Progress in Physical Geography 23,3 (1999) pp. 329–357

The weather generation game: a review of stochastic weather models

D.S. Wilks[1] and R.L. Wilby[2]

[1]Atmospheric Science Group, Cornell University, Ithaca, New York, NY 14853, USA
[2]National Center for Atmospheric Research, Boulder, CO 80307, USA, and Division of Geography, University of Derby, Kedleston Road, Derby DE22 1GB, UK

Abstract: This article reviews the historical development of statistical weather models, from simple analyses of runs of consecutive rainy and dry days at single sites, through to multisite models of daily precipitation. Weather generators have been used extensively in water engineering design and in agricultural, ecosystem and hydrological impact studies as a means of in-filling missing data or for producing indefinitely long synthetic weather series from finite station records. We begin by describing the statistical properties of the rainfall occurrence and amount processes which are necessary precursors to the simulation of other (dependent) meteorological variables. The relationship between these daily weather models and lower-frequency variations in climate statistics is considered next, noting that conventional weather generator techniques often fail to capture wholly interannual variability. Possible solutions to this deficiency – such as the use of mixtures of slowly and rapidly varying conditioning variables – are discussed. Common applications of weather generators are then described. These include the modelling of climate-sensitive systems, the simulation of missing weather data and statistical downscaling of regional climate change scenarios. Finally, we conclude by considering ongoing advances in the simulation of spatially correlated weather series at multiple sites, the downscaling of interannual climate variability and the scope for using nonparametric techniques to synthesize weather series.

Key words: climate change, impact assessment, stochastic model, time series, weather generator.

I Introduction

Models of observed daily weather sequences are frequently used in water engineering design, and agricultural, ecosystem or climate change simulations because observed ground-based meteorological data are often inadequate in terms of their length, completeness or spatial coverage. These statistical models are also known as 'weather generators' since they can in-fill missing data or produce indefinitely long synthetic

 0309–1333(99)PP227RA

weather series by simulating key properties of observed meteorological records (i.e., daily means, variances and covariances, frequencies, extremes, etc.). Daily weather simulators are by far the most common, both because of the wide availability of weather data on this timescale, and (relatedly) the abundance of impacts models that are driven by daily weather inputs. To date, the majority of weather generators have focused on the precipitation process in recognition of the dominant control exerted by rainfall on many environmental processes, and due to the complexity of building internally consistent, multivariable models (Hutchinson, 1995). However, companion algorithms that simulate other meteorological variables are also in routine use.

There are two complementary ways in which these models can be viewed. In the first instance they are stochastic models for day-to-day (and, by extension, longer-period) variations in the weather. From this perspective the parameters of a stochastic weather model comprise a concise distillation of certain aspects of the local climate. Secondly, when these models are used for Monte-Carlo simulation (i.e., 'weather generation'), they can be regarded as elaborate random number generators whose outputs statistically resemble daily weather data at a location. It is important to note that weather generators are not weather forecasting algorithms, and thus are quite different from deterministic weather models, which operate by numerically integrating the partial differential equations describing fluid flows. A major implication of this distinction is that, while stochastic model outputs behave statistically like weather data, it is not expected that any particular simulated weather sequence will be duplicated in weather observations at a given time in either the past or future.

In this article we provide a brief history of weather generators: from simple descriptions of 'runs' of wet and dry days at single sites, through multivariate models, and on to the relationships between the daily simulations and the longer-term climate. We then sketch some of the applications of stochastic weather models that have been made in agriculture, ecology, hydrology and simulations of regional climate change. Finally, we speculate about future directions in weather generation techniques.

II A brief history and exposition

As mentioned in the introduction, most effort in the construction of weather generators has been devoted to precipitation processes. Not only is precipitation the most critical meteorological variable for many applications, but the presence or absence of precipitation also typically affects the statistics of many nonprecipitation variables to be simulated. Precipitation data exhibit distinctive and difficult characteristics which complicate the statistical models needed to describe them. In addition to exhibiting the correlation between values at successive time periods that is typical of all weather variables, precipitation is unique in its mixed character as both a discrete and continuous variable. That is, precipitation is very often exactly zero, and hence there is a discontinuity in the probability distribution of precipitation data between the zero and the nonzero observations.

Accordingly, most weather generators contain separate treatments of the precipitation occurrence and intensity processes. The precipitation occurrence process manifests itself in two weather states, wet or dry. A key aspect of stochastic weather models is their representation of the tendency of wet and dry days to exhibit persistence, or positive serial (i.e., auto-) correlation, so that wet and dry runs tend to clump together in time more strongly than could be expected by chance. The precipitation intensity process pertains to the modelling and simulation of the nonzero precipitation amounts. These are typically strongly skewed to the right, with many small values and few but quite important large precipitation amounts. Although these concepts appear straightforward it has taken more than a century to formalize many of the processes within stochastic precipitation models.

1 The precipitation occurrence process

Apparently the earliest published work on probabilistic modelling of precipitation occurrence was that of Quetelet, who reported in 1852 (see Katz, 1985) that runs of consecutive rainy and dry days at Brussels for 1833–50 exhibited persistence. Another expression of this tendency for wet and dry weather to persist was provided by Newnham (1916), who used daily rainfall records at Kew, Aberdeen, Valencia and Greenwich, UK, to demonstrate that the probability of a 'rain day' is greater if the preceding day was wet rather than dry. These two approaches – considering run lengths and day-to-day statistical dependence – to describing the temporal dependence of precipitation occurrences were pursued further by Besson (1924), Gold (1929) and Cochran (1938). Williams (1952) used geometric series to model dry and wet-spell (i.e., consecutive runs of dry or wet days) lengths at Rothamsted Experimental Station, Harpenden, UK (see Figure 1). The greater probability of long dry (as opposed to wet) spells is clearly evident from Figure 1. Longley (1953) subsequently improved geometric-series fits to observed wet and dry-spell lengths in five Canadian cities by differentiating between the months in which the spells fell.

Gabriel and Neumann (1962) are generally credited with presenting the first statistical model of daily rainfall occurrence. In their seminal work using rainfall data

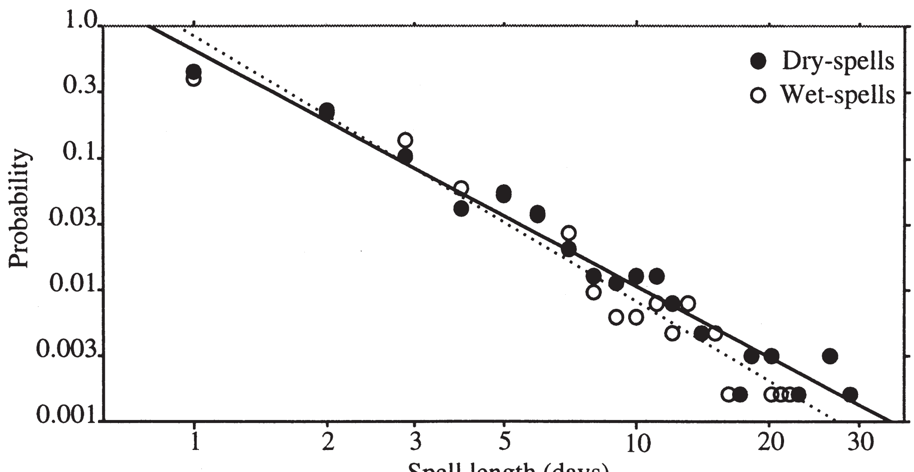

Figure 1 The distribution of wet and dry-spell lengths at Rothamsted Experimental Station, Harpenden, UK, 1938–47 (after Williams, 1952). The dashed line is the best- fit for wet spells; the solid line for dry spells

for Tel Aviv, Israel, the authors recognized that the frequency distributions for wet and dry spell length of the types identified by Williams (1952) and Longley (1953) may arise from a simple Markov chain model. In particular, Gabriel and Neumann (1962) proposed the use of a first-order Markov chain for precipitation occurrence, assuming that the probability of rainfall on any day depends only on whether the previous day was wet or dry. This model can be fully defined by the two conditional probabilities

$$p_{01} = \Pr \{\text{precipitation on day } t \mid \text{no precipitation on day } t - l\} \tag{1a}$$

and

$$p_{11} = \Pr \{\text{precipitation on day } t \mid \text{precipitation on day } t - l\} \tag{1b}$$

which are called transition probabilities. Here the vertical bar symbol '|' is read as 'given' or 'conditional on'. Since there are only two possible states on a given day, the two complementary transition probabilities are $p_{00} = 1 - p_{01}$ (dry day following a dry day) and $p_{10} = 1 - p_{11}$ (dry day following a wet day).

It was noted by Gabriel and Neumann (1962) that this simple model for rainfall occurrence was able to describe closely the persistent nature of daily precipitation occurrence patterns, and that certain other properties of the occurrence series could be derived from the two transition probabilities. Of particular importance are the long-run (i.e., climatological) relative frequency of precipitation days

$$\pi = \frac{p_{01}}{1 + p_{01} - p_{11}} \tag{2a}$$

and the first-lag autocorrelation of the precipitation occurrence series

$$r_1 = p_{11} - p_{01} \tag{2b}$$

Because of the persistent nature of daily rainfall occurrence one finds in practice that the conditional probability in (1b) is larger than that in (1a), so that $p_{01} < \pi < p_{11}$ and $r_1 > 0$. Furthermore, the lengths of the alternating wet and dry spells produced by the first-order Markov model are independent, with those lengths distributed according to the geometric distribution

$$\Pr\{X = x\} = p(1 - p)^{x-1}, \; x = 1, 2, 3, \ldots \tag{3}$$

where $p = p_{01}$ for dry spells and $p = 1 - p_{11}$ for wet spells. These authors pointed out further that the average number of wet days, and the variance of the number of wet days, within a string of *T* consecutive days under the first-order Markov model can be computed as

$$E[N(T)] = \pi T \tag{4a}$$

and

$$\mathrm{Var}[N(T)] \approx \pi (1 - \pi) T \frac{1 + r_1}{1 - r_1} \tag{4b}$$

respectively. Equations (4a) and (4b) can thus pertain to rainfall occurrence statistics for monthly ($T \approx 30$) or seasonal ($T \approx 90$) periods. Usually the parameters of the precipitation generation models are fit separately for the 12 calendar months, in order to allow for seasonal cycles in the precipitation statistics. Equation (4b) is an approximation that holds for large T, although this approximation was found to be very close for monthly totals in the Tel Aviv data, and subsequently for other locations as well (e.g., Gregory *et al*., 1993; Wilks, 1999a).

Equations (2)–(4) illustrate the use of the first-order Markov model to characterize important aspects of the precipitation occurrence climate. This model also provides a very convenient and efficient means of generating sequences of random numbers that resemble (i.e., simulate) the corresponding real weather data. For each simulated day, a random number u is drawn from the interval [0,1] in a way that any real number (up to the precision of the computer) in that interval is equally likely to be picked. In practice these are usually produced by widely available computer algorithms called uniform pseudo-random number generators (e.g., Press *et al*., 1986; Bratley *et al*., 1987). Note, however, that there are a number of surprising and subtle pitfalls in the generation of uniform random numbers, and in particular that the simplest of these algorithms (as might be implemented as part of a computer's operating system) can have serious deficiencies that may compromise the results of stochastic simulations. See, for example, Press *et al*. (1986) for a brief exposition and guide to improved algorithms, or Knuth (1997), which is the standard reference on this subject.

