Vol. 11: 125–136, 1999 | CLIMATE RESEARCH Clim Res | Published March 22

Multisite downscaling of daily precipitation with a stochastic weather generator

D. S. Wilks*

Atmospheric Science Group, Cornell University, Ithaca, New York 14853, USA

ABSTRACT: Stochastic models of daily precipitation are useful both for characterizing different precipitation climates and for stochastic simulation of these climates in conjunction with agricultural, hydrological, or other response models. A simple stochastic precipitation model is used to downscale—i.e. disaggregate from area-average to individual station—precipitation statistics for 6 groups of 5 U.S. stations, in a way that is consistent with observed relationships between the area-averaged series and their constituent station series. Each group of stations is located within a General Circulation Model grid-box-sized area, and collectively they exhibit a broad range of precipitation climates. The downscaling procedure is validated using natural climate variability in the observed precipitation records as an analog for climate change, by alternately considering collections of the driest and wettest seasons as 'base' and 'future' climates, and comparing the 2 sets of downscaled station parameters to those fit directly to the respective withheld observations. The resulting downscaled stochastic model parameters can be readily used for local-scale simulation of climate-change impacts.

KEY WORDS: Downscaling · Climate change · Precipitation · Stochastic modelling · United States

1. INTRODUCTION

Global, coupled atmosphere-ocean General Circulation Models (GCMs) are the most powerful tools available to simulate evolving and future changes in the climate system. While these models are most accurate at large (continental, hemispheric, and global) spatial scales (Gates et al. 1996), smaller-scale (at or near the spatial resolution of the GCMs) climatic details are less well portrayed (Mearns et al. 1990, Grotch & MacCracken 1991, Kattenberg et al. 1996). However, it is the changes in near-surface local climates that will determine the consequences for agricultural, ecological, hydrological, and other life-sustaining systems.

Accordingly, a variety of approaches to the 'downscaling' of grid-scale (hundreds of km) GCM information to local-scale surface climate have been devised (e.g. reviews by Giorgi & Mearns 1991, Wilby & Wigley 1997). These range from quite simple to extremely elaborate. A very simple but widely used approach (e.g. Cohen 1990) is direct adjustment of instrumental weather records at a location according to an assumed climate change. For example, all measured temperatures might be increased by 2°C, and all observed precipitation amounts multiplied by 1.10. Other investigators have devised more elaborate schemes based on statistical relationships between large-scale atmospheric circulation patterns and local surface weather variables (e.g. Wigley et al. 1990, Bardossy & Plate 1992, von Storch et al. 1993, Burger 1996, Lettenmaier 1995). At the extreme of complexity and computational demands are mesoscale models nested within (i.e. provided with boundary conditions by) a GCM (Giorgi 1990, Mearns et al. 1995).

A fairly simple but flexible and computationally economical approach to producing local climate-change 'scenarios' is through the use of stochastic weather models, or 'weather generators' (Wilks 1992, Woo 1992, Semenov & Barrow 1997). These are statistical models for daily weather data at a single location (e.g. Richardson 1981), which can be regarded alternatively as statistical characterizations of the local climate, or as elaborate random number generators whose output resembles real weather data. Their use in climate-change studies involves perturbing the stochastic model parameters to reflect a changed climate, and

*E-mail: dsw5@cornell.edu

then generating synthetic weather series consistent with this new climate for use with impact (e.g. agricultural or hydrological) models.

Differences in climate statistics between local and area-averaged (presumably corresponding to GCM grid cell) series are quite large, especially for precipitation. In particular, the probability of precipitation in area-averaged series is necessarily larger than the corresponding probabilities at any of the constituent stations, while the average precipitation amounts are smaller (e.g. Osborn & Hulme 1997). While procedures for adjusting the stochastic model parameters to yield desired altered climates have been published previously (Wilks 1992, Katz 1996), the change of spatial scale between the GCM grid and the local stations has not yet been considered explicitly in this context. In the following, the method of Wilks (1992) for construction of climate-change scenarios with stochastic weather models is extended to accommodate the differences in local and area-averaged weather statistics, and then validated using observed interannual climatic variability as an analog for climate change. While the emphasis is on precipitation modeling and precipitation statistics, the procedure could be applied also to temperature and other surface-weather variables.

2. METHODS

2.1. Study areas and stations. Daily precipitation data from 30 U.S. stations, listed in Table 1, are used in the following. Table 1 shows that these stations are grouped into 6 areas, denoted OR (stations in the state of Oregon), CA/NV (stations primarily in California and Nevada), MN (stations in Minnesota), OK (stations primarily in Oklahoma), MS (stations primarily in Mississippi), and CNE (stations primarily on or near the coast of the northeastern U.S.). As indicated by the average seasonal precipitation values in Table 1, the stations span a wide range of precipitation climates. These stations and areas have been chosen as part of a larger project (Wilby et al. 1996, 1998, Wilby & Wigley

Table 1. Time period, location, and precipitation data for the 30 stations (6 areas) used in this study. OR: Oregon; CA/NV: primarily California and Nevada; MN: Minnesota; OK: primarily Oklahoma; MS: primarily Mississippi; CNE: primarily on or near coast of northeastern U.S.

Area	Time period (mo/yr)	Stn	Location name	Elevation (m)	Latitude (°N)	Longitude (°W)	Average precipitation (mm)			
							DJF	MAM	JJA	SON
OR	(9/48–2/98)	1	Eugene, OR	109	44.12	123.22	556	273	69	326
		2	Portland, OR	6	45.60	122.60	396	222	85	262
		3	Redmond, OR	933	44.27	121.15	72	53	43	51
		4	Salem, OR	60	44.90	123.00	462	223	64	283
		5	Astoria, OR	2	46.15	123.88	721	383	129	495
CA/NV	(12/33–2/98)	1	Medford, OR	396	42.37	122.87	215	105	36	135
		2	Winnemucca, NV	1310	40.90	117.80	66	65	34	49
		3	Reno, NV	1342	39.50	119.78	80	46	23	38
		4	Sacramento, CA	5	38.52	121.50	267	118	6	90
		5	Red Bluff, CA	104	40.15	122.25	316	144	17	127
MN	(3/48–11/97)	1	Alexandria, MN	432	45.87	95.38	56	165	277	140
		2	Minneapolis, MN	254	44.88	93.22	69	182	297	158
		3	Rochester, MN	395	44.00	92.45	65	204	306	172
		4	St. Cloud, MN	313	45.58	94.18	57	174	306	156
		5	Redwood Falls, MN	312	44.55	95.08	48	178	290	141
OK	(3/48–11/97)	1	Oklahoma City, OK	392	35.40	97.60	101	273	250	217
		2	Tulsa, OK	198	36.20	95.90	135	313	278	262
		3	McAlester, OK	232	34.92	95.77	183	358	269	314
		4	Hobart, OK	168	35.00	99.05	69	218	203	173
		5	Wichita Falls, TX	303	33.98	98.52	95	231	195	187
MS	(12/48–11/97)	1	Jackson, MS	87	32.32	90.08	387	423	302	292
		2	Meridian, MS	88	32.33	88.75	401	412	323	266
		3	Greenwood, MS	47	33.50	90.20	383	405	278	281
		4	McComb, MS	126	31.25	90.47	442	438	381	299
		5	Monroe, LA	24	32.52	92.05	369	388	279	272
CNE	(6/48–11/97)	1	Williamsport, PA	160	41.25	76.92	216	263	290	263
		2	New York City, NY	4	40.65	73.78	242	285	276	248
		3	Baltimore, MD	45	39.18	76.67	247	280	294	254
		4	Atlantic City, NJ	20	39.45	74.58	250	253	267	226
		5	Philadelphia, PA	2	39.88	75.23	236	272	295	237

