Mixtures of stochastic processes: application to statistical downscaling

Richard W. Katz[1,*], Marc B. Parlange[2,***]

[1]Environmental and Societal Impacts Group, National Center for Atmospheric Research[**], Boulder, Colorado 80307, USA
[2]Hydrologic Science, Biological and Agricultural Engineering, University of California, Davis, California 95616, USA

ABSTRACT: Analyses of mixtures of stochastic processes have begun to appear in climate research in recent years. Some general properties of mixtures that are well known within statistics, but not ordinarily utilized in complete generality in climate applications, are reviewed. How these issues apply in certain types of statistical downscaling is described. An important distinction is drawn between 'conditional' models, sometimes utilized in downscaling, and 'unconditional' models, utilized in more traditional approaches. Through a combination of the individual conditional models, a single overall (or 'induced') model is obtained. Among other things, the mixture concept suggests physically plausible mechanisms by which more complex stochastic models could arise in climate applications. As an application, the stochastic modeling of time series of daily precipitation amount conditional on a monthly index of large- (or regional) scale atmospheric circulation patterns is considered. Chain-dependent processes are used both as conditional and unconditional models of precipitation. For illustrative purposes, precipitation measurements for a site in California, USA, were fitted. How the mixture approach can aid in determining the properties of climate change scenarios produced by downscaling is demonstrated in this example. In particular, changes in the relative frequency of occurrence of the states of the circulation index would be associated not just with changes in mean precipitation, but with changes in its variance as well.

KEY WORDS: Atmospheric circulation index · Chain-dependent process · Climate change scenarios · Daily precipitation · Overdispersion

1. INTRODUCTION

Climate research rarely considered the topic of statistical mixtures until recently, despite a long history of study within the field of statistics (e.g. Everett & Hand 1981, Titterington et al. 1985). One instance where statistical mixtures are needed is termed statistical 'downscaling' (von Storch et al. 1993). Impact assessments require climate change scenarios on small spatial and short temporal scales and for variables, especially precipitation, that are not predicted well by physical/numerical models of climate. On the other hand, these models do have substantial skill in predicting large- (or

regional) scale patterns in atmospheric circulation that are long-lived. Statistical downscaling attempts to empirically relate these 2 seemingly disparate forms of climate information. Lettenmaier (1995) reviewed the recent literature on statistical downscaling as applied to precipitation.

In statistical terminology, a 'mixture' refers to the representation of the marginal distribution of a random variable as a weighted average of component (or conditional) distributions. More generally, a mixture could consist of a weighted average of joint distributions, such as those that characterize a stochastic process. Most often the concept of a mixture of distributions or of stochastic processes is not explicitly recognized as such in climate research. Instead, it only appears indirectly through consideration of conditional distributions for a climate variable (e.g. the distribution of monthly or seasonal total precipitation is specified conditional on whether or not an El Niño event occurs) or of conditional stochastic processes as a model for

climate time series (e.g. as sometimes arises in statistical downscaling).

In this study, some well known statistical properties of mixtures of distributions and of stochastic processes are reviewed, properties that have usually not been utilized in their full generality in applications to climate. A single overall (or 'induced') stochastic process is obtained through a combination of the individual conditional stochastic processes. The probabilistic structure of such an induced model is compared with that for a single 'unconditional' model of the same form as the conditional models. In this way, the properties of climate change scenarios generated through statistical downscaling are determined.

Katz & Parlange (1993) fit stochastic models for local daily precipitation conditional on a monthly index of large- (or regional) scale atmospheric circulation. Called 'chain-dependent processes', these stochastic models allow for the intermittency of precipitation. This application of statistical downscaling serves as motivation for the present paper. In particular, the same precipitation measurements for a site in California, USA, previously modeled in Katz & Parlange (1993) are considered for illustrative purposes.

2. DOWNSCALING AS A STATISTICAL MIXTURE

In this section, the statistical properties for the simpler situation of a mixture of distributions are reviewed first. Then the more complex situation of a mixture of stochastic processes is considered. We note that Katz (1996) made use of some of these same statistical characteristics of mixtures in a context somewhat different from downscaling, namely, stochastic models of time series of daily weather in which one variable is specified conditional on another.