Once the random number u has been generated, whether the next day in the sequence is wet or dry is determined using (1). If the previous day ($t - 1$) was dry, then day t is simulated to be wet if $u \leq p_{01}$, and otherwise it is also dry. If the previous day was wet, then the current day is simulated to be wet if $u \leq p_{11}$, and is dry otherwise (cf. Figure 2a). Because first-order Markov models fit to daily precipitation data yield $p_{01} < \pi < p_{11}$, the simulations yield sequences of wet and dry days that are more persistent than independent draws according to the climatological probability π.

For some climates it has been found that the simple first-order Markov model generates synthetic rainfall series with too few long dry spells (e.g., Buishand, 1977; 1978; Racsko *et al*., 1991; Guttorp, 1995). Dennett *et al*. (1983), Singh and Kripalani (1986), Jones and Thornton (1997) and Wilks (1999a) addressed this deficiency by considering Markov chains of higher order. These techniques increase the length of the Markov model's 'memory' of antecedent wet and dry days. For example, second-order Markov chains use the wet/dry state on both the preceding day, and two days prior, such that eight transition probabilities p_{ijk} must be defined. Here each of the indices i, j, k may be either one (wet) or zero (dry). Hence, p_{101} would be the probability of a wet day given that the previous day was dry, and the day before that was wet. Third and higher-order Markov chains can be similarly defined, although the number of parameters (i.e., transition probabilities) required increases exponentially as the order increases, being 2^k for a k^{th}-order chain. When only the dry spells are not adequately modelled by the first-order Markov model it is possible to improve the statistics of the simulated dry spells using 'hybrid-order' Markov models, in which the Markov 'memory' extends further back in time for the dry spells only (Stern and Coe, 1984; Wilks, 1999a).

When deciding among models having different degrees of complexity, one must judge how elaborate a model is justified by the data. Gabriel and Neumann (1962) compared the first-order Markov model for precipitation occurrence at Tel Aviv with the next simplest model, namely, independent Bernoulli (i.e., binomial) occurrences,

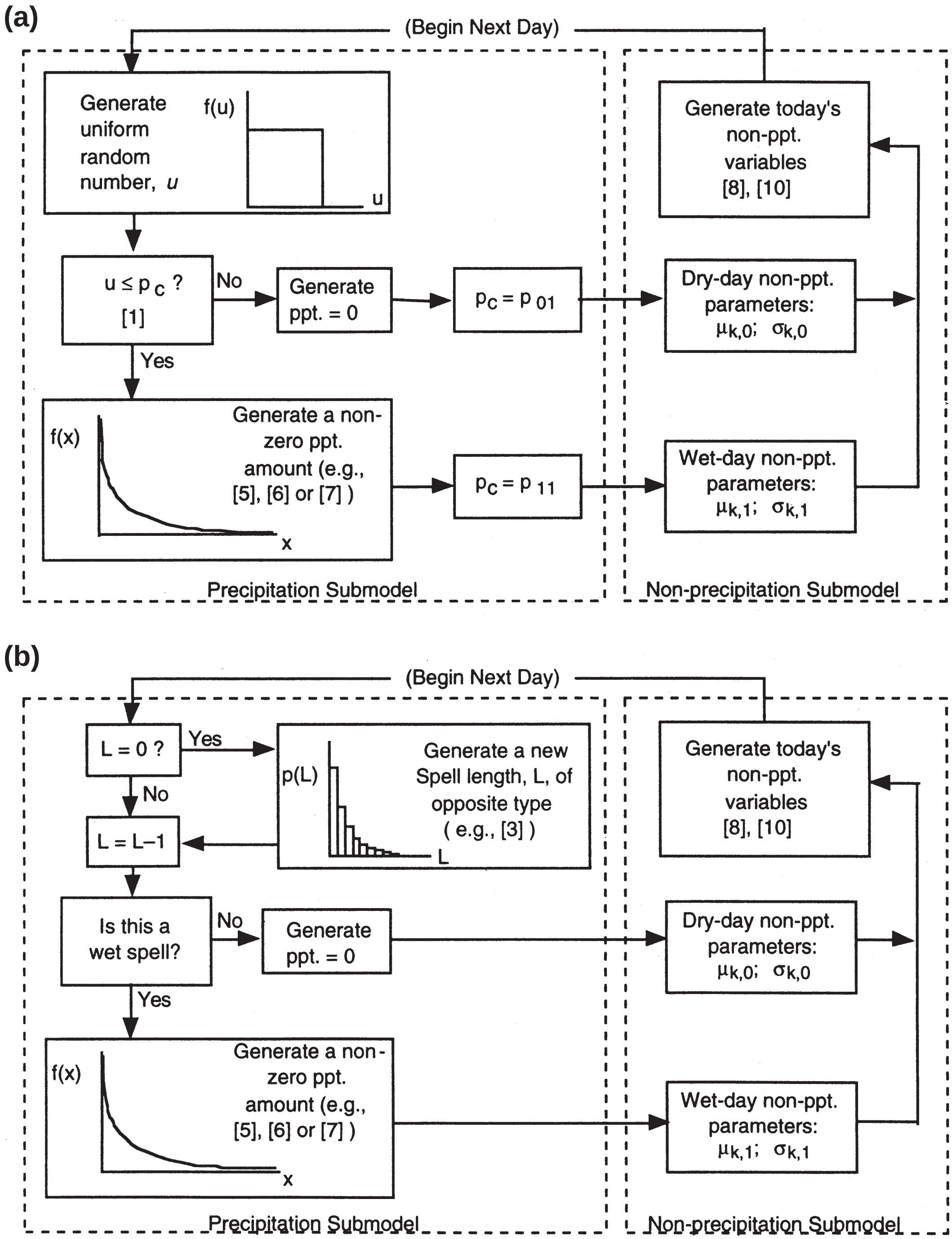

Figure 2 Flowcharts for daily weather generation using the WGEN framework and (a) Markov chain and (b) spell-length models for the precipitation component. Numbers in square brackets refer to equations in the text

using a Chi-square goodness-of-fit test. While this is a reasonable approach when only two alternatives are being considered, ambiguities in such statistical tests arise when multiple comparisons are made, for example when choosing among Markov models of zeroth (Bernoulli distribution), first and second orders. The usual approach in circumstances like this is to employ an objective order-selection criterion such as Akaike's information criterion (AIC – Akaike, 1974) or the Bayesian information criterion (BIC – Schwarz, 1978). Both AIC and BIC are likelihood-based criteria, in that they choose the model having the largest maximized likelihood, after application of a penalty that increases with the number of free parameters allowed by each of the models considered. (The likelihood function is notationally analogous to the probability distribution function or the probability density function; but the data are considered as fixed, while values of the parameters associated with the global maximum of this function are the fitted maximum likelihood estimates.) The AIC and BIC differ only in the forms of their penalty functions. Gates and Tong (1976) concluded that second-order Markov dependence was justified according to the AIC for the Tel Aviv precipitation data considered by Gabriel and Neumann (1962). However, Katz (1981) concluded first-order dependence for these data was adequate on the basis of the BIC, which he also showed to be asymptotically consistent (i.e., the BIC is correct on average for sufficiently large data samples). Since many years of meteorological data are typically available for fitting stochastic weather models, this consistency result is a strong argument for use of the BIC in this context.

An alternative to Markov chain models for simulating precipitation occurrences is the use of spell-length models. Rather than simulating rainfall occurrences day by day, spell-length models operate by fitting probability distributions such as (3) to observed relative frequencies of wet and dry-spell lengths. This kind of model is sometimes called an 'alternating renewal process' (Buishand, 1977; 1978; Roldan and Woolhiser, 1982), in that random numbers are generated alternately from the wet and dry spell-length distributions (Figure 2b). That is, a new spell length L is generated only when a run of consecutive wet or dry days has come to an end, at which point a new spell of the opposite type is simulated. Of course, if geometric distributions (3) are used to model the lengths of wet and dry spells the resulting synthetic series will exhibit the same characteristics as the equivalent first-order Markov process (1). Higher-order Markov chains have spell-length distributions associated with them that are generalizations of the geometric distribution. Precipitation occurrence sequences with different statistical characteristics can be obtained using different distributions for the frequencies of spell lengths. Such distributions include the truncated negative binomial (Buishand, 1977; 1978; Roldan and Woolhiser, 1982), the negative binomial distribution (Wilby *et al*., 1998; Wilks, 1999a), and the mixed geometric distribution (Racsko *et al*., 1991). For climates where (1) or (3) yield very long dry spells with insufficient frequency, these more elaborate choices for modelling precipitation occurrence have been found to yield more realistic results in this regard (Buishand, 1977; 1978; Racsko *et al*., 1991; Wilks, 1999a). However, this method can be susceptible to poor parameter estimates in arid regions or wherever less than 25 years of observations is available (Roldan and Woolhiser, 1982).

2 The precipitation amount process

The next necessary element of a weather generator is a model for the nonzero precipitation amounts on wet days. The most prominent statistical feature of daily precipitation amounts is that their distribution is strongly skewed to the right. That is, very small daily precipitation amounts are quite common, while the large daily precipitation amounts that are most important to hydrological, agricultural and engineering impacts are comparatively rare. Todorovic and Woolhiser (1975) were the first to produce a daily stochastic precipitation generator, by combining the first-order Markov model for daily precipitation occurrence described above with a statistical model for daily nonzero precipitation amounts. Their choice for modelling the amounts was the exponential distribution, whose probability density function is

$$f(x) = \frac{1}{\mu} \exp\left[\frac{-x}{\mu}\right] \tag{5}$$

The exponential distribution is probably the simplest reasonable model for daily precipitation amounts, as it requires specification of only one parameter, μ, yet reproduces qualitatively the strong positive skewness exhibited by daily precipitation data. The average nonzero precipitation amount according to (5) is μ, and the corresponding variance $\sigma^2 = \mu^2$. Exponential distributions have also been used by Richardson (1981) and Wilby (1994), among others.

A number of more elaborate models have also been proposed for the distribution of daily precipitation amounts given the occurrence of a wet day. The two-parameter gamma distribution has been the most popular choice (e.g., Thom, 1958; Katz, 1977; Buishand, 1977; 1978; Stern and Coe, 1984; Wilks, 1989; 1992), and has the probability density function

$$f(x) = \frac{(x/\beta)^{\alpha-1} \exp[-x/\beta]}{\beta\, \Gamma(\alpha)} \tag{6}$$

This distribution involves two parameters: the shape parameter α and the scale parameter β. The factor $\Gamma(\alpha)$ is the gamma function (see, e.g., Abramowitz and Stegun, 1984, or Wilks, 1995) evaluated at α. This distribution has mean $\mu = \alpha\beta$, and variance $\sigma^2 = \alpha\beta^2$. For $\alpha \leq 1$ gamma distributions are qualitatively similar to exponential distributions (5) in concentrating most of the probability near zero and producing large precipitation amounts only rarely. For $\alpha = 1$ gamma distributions (6) reduce to exponential distributions (5), but in general the additional parameter in (6) allows more flexible accommodation of rainfall amount frequencies, and so improves the realism of stochastic precipitation models.