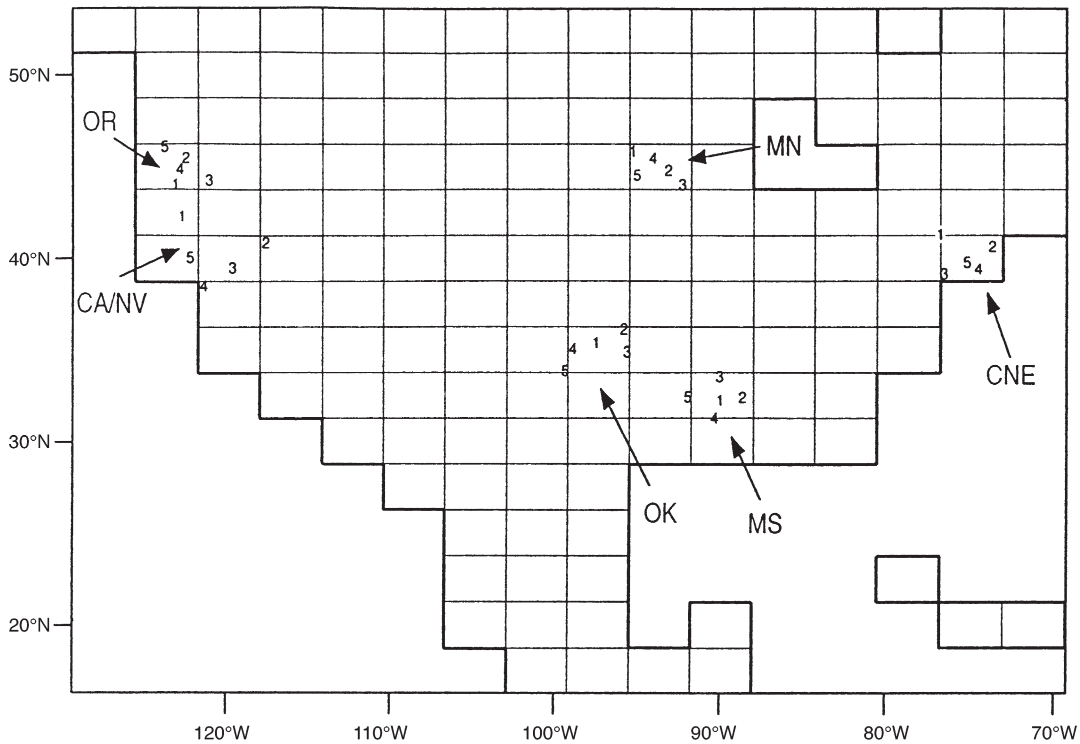

Fig. 1. Locations of the 6 areas in Table 1 and their constituent stations superimposed on the 2.50° × 3.75° HadCM2 grid over a portion of North America

1997), to correspond to 6 grid-boxes of the HadCM2 coupled atmosphere-ocean GCM (Johns et al. 1997, Mitchell & Johns 1997). Fig. 1 shows the 2.50° latitude × 3.75° longitude grid boxes of the HadCM2 model over a portion of North America, with the locations of the 30 stations in Table 1 indicated as well. It is evident from Fig. 1 that each of the 6 groups of 5 stations occupies an area that is comparable in size to 1 HadCM2 grid box, and that the station groups are located within or very near 1 of 6 particular grid boxes.

The geographic correspondence between the study areas and the HadCM2 grid will be taken as only incidental here. Rather, the focus below will be on the relationships between precipitation statistics for the individual daily station series and their counterparts for the corresponding area averages, which represent information on the spatial scale available from a GCM. For simplicity it will be assumed that each of the 6 area-averaged daily precipitation series is comprised of an unweighted average of the non-missing daily precipitation amounts for its 5 constituent stations, although extension of the methods presented below to weighted area averages would be straightforward. The analyses are stratified into the 4 standard climatological seasons, DJF, MAM, JJA and SON. The periods of data availability for the 6 areas, which are also indicated in Table 1, have been chosen so that at least 4 of each set of 5 stations are active throughout.

2.2. Stochastic models. The basic model for daily precipitation employed here is the well-known and widely used 'chain-dependent process' (Todorovic & Woolhiser 1975, Katz 1977). In this model, daily precipitation occurrence is governed by a 2-state (either precipitation occurs or it does not), first-order (the probability of precipitation depends only on whether or not precipitation occurred on the previous day) Markov chain. This process can be characterized by the 2 conditional probabilities

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

and

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

which are known as transition probabilities. Because only 2 precipitation occurrence states are defined, the complementary probabilities for dry-day occurrences are given by $p_{00} = 1 - p_{01}$ and $p_{10} = 1 - p_{11}$. The transition probabilities are very convenient for stochastic simulation (see Appendix 1). For the purpose of statistically characterizing a given precipitation climate it is often more convenient to define the Markov chain using the 2 parameters

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

and

$$r = p_{11} - p_{01} \quad (2b)$$

which are the unconditional probability (i.e. climatological relative frequency) of daily precipitation occurrence and the lag-1 autocorrelation of the daily precipitation occurrence series, respectively. The 2 parameter sets $\{p_{01}, p_{11}\}$ and $\{\pi, r\}$ clearly embody equivalent information, as one set can be computed from the other using Eq. (2) (see Eq. A1). Typically one finds positive serial dependence in daily precipitation occurrence data (wet and dry days tend to occur in runs or 'clumps'), so that $r > 0$ and $p_{01} < \pi < p_{11}$.

The remainder of the chain-dependent process consists of a statistical model for the nonzero precipitation amounts. Most commonly the gamma distribution is chosen for this purpose (e.g. Katz 1977, Richardson & Wright 1984, Wilks 1989, 1992, Gregory et al. 1993), although some investigators have found that the mixed exponential distribution provides a much better fit to daily nonzero precipitation amounts (Woolhiser & Roldan 1982, Foufoula-Georgiou & Lettenmaier 1987, Wilks 1998a). Use of these 2 distributions for both downscaling and stochastic simulation is detailed in Appendix 1. For now, however, the choice of a particular probability model for the precipitation amounts can remain implicit, since in the following it will be necessary to know only the mean daily nonzero precipitation amount, μ, and the corresponding variance, σ^2.