2.1. Mixtures of distributions

In some downscaling applications, the conditional distribution of a climate variable, X (e.g. monthly or seasonal total precipitation, ignoring the possibility of no precipitation), is specified given a discrete random variable, I (e.g. an index of atmospheric circulation), with density function $f_{X|I}(x|i)$ where $i = 0, 1, \ldots$. The marginal distribution of X, with density function $f_X(x)$, can be obtained from these conditional distributions by

$$f_X(x) = \sum_i f_{X|I}(x|i)\Pr\{I = i\} \tag{1}$$

That is, the marginal density function is a weighted average of the conditional densities. If the possible states of the index I are finite, then Eq. (1) is termed a 'finite' mixture (Everett & Hand 1981, Titterington et

al. 1985). Because of the intermittency of precipitation, daily amount has itself a mixed continuous-discrete distribution, whether modeled conditionally or unconditionally.

Much research in statistics has dealt with the problem of estimating the components on the right-hand side of Eq. (1) when the index I is not actually observed or is even only imagined to exist, the case of a 'hidden' mixture (e.g. Zucchini & Guttorp 1991). Nevertheless, climate applications have not tended to make use of these methods, but to focus instead on related, but not equivalent, problems like detecting multimodality (e.g. Nitsche et al. 1994). Among other things, the marginal distribution in Eq. (1) can still be unimodal (see Katz 1996 for a climate example illustrating this point). In applications such as statistical downscaling for which the index I is observed, only the conditional distributions (not the equivalent form of marginal distribution in Eq. 1) are typically explicitly examined.

We now focus on the relationship between the moments of the marginal distribution and the moments of the corresponding conditional distributions in Eq. (1). The definition of conditional moments is reviewed in Appendix 1. A well-known property of conditional expectation, denoted by $E(X|I)$, is

$$E(X) = E[E(X|I)] = \sum_i E(X|I = i)\Pr\{I = i\} \tag{2}$$

This relationship of the mean of the marginal distribution of X being just a weighted average of conditional means follows directly from Eq. (1). The conditional variance, denoted by $\text{var}(X|I)$, has the property

$$\text{var}(X) = E[\text{var}(X|I)] + \text{var}[E(X|I)] \tag{3}$$

That is, the variance of the marginal distribution of X is not simply a weighted average of conditional variances, but reflects the variation in the conditional mean as well.

2.2. Mixtures of stochastic processes

Appendix 1 includes the definition of conditional covariance, as is necessary to determine the autocorrelation function of a mixture of stochastic processes. For the following, let Y denote another arbitrary climate variable (e.g. X and Y might represent observations of the same climate variable at consecutive time periods). The covariance of the joint distribution of X and Y can be expressed in terms of the covariance of the conditional joint distribution of X and Y given the index I, denoted by $\text{cov}(X, Y|I)$, and the 2 conditional means as

$$\text{cov}(X, Y) = E[\text{cov}(X, Y|I)] + \text{cov}[E(X|I), E(Y|I)] \tag{4}$$

Being a natural generalization of Eq. (3), the relationship in Eq. (4) indicates that the covariance is likewise

not simply a weighted average of conditional covariances, but also depends on the covariation of the 2 conditional means.

Eq. (4) has significant implications for the probabilistic structure of the stochastic process induced by a mixture of conditional stochastic processes. For instance, it indicates that a mixture of first-order Markov chains (a model for daily precipitation occurrence) will have a more complex dependence structure than that for a first-order chain, with an autocorrelation function that resembles a higher-order chain. In other words, directly fitting an unconditional first-order chain is not necessarily equivalent to (or consistent with) assuming the same form of model conditionally. Similar comments would apply to a mixture of first-order autoregressive [AR(1)] processes.