Another natural generalization of the exponential distribution (5) is the mixed exponential distribution, which is simply a probability mixture of two one-parameter exponential distributions. Its probability density function is

$$f(x) = \frac{\alpha}{\mu_1} \exp\left[\frac{-x}{\mu_1}\right] + \frac{1-\alpha}{\mu_2} \exp\left[\frac{-x}{\mu_2}\right] \tag{7}$$

Mathematically, (7) indicates a superposition of two ordinary exponential distributions (5) whose respective means are μ_1 and μ_2. From the standpoint of simulation, (7) indicates that the first exponential distribution is used to generate the precipitation amounts with probability α, and the second is used with probability $1 - \alpha$, where the third parameter α is called the mixing parameter. The mixed exponential distribution has mean $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$ and variance $\sigma_2 = \alpha\mu_1{}^2 + (1 - \alpha)\mu_2{}^2 + \alpha\,(1 - \alpha)(\mu_1 - \mu_2)^2$. First suggested as a model for daily precipitation amounts by Woolhiser and Pegram (1979), the mixed exponential distribution has been used comparatively rarely. However, it has been reported to provide substantially better overall fits to daily precipitation data than the gamma distribution (Woolhiser and Roldan, 1982; Foufoula-Georgiou and Lettenmaier, 1987; Wilks, 1998; 1999a), and in particular Wilks (1999a) reports better representation of the frequencies of the very largest precipitation amounts.

Figure 3 illustrates the differences between the three probability distributions (5)–(7) for the case of daily January (liquid-equivalent) precipitation at Ithaca, New York, for the years 1900–98, by comparing the respective fitted probability density functions with a histogram of the data. There were a total of 1499 January wet days during this period. All three distributions capture the most prominent characteristic of this data, namely, its very strong positive skewness. Because of this skewness the larger precipitation amounts that are most important in many applications are comparatively rare, so that the portion of the graph depicting daily precipitation greater than about 15 mm has been enlarged vertically in the inset for clarity. It is evident that the exponential distribution (5) (dashed line) underestimates the frequencies of both the very small and very large precipitation amounts, while overestimating probabilities for amounts near 5 mm. The gamma distribution (6) (solid line) improves the representation of the small amounts substantially, but improves the representation of large amounts only slightly, so that frequencies of amounts near 5 mm are also over-represented by this gamma distribution. The mixed exponential distribution (7) (grey line) is clearly the best of the three over the full range of the data. This qualitative comparison of the fits of these three distributions is consistent both with AIC (Akaike, 1974) and BIC (Schwarz, 1978) statistics. Values of the maximized (log-) likelihoods minus the corresponding BIC penalty functions are 764.6 for the exponential distribution, 784.4 for the gamma distribution and 836.6 for the mixed exponential distribution.

Given a distribution to represent the nonzero precipitation amounts, simulations are accomplished through computer algorithms that generate random numbers according to the fitted distribution. For each day the precipitation occurrence model simulates wet conditions, a new random variate from the fitted distribution for nonzero precipitation amounts is generated (Figure 2). Much more on these algorithms can be found in Devroye (1986) or Bratley *et al*. (1987).

Most stochastic weather generators make the assumption that precipitation amounts on wet days are independent, and follow the same distribution. Allowing different probability distributions for precipitation amounts depending on that day's position in a wet spell (e.g., the mean rainfall on a wet day following a wet day might be greater than on a wet day following a dry day) has been considered by Katz (1977), Buishand (1977; 1978), Chin and Miller (1980) and Wilks (1999a), but allowing this extra complexity often makes little difference to the result. Similarly, the autocorrelation between successive nonzero precipitation amounts in daily series is sometimes (statistically) significantly different from zero, but is typically quite small and usually of little practical importance (Katz, 1977; Buishand, 1977; 1978; Foufoula-Georgiou and Lettenmaier, 1987). In contrast, accounting for serial correlation of nonzero precipitation amounts is essential if the precipitation model has an hourly (or smaller) rather than a daily time step (Katz and Parlange, 1995).

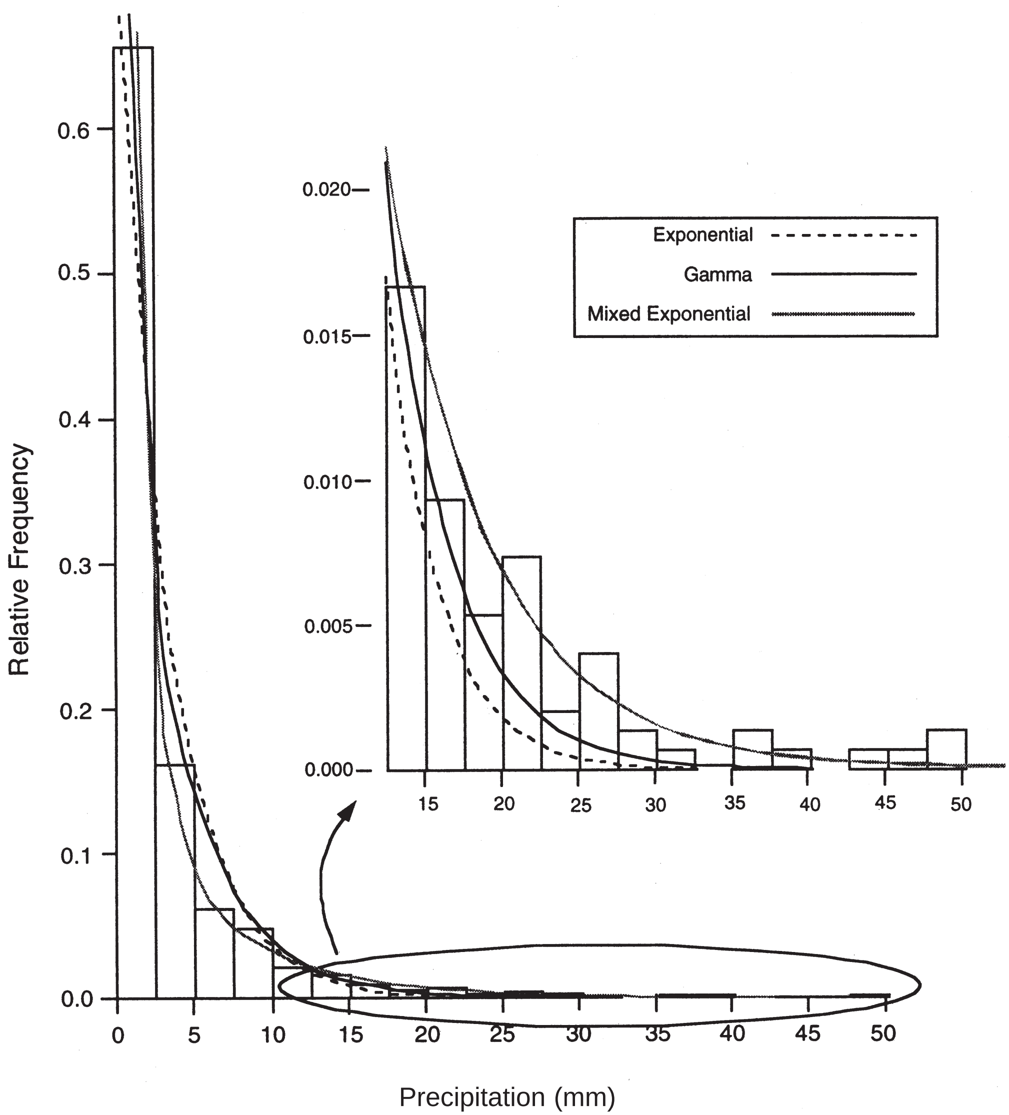

Figure 3 Comparison of the histogram for the 1499 January wet-day (liquid-equivalent) precipitation amounts at Ithaca, New York, for 1900–98, with corresponding fitted exponential (5) (dashed line), gamma (6) (solid line) and mixed exponential (7) (grey line) probability density functions. Inset shows enlargement of the graph for daily precipitation greater than about 15 mm

Another approach to accounting for correlation in the nonzero precipitation amounts is the use of multistate (i.e., greater than the 2 states in Equation 1) Markov models. These Markov models simulate both precipitation occurrence and amounts, by defining different ranges of precipitation amounts as constituting distinct states. Transition probabilities among all possible pairs of states are then estimated from data, and used in simulation. For example, for southeast England in spring, Gregory *et al*. (1993) found that there is a 16% probability that a rainfall total of greater than 6.62 mm will be equalled or exceeded on the following day, but only a 5.7% chance that such a day will be followed by a dry day. Gregory *et al*. (1993) also found that first-order Markov models yielded smaller discrepancies in seasonal precipitation totals if specified ranges of daily precipitation totals are used to condition the rainfall amounts on the following day. Haan *et al*. (1976) followed a similar approach. The validity of this multistate Markov approach clearly rests on the choice of the number of states and their ranges (i.e., the upper and lower rainfall thresholds) and on the distributions used for wet-day amounts in any given state. These models involve comparatively large numbers of parameters, and thus require quite long data records in order to be estimated well.

3 Other meteorological variables

Until now the discussion has been restricted to daily stochastic precipitation models. However, it was noted at the outset that for many practical applications weather generators must produce additional meteorological variables. For example, sensitivity analyses of crop models have revealed that inputs of mean temperature values consistently result in overstimated yields. Therefore, accurate crop productivity estimates require synthetic data that reproduce observed daily variability in driving meteorological variables other than precipitation as well (Richardson, 1985; Nonhebel, 1994; Semenov and Porter, 1995; Semenov *et al*., 1998). The most widely cited model of daily weather variables (Richardson, 1981), developed to support models of crop development and yield, also includes simulations of daily maximum temperature, minimum temperature and solar radiation. Stochastic weather models of the type originally proposed by Richardson (1981), and now widely referred to as WGEN (for 'Weather Generator' as in Richardson and Wright, 1984), condition the statistics of the daily nonprecipitation variables on occurrence or nonoccurrence of precipitation. However, this conditioning is weakly defined and is only a proxy for other unspecified processes such as cloud cover, which affect temperature and solar radiation (Hutchinson, 1995).