The climatological statistics of seasonal-total precipitation depend on the statistics of the daily precipitation amounts that comprise the seasonal total. Similarly, the seasonal statistics of synthetic precipitation generated by a daily stochastic model depend on the parameters governing the daily precipitation occurrences and amounts. Let $S(T)$ be the sum of T daily precipitation amounts. Since seasonal precipitation is considered here, $T \approx 90$. The average seasonal precipitation, $E[S(T)]$, and its variance, $Var[S(T)]$ (which characterizes the interannual variability of seasonal precipitation), can be expressed as (e.g. Gregory et al. 1993, Katz & Parlange 1998)

$$E[S(T)] = E[N(T)]\,\mu \quad (3a)$$

and

$$Var[S(T)] = E[N(T)]\,\sigma^2 + Var[N(T)]\,\mu^2 \quad (3b)$$

As noted above μ and σ^2 are the mean and variance of the nonzero precipitation amounts, regardless of the particular distribution chosen to represent them. The quantity $E[N(T)]$ is the average number of days with nonzero precipitation in a period of T consecutive days and $Var[N(T)]$ is the interannual variance of the number of wet days. For first-order Markov dependence, these latter 2 quantities can be expressed as

$$E[N(T)] = T\pi \quad (4a)$$

and

$$Var[N(T)] \approx \frac{T\pi(1-\pi)(1+r)}{(1-r)} \quad (4b)$$

Eq. (4b) is an approximation to the true variance, although for monthly and longer totals this approximation is found to be very close (Gabriel & Neumann 1962, Gregory et al. 1993).

2.3. Downscaling. As conventionally understood, the process of 'downscaling' GCM information to produce scenarios of climate change at local scales is actually a combination of 2 operations. First, because of approximations necessarily made in the formulation of any GCM, the 'control' GCM climate is to a greater or lesser degree different from the observed climate, even when it is compared to area-averaged observations at a comparable scale. Accordingly, results from changed-climate GCM integrations are not downscaled directly, but rather some adjustment is made to the climatological observations or to the statistics of the observations which reflects relative changes between the control and changed-climate GCM integrations. The result, either tacitly or explicitly, is the specification of a changed climate for an area average at the scale of the GCM. Recognizing that this first step does not actually involve a change of spatial scale, it might better be referred to as 'extrapolation' (Wilks 1988, Wilby et al. 1998).

The focus of this paper is on the second part of this process, in which the area-averaged (extrapolated) climates or climate statistics are downscaled, i.e. disaggregated to the individual station level. It will therefore be assumed that the extrapolated area-average stochastic model parameter set $\{\pi_{ex}, r_{ex}, \mu_{ex}, \sigma^2_{ex}\}$ for the changed climate is available. There are a number of ways that these extrapolated parameters might be arrived at. One possibility (Wilby et al. 1998) is related mathematically and conceptually to the downscaling procedure outlined below. Alternatively, one could imagine that changes in the 4 quantities $E[S(T)]$, $Var[S(T)]$, $E[N(T)]$ and $Var[N(T)]$ might be computed from many months or seasons of GCM output, with the corresponding extrapolated $\{\pi_{ex}, r_{ex}, \mu_{ex}, \sigma^2_{ex}\}$ then calculated through Eqs. (3) & (4). It might also be possible to base the extrapolation of these parameters on changes in atmospheric circulation statistics, following approaches similar to Wigley et al. (1990) or von Storch et al. (1993). Any of these procedures place great reliance on particular GCM results, however, and for some purposes it could be desirable instead to extrapolate more generalized climate-change scenarios (Wilks 1992, Hulme et al. 1995).

The basic downscaling problem, then, is the following. One has from observations the parameter sets $\{\pi_{station}, r_{station}, \mu_{station}, \sigma^2_{station}\}$ fit to each of the (for the present data, 5) observed station-level daily precipitation series within an area, and the set $\{\pi_{area}, r_{area}, \mu_{area}, \sigma^2_{area}\}$ fit to the corresponding time series of area-averaged daily precipitation amounts. One also has an extrapolated changed precipitation climate, at the same spatial scale as the area-averaged time series, characterized by the parameter set $\{\pi_{ex}, r_{ex}, \mu_{ex}, \sigma^2_{ex}\}$. The task of the downscaling procedure is to disaggregate the extrapolated changed climate to produce the (again, 5) downscaled parameter sets $\{\pi_{down}, r_{down}, \mu_{down}, \sigma^2_{down}\}$. Of course this process is not uniquely defined, and the procedure presented below is but one of very many alternatives that could be imagined.

Fundamental to good statistical representation of a precipitation climate is the accurate portrayal of the climatological wet-day probability, π, since this parameter governs the average number of wet days in a month or season (Eq. 4a), and through this also strongly affects the average monthly or seasonal precipitation (Eq. 3a). Theoretical relationships between π_{area} and $\pi_{station}$ have been developed (Epstein 1966, Osborn & Hulme 1997), but their application here is problematic because of the extremely varied precipitation climates within some of the present study areas. In Area OR for example, there are 3 rather different precipitation climates represented by the 5 stations which result from 2 north-south topographic barriers interacting with the prevailing westerly atmospheric flow. Astoria (Stn 5) is a very wet location on the Pacific coast of Oregon. The 3 stations Eugene (Stn 1), Portland (Stn 2) and Salem (Stn 4) are located in the somewhat drier Willamette Valley, which is separated from the ocean by the Coast Range mountains (typical elevations about 750 m). The semidesert station Redmond (Stn 3) is located still further east, in the rain shadow of the Cascade Range mountains (typical elevations about 2000 m).

The climatological wet-day probability is downscaled here by adjusting each of the observed $\pi_{station}$ values according to differences between the observed π_{area} and the imposed π_{ex}. Direct additive or proportional adjustments have the potential to yield misleading or even nonsense results (e.g. $\pi_{down} < 0$), particularly for stations like Redmond, for which $\pi_{station} \ll \pi_{area}$. This problem is avoided by first transforming to the log-odds scale, which is defined by

$$L(\pi) = \ln\left[\frac{\pi}{1-\pi}\right] \quad (5)$$

The downscaled climatological probability is then computed by adjusting the log-odds transformed $\pi_{station}$ by the difference of the log-odds transforms of π_{ex} and π_{area},

$$\pi_{down} = L^{-1}[L(\pi_{station})][L(\pi_{station}) + L(\pi_{ex}) - L(\pi_{area})] \quad (6a)$$

$$= \frac{\exp[L(\pi_{station}) + L(\pi_{ex}) - L(\pi_{area})]}{1+\exp[L(\pi_{station}) + L(\pi_{ex}) - L(\pi_{area})]} \quad (6b)$$

Since Eq. (5) transforms the probabilities π from the unit interval to the real line, Eq. (6) necessarily produces a properly bounded result ($0 < \pi_{down} < 1$).

Similarly, the parameter r is a correlation and must be bounded by $-1 < r < 1$. Adjustments to this parameter can be made on the scale of the Fischer Z-transform (i.e. the inverse hyperbolic tangent),

$$Z(r) = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] \quad (7)$$

which transforms correlations from the interval [–1, 1] to the real line. Following the same idea as for the wet-day probability, the correlation r can be downscaled by making an additive adjustment on this transformed scale, using

$$r_{down} = Z^{-1}[Z(r_{station}) + Z(r_{ex}) - Z(r_{area})] \quad (8a)$$

$$= \frac{\exp\{2[Z(r_{station}) + Z(r_{ex}) - Z(r_{area})]\} - 1}{\exp\{2[Z(r_{station}) + Z(r_{ex}) - Z(r_{area})]\} + 1} \quad (8b)$$

which will necessarily yield $-1 < r_{down} < 1$.