3. APPLICATION TO DAILY PRECIPITATION

In this section, we abstract the type of modeling approach used by Katz & Parlange (1993) in stochastic downscaling applied to precipitation, making clear how the methodology of mixtures is relevant. First the definition and properties of a chain-dependent process, a relatively simple stochastic model for time series of daily precipitation amount, are reviewed. Then the situation of such a chain-dependent process being fitted conditional on an index of large- (or regional) scale atmospheric circulation is introduced. Chain-dependent processes are frequently used to model daily precipitation; for instance, in the popular stochastic weather generator WGEN (Richardson & Wright 1984), this process is adopted for its precipitation component.

3.1. Unconditional chain-dependent process

A chain-dependent process (Todorovic & Woolhiser 1975, Katz 1977) has the desirable feature of requiring only a relatively small number of parameters, while still accounting for the most important statistical features of precipitation time series. These features include the intermittency of the precipitation process, the tendency of wet spells (i.e. runs of consecutive days on which precipitation occurs) or of dry spells to persist, and the positive skewness of the distribution of daily precipitation 'intensity' (i.e. precipitation amount given its occurrence). Because the statistical downscaling approach can have a multiplicative effect on the total number of parameters required to be estimated, parsimonious modeling is an especially important consideration.

Let $\{J_t: t = 1, 2, ...\}$ denote the sequence of daily precipitation occurrence (i.e. $J_t = 1$ indicates a 'wet day'

and $J_t = 0$ a 'dry day'). It is assumed that this process is a first-order Markov chain, a model completely characterized by the transition probabilities

$$P_{ij} = \Pr\{J_{t+1} = j \mid J_t = i\} \qquad i, j = 0, 1 \qquad (5)$$

It is convenient to reparameterize the Markov chain in terms of the probability of a wet day, $\pi = \Pr\{J_t = 1\}$, and the first-order autocorrelation coefficient (or persistence parameter), $d = \mathrm{corr}(J_t, J_{t+1})$. Now it is straightforward to show that

$$\pi = P_{01}/(P_{01} + P_{10}) \qquad d = P_{11} - P_{01} \qquad (6)$$

Note that $d > 0$ for time series of daily precipitation occurrence.

Let $\{X_t: t = 1, 2, ...\}$ denote the time series of daily precipitation amounts. The intensities $X_t > 0$ (i.e. days for which $J_t = 1$) are taken to be conditionally independent and identically distributed with mean $\mu = E(X_t \mid J_t = 1)$ and $\sigma^2 = \mathrm{var}(X_t \mid J_t = 1)$. That is, given the states of the Markov chain model for daily precipitation occurrence, the intensities are independently generated from a common distribution. In Katz & Parlange (1993), this daily intensity distribution was based on a power transformation to normality.

For this basic form of chain-dependent process, the mean, variance, and autocorrelation function of daily precipitation amount (i.e. all data including both zero and nonzero measurements) can be expressed as (e.g. Katz & Parlange 1995)

$$E(X_t) = \pi\mu \qquad \mathrm{var}(X_t) = \pi\sigma^2 + \pi(1 - \pi)\mu^2 \qquad (7)$$

$$\mathrm{corr}(X_t, X_{t+l}) = \pi(1 - \pi)d^l\mu^2/\mathrm{var}(X_t) \qquad l = 1, 2, ...$$

The expressions in Eq. (7) for the mean and variance of daily precipitation amount can be straightforwardly derived. Despite the assumed conditional independence of the daily intensities, precipitation amount is dependent, with its autocorrelation function being directly related to that of the first-order Markov chain for daily precipitation occurrence, $\mathrm{corr}(J_t, J_{t+l}) = d^l$ where $l = 1, 2, ...$ This autocorrelation function for daily precipitation amount (Eq. 7) resembles that for an autoregressive-moving average (ARMA) process of order (1, 1) (Box & Jenkins 1976: Chap. 3), with the autocorrelations decaying exponentially from a lag of 1 d [as opposed to a lag of 0 d for an AR(1) process].