In WGEN models, the statistical process underlying the nonprecipitation variables is a first-order vector (i.e., multiple variables modelled simultaneously) autoregression. Here the meaning of 'first order' is the same as for the Markov chains described above: the statistics of the current day's values are fully defined by the values on the previous day. The equation governing this process can be written

$$\mathbf{z}(t) = [\mathbf{A}]\, \mathbf{z}\,(t-1) + [\mathbf{B}]\, \varepsilon(t) \tag{8}$$

where $\mathbf{z}(t)$ is a K-dimensional vector of standard Gaussian (i.e., normally distributed, with zero mean and unit variance) values for today's nonprecipitation variables, $\mathbf{z}(t-1)$ is the corresponding vector for the previous day, and $[\mathbf{A}]$ and $[\mathbf{B}]$ are $K \times K$

matrices of parameters. The *K*-dimensional vector $\varepsilon(t)$ of independent standard normal variables is known alternatively as 'error', or white-noise forcing. Here *K* is the number of nonprecipitation variables being simulated, so that $K = 3$ in the Richardson model for maximum temperature, minimum temperature and solar radiation. Decomposing (8) into scalar notation, each of the *K* elements of $\mathbf{z}(t)$ is specified as a linear combination of all *K* values for the previous day $\mathbf{z}(t-1)$, and the three random quantities $\varepsilon(t)$. In the original Richardson (1981) model, for example, $K = 3$ and

$$\begin{aligned} z_k(t) = {} & a_{k,1}\, z_1\,(t-1) + a_{k,2}\, z_2\,(t-1) + a_{k,3}\, z_3(t-1) \\ & + b_{k,1}\, \varepsilon_1(t) + b_{k,2}\, \varepsilon_2(t) + b_{k,3}\, \varepsilon_3(t) \end{aligned} \tag{9}$$

where z_1, z_2 and z_3 correspond to maximum temperature minimum temperature, and solar radiation, respectively. The first line in (9) corresponds to the first term in (8), in which today's value is expressed as a function of all three of yesterday's values, as in an ordinary regression equation. The second line in (9), corresponding to the second term in (8), provides the random variation that allows the generator to produce different weather sequences. For each new day to be simulated, a realization for each of the ε_k is produced by a standard normal random number generator (e.g., Devroye, 1986; Press *et al*., 1986; Bratley *et al*., 1987), and then the linear combination of these (through the *b* coefficients) provides properly correlated random contributions to the new z_k.

Specific parameter values in the [**A**] and [**B**] matrices are computed from the simultaneous and lagged (by one day) correlations among the nonprecipitation variables considered. In the conventional implementation of WGEN (Richardson and Wright 1984), average values of these correlations – computed by combining data from 31 USA stations and across all seasons – are used to compute fixed [**A**] and [**B**] matrices that are used for any location and in any season. Hayhoe (1998), working with Canadian data, has found that geographic and especially seasonal departures from these commonly assumed correlations can be quite large, particularly for the correlations involving solar radiation. At high latitudes these correlations tend to be high in summer, and low or even negative in winter. For some applications these discrepancies may be only of minor importance. For example, Richardson (1985) found that the output from a wheat model was not sensitive to daily variations in the solar radiation. However, many users would be well advised to compute their own location and season-specific [**A**] and [**B**] matrices.

Once they have been generated using (8), the $\mathbf{z}(t)$ are transformed to weather variables in a way that depends on whether the day has been simulated to be wet or dry. Most WGEN implementations assume first-order Markov dependence (1) for precipitation occurrence (Figure 2a), although direct simulation of wet and dry spells (Figure 2b) has also been employed (e.g., Semenov and Porter, 1995). In either case, each of the *K* nonprecipitation weather variables is computed for each day according to

$$T_k(t) = \begin{cases} \mu_{k,0}(t) + \sigma_{k,0}(t)\, z_k(t) & \text{if day } t \text{ is dry} \\ \mu_{k,1}(t) + \sigma_{k,1}(t)\, z_k(t) & \text{if day } t \text{ is wet} \end{cases} \tag{10}$$

where each T_k is any of the nonprecipitation variables, $\mu_{k,0}$ and $\sigma_{k,0}$ are its mean and standard deviation for dry days, and $\mu_{k,1}$ and $\sigma_{k,1}$ are its mean and standard deviation for wet days. The seasonal dependence of the means and standard deviations in (10) is usually achieved through Fourier harmonics (i.e., sine and cosine functions) that vary smoothly through the year, simulating the annual cycle for each of these parameters. Note that even though the $\mathbf{z}(t)$ as produced by (8) have Gaussian distributions, the resulting nonprecipitation weather variables T are not necessarily also Gaussian because of the random application of dry or wet-day means and variances in (10) (Katz 1996), although in practice the simulated distributions are often close to Gaussian. The means and variances of the Ts in (10) depend not only on the dry and wet-day means and variances on the right-hand side of (10) but also on the unconditional rain-day probability (Katz, 1996), which for first-order Markov dependence is given by (2a).

A number of formulations for producing more than the basic set of three nonprecipitation variables in the WGEN framework have been proposed. In addition to maximum temperature, minimum temperature and solar radiation, the model of Wallis and Griffiths (1997) also simulates daytime winds, night-time winds and daily dew point through a straightforward extension of (8) to include $K > 3$ nonprecipitation variables. Parlange and Katz (1999) followed a similar approach to simulate wind speed and dew point in addition to temperatures and radiation, and furthermore subjected the wind speed data to a square-root transformation in order to simulate better its skewed distribution. Less comprehensive algorithms were used by Bruhn *et al*. (1980) to add simulation of daily minimum relative humidity; and by Richardson and Nicks (1990) also to simulate wind speed and wind direction.

Figure 2 shows flowcharts for the basic WGEN structure when 1) the precipitation submodel is based on a first-order Markov chain; and 2) when it is based on direct simulations of spell lengths. In this figure, numbers in square brackets refer to equation numbers in the foregoing text. Regardless of how the precipitation occurrences are handled, the nonprecipitation submodels (right-hand sides of the panels) are identical.

III Relationships between daily models and their climatic statistics

While most weather generators operate on a daily time step, their output nevertheless exhibits longer-term variations that are the counterparts of lower-frequency variations in real weather data. For example, both weather generator output and observed weather data will exhibit different monthly mean temperature or monthly total precipitation in different years. A natural way to characterize these variations is to compute the variance of these monthly quantities, which is often referred to as the interannual variance. Of course one would like the interannual variability in weather generator output to match closely the corresponding variability in the observed data, and it has been suggested (Gregory *et al*., 1993) that this comparison provides a crucial test of the similarity of synthetic and real climates.

Comparison of the precipitation statistics is particularly informative. Let $S(T)$ be the sum of T consecutive daily precipitation amounts, which will correspond to a monthly precipitation total if $T \approx 30$. The statistics of the synthetic $S(T)$ climate can be expressed in terms of its long-term mean $E[S(T)]$ and interannual variance $Var[S(T)]$ according to (e.g., Gregory *et al*., 1993; Katz and Parlange, 1998):

$$E[S(T)] = E[N(T)]\ \mu = T\ \pi\ \mu \tag{11a}$$

and

$$Var[S(T)] = E[N(T)]\ \sigma^2 + Var[N(T)]\ \mu^2 \tag{11b}$$

Here N(T) is the number of wet days during the T consecutive days, E[N(T)] is the climatological average of this quantity (4a), and Var[N(T)] is the corresponding variance (given by (4b), under first-order Markov dependence). The N(T) statistics depend only on the model chosen to represent the precipitation occurrences. The quantities μ and σ^2 in (11) are respectively the mean and variance of the daily nonzero precipitation amounts, expressions for which are given in the text above, following each of the three probability density functions (5)–(7).

Markov-chain and spell-length precipitation occurrence models both yield practically exact reproductions of the average number of wet days E[N(T)], and the probability density functions that are commonly used to represent daily nonzero precipitation likewise recover the average wet-day amount μ essentially exactly (Wilks, 1999a). Thus, (11a) indicates that weather generator models will necessarily reproduce the climatological precipitation averages. The situation for simulated temperature and solar radiation averages is similar: the average **z** produced by (8) is **0**, so that (because π is simulated correctly) the results from (10) will exhibit the correct monthly means to the extent that the conditional means $\mu_{k,0}$ and $\mu_{k,1}$ are correctly specified. These considerations imply that 'validating' a weather generator by examining only its reproduction of monthly or seasonal mean values is naive. That is, simulated and observed time-mean values will be different only to the extent that: 1) the comparison is made using a comparatively short simulated record from the generator, so that differences can be ascribed to sampling variations; 2) the specifications for within-year variations of the parameters (particularly the harmonic functions specifying the annual cycles of $\mu_{k,0}$ and $\mu_{k,1}$) require refinement; or 3) the generator has been implemented incorrectly.

The issue of simulated interannual variability is more difficult. It seems to be a general characteristic of weather generator models that their interannual variability is smaller than that of the corresponding observed data (e.g., Buishand, 1977; 1978; Gregory *et al*., 1993; Katz and Parlange, 1993; 1998; Wilks, 1989; 1999a). For the case of precipitation, (11b) indicates that this deficiency must arise from shortcomings in simulation of Var[N(T)], or σ^2, or both; because the average quantities E[N(T)] and μ are well simulated in any properly constructed generator. One possible cause for the behaviour could be that the component submodels are not sufficiently complex. Katz and Parlange (1998) and Wilks (1999a) investigated the effects of progressively more elaborate models for precipitation occurrence (second and third-order Markov chains and progressively more elaborate nonzero precipitation amount models), and found that the synthetic interannual variances could be increased but still fell short of observed climatic variability on average.

Probably the most appealing explanation for this 'missing variance' is that climate statistics in the real world change somewhat from year to year, and that the simpler weather generator forms discussed thus far cannot capture this contribution to weather variability because their statistics vary only through a fixed annual cycle. Klugman and Klugman (1981) were among the first to examine this issue, and attributed model nonstationarity to climatic change. A more common view currently is that the nonstationarity results from interannual climate variations. For example, Figure 4 shows interannual variations in three climatic quantities that are also stochastic rainfall parameters, namely: the conditional dry and wet-day probabilities $1 - p_{01}$ and p_{11}, and the mean wet-day amount μ. For this site, it is evident that year-to-year precipitation variability includes contributions both from random within-year fluctuations, and from contributions deriving from more systematic longer-period variations in these three parameters.

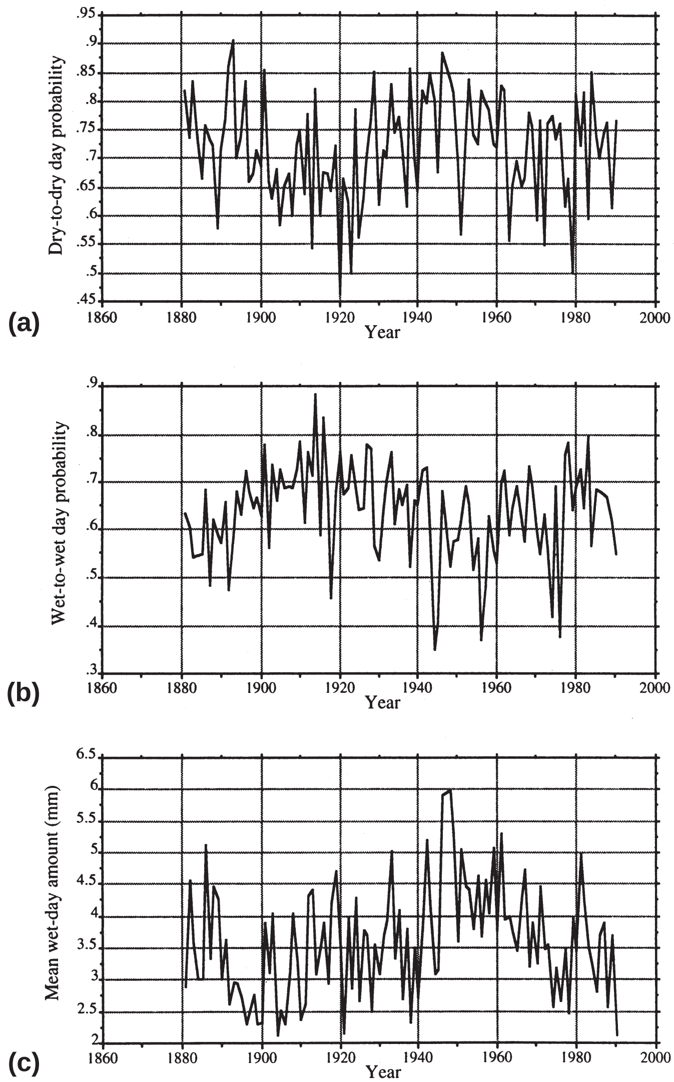

Figure 4 Annual time-series of the spring (a) conditional dry-day probabilities, p_{00}, (b) conditional wet-day probabilities, p_{11}, and (c) mean wet-day amounts, μ at Kempsford, UK, 1881–1990

It has been found that a relatively simple and straightforward approach to simulating interannual variability better is to condition the weather generator parameter set on a covariate. That is, different parameter sets are chosen for simulation in different years, depending on the value of some external variable. For example, much of the variation evident in Figure 4 is strongly correlated with large-scale, free-atmosphere vorticity over Britain (Wilby, 1997). Simple covariates that have been used successfully to condition the choice of stochastic model parameters include the monthly statistics themselves (Wilks, 1989), long-range forecasts of the monthly statistics (Briggs and Wilks, 1996), or even random numbers (Jones and Thornton, 1997). Katz and Zheng (1999) used a ‘hidden’ mixture approach to capture some interannual variability not exhibited by a stationary weather generator formulation.

