected daily precipitation amount (\hat{R}) for the given value of Z is then determined using (19). Finally, the actual wet-day precipitation amount (\hat{R}) is determined using

$$R = -\phi\hat{R}(r_2) \quad (20)$$

where r_2 is a second uniformly distributed random number ($0 < r_2 \leq 1$) and ϕ is a random scaling factor used to inflate the variance of R to accord better with observations. If the mean of ϕ is one, the equality of means of R and \hat{R} is preserved. Present and future daily precipitation characteristics (occurrence and amounts) are calculated using either observed or GCM-derived daily vorticities respectively.

B4. Vorticity Binning Method (B-Circ)

The binning method consists of generating distributions of rainfall (or absence of rainfall) events, for categories of air flow indices from the historical record [Conway *et al.*, 1996]. Distributions are generated for each of the four seasons (DJF, MAM, JJA, SON) and for eight categories of vorticity ranging in the present case from less than -4 to over $+8$ in steps of 2 units. (Note that the vorticity values we used involved scaling factors that depended on the particular grid spacing and so cannot be directly compared with conventional numerical values.) In addition, although there is a certain degree of persistence in vorticity, in order to fully reproduce the persistence of wet and dry periods in the observed record two additional categories are defined, days when the previous day was wet and days when it was dry. This leads to a total of $8 \times 4 \times 2 = 64$ distributions which are then resampled randomly to generate a new series of daily rainfall from either observed or GCM-derived series of vorticity.

Acknowledgments. This research was sponsored by the Electric Power Research Institute, Palo Alto, CA, USA, under contract number WO8020-19. We thank Charles Hakkarinen, the EPRI ACACIA Technical Manager, for his advocacy and enthusiastic support of the research. We are also grateful to David Viner of the Climate Impacts LINK Project (UK Department of the Environment contract EPG1/1/16) for supplying us with the HadCM2 data on behalf of the Hadley Centre and UK Meteorological Office. Similarly, we thank Doug Lind-

holm of CGD, UCAR/NCAR for his assistance in acquiring the NCEP reanalysis data, and Dick Knight at the National Climatic Data Center for providing the station climate data. Finally, we are indebted to Lisa Butler and Christy Tidd for their logistical support. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