It is also of interest to treat aggregated statistics such as the number of wet days, $N(T)$, and the total precipitation, $S(T)$, over some time period of length T days:

$$N(T) = \sum_{t=1}^{T} J_t \qquad S(T) = \sum_{t=1}^{T} X_t \qquad (8)$$

The mean and approximate variance (for the large number of days T) of the number of wet days for

a first-order Markov chain are given by (Gabriel 1959)

$$E[N(T)] = T\pi \tag{9}$$
$$\text{var}[N(T)] \approx T\pi(1 - \pi)\,[(1 + d)/(1 - d)]$$

It is evident that the variance of the number of wet days is inflated by the tendency of daily precipitation occurrences to persist {i.e. the factor $[(1 + d)/(1 - d)]$ in Eq. (9)}. The mean and approximate variance of total precipitation are given by (Katz & Parlange 1993)

$$E[S(T)] = T\pi\mu \tag{10}$$
$$\text{var}[S(T)] \approx T\,(\pi\sigma^2 + \pi(1 - \pi)\,[(1 + d)/(1 - d)]\,\mu^2)$$

Comparing Eqs. (9) and (10), the variance of total precipitation is evidently directly related to the variance of the number of wet days.

3.2. Conditional chain-dependent process

For simplicity, it is assumed that the monthly index of large- (or regional) scale atmospheric circulation takes on only 2 values, $I = 0$ or 1 (e.g. $I = 1$ might denote the occurrence of an El Niño event and $I = 0$ its nonoccurrence). Let this index I have probability distribution denoted by

$$w = \Pr\{I = 1\} \qquad 1 - w = \Pr\{I = 0\} \tag{11}$$

Further, let π_i, d_i, μ_i, and σ_i^2 denote the parameters of the conditional chain-dependent process fitted only to time series of daily precipitation amounts for months during which $I = i$ where $i = 0$, 1. In other words, the precipitation data set would be divided into 2 subsets to which the chain-dependent process is separately fitted.

Because the circulation index I assumes only 2 values, Eqs. (2) to (4) involving the conditional mean, variance, and covariance simplify somewhat. For a generic precipitation statistic, X, Eq. (2) for the induced mean reduces to

$$E(X) = (1 - w)\,E(X|I = 0) + w\,E(X|I = 1) \tag{12}$$

Not surprisingly, the mean is a linear function of the circulation state probability, w [which can also be interpreted as the mean index, because $E(I) = w$]. For instance, use of Eq. (10) yields an expression for the induced mean monthly total precipitation of

$$E[S(T)] = T\,[(1 - w)\pi_0\mu_0 + w\,\pi_1\mu_1] \tag{13}$$

Likewise, Eq. (3) for the induced variance reduces to

$$\text{var}(X) = (1 - w)\,\text{var}(X|I = 0) + w\,\text{var}(X|I = 1) +$$
$$w(1 - w)\,[E(X|I = 1) - E(X|I = 0)]^2 \tag{14}$$

It is evident from Eq. (14) that the variance is a nonlinear (quadratic) function of the circulation state probability w [from Eq. (11), it follows that $\text{var}(I) = w(1 - w)$, as explicitly appears in Eq. (14)]. For example, the induced variance of monthly total precipitation (using Eq. 10) can be expressed as

$$\text{var}[S(T)] \approx T\,[(1 - w)\,(\pi_0\sigma_0^2 + \pi_0(1 - \pi_0)\,[(1 + d_0)/$$
$$(1 - d_0)]\,\mu_0^2) + w\,(\pi_1\sigma_1^2 + \pi_1(1 - \pi_1)\,[(1 + d_1)/$$
$$(1 - d_1)]\,\mu_1^2) + Tw(1 - w)(\pi_1\mu_1 - \pi_0\mu_0)^2] \tag{15}$$

Finally, for 2 generic precipitation statistics, X and Y (e.g. precipitation amount on consecutive days), Eq. (4) for the induced covariance reduces to

$$\text{cov}(X, Y) = (1 - w)\,\text{cov}(X, Y|I = 0) +$$
$$w\,\text{cov}(X, Y|I = 1) + w(1 - w)\,[E(X|I = 1) -$$
$$E(X|I = 0)]\,[E(Y|I = 1) - E(Y|I = 0)] \tag{16}$$