More commonly, the stochastic model parameters are conditioned on some aspect of large-scale atmospheric circulation (e.g., Hay *et al*., 1991; Bardossy and Plate, 1992; Woolhiser *et al*., 1993; Hughes and Guttorp 1994a; Katz and Parlange, 1993; 1996; Wallis and Griffiths, 1997; Kiely *et al*., 1998) or by a hierarchy of precipitation mechanisms (e.g., Sansom and Thomson, 1992). For example, Wilby (1995) simulated daily precipitation for sites in the UK using the Lamb Weather Type (Lamb, 1972) (LWT) categories, subclassified by the presence or absence of weather fronts. Table 1 compares the probability of a frontal system under anticyclonic, cyclonic and purely directional airflows, as well as the corresponding likelihood of precipitation occurrence at Kempsford, UK. From such information LWTs may be used to condition both the probability of frontal weather (over the British Isles as a whole) and the occurrence of precipitation at individual stations. However, the veracity of weather pattern models depends upon the chosen weather classification system and on the observed links between model parameters and circulation classes remaining constant through time (Wilby, 1994; 1997). More recent examples of the circulation-based weather generators have favoured the use of continuous atmospheric variables as an alternative to conditioning by discrete weather categories (e.g., Conway *et al*., 1996).

In a comparable study of interannual precipitation variability at selected sites in the UK, Wilby (1998) employed the North Atlantic Oscillation Index (NAOI) and North Atlantic sea surface temperature (SST) anomalies as conditioning variables. For example, Figure 5 demonstrates that weak but statistically significant correlations may be detected between p_{11}, p_{00} and μ, and spring (March–May) North Atlantic SST anomalies, 1901–90. It is noteworthy that for this site – with the exception of the spring p_{11} parameter – the correlation strengths were not generally improved by lagging the SST anomalies (cf. Colman, 1997). In this unusual case, spring wet-spell persistence p_{11} at Kempsford, Cotswolds, UK was most strongly correlated with SST anomalies in the preceding winter (November–December). However, the over-riding implication of the

Table 1 The probability of frontal weather ($F > 0$) and of precipitation occurrence (π) by circulation type at Kempsford in the Cotswolds, UK, 1970–90

Circulation type	Probability ($F > 0$)	Probability (π) ($F > 0$)	Probability (π) ($F = 0$)
Anticyclonic	0.58	0.15	0.10
Cyclonic	0.75	0.74	0.65
Directional	0.70	0.53	0.40

study was that the realism of stochastic rainfall models may be enhanced using mixtures of slowly and rapidly varying conditioning variables (in this case, monthly SST anomalies and daily vorticity, respectively).

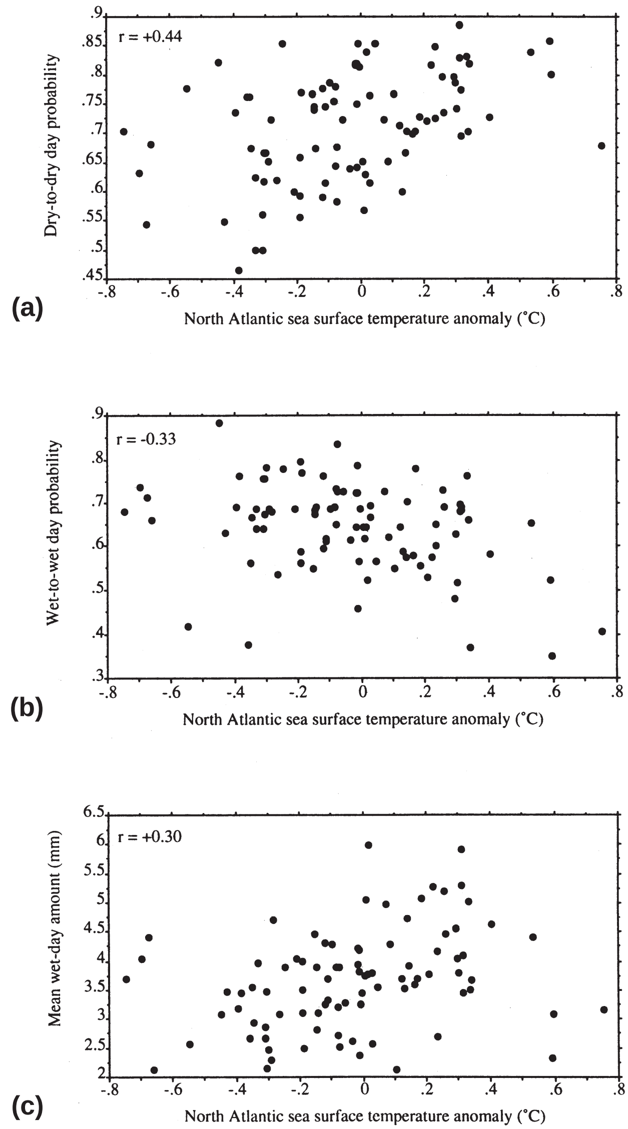

Figure 5 The relationship between spring North Atlantic sea surface temperatures and spring (a) conditional dry-day probabilities, p_{00}, (b) conditional wet-day probabilities, p_{11}, and (c) mean wet-day amounts, μ at Kempsford, UK, 1901–90

IV Weather generator applications

1 Modelling of weather and climate-sensitive systems

As mentioned above, much of the motivation for the development of weather generator models has been to support computer models of crop growth and development. Widely distributed crop modelling packages include built-in weather generation algorithms as a standard component (e.g., Richardson and Nicks, 1990). Richardson (1985) outlines the use of long synthetic weather series to estimate more smoothly (than with a comparatively short observed climate record) the frequency distributions of simulated wheat yields. Long synthetic weather series have also been used to examine the impact of extreme weather events or severe droughts on crop behaviour (e.g., Mearns *et al*., 1984) or long-term rates of soil erosion (e.g., Favis-Mortlock *et al*., 1991; 1997). Stern *et al*. (1982) computed weather statistics of relevance to a particular agricultural setting, by deriving them from a fitted weather generator model without having to rely directly on their occurrence frequencies in a short observed record. WGEN models have also found application in sensitivity studies of crop model responses to changes in climatic variability that can be controlled through judicious adjustments of the generator parameters (Nonhebel, 1994; Semenov and Porter, 1995; Mearns *et al*., 1996; Riha *et al*., 1996).

Weather generators also provide an attractive way to provide weather inputs for ecological models. Friend *et al*. (1997) obtained principal driving variables for a comprehensive terrestrial ecosystem dynamics model from a WGEN-type generator. WGEN techniques were also used to construct high-resolution, gridded daily bioclimatic data sets using monthly station data in the Vegetation/Ecosystem Modeling and Analysis Project, VEMAP (Kittel *et al*., 1995). This data set has facilitated analyses of the spatial variability in ecosystem processes at the continental scale, as well as ecosystem sensitivity to climate change (Schimel *et al*., 1997). Similarly, He (1997) used simulated weather to study hydrologic effects on the Great Lakes of increasing water withdrawals for the irrigation of nearby crop lands. Pickering *et al*. (1988) used generated weather series to investigate effects of weather variability on pollution movement through hydrologic systems.

2 Simulation of missing weather data

Another common use of stochastic weather models is the estimation of missing meteorological data. One pervasive problem in crop modelling is that very few locations report daily solar radiation data, but this meteorological variable is critically important for simulation of both photosynthesis and evaporation, and thus all aspects of plant growth and development. Richardson and Wright (1984) and Hanson *et al*. (1994) provide maps for the USA on which the radiation statistics $\mu_{3,0}$, $\mu_{3,1}$, $\sigma_{3,0}$ and $\sigma_{3,1}$ (cf. Equation 10) are interpolated as smooth contour plots. Similarly, Dennett *et al*. (1983) describe the spatial interpolation of weather generator parameters to evaluate agricultural potential at a site between two locations where weather observations are available, and Hutchinson (1986; 1995) describes spatial interpolation of weather generator parameters more generally.

A different kind of missing-data problem occurs when only monthly or seasonal statistics are available for a site, and weather generator parameters are needed for simulation at the daily timescale. Since the monthly statistics (Equation 11, for precipitation) depend on the daily parameters, one can exploit these relationships to develop specifications for the daily parameters in terms of the known monthly statistics. The accuracy of this process is improved for locations at which the average number of wet days per month (and thus the unconditional probability of a wet day, π) is available, in addition to the monthly precipitation totals from which the monthly precipitation mean and variance can be computed. Hershfield (1970) found approximately linear relationships for selected USA locations between π and the two Markov transition probabilities (1), and Hutchinson (1986) confirmed (different) linear relationships for these variables in Australian data. Geng *et al*. (1986) extended this kind of empirical linear 'inverting' of (11) to include the gamma distribution (6) parameters α and β, and for varied climates in different parts of the world. However, Hutchinson (1995) has pointed out that such linear, empirical specification equations are valid only over the range of their calibration data, and thus can lead to serious errors or even nonsense (e.g., negative probabilities) if extrapolated too far. A more theoretically satisfying approach to this problem is to use (11) directly to define the relationships between the daily parameters on the right-hand sides, in terms of the monthly or seasonal statistics on the left-hand sides (e.g., Katz, 1996; Wilks, 1992; 1999b).

3 Downscaling, and regional climate change scenarios

Another application where weather generators have found much use is in the construction of 'scenarios' of climate change, in order to investigate the impacts of those changes as portrayed by response models. Weather generators are attractive here because long samples of future climatic data are clearly not available, but estimates of the consequences of prospective climate changes are needed with some urgency. In this context, the relationships between daily weather generator parameters and climatic averages such as (11) can be used to characterize the nature of future daily statistics on the basis of more readily available time-averaged climate-change information. The resulting weather generator models are then used to simulate daily series of indefinite lengths representative of altered climates (Wilks, 1992; Katz, 1996). Another appealing characteristic of using weather generators for climate-change studies is that changes in climatic variability, in addition to changes in climatic means, can be easily simulated.