It remains to downscale the 2 parameters controlling nonzero precipitation amounts, μ and σ^2. While there are potentially many ways to approach this task, it seems clear that any reasonable procedure should preserve consistent relationships between the statistics of local and area-averaged seasonal precipitation, $S(T)$. The (time) average of each area-averaged precipitation series is just the average (time) mean over each of the n stations comprising the area average,

$$E[S(T)_{area}] = \frac{1}{n}\sum_{station=1}^{n} E[S(T)_{station}] \quad (9a)$$

Similarly, for the variances,

$$Var[S(T)_{area}] \propto \frac{1}{n}\sum_{station=1}^{n} Var[S(T)_{station}] \quad (9b)$$

holds (Kagan 1966, Jones et al. 1997). The constant of proportionality in Eq. (9b) depends on the average interstation correlation, $\bar{c}$, among the n stations comprising the area average, and is equal to $[1 + (n-1)\bar{c}]/n$. The extent to which this correlation structure might change in a changing climate would not be known on the basis of a GCM integration, which could not resolve it, although changes in the interstation precipitation correlations would influence the interannual variance of the area-averaged series (Osborn 1997). While results from a mesoscale model nested in the GCM could suggest changes in this correlation structure, their existence might also obviate the need for statistical downscaling. In the following it will be assumed that the proportionality in Eq. (9b) remains unchanged in a changed climate, although a different proportionality could also be assumed.

There are many ways in which the n station means and variances on the right-hand sides of Eq. (9) might change to yield particular extrapolated values $E[S(T)_{ex}]$ and $Var[S(T)_{ex}]$. As before, these would not be known from GCM output, and results from a nested mesoscale model suggesting a particular set of changes might also render statistical downscaling unnecessary. Here it will be assumed that changes in the station-series means and variances will be proportional to the changes in the respective area-averaged moments, yielding

$$E[S(T)_{down}] = E[S(T)_{station}]\frac{E[S(T)_{ex}]}{E[S(T)_{area}]} \quad (10a)$$

and

$$Var[S(T)_{down}] = Var[S(T)_{station}]\frac{Var[S(T)_{ex}]}{Var[S(T)_{area}]} \quad (10b)$$

Using Eqs. (3) & (4), together with the downscaled occurrence parameters from Eqs. (6) & (8), yields

$$\mu_{down} = \frac{E[S(T)_{down}]}{T\,\pi_{down}} \quad (11a)$$

and

$$\sigma^2_{down} = \frac{Var[S(T)_{down}]}{T\,\pi_{down}} - \frac{(1-\pi_{down})(1+r_{down})}{(1-r_{down})}\mu^2_{down} \quad (11b)$$

for the downscaled precipitation-amount parameters.

3. RESULTS AND VALIDATION

The downscaling procedure just described relies on a number of somewhat arbitrary choices and assumptions. It is natural, therefore, that one would like to validate the scheme in a setting where the correct results are known and can be compared to the downscaled values. Validation exercises of this kind have apparently not been previously applied to climate-change downscaling procedures, presumably because the future local-scale climates to which they pertain are not known, and downscaling of any kind would be moot if they were known.

However, it is possible to validate a downscaling procedure, at least in a limited way, by using observed interannual climatic variability as an analog for climate change. Analog methods have certain limitations for climate-change studies (e.g. Giorgi & Mearns 1991), and in the present context require the assumption that relationships between local and area-averaged statistics will remain stable, even though future changes in the climatic statistics may be driven by physical processes different from those responsible for the climate variations in the instrumental record. Here climate-change analogs have been constructed by stratifying the data, separately for each of the 4 seasons and each of the 6 areas, according to the total area-averaged seasonal precipitation. Data from years comprising the driest 40% of each of the 4 seasons (20 yr for all areas except Area CA/NV; 25 yr for Area CA/NV) were defined to constitute the dry climates, and data from the wettest 40% constituted the wet climates. The remaining 20% of the data were not used. These percentages are subjective choices meant to balance clear separation between the wet and dry climates, with the need for adequate sample size.

Separate parameter sets $\{\pi, r, \mu, \sigma^2\}$ were computed for the dry years and the wet years, for both the station and area-averaged series. The downscaling procedure described in Section 2.3 was then applied twice for each area and for each of the 4 seasons. In the first application, the wet years were considered to be the base, or present, climate. The dry years then play the role of the changed future climate, so that the parameters for the dry area-averaged series correspond to the extrapolated climate at the area-average scale. The parameter set $\{\pi_{down}, r_{down}, \mu_{down}, \sigma^2_{down}\}$ produced by the downscaling procedure should then correspond to, and can be compared with, the observed station-level parameters from the dry years. In the second application of the downscaling procedure the roles of the wet and dry years are reversed, so that the dry years are considered to represent the base climate, the wet years are the changed climate, and the correct values of the downscaled parameters are the observed station parameters in the wet years. This procedure is a fairly severe test of the downscaling algorithm, but is an appropriate simulation of its potential applications.

Figs. 2 & 3 illustrate the results for winter in the OR Area and summer in Area CNE, respectively. As noted previously, the topographic variations in OR area might be expected to challenge the disaggregation procedure. Summer precipitation in Area CNE provides an example of the performance of the scheme for typically small-scale convective precipitation. The panels (a) to (d) in these figures show results for π, r, μ, and σ (not σ^2, for ease of dimensional comparison), respectively. Both the wet to dry, and dry to wet, climate changes for each parameter are shown on the same panel. The arrows point from the base-climate station parameter value to the downscaled parameter value, with the horizontal plotting positions of the arrows indicating the true station-level parameters toward which the downscaling is targeted. Thus, the magnitudes of the downscaled climate changes are proportional to the lengths of the arrows, and perfect downscaling is indicated by the tips of the arrows touching the 1:1 lines exactly. The horizontal separa-