Like the variance, the covariance is a quadratic function of the circulation state probability, w. For instance, the induced autocovariance function of daily precipitation amount (using Eq. 7) is given by

$$\text{cov}(X_t, X_{t+l}) = (1 - w)\pi_0(1 - \pi_0)d_0^l\mu_0^2 + w\,\pi_1(1 - \pi_1)d_1^l\mu_1^2$$
$$+ w(1 - w)(\pi_1\mu_1 - \pi_0\mu_0)^2 \quad l = 1, 2, \dots \tag{17}$$

In other words, the autocorrelation function is more complex than that for an unconditional chain-dependent process (Eq. 7), resembling in some respects that for an ARMA(2, 2) process (Box & Jenkins 1976: Chap. 3).

4. EXAMPLE

The relationship between daily precipitation at individual gauges in California and gridded monthly mean sea level pressure (SLP) off the coast in the Pacific Ocean in January, the peak of the wet season, was studied in Katz & Parlange (1993). The methodology involves fitting separate conditional chain-dependent processes to daily time series of precipitation amount, depending on whether or not the monthly mean pressure was above or below normal. These 2 conditional models are compared with a single unconditional chain-dependent process fitted to all of the years of daily precipitation data.

Formally, the index I is defined by

$$\begin{array}{l} I = 0 \text{ if monthly SLP} < \text{mean} \\ I = 1 \text{ if monthly SLP} > \text{mean} \end{array} \tag{18}$$

Here 'monthly SLP' refers to the observed January mean sea level pressure for a given year at a single grid point (40° N, 130° W), and 'mean' refers to the

Table 1. Parameter estimates of unconditional and conditional stochastic models fit to time series of daily precipitation amount in January at Chico, California, USA. SLP: mean sea level pressure at 40° N, 130° W

Stochastic model	Precipitation probability $\hat{\pi}$	Persistence parameter $\hat{d}$	Intensity	
			mean $\hat{\mu}$ (mm)	SD $\hat{\sigma}$ (mm)
Unconditional	0.329	0.359	13.36	14.68
Conditional on low SLP $(I = 0)$[a]	0.413	0.334	15.16	15.50
Conditional on high SLP $(I = 1: \hat{w} = 0.526)$[a]	0.253	0.349	10.75	12.99
[a]According to Akaike's and the Bayesian information criteria: parameters π, μ, and σ (but not d) should vary with index I				

long-term sample average over the entire record of 1899 to 1988. Although this index is based on the pressure at a single grid point, its location was chosen as representative of the pressure within a region of the Pacific Ocean for which Cayan & Peterson (1989) found a significant link to monthly or seasonal total precipitation during winter over much of California. Hence, this index can be viewed as reflecting large- (or regional) scale atmospheric circulation patterns.

A time series of daily precipitation amount in January for 78 yr at Chico, California, was analyzed in Katz & Parlange (1993). According to the index (Eq. 18), 37 of these years are classified as being 'low pressure' $(I = 0)$, with the remaining 41 being 'high pressure' $(I = 1)$. The January monthly total precipitation has a conditional mean of 190.9 mm for the sample of 37 low pressure years, much higher than the conditional mean of 84.1 mm for the 41 high pressure years (unconditional mean of 134.8 for all 78 years). The conditional standard deviation of total precipitation ranges from 82.4 mm for low pressure years to 58.9 mm for high pressure years (both lower than the unconditional standard deviation of 88.6 mm). Given the substantial distance between the location on which the index is based and the site of the precipitation measurements, these results demonstrate that the index does indeed represent features of large- (or regional) scale circulation.

Table 1 includes the parameter estimates (denoted by carets) for the unconditional chain-dependent process fitted to the time series of daily precipitation amount for all of the months of January at Chico. The observed relative frequency of a wet day was about 0.33, with the conditional probability of a wet day ranging from about 0.21 if the previous day was dry to about 0.57 if the previous day was wet (i.e. a difference of roughly 0.36 reflecting the tendency of wet or dry spells to persist). The mean of precipitation amounts on wet days was about 13.4 mm, with a standard deviation of about 14.7 mm.