The relationships between the parameters characterizing weather series at different spatial scales can also be used to 'downscale'; i.e., disaggregate from large (climate model) to local scales before impact simulation (Dubrovsky, 1997; Semenov and Barrow, 1997; Wilks, 1999b). Most such climate-change impacts assessments have been concerned with agriculture (e.g., Wilks, 1988; Kaiser *et al*., 1993; Mearns *et al*., 1997; Semenov and Barrow, 1997) although weather-generator methods have also been used to simulate effects of climate changes on ecosystem dynamics (Strandman *et al*., 1993), erosion potential (Favis-Mortlock and Boardman, 1995) and other aspects of hydrology (Valdes *et al*., 1994; Wilby *et al*., 1994).

V Future directions

1 Simultaneous weather simulation at multiple sites

The stochastic weather models described so far are of limited value for modelling certain spatially distributed processes. That is, using these relatively simple single-site models for simultaneous simulation of weather sequences at multiple points, for example to evaluate regional hydrological or agricultural behaviour, requires that the quite strong spatial correlation in weather be ignored.

One approach to simultaneous weather simulation at multiple locations is through the multivariate normal distribution (e.g., Wilks, 1995), the properties of which are very well known, and which is very convenient conceptually and computationally. Bardossy and Plate (1992) and Hutchinson (1995) have described multisite precipitation simulation based on the multivariate normal, which relies on transformation of distributions of daily precipitation at the sites to normal distributions. This model tacitly assumes that the same characteristic spatial scale applies to both precipitation occurrence and precipitation amount, and also has some difficulty capturing the intermittent (mixed discrete and continuous) nature of precipitation fields.

Another approach to simultaneous weather simulation at multiple sites is through models that explicitly (but stochastically) simulate the spatially distributed physical phenomena that generate weather. These have been primarily models of precipitation (e.g., Waymire *et al.*, 1984; Cox and Isham 1988). A somewhat similar approach is the use of 'hidden' Markov models to simulate stochastically the effects of large-scale atmospheric circulation on local weather (e.g., Hughes and Guttorp, 1994b; Hughes *et al.*, 1999).

Recently Wilks (1998) described a generalization of the precipitation part of simple WGEN-type models, described in previous sections, to simultaneous weather simulation at multiple sites. Wilks (1999c) extends this approach to include nonprecipitation variables also. Here the idea is to retain the single-site model parameters at each location to be simulated, and to drive each of these local WGEN models with random numbers that are spatially correlated. The key step in constructing such a model is then specification of the spatial correlation structure of the random numbers, which can be done using a straightforward, empirical algorithm.

2 Modelling interannual climate variability

As noted previously, seasonal and interannual climate variability is not always captured adequately by conventional weather generators. This deficiency is an artifact of the simplifying assumptions made by the models and/or the absence of low-frequency predictor variables. Further to Wilby's (1998) analysis of precipitation data for central England, Figure 6 shows spatial variations in the correlation strengths between the NAOI and seasonal p_{11} series for the period 1961–90 using a network of over 90 precipitation stations distributed throughout the British Isles. It is evident that the strongest positive correlations actually occurred in northern and western Scotland, the English Lake District and Wales, with a marked west–east gradient in the correlation field for winter (DJF). The same season had weak (and statistically insignificant) negative correlations in central and south-east England. During the transitions

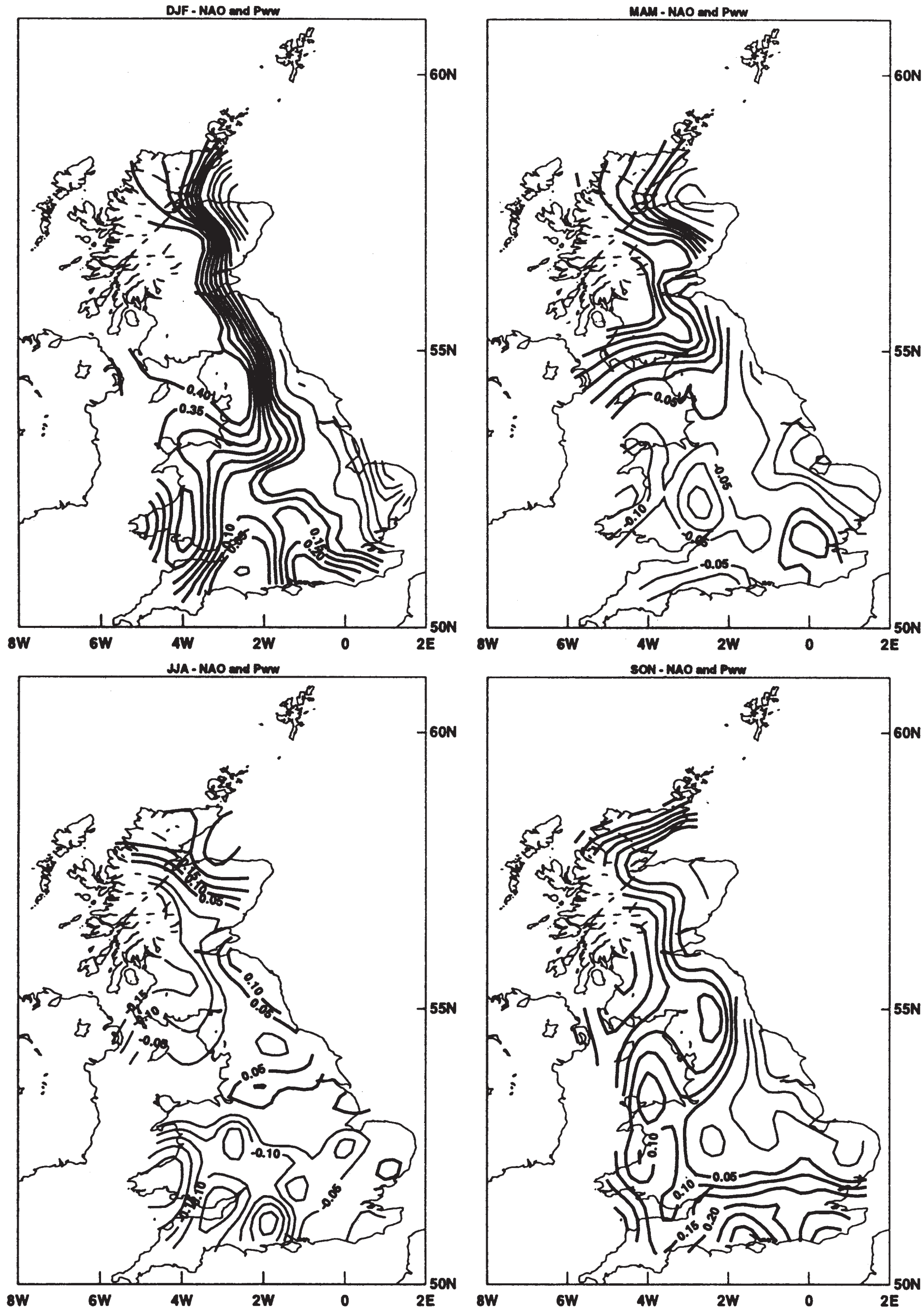

Figure 6 Correlations between the seasonal North Atlantic Oscillation Index and conditional wet-day probabilities, p_{11} across the British Isles, 1961–90. Correlation coefficients of ± 0.35 are significant at $p = 0.05$
Source: Wilby *et al*. (1999)

between winter and spring (MAM), and then from spring to summer (JJA), there were progressive retreats of the area of significant positive correlations and an expansion of the negative correlation field. For summers there was a breakdown of the dipole pattern with statistically insignificant correlations over all the British Isles. However, by autumn (SON) significant positive correlations emerge once again in northern and western Scotland.

The correlation fields shown in Figure 6 have two implications for weather generation. First, it is clear that the modest gains in the simulation of seasonal precipitation reported for Wilby's (1998) model reflect the location of the two test sites (i.e., Kempsford, central England, and Durham, northeast England) with respect to the regions of strongest forcing. Had sites been chosen in highland Scotland, for example, it is probable that the incorporation of the slowly varying predictor (i.e., the NAOI) would have yielded more significant gains in model performance relative to the vorticity-only model, particularly in winter. Secondly, it is evident that the strength of NAOI forcing of WGEN parameters is highly seasonal, accounting for as much as 40% of the interannual variability in wintertime p_{11} over parts of western Scotland, but having little or no influence in summer at the same sites. Therefore improved representations of interannual precipitation variability in downscaling schemes by means of low-frequency predictor variables (such as the NAOI) will be spatially and seasonally specific.

The strength of correlations between slowly varying predictors and WGEN parameters at individual sites may also be increased by optimizing the predictor domain. This is accomplished by inverting the correlation search, using a fixed target site (e.g., Kempsford) and searching a global predictor set for the spatial domain of strongest correlation(s). For example, the weak positive correlation (shown in Figure 5a) between spring p_{00} and North Atlantic SST anomalies was obtained using sea-surface temperature anomalies averaged over the domain 40° N – 60° N, 35° W – 5° E. However, as Figure 7 indicates, this domain (shown by the grey shading) only partially samples the region of significant correlations. Given the correlation field for spring, it would be legitimate to employ a predictor domain that extends further south, even as far as 20° N. Alternatively, the shape of the predictor domain might be modified in the case of autumn (when weak and/or negative correlations at the western edge of the domain 'dilute' the strongly positive correlations to the east). In the case of summer and autumn it might even be reasonable to employ an additional predictor domain such as the ocean region off the eastern seaboard of the USA.

The preceding examples suggest that slowly varying predictors have explanatory power for certain meteorological variables, in certain regions and seasons, under present climate conditions. However, the validity of using such empirical relationships for statistically downscaling future precipitation will depend on 1) the ability of GCMs faithfully to reproduce large-scale forcing patterns (such as the NAO) under current climate conditions (see, for example, Davies *et al*., 1997; Osborn *et al*., 1999); 2) the extent to which global warming will affect future interdecadal climate variability (e.g., Trenberth and Hoar, 1997); and 3) the stationarity of the empirical relationships between mesoscale forcing and local-scale meteorological responses (Wilby, 1997).