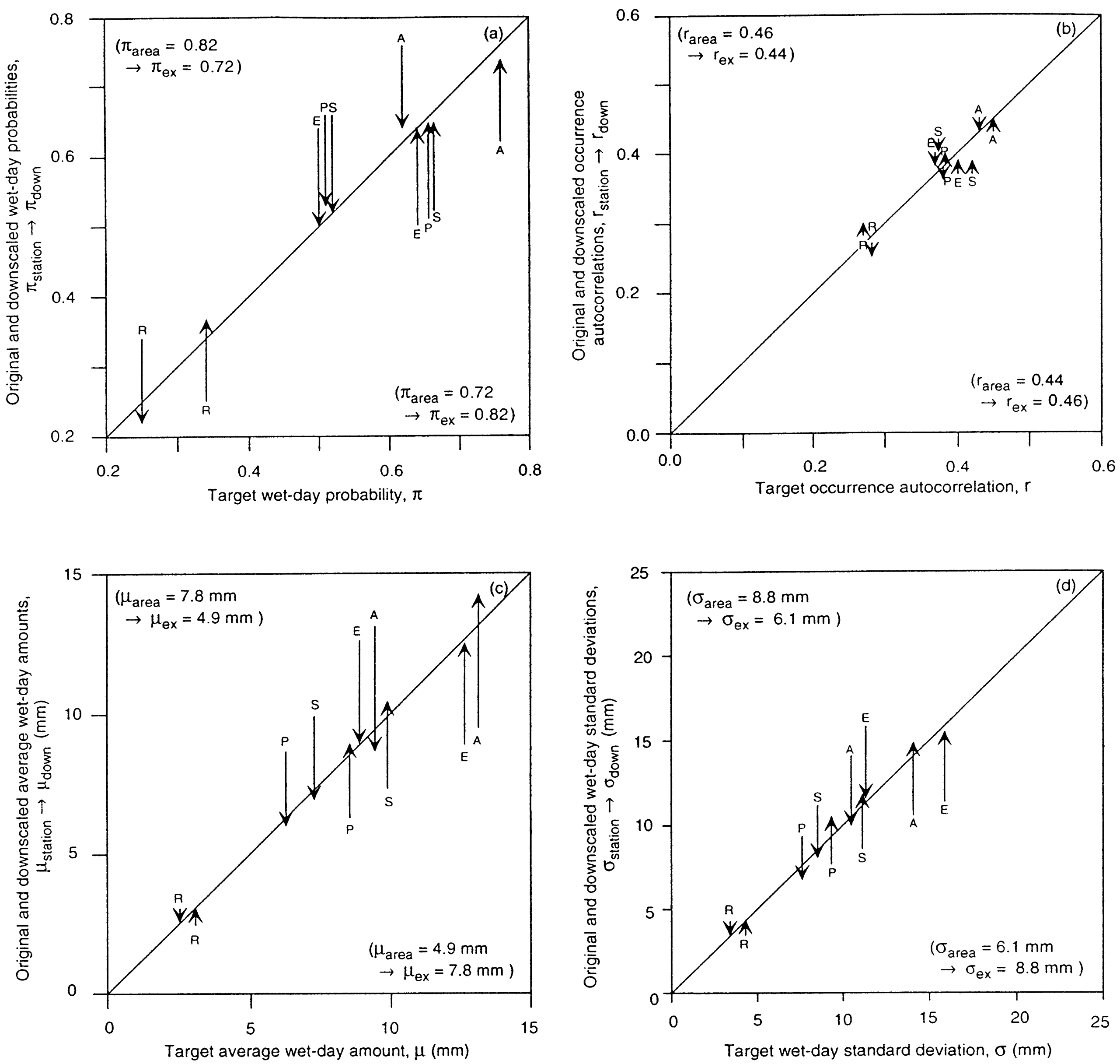

Fig. 2. Downscaling results for the 5 stations of Area OR during the winter (DJF) season, for (a) climatological wet-day probabilities, (b) daily precipitation occurrence autocorrelations, (c) average wet-day precipitation amounts, and (d) standard deviation of wet-day precipitation amounts. Arrows point from the original station parameter to the downscaled station parameter, with the horizontal plotting positions indicating the true station parameters toward which the downscaling is targeted. Letters indicate station names (cf. Table 1). Parameter values in the panel corners indicate original and changed-climate ('extrapolated') parameters for the area-averaged series

tions between pairs of arrows pertaining to the same station indicate the differences in station parameters between the dry and wet years. Parameters in the corners of each panel indicate the wet and dry values for the area-averaged series. In both figures, all 4 of the parameters are larger in the wet years and smaller in the dry years, both at the station and area-average scales. For example, in Fig. 2a, π_{area} = 0.82 for the wet winters and 0.72 for the dry winters. The 5 downward pointing arrows show the downscaling results for the wet base to the dry changed climate at each of the stations, and the 5 upward arrows indicate the downscaling for the dry base to the wet changed climate.

Fig. 2 shows very good results for the downscaling of all 4 parameters, particularly considering the extremely varied precipitation climates within Area OR. There are large changes in both π and μ, and moderate changes in σ, all of which are well represented. Even