Table 1 also lists the parameter estimates for the 2 conditional chain-dependent processes fit to only a subset of the months of January, depending on whether monthly SLP was below or above normal (Eq. 18). In Katz & Parlange (1993), Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) were applied to determine the form of model that best fits the time series of daily precipitation amount. These criteria both involve a comparison of the maximized likelihood function, adjusted for the number of parameters required to be estimated by each model. For the Chico data set, both the AIC and BIC indicate that the best model is obtained when the probability of a wet day π, the intensity mean μ, and the intensity standard deviation σ, but not necessarily the persistence parameter d, are varied with the index I. In particular, low pressure years appeared to be associated with more frequent wet days (the odds of a wet day are roughly twice as large in low pressure years than in high pressure years), and with a distribution of precipitation amount on wet days that has not only a higher mean (roughly 40% larger), but also a higher standard deviation (roughly 20% larger). In subsequent calculations, the persistence parameter d will be varied with the index I as well.

Table 2 compares the observed daily and monthly precipitation statistics with those derived from the unconditional model and from the induced model (i.e.

Table 2. Derived daily and monthly precipitation statistics, along with observed values, for stochastic models fit to time series of daily amount in January at Chico

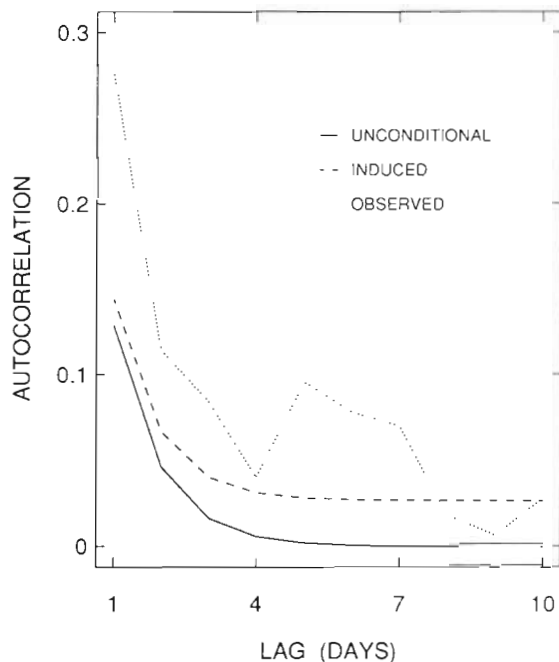| Precipitation statistic		Unconditional model	Induced model	Observed
Daily amount (mm)	Mean	4.40	4.40	4.35
	SD	10.50	10.51	10.45
Autocorrelation daily amount	lag 1	0.128	0.144	0.276
	lag 2	0.046	0.068	0.114
Monthly total (mm)	Mean	136.3	136.4	134.8
	SD	69.2	86.8	88.6
Monthly no. wet days	Mean	10.20	10.20	10.09
	SD	3.81	4.43	4.33

Fig. 1 Autocorrelation functions for time series of daily precipitation amount in January at Chico, California, USA

obtained through the combination of the 2 conditional models) using the parameter estimates listed in Table 1. Although any uncertainties in these parameter estimates are ignored, the large number of daily observations of precipitation (i.e. $1147 = 37 \times 31$ for low pressure years and $1271 = 41 \times 31$ for high pressure years) would indicate that these errors are relatively small.

Because the daily precipitation observations are directly fitted, the mean and standard deviation of daily precipitation amount (obtained via Eq. 7) for the unconditional and for the 2 conditional chain-dependent processes are constrained to either exactly or approximately (depending on the method of parameter estimation) reproduce the corresponding observed statistics. Consequently, the induced process also shares this property. The slight discrepancies evident between these statistics in Table 2 are attributable to the fact that the probability of precipitation π is indirectly derived from the estimated transition probabilities of the Markov chain (Eq. 6). Additional discrepancies would have been introduced if the mean and standard deviation of the intensities were estimated on the basis of the power-transformed distribution approach (as in Katz & Parlange 1993), as the nonlinear transformation introduces some bias.