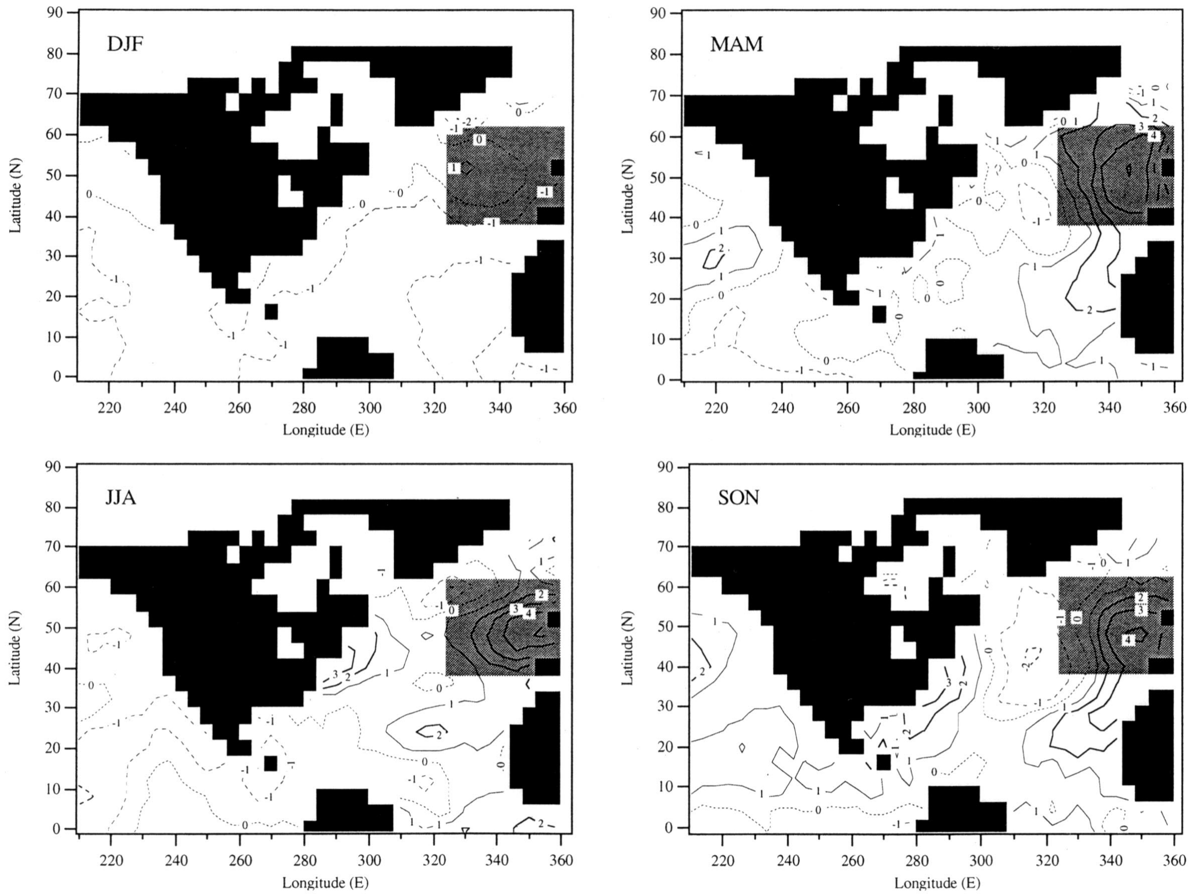

Figure 7 Correlations between seasonal SSTs and conditional dry-day probabilities (p_{00}) at Kempsford, UK, for the period 1881–1990. Positive correlations are indicated by solid isolines, a dotted line is the zero correlation isoline and negative correlations are indicated by dashed isolines. The contour interval is 0.1 and correlation coefficients of ± 0.2 (thicker isolines) are significant at p = 0.05. The predictor domain used by Wilby's (1998) downscaling model is shown in grey

3 Nonparametric approaches

Finally, we conclude with a brief description of nonparametric (i.e., not requiring that particular theoretical probability distributions be assumed) approaches to stochastic simulation of weather series. Many modern nonparametric statistical methods involve 'resampling' (e.g., Efron, 1982), which means that large ersatz samples are synthesized by repeatedly but randomly copying data values from an existing finite record. For the case of weather series, an important complication that usually invalidates simple resampling schemes is the prominent time correlation evident in these data, the modelling of which is the subject of much of this review. That is, simply concatenating a random sample of daily weather data (even if restricted to a single month or season) will simulate the original series poorly because the prominent time correlation will be destroyed by the resampling process.

More complex procedures to resample weather series in ways that also capture the important time correlations have been described by Young (1994), Lall and Sharma (1996), Lall *et al*. (1996) and Rajagopalan *et al*. (1997). Recently, Bardossy (1998) has proposed the use of an algorithm known as simulated annealing to reshuffle simple resampled data in a way that important statistical properties of the original series (especially, but not limited to, the time correlation) are reconstituted also. The simulated annealing approach is quite intensive computationally, but may prove to be especially useful for simulating series with very short (minutes or hours) time steps.

These nonparametric methods are attractive in that subjective judgements about most appropriate model forms and probability distributions are avoided, and furthermore as data-based methods they can capture deviations from theoretical probability distributions for the individual variables, and nonlinearities in the relationships among variables. On the other hand, being based wholly on the observed data they are somewhat limited in the range of extreme values that can be generated, although 'smoothed' resampling schemes (e.g., Rajagopalan *et al*., 1997) ameliorate this problem somewhat. Another difficulty with the nonparametric methods for some applications is that future climate regimes cannot be easily constructed through simple parameter adjustments.

VI Concluding remarks

From the preceding review it is evident that there have been considerable advances in weather generation research in the 150 years since Quetelet's pioneering study. As in many other branches of meteorology there is an ongoing need for critical assessment of emergent techniques as well as for comparisons between conventional and refined model formulations. The widespread and growing use of regionally specific, environmental impacts models implies a corresponding demand for high integrity weather data. For as long as there remains a mismatch between the data that our limited observational networks can supply and that which the impacts community requires, there will be a continuing need for statistical weather generation.

Acknowledgements

We thank Rick Katz and an anonymous reviewer for constructive suggestions that have improved this review. This work was sponsored in part by ACACIA (A Consortium for the Application of Climate Impact Assessments), the National Center for Atmospheric Research and by ACCORD (Atmospheric Circulation Classification and Regional Downscaling) under the European Community Environment Research Programme (contract: ENV4-CT97-0530). NCAR is sponsored by the National Science Foundation.

References

Abramowitz, M. and **Stegun, I.A.**, editors, 1984: *Pocketbook of mathematical functions*. Frankfurt: Verlag Harri Deutsch.

Akaike, K. 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–23.

Bardossy, A. 1998: Generating precipitation time series using simulated annealing. *Water Resources Research* 34, 1737–44.

Bardossy, A. and **Plate, E.J.** 1992: Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research* 28, 1247–59.

Besson, L. 1924: On the probability of rain. *Monthly Weather Review* 52, 308.

Bras, R. and **Rodriguez-Iturbe, I.** 1976: Rainfall generation: a nonstationary time varying multidimensional model. *Water Resources Research* 12, 450–56.

Bratley, P., Fox, B.L. and **Schrage, L.E.** 1987: *A guide to simulation*. New York: Springer-Verlag.

Briggs, W.M. and **Wilks, D.S.** 1996: Extension of the CPC long-lead temperature and precipitation outlooks to general weather statistics. *Journal of Climate* 9, 3496–504.

Bruhn, J.A., Fry, W.E. and **Fick, G.W.** 1980: Simulation of daily weather data using theoretical probability distributions. *Journal of Applied Meteorology* 19, 1029–36.

Buishand, T.A. 1977: *Stochastic modeling of daily rainfall sequences. Mededlingen Landbouwhogeschool Wageningen* 77–3.

—— 1978: Some remarks on the use of daily rainfall models. *Journal of Hydrology* 36, 295–308.

Chin, E.H. and **Miller, J.F.** 1980: On the conditional distribution of daily precipitation amounts. *Monthly Weather Review* 108, 1462–64.

Cochran, W.G. 1938: An extension of Gold's method of examining the apparent persistence of one type of weather. *Quarterly Journal of the Royal Meteorological Society* 64, 631–34.

Colman, A. 1997: Prediction of summer Central England Temperature from preceding North Atlantic winter sea surface temperature. *International Journal of Climatology* 17, 1285–1300.

Conway, D., Wilby, R.L. and **Jones, P.D.** 1996: Precipitation and air flow indices over the British Isles. *Climate Research* 7, 169–83.

Cox, D.R. and **Isham, V.** 1988: A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society London* A415, 317–28.

Davies, J.R., Rowell, D.P. and **Folland, C.K.** 1997: North Atlantic and European seasonal predictability using an ensemble of multi-decadal AGCM simulations. *International Journal of Climatology* 17, 1263–84

Dennett, M.D.J., Rodgers, J.A. and **Keatinge, J.D.H.** 1983: Simulation of a rainfall record for a new site of a new agricultural development: an example from northern Syria. *Agricultural Meteorology* 29, 247–58.

Devroye, L. 1986: *Non-uniform random variate generation*. New York: Springer.

Dubrovsky, M. 1997: Creating daily weather series with use of the weather generator. *Environmetrics* 8, 409–24.

Efron, B. 1982: *The Jackknife, the Bootstrap, and other resampling plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Favis-Mortlock, D.T. and **Boardman, J.** 1995: Nonlinear responses of soil erosion to climate change: a modelling study on the UK South Downs. *Catena* 25, 365–87.

Favis-Mortlock, D.T., Boardman, J. and **Bell, M.** 1997: Modelling long-term anthropogenic erosion of a loess cover: South Downs, UK. *The Holocene* 7, 79–89.

Favis-Mortlock, D.T., Evans, R., Boardman, J. and **Harris, T.M.** 1991: Climate change, winter wheat yield and soil erosion on the

English South Downs. *Agricultural Systems* 37, 415–33.

Foufoula-Georgiou, E. and **Lettenmaier, D.P.** 1987: A Markov renewal model for rainfall occurrences. *Water Resources Research* 23, 875–84.

Friend, A.D., Stevens, A.K., Knox, R.G. and **Cannell, M.G.R.** 1997: A process-based terrestrial biosphere model of ecosystem dynamics. *Ecological Modelling* 95, 249–87.

Gabriel, K.R. and **Neumann, J.** 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society* 88, 90–95.

Gates, P. and **Tong, H.** 1976: On Markov chain modeling to some weather data. *Journal of Applied Meteorology* 15, 1145–51.

Geng, S., Penning de Vries, F.T.W. and **Supit, I.** 1986: A simple method for generating daily rainfall data. *Agricultural and Forest Meteorology* 36, 363–76.

Gold, E. 1929: Note on the frequency of occurrence of sequences in a series of events of two types. *Quarterly Journal of the Royal Meteorological Society* 55, 307–309.

Gregory, J.M., Wigley, T.M.L. and **Jones, P.D.** 1993: Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. *Climate Dynamics* 8, 299–310.

Guttorp, P. 1995: *Stochastic modeling of scientific data*. London: Chapman & Hall.

Haan, C.T., Allen, D.M. and **Street, J.O.** 1976: A Markov chain model of daily rainfall. *Water Resources Research* 12, 443–49.

Hanson, C.L., Cumming, K.A., Woolhiser, D.A. and **Richardson, C.W.** 1994: *Microcomputer program for daily weather simulation in the contiguous United States* (ARS-114). Washington, DC: US Department of Agriculture.

Hay, L.E., McCabe, G.J., Wolock, D.M. and **Ayers, M.A.** 1991: Simulation of precipitation by weather type analysis. *Water Resources Research* 27, 493–501.

Hayhoe, H.N. 1998: Relationship between weather variables in observed and WXGEN generated data series. *Agricultural and Forest Meteorology* 90, 203–14.

He, C. 1997: Modeling hydrologic impact of withdrawing the Great Lakes water for agricultural irrigation. *Journal of the American Water Resources Association* 33, 1055–68.

Hershfield, D.M. 1970: A comparison of conditional and unconditional probabilities for wet- and dry-day sequences. *Journal of Applied Meteorology* 9, 825–27.

Hughes, J.P. and Guttorp, P. 1994a: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research* 30, 1535–46.

—— 1994b: Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology* 33, 1503–15.