- Bardossy, A., and E. J. Plate, Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, **28**, 1247–1259, 1992.
- Bass, B., Interim Report on The Weather Generator Project, Focus 4 of IGBP Biospheric Aspects of the Hydrological Cycle (BAHC), Environ. Adapt. Res. Group, Atmos. Environ. Serv., Ontario, Canada, 1996.
- Carter, T. R., M. L. Parry, H. Harasawa, and S. Nishioka, IPCC technical guidelines for assessing climate change impacts and adaptations, special report to Working Group II, Intergov. Panel on Climate Change, 59 pp., 1994.
- Conway, D., and P. D. Jones, POPSICLE—Production of precipitation scenarios for climate impacts in Europe, final report, 50 pp., Eur. Commun. Environ. Res. Programme, Clim. Res. Unit 1996.
- Conway, D., and P. D. Jones, The use of weather types and air flow indices for GCM downscaling, *J. Hydrol.*, in press, 1998.
- Conway, D., R. L. Wilby, and P. D. Jones, Precipitation and air flow indices over the British Isles, *Clim. Res.*, **7**, 169–183, 1996.
- Crane, R. G., and B. C. Hewitson, Doubled CO₂ precipitation changes for the Susquehanna Basin: Downscaling from the GENESIS general circulation model, *Int. J. Climatol.*, **17**, 1–12, 1997.
- Daly, C., R. P. Neilson, and D. L. Phillips, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, **33**, 140–158, 1994.
- Duckstein, L., A. Bardossy, and I. Bogardi, Linkage between the occurrence of daily atmospheric circulation patterns and floods: An Arizona case study, *J. Hydrol.*, **143**, 413–428, 1993.
- Favis-Mortlock, D. T., R. Evans, J. Boardman, and T. M. Harris, Climate change, winter wheat yield and soil erosion on the English South Downs, *Agri. Syst.*, **37**, 415–433, 1991.
- Foufoula-Georgiou, E., and D. P. Lettenmaier, A Markov renewal model for rainfall occurrences, *Water Resour. Res.*, **23**, 875–884, 1987.
- Giorgi, F., and L. O. Mearns, Approaches to the simulation of regional climate change: A review, *Rev. Geophys.*, **29**, 191–216, 1991.
- Gleick, P. H., Regional hydrologic consequences of increases in atmospheric CO₂ and other trace gases, *Clim. Change*, **10**, 137–161, 1987.
- Gregory, J. M., T. M. L. Wigley, and P. D. Jones, Application of Markov models to area-average daily precipitation series and inter-annual variability in seasonal totals, *Clim. Dyn.*, **8**, 299–310, 1993.
- Grotch, S. L., and M. C. MacCracken, The use of general circulation models to predict regional climatic change, *J. Clim.*, **4**, 286–303, 1991.
- Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers, Simulation of precipitation by weather type analysis, *Water Resour. Res.*, **27**, 493–501, 1991.
- Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers, Use of weather types to disaggregate general circulation model predictions, *J. Geophys. Res.*, **97**, 2781–2790, 1992.
- Henderson-Sellers, A., Continental vegetation as a dynamic component of a global climate model: A preliminary assessment, *Clim. Change*, **23**, 337–377, 1993.
- Hewitson, B. C., and R. G. Crane, Large-scale atmospheric controls on local precipitation in tropical Mexico, *Geophys. Res. Lett.*, **19**, 1835–1838, 1992a.
- Hewitson, B. C., and R. G. Crane, Regional climate prediction from the GISS GCM, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **97**, 249–267, 1992b.
- Hewitson, B. C., and R. G. Crane (Eds.), *Neural Nets: Looks and Uses, in Applications in Geography*, Kluwer Acad., Norwell, Mass., 1994.
- Hewitson, B. C., and R. G. Crane, Climate downscaling: Techniques and application, *Climate Res.*, **7**, 85–95, 1996.
- Heyen, H., E. Zorita, and H. von Storch, Statistical downscaling of monthly mean N. Atlantic air-pressure to sea level anomalies in the Baltic Sea, *Tellus, Ser. A*, **48**, 312–323, 1996.
- Hostetler, S. W., Hydrologic and atmospheric models: The (continuing) problem of discordant scales, *Clim. Change*, **27**, 345–350, 1994.
- Hughes, J. P., and P. Guttorp, A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res.*, **30**, 1535–1546, 1994.
- Hulme, M., K. R. Briffa, P. D. Jones, and C. A. Senior, Validation of GCM control simulations using indices of daily airflow types over the British Isles, *Clim. Dyn.*, **9**, 95–105, 1993.
- Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, The second Hadley Center coupled model ocean-atmosphere GCM: Model description, spinup, and validation, *Clim. Dyn.*, **13**, 103–134, 1997.
- Johnson, N. L., S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, 2nd Ed., 565 pp., Wiley-Interscience, New York, 1992.
- Jones, P. D., M. Hulme, and K. R. Briffa, A comparison of Lamb circulation types with an objective classification scheme, *Int. J. Climatol.*, **13**, 655–663, 1993.
- Kalnay, E., et al., The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471, 1996.
- Karl, T. R., W. C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, A method of relating general circulation model simulated climate to the observed local climate, I, Seasonal statistics, *J. Clim.*, **3**, 1053–1079, 1990.
- Kattenberg, A., F. Giorgi, H. Grassl, G. A. Meehl, J. F. B. Mitchell, R. J. Stouffer, T. Tokioka, A. J. Weaver, and T. M. L. Wigley, Climate models projections of future climate. In *Climate Change 1995: The Science of Climate Change, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., pp. 285–357, Cambridge Univ. Press, New York, 1996.
- Katz, R. W., An application of chain-dependent processes to meteorology, *J. Appl. Prob.*, **14**, 598–603, 1977a.
- Katz, R. W., Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, **16**, 671–676, 1977b.
- Kim, J. W., J. T. Chang, N. L. Baker, D. S. Wilks, and W. L. Gates, The statistical problem of climate inversion: Determination of the relationship between local and large-scale climate, *Mon. Weather Rev.*, **112**, 2069–2077, 1984.
- Leggett, J. A., W. J. Pepper, and R. J. Swart, Emissions scenarios for IPCC: An update, In *Climate Change, 1992. The Supplementary Report to the IPCC Scientific Assessment*, edited by J. T. Houghton, B. A. Callander, and S. K. Varney, pp. 69–95, Cambridge Univ. Press, New York, 1992.
- Levitus, S., *Climatological Atlas of the World Ocean*. NOAA Prof. Pap. 13, U.S. Gov. Print. Off., Washington, D. C., 1982.
- Maochang, C., H. von Storch, and E. Zorita, Coastal sea level and large-scale climate state: A downscaling exercise for the Japanese Islands, *Tellus, Ser. A*, **47**, 132–144, 1995.
- Matyasowszky, I., I. Bogardi, A. Bardossy, and L. Duckstein, Space-time precipitation reflecting climate change, *Hydrol. Sci. J.*, **38**, 539–558, 1993.
- McCabe, G. J., and R. A. Muller, Synoptic weather types: An index of evaporation in southern Louisiana, *Phys. Geogr.*, **8**, 99–112, 1987.
- Mearns, L. O., C. Rosenzweig, and R. Goldberg, The effect of changes in daily and interannual climatic variability on CERES-wheat: A sensitivity study, *Clim. Change*, **32**, 257–292, 1996.
- Mitchell, J. F. B., and T. C. Johns, On modification of global warming by sulfate aerosols, *J. Clim.*, **10**, 245–267, 1997.
- Pittock, A. B., On causes of local precipitation anomalies, with special reference to precipitation in Washington State, *J. Appl. Meteorol.*, **16**, 223–230, 1977.
- Richardson, C. W., Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, **17**, 182–190, 1981.
- Rind, D., C. Rosenzweig, and R. Goldberg, Modelling the hydrological cycle in assessments of climate change, *Nature*, **358**, 119–122, 1992.
- von Storch, H., E. Zorita, and U. Cubasch, Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime, *J. Clim.*, **6**, 1161–1171, 1993.
- Wigley, T. M. L., P. D. Jones, K. R. Briffa, and G. Smith, Obtaining sub-grid scale information from coarse resolution general circulation model output, *J. Geophys. Res.*, **95**, 1943–1953, 1990.
- Wilby, R. L., The influence of variable weather patterns on river water quantity and quality regimes, *Int. J. Climatol.*, **13**, 447–459, 1993.
- Wilby, R. L., Non-stationarity in daily precipitation series: Implications for GCM downscaling using atmospheric circulation indices, *Int. J. Climatol.*, **17**, 439–454, 1997.
- Wilby, R. L., Modelling high magnitude rainfall events using weather pattern and frontal analysis, *J. Hydrol.*, in press, 1998.