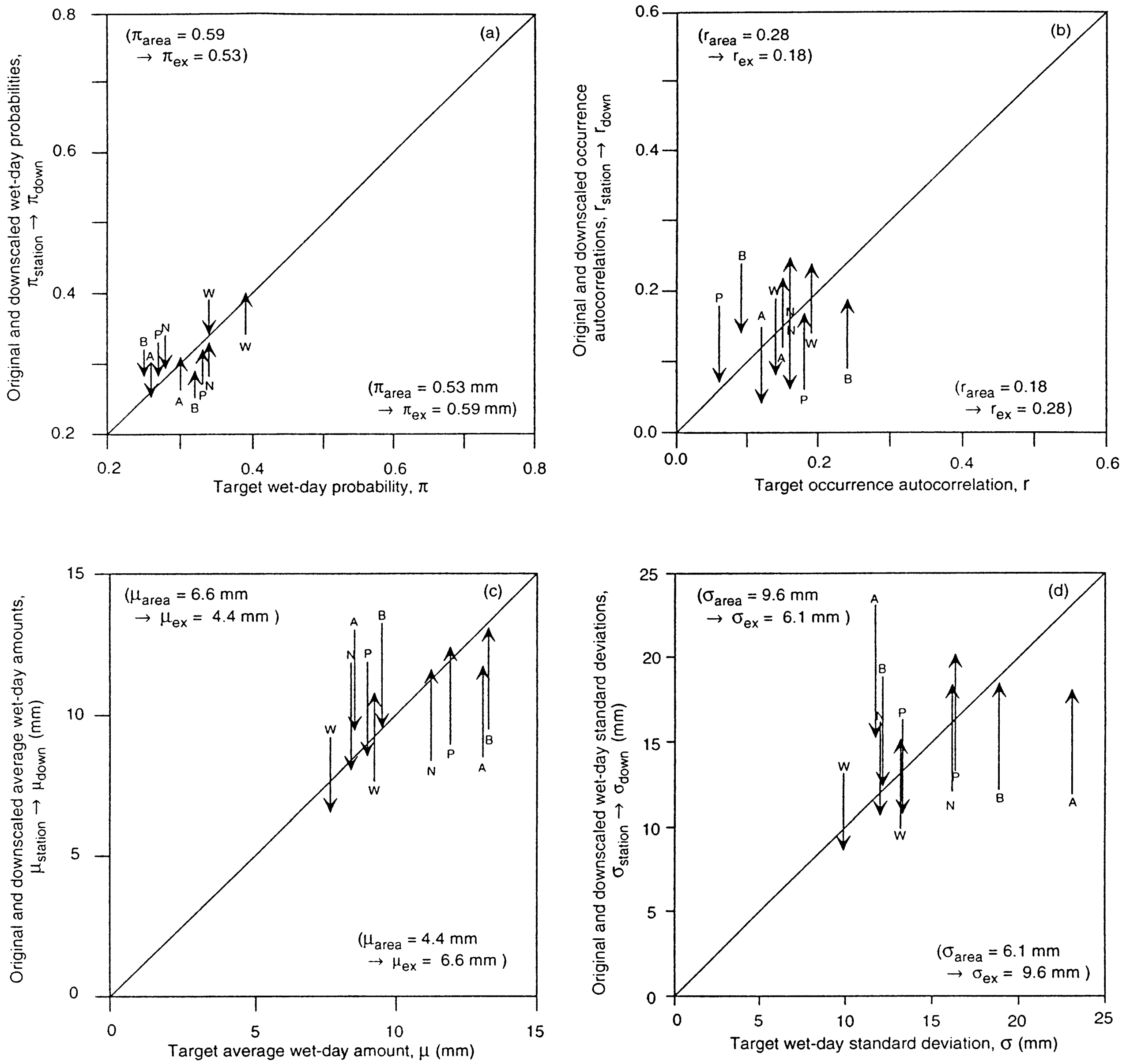

Fig. 3. Downscaling results for the 5 stations of Area CNE during the summer (JJA) season, for (a) climatological wet-day probabilities, (b) daily precipitation occurrence autocorrelations, (c) average wet-day precipitation amounts, and (d) standard deviation of wet-day precipitation amounts. Arrows point from the original station parameter to the downscaled station parameter, with the horizontal plotting positions indicating the true station parameters toward which the downscaling is targeted. Letters indicate station names (cf. Table 1). Parameter values in the panel corners indicate original and changed-climate ('extrapolated') parameters for the area-averaged series

the very small differences in μ and σ for Redmond are captured by the downscaling, despite the fact that the differences between the corresponding area-average parameters are very substantial. The occurrence autocorrelations change very little at either the station or area-averaged scales.

Fig. 3 shows quite good results for the modest changes in π and large changes in μ for summer precipitation in the northeastern U.S. The comparatively large changes in *r* are downscaled less satisfactorily, although the magnitudes of this parameter for summer precipitation in this area are small. Fig. 3d shows that the downscaling errors for *r* in Fig. 3b are propagated into the downscaled precipitation amount variances, which follows from Eqs. (10b) & (11b). The overall results are nevertheless reasonably good.

Finally, Figs. 4 & 5 show boxplots sketching the full distributions of the downscaling errors, with the

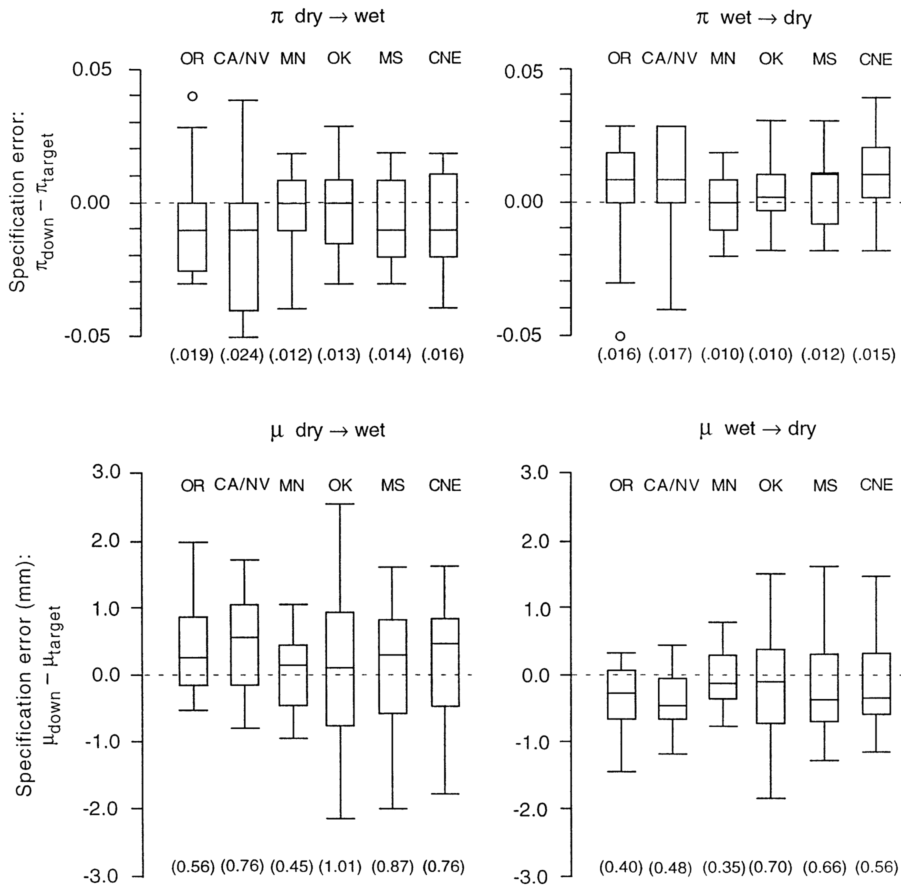

Fig. 4. Distributions of downscaling errors for the parameters π and μ, aggregating the 5 stations and all 4 seasons for each area. Left panels: dry base climate to wet changed climate; right panels: wet base climate to dry changed climates. Mean absolute errors are shown parenthetically below each boxplot

errors for the 5 stations and all 4 seasons aggregated within each of the 6 areas. Fig. 4 shows the errors for π and μ (which jointly determine the mean precipitation, $E[S(T)]$), and Fig. 5 shows the errors for r and σ (which influence only the variance, $Var[S(T)]$). There appears to be a slight bias in the downscaling of π, in the sense that the downscaled values tend on average not to change the original $\pi_{station}$ by quite enough. This results in a compensating small bias in the opposite direction for the downscaled μ (cf. Eq. 11a). Overall, however, the magnitudes of the errors are rather small in most cases, and these figures indicate that the scheme is quite workable overall.

4. SUMMARY AND CONCLUSIONS

This paper has detailed the construction and validation of a statistical downscaling procedure suitable for use with information regarding changes in the climatology of daily precipitation at the scale of a GCM grid box. The procedure is based on characterization of the precipitation climate in terms of a few parameters defining a stochastic weather model, which is convenient also for eventual stochastic simulation of the changed climate and its impacts on agricultural, ecological or hydrological processes.

An important aspect of this study is that the procedure has been validated for 6 diverse regions across the U.S., by using natural climatic variability as a proxy for climate change. For this purpose the wettest 40% and the driest 40% of years were alternately regarded as baseline- and changed climates. This approach might profitably be adopted in other downscaling studies as well, regardless of whether they are statistically or physically based. In order to achieve statistically stable parameter estimates for the validation comparisons, the daily precipitation data has been stratified into the standard 3 mo seasons, DJF, MAM, JJA and SON. However, noticeable variations of the parameters within these periods do occur in some instances, and in practice it might be better to consider aggregation over 1 or 2 mo periods.