Hughes, J.P., Guttorp, P. and **Charles, S.P.** 1999: A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, in press.

Hutchinson, M.F. 1986: Methods of generation of weather sequences. In: Bunting, A.H., editor, *Agricultural evironments*, Wallingford: CAB International, 149–57.

—— 1995: Stochastic space-time models from ground-based data. *Agricultural and Forest Meteorology* 73, 237–64.

Jones, P.G. and **Thornton, P.K.** 1997: Spatial and temporal variability of rainfall related to a third-order Markov model. *Agricultural and Forest Meteorology* 86, 127–38.

Kaiser, H.M., Riha, S.J., Wilks, D.S., Rossiter, D.G. and **Sampath, R.** 1993: A farm-level analysis of the economic and agronomic impacts of gradual climate warming. *American Journal of Agricultural Economics* 75, 387–98.

Katz, R.W. 1977: Precipitation as a chain-dependent process. *Journal of Applied Meteorology* 16, 671–76.

—— 1981: On some criteria for estimating the order of a Markov chain. *Technometrics* 23, 243–49.

—— 1985: Probabilistic models. In: Murphy, A.H. and Katz, R.W., editors, *Probability, statistics, and decision making in the atmospheric sciences*, Boulder, CO: Westview Press, 261–88.

—— 1996: Use of conditional stochastic models to generate climate change scenarios. *Climatic Change* 32, 237–55.

Katz, R.W. and **Parlange, M.B.** 1993: Effects of an index of atmospheric circulation on stochastic properties of precipitation. *Water Resources Research* 29, 2335–44.

—— 1995: Generalization of chain-dependent processes: application to hourly precipitation. *Water Resources Research* 31, 1331–41.

—— 1996: Mixtures of stochastic processes: application to statistical downscaling. *Climate Research* 7, 185–93.

—— 1998: Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate* 11, 591–601.

Katz, R.W. and **Zheng, X.** 1999: Mixture model for overdispersion of precipitation. *Journal of Climate*, in press.

Kiely, G., Albertson, J.D., Parlange, M.B. and **Katz, R.W.** 1998: Conditioning stochastic properties of daily precipitation on indices of atmospheric circulation. *Meteorological Applications* 5, 75–87.

Kittel, T.G.F., Rosenbloom, N.A., Painter, T.H., Schimel, D.S. and **VEMAP Modeling Participants** 1995: The VEMAP integrated database for modeling United States ecosystem/vegetation sensitivity to climate change. *Journal of Biogeography* 22, 857–62.

Klugman, M.R. and **Klugman, S.A.** 1981: A method for determining change in precipitation data. *Journal of Applied Meteorology* 20, 1506–509.

Knuth, D.E. 1997: *The art of computer programming, v 2, semi-numerical algorithms*. Reading, MA: Addison-Wesley.

Lall, U., Rajagopalan, B. and **Tarboton, D.G.** 1996: A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resources Research* 32, 2803–23.

Lall, U. and **Sharma, A.** 1996: A nearest-neighbor bootstrap for resampling hydrological time series. *Water Resources Research* 32, 679–93.

Lamb, H.H. 1972: *British Isles weather types and a register of the daily sequence of circulation patterns, 1861–1971. Meteorological Office, Geophysical Memoir* 116. London: HMSO.

Longley, R.W. 1953: The length of dry and wet periods. *Quarterly Journal of the Royal Meteorological Society* 79, 520–27.

Mearns, L.O., Katz, R.W. and **Schneider, S.H.** 1984: Extreme high-temperature events: changes in their probabilities with changes in mean temperature. *Journal of Applied Meteorology* 23, 1601–13.

Mearns, L.O., Rosenzweig, C. and **Goldberg, R.** 1996: The effect of changes in daily and interannual climatic variability on CERES-wheat yields. A sensitivity study. *Climatic Change* 32, 257–92.

—— 1997: Mean and variance changes in climate scenarios: methods, agricultural applications, and measures of uncertainty. *Climatic Change* 35, 367–96.

Newnham, E.V. 1916: The persistence of wet and dry weather. *Quarterly Journal of the Royal Meteorological Society* 42, 153–62.

Nonhebel, S. 1994: The effects of use of average instead of daily weather data in crop growth simulation models. *Agricultural Systems* 44, 377–96.

Osborn, T.J., Briffa, K.R., Tett, S.F.B., Jones, P.D. and **Trigo, R.M.** 1999: Evaluation of the North Atlantic Oscillation as simulated by a climate model. *Climate Dynamics*, in press.

Parlange, M.B. and **Katz, R.W.** 1999: An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology*, in press.

Pickering, N.B., Stedinger, J.R. and **Haith, D.A.** 1988: Weather input for nonpoint source pollution models. *Journal of Irrigation and Drainage Engineering* 114, 674–90.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and **Vetterling, W.T.** 1986: *Numerical recipes, the art of scientific computing*. Cambridge: Cambridge University Press.

Racsko, P., Szeidl, L. and **Semenov, M.** 1991: A serial approach to local stochastic weather models. *Ecological Modelling* 57, 27–41.

Rajagopalan, B., Lall, U., Tarboton, D.G. and **Bowles, D.S.** 1997: Multivariate nonparametric resampling scheme for generation of daily weather variables. *Stochastic Hydrology and Hydraulics* 11, 65–93.

Richardson, C.W. 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* 17, 182–90.

—— 1985: Weather simulation for crop management models. *Transactions of the American Society of Agricultural Engineers* 28, 1602–606.

Richardson, C.W. and **Nicks, A.D.** 1990: Weather generator description. In Sharpley, A.N. and Williams, J.R., editors, *Erosion/productivity impact calculator. 1. Model documentation*. *Agricultural Research Service Technical Bulletin* 1768, Washington, DC: US Department of Agriculture, 93–104.

Richardson, C.W. and **Wright, D.A.** 1984: *WGEN: a model for generating daily weather variables* (ARS-8). US Department of Agriculture.

Riha, S.J., Wilks, D.S. and **Simoens, P.** 1996: Impact of temperature and precipitation variability on crop model predictions. *Climatic Change* 32, 293–311.

Roldan, J. and **Woolhiser, D.A.** 1982: Stochastic daily precipitation models. 1. A comparison of occurrence processes. *Water Resources Research* 18, 1451–59.

Sansom, J. and **Thomson, P.J.** 1992: Rainfall classification using breakpoint pluviograph data. *Journal of Climate* 5, 755–64.

Schimel, D.S., VEMAP Participants and **Braswell, B.H.** 1997: Spatial variability in ecosystem processes at the continental scale: models, data and the role of disturbance. *Ecological Monographs* 67, 251–71.
Schwarz, G. 1978: Estimating the dimension of a model. *Annals of Statistics* 6, 461–64.
Semenov, M.A. and **Barrow, E.M.** 1997: Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change* 35, 397–414.
Semenov, M.A., Brooks, R.J., Barrow, E.M. and **Richarrdson, C.W.** 1998: Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research* 10, 95–107.
Semenov, M.A. and **Porter, J.R.** 1995: Climatic variability and the modelling of crop yields. *Agricultural and Forest Meteorology* 73, 265–83.
Singh, S.V. and **Kripalani, R.H.** 1986: Analysis of persistence in daily monsoon rainfall over India. *Journal of Climatology* 6, 625–39.
Stern, R.D. and **Coe, R.** 1984: A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society Series A* 147, 1–34.
Stern, R.D., Dennett, M.D. and **Dale, I.C.** 1982: Analysing daily rainfall measurements to give agronomically useful results. II. A modelling approach. *Experimental Agriculture* 18, 237–53.
Strandman, H., Vaisanen, H. and **Kellomaki, S.** 1993: A procedure for generating synthetic weather records in conjunction of climate scenario for modelling of ecological impacts of changing climate in boreal conditions. *Ecological Modelling* 70, 195–220.

Thom, H.C.S. 1958: A note on the gamma distribution. *Monthly Weather Review* 86, 117–22.
Todorovic, P. and **Woolhiser, D.A.** 1975: A stochastic model of n-day precipitation. *Journal of Applied Meteorology* 14, 17–24.
Trenberth, K.E. and **Hoar, T.J.** 1997: El Nino and climate change. *Geophysical Research Letters* 24, 3057–60.

Valdes, J.B., Seoane, R.S. and **North, G.R.** 1994: A methodology for the evaluation of global warming impact on soil moisture and runoff. *Journal of Hydrology* 161, 389–413.

Wallis, T.W.R. and **Griffiths, J.F.** 1997: Simulated meteorological input for agricultural models. *Agricultural and Forest Meteorology* 88, 241–58.
Waymire, E., Gupta, V.K. and **Rodriguez-Iturbe, I.** 1984: A spectral theory of rainfall at the meso-β scale. *Water Resources Research* 20, 1453–65.
Wilby, R.L. 1994: Stochastic weather type simulation for regional climate change impact assessment. *Water Resources Research* 30, 3395–403.
—— 1995: Simulation of precipitation by weather pattern and frontal analysis. *Journal of Hydrology* 173, 91–109.
—— 1997: Nonstationarity in daily precipitation series: implications for GCM downscaling using atmospheric circulation indices. *International Journal of Climatology* 17, 439–54.
—— 1998: Statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices. *Climate Research* 10, 163–178.
Wilby, R.L., Conway, D. and **Jones, P.D.** 1999: Prospects for downscaling seasonal precipitation variability using conditioned weather generator parameters. *Hydrological Processes*, submitted.
Wilby, R.L., Greenfield, B. and **Glenny, C.** 1994: A coupled synoptic-hydrological model for climate change impact assessment. *Journal of Hydrology* 153, 265–90.
Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J. and **Wilks, D.S.** 1998: Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research* 34, 2995–3008.
Wilks, D.S. 1988: Estimating the consequences of CO_2-induced climatic change on North American grain agriculture using general circulation model information. *Climatic Change* 13, 19–42.
—— 1989: Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research* 25, 1429–39.
—— 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change* 22, 67–84.
—— 1995: *Statistical methods in the atmospheric sciences*. San Diego, CA: Academic Press.
—— 1998: Multisite generalizations of a daily stochastic precipitation generation model. *Journal of Hydrology* 210, 178–91.
—— 1999a: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology* 93, 153–69.
—— 1999b: Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research* 11, 125–36.
—— 1999c: Simultaneous stochastic simulation of daily precipitation, temperature, and solar radiation at multiple sites in complex terrain. *Agricultural and Forest Meteorology*, in press.
Williams, C.B. 1952: Sequences of wet and of dry

days considered in relation to the logarithmic series. *Quarterly Journal of the Royal Meteorological Society* 78, 91–96.

Woolhiser, D.A., Keefer, T.O. and **Redmonds, K.T.** 1993: Southern Oscillation effects on daily precipitation in the southwestern United States. *Water Resources Research* 29, 1287–95.

Woolhiser, D.A. and **Pegram, G.G.S.** 1979: Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *Journal of Applied Meteorology* 18, 34–42.

Woolhiser, D.A. and **Roldan, J.** 1982: Stochastic daily precipitation models. 2. A comparison of distribution of amounts. *Water Resources Research* 18, 1461–68.

Young, K.C. 1994: A multivariate chain model for simulating climatic parameters from daily data. *Journal of Applied Meteorology* 33, 661–71.