- Wilby, R. L., and T. M. L. Wigley, Downscaling general circulation model output: A review of methods and limitations, *Prog. Phys. Geogr.*, 21, 530–548, 1997.
- Wilby, R. L., B. Greenfield, and C. Glenny, A coupled synoptic-hydrological model for climate change impact assessment, *J. Hydrol.*, 153, 265–290, 1994.
- Wilby, R. L., H. Hassan, and K. Hanaki, Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, 205, 1–19, 1998.
- Wilby, R. L., T. M. L. Wigley, D. S. Wilks, B. C. Hewitson, D. Conway, and P. D. Jones, Statistical downscaling of general circulation model output, a report prepared by the National Center for Atmospheric Research, 730 pp., Electr. Power Res. Inst., Palo Alto, Calif., 1996.
- Wilks, D. S., Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resour. Res.*, 25, 1429–1439, 1989.
- Wilks, D. S., Adapting stochastic weather generation algorithms for climate change studies, *Clim. Change*, 22, 67–84, 1992.
-
- D. Conway and P. D. Jones, Climatic Research Unit, University of East Anglia, Norwich, England NR4-7TJ.
- B. C. Hewitson and J. Main, Environmental and Geophysical Science, University of Cape Town, Rondebosch 7700 South Africa.
- T. M. L. Wigley, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
- D. S. Wilks, Atmospheric Science Group, Cornell University, Ithaca, NY 14853.

(Received February 17, 1998; revised July 1, 1998; accepted August 3, 1998.)