Of the basic downscaling equations proposed here, only those for the precipitation amount parameters (Eq. 11) have any theoretical basis (i.e. Eq. 9). Eqs. (6) & (8) for downscaling the precipitation occurrence parameters are ad hoc choices, although they satisfy sensible mathematical constraints, and produce gener-

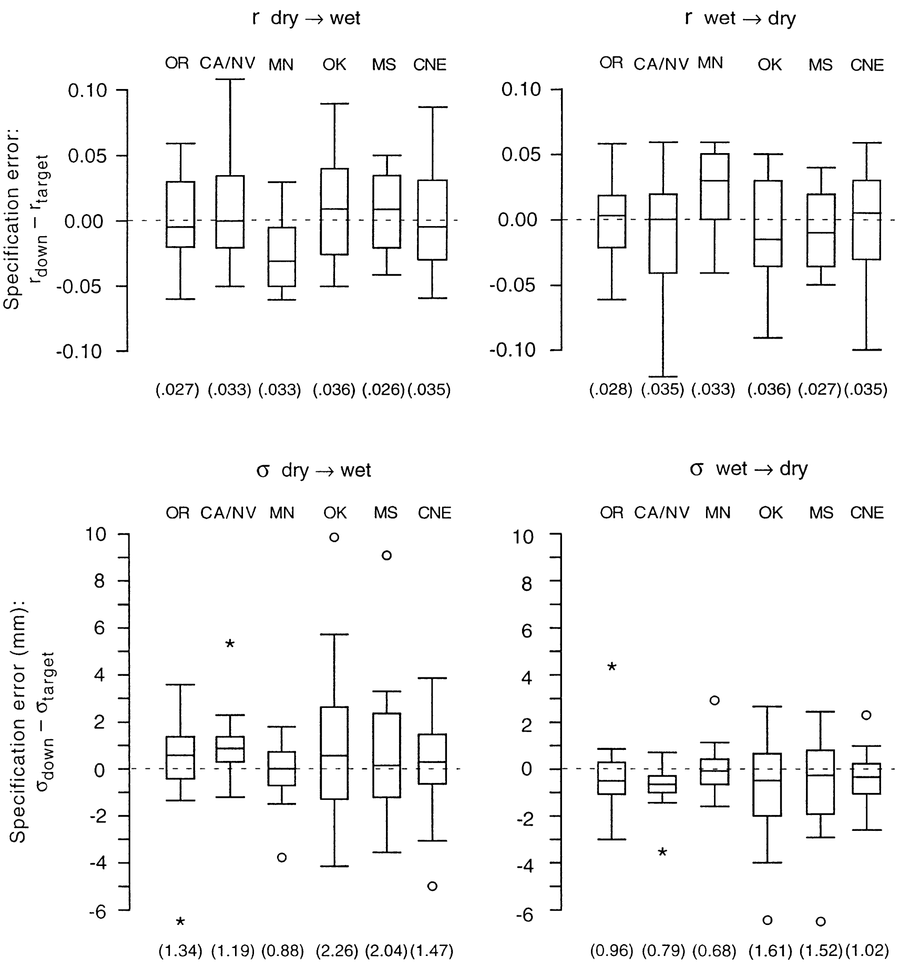

Fig. 5. Distributions of downscaling errors for the parameters r and σ, aggregating the 5 stations and all 4 seasons for each area. Left panels: dry base climate to wet changed climate; right panels: wet base climate to dry changed climates. Mean absolute errors are shown parenthetically below each boxplot

ally good results in the validation comparisons. More satisfying theoretically would be to downscale these parameters using explicit information regarding the physical climatic controls within each grid-box area, such as topography, distance to water bodies, etc. For the present data, attempts to downscale the wet-day probabilities using statistical specifications for individual $\pi_{station}$ in terms of π_{area} and various geographical predictor variables (Wilks 1998b) yielded validation results comparable to those from Eq. (6) only for Area OR (e.g. mean absolute error of 0.020 for the dry to wet climate change; cf. Fig. 4), but decidedly biased and inferior downscaling specifications for the remaining areas (mean absolute errors of 0.036, 0.028, 0.036, 0.062, and 0.022, for the dry to wet change in Areas CA/NV, MN, OK, MS, and CNE, respectively). As noted previously, biased specification of π also leads, through Eq. (11a), to biased downscaling of μ.

Attention has been confined here to downscaling stochastic model parameters for daily precipitation, although in principle similar approaches could be applied for temperature or solar radiation using well-known stochastic models for these quantities (e.g. Richardson 1981). It is possible that this could be a simpler process, since Eqs. (9) & (10) alone might be sufficient to downscale the critical parameters. Attempting this in future work would probably be worthwhile. However, it should be remembered that unexpected results can be obtained when simultaneously altering the parameters for the temperature and precipitation processes in models of this kind (Katz 1996).

One very attractive feature of using stochastic model parameters for downscaling is that the results are immediately in a convenient form for stochastic simulation ('weather generation'), in order to provide input for various impacts models. Appendix 1 outlines procedures for implementing these simulations for a location once the parameters $\{\pi_{down}, r_{down}, \mu_{down}, \sigma^2_{down}\}$ have been computed. It is also possible to extend these simulations to the generation of spatially coherent time series at multiple locations within a grid-box-sized area (Wilks 1998a), which could be particularly useful for investigation of such issues as distributed hydrologic responses to changes in climate.

Appendix 1. Stochastic simulation with the downscaled parameters

Once the model parameters $\{\pi_{down}, r_{down}, \mu_{down}, \sigma^2_{down}\}$ have been downscaled, their use in stochastic simulation of the inferred changed precipitation climates at individual stations is straightforward. First, since use of the transition probabilities (Eq. 1) is the most convenient approach to simulating daily precipitation occurrences, they need to be recovered from π_{down} and r_{down}, by inverting Eq. (2):

$$p_{01} = \pi(1-r) \quad \text{(A1a)}$$

and

$$p_{11} = \pi + r(1-\pi) \quad \text{(A1b)}$$

The transition probabilities can then be used to simulate sequences of wet and dry days, by generating a uniform [0,1] variate, u (e.g. Press et al. 1986, Bratley et al. 1987), for each day, and comparing it with the appropriate transition probability. The current day is simulated to be wet when

$$u \leq \begin{cases} p_{01}, & \text{if the previous day was dry} \\ p_{11}, & \text{if the previous day was wet} \end{cases} \quad \text{(A2)}$$

and simulated to be dry otherwise.

When a wet day is simulated, a precipitation amount must also be generated. Most commonly these are assumed to follow a gamma distribution, the probability density function for which is

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-x/\beta)}{\beta\Gamma(\alpha)} \quad \text{(A3)}$$

Since the mean and variance of the gamma distribution are $\alpha\beta$ and $\alpha\beta^2$, respectively, the 2 distribution parameters can be simply obtained from μ_{down} and σ^2_{down} using

$$\alpha = \mu^2/\sigma^2 \quad \text{(A4a)}$$

and

$$\beta = \sigma^2/\mu^2 \quad \text{(A4b)}$$

The gamma distribution parameters can then be used with readily available computer codes (e.g. Press et al. 1986, Bratley et al. 1987) to simulate the nonzero precipitation amounts.

Alternatively, it has been found in at least some cases (Woolhiser & Roldan 1982, Foufoula-Georgiou & Lettenmaier 1987, Wilks 1998a) that the distributions of daily nonzero precipitation amounts are better represented by the mixed exponential distribution. This is simply a probability mixture of 2 ordinary 1-parameter exponential distributions, the probability density function of which is

$$f(x) = \frac{\alpha}{\beta_1}\exp\left(\frac{-x}{\beta_1}\right) + \frac{1-\alpha}{\beta_2}\exp\left(\frac{-x}{\beta_2}\right) \quad \text{(A5)}$$

Using the mixed exponential distribution in the present context is slightly more complicated, because only the 2 quantities μ_{down} and σ^2_{down} are supplied by the downscaling procedure, while the mixed exponential distribution has 3 parameters. These are the mixing parameter, α, and the means β_1 and β_2 of the 2 exponential distributions. The most straightforward solution to this problem is to first determine a value for the mixing probability, perhaps by assuming that it is unchanged from the base climate, or by downscaling it using Eq. (6) (Wilks 1998b). Then, consistent values for the 2 means can generally be obtained using

$$\beta_1 = \mu + \frac{\sqrt{2\alpha(1-\alpha)(\sigma^2-\mu^2)}}{2\alpha} \quad \text{(A6a)}$$

and

$$\beta_2 = \frac{\mu - \alpha\beta_1}{1-\alpha} \quad \text{(A6b)}$$

Simulation of a mixed exponential variate is a simple 2-step process. First, either β_1 or β_2 is chosen by comparing a new uniform [0,1] variate u to the mixing parameter, α. If $u \leq \alpha$ then β_1 is chosen, and β_2 is chosen otherwise. Finally, a precipitation amount x is generated using

$$x = x_{min} - \beta\ln(u) \quad \text{(A7)}$$

where x_{min} is the daily precipitation amount below which a day is recorded as being dry, β is either β_1 or β_2 (determined as just described), and u is a third uniform [0,1] variate.

Acknowledgements. I thank Tom Wigley for motivating this work, and Wigley and Rob Wilby for useful discussions. This research was supported by the Electric Power Research Institute, under UCAR-EPRI Agreement WO8020-19.

LITERATURE CITED

Bardossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 28:1247–1259
Bratley P, Fox BL, Schrage LE (1987) A guide to simulation. Springer, New York
Burger G (1996) Expanded downscaling for generating local weather scenarios. Clim Res 7:111–128
Cohen SJ (1990) Bringing the global warming issue closer to home: the challenge of regional impact studies. Bull Am Meteorol Soc 71:520–526
Epstein ES (1966) Point and area precipitation probabilities. Mon Weather Rev 94:595–598
Foufoula-Georgiou E, Lettenmaier DP (1987) A Markov renewal model for rainfall occurrences. Water Resour Res 23: 875–884
Gabriel KR, Neumann J (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. Q J R Meteorol Soc 88:90–95
Gates WL, Henderson-Sellers A, Boer GJ, Folland CK, Kitoh A, McAvaney JB, Semazzi F, Smith N, Weaver AJ, Zeng QC (1996) Climate models—evaluation. In: Houghton JT et al. (eds) Climate change 1995. Cambridge University Press, Cambridge, p 229–284
Giorgi F (1990) On the simulation of regional climate using a limited area model nested in a general circulation model. J Clim 3:941–963
Giorgi F, Mearns LO (1991) Approaches to simulation of regional climate change: a review. Rev Geophys 29:191–216
Gregory JM, Wigley TML, Jones PD (1993) Application of Markov models to area-average daily precipitation series and interannual variability of seasonal totals. Clim Dyn 8: 299–310
Grotch SL, MacCracken MC (1991) The use of general circulation models to predict regional climate change. J Clim 4: 286–303
Hulme M, Jiang T, Wigley TML (1995) SCENGEN, a climate change scenario generator: user manual. Climatic Research Unit, Univ of East Anglia, Norwich

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre coupled ocean-atmosphere GCM: model description, spinup and validation. Clim Dyn 9:95–105

Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. J Clim 10: 2548–2568

Kagan RL (1966) An evaluation of the representativeness of precipitation data. Gidrometeoizdat, Leningrad (in Russian)

Kattenberg A, Giorgi F, Grassl H, Meehl GA, Mitchell JFB, Stouffer RJ, Tokioka T, Weaver AJ, Wigley TML (1996) Climate models—projections of future climate. In: Houghton JT et al. (eds) Climate change 1995. Cambridge University Press, Cambridge, p 285–357

Katz RW (1977) Precipitation as a chain-dependent process. J Appl Meteorol, 16:671–676

Katz RW (1996) Use of conditional stochastic models to generate climate change scenarios. Clim Change 32:237–255

Katz RW, Parlange MB (1998) Overdispersion phenomenon in stochastic modeling of precipitation. J Clim 11:591–601

Lettenmaier D (1995) Stochastic modeling of precipitation with applications to climate model downscaling. In: von Storch H, Navarra A (eds) Analysis of climate variability. Springer, Berlin, p 197–212

Mearns LO, Schneider SH, Thompson SL, McDaniel LR (1990) Analysis of climate variability in general circulation models: comparison with observations and changes in variability in 2 × CO_2 experiments. J Geophy Res 95: 20469–20490

Mearns LO, Giorgi F, McDaniel L, Shields C (1995) Analysis of daily variability of precipitation in a nested regional climate model: comparison with observations and doubled CO_2 results. Global Planet Change 10:55–78

Mitchell JFB, Johns TC (1997) On modification of global warming by sulphate aerosols. J Clim 10:245–267

Osborn TJ (1997) Areal and point precipitation intensity changes: implications for the application of climate models. Geophys Res Lett 24:2829–2832

Osborn TJ, Hulme M (1997) Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. J Clim 10:1885–1908

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes. Cambridge University Press, Cambridge

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Richardson CW, Wright DA (1984) WGEN: a model for generating daily weather variables. USDA/ARS, ARS-8, Washington, DC

Semenov MA, Barrow EM (1997) Use of a stochastic weather generator in the development of climate change scenarios. Clim Change 35:397–414

Todorovic P, Woolhiser DA (1975) A stochastic model of n-day precipitation. J Appl Meteorol 14:17–24

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6: 1161–1171

Wigley TML, Jones PD, Briffa KR, Smith G (1990) Obtaining sub-grid-scale information from coarse-resolution general circulation model output. J Geophys Res 95:1943–1953

Wilby RL, TML Wigley (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21:530–548

Wilby RL, Wigley TML, Wilks DS, Hewitson BC, Conway D, Jones PD (1996) Statistical downscaling of general circulation model output. A report prepared by the National Center for Atmospheric Research on behalf of the Electric Power Research Institute, Palo Alto, CA

Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. Water Resour Res 34:2995–3008

Wilks DS (1988) Estimating the consequences of CO_2-induced climatic change on North American grain agriculture using general circulation model information. Clim Change 13:19–42

Wilks DS (1989) Conditioning stochastic daily precipitation models on total monthly precipitation. Water Resour Res 23:1429–1439

Wilks DS (1992) Adapting stochastic weather generation algorithms for climate change studies. Clim Change 22: 67–84

Wilks DS (1998a) Multisite generalization of a daily stochastic precipitation generation model. J Hydrol 210:178–191

Wilks DS (1998b) Precipitation simulation and statistical downscaling using 'WGEN' and spell length models. Final Report prepared for the Electric Power Research Institute, Palo Alto, CA

Woo MK (1992) Application of stochastic simulation to climate-change studies. Clim Change 20:313–330

Woolhiser DA, Roldan J (1982) Stochastic daily precipitation models. 2. A comparison of distribution of amounts. Water Resour Res 18:1461–1468

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: June 15, 1998; Accepted: October 15, 1998
Proofs received from author(s): January 25, 1999