No such constraint exists on the autocorrelation of daily precipitation amount. It is evident in Table 2 that the theoretical autocorrelations (obtained via Eq. 7)

for the unconditional chain-dependent process substantially underestimate the observed values at lags 1 and 2 d. Such an underestimation has been previously noted by Gregory et al. (1993). Fig. 1 shows the autocorrelation functions out to lag 10 d for both the unconditional and induced processes (based on Eq. 17), along with that for the observations. Although the induced process produces only a slight improvement at lags 1 and 2 d (Table 2), its autocorrelation function decays at a slower rate, better matching the observed values at higher lags. The apparent increase in autocorrelation at lag 5 d is not necessarily real.

Because the mean daily amount is reproduced and because expectation is a linear operator, the mean of monthly total precipitation is automatically reproduced to a close approximation by both the unconditional and induced processes (recall Eqs. 10 & 13). However, the variance of monthly total precipitation is substantially underestimated by the unconditional chain-dependent process, with a standard deviation of 69.2 mm produced by Eq. (10) as compared to an observed value of 88.6 mm (see Table 2). This phenomenon is termed 'overdispersion' in the statistics literature (e.g. Cox 1983) and has been noted in the context of stochastic modeling of precipitation by Wilks (1989), among others.

For the induced process, the variance of monthly total precipitation is estimated by Eq. (15) as 86.8 mm, implying that the overdispersion phenomenon has essentially disappeared (see Table 2). But what is the explanation for this improvement? Because the variance of any sum is essentially determined by the autocorrelation function, Fig. 1 provides some indication that the induced monthly variance ought to constitute an improvement.

To further diagnose the mechanism by which this improvement occurs, the daily precipitation occurrence process is separately examined. This process is the sole source of any autocorrelation in daily amount for a chain-dependent process (Eq. 7). Table 2 includes, as additional statistics, the mean and standard deviation of the monthly number of wet days (Eq. 9). As with other mean precipitation statistics, the mean monthly number of wet days is approximately reproduced by both processes, but the unconditional process underestimates the variance of the monthly number of wet days (i.e. 3.81 vs 4.33 d) whereas the induced process reproduces the observed value to a close approximation (i.e. 4.43 vs 4.33 d). A mixture of 2 conditional first-order, 2-state Markov chains has an autocorrelation function resembling that for a second-order chain (Lloyd 1974). Evidently, the apparent elimination of the overdispersion of monthly total precipitation is attributable to this additional complexity in the model of the occurrence process, as induced by the statistical downscaling approach.

5. CLIMATE CHANGE SCENARIOS

The way in which the mixture approach helps in interpreting the properties of climate change scenarios is illustrated in this section, making use of the same conditional stochastic models previously fit to the time series of daily precipitation amount in January at Chico. A case of hypothetical climate change is considered in which the atmospheric circulation state probability (i.e. the parameter $w = \Pr\{I = 1\}$) changes. How the induced precipitation statistics would be anticipated to vary along with the probability w is examined under the assumption that the parameters of the 2 conditional chain-dependent processes for daily precipitation amount are held fixed (Table 1).

Fig. 2 shows the induced mean of January total precipitation as a function of the circulation state probability w. As noted previously, this relationship (Eq. 13) is linear, with the slope being negative because the mean daily precipitation amount in low pressure years was greater than that in high pressure years [i.e. because $\pi_0 \mu_0 > \pi_1 \mu_1$ (see Table 1)]. The mean monthly total precipitation for the present climate is roughly in the center of the figure (i.e. w is about 0.53), with the 2 extremes of $w = 0$ and $w = 1$ representing the limiting situations in which only 1 of the 2 conditional chain-dependent processes ever arises. Consequently, these limiting mean values of about 194.1 and 84.3 mm essentially agree with the observed conditional monthly means (whose values were given earlier).

Fig. 2 also shows the induced standard deviation of January total precipitation as a function of the probability w. From Eq. (15), it is known that this relationship is nonlinear (note that the standard deviation, not the variance, is plotted). This standard deviation reaches a maximum of about 89.1 mm near $w = 0.34$ and relative minima of about 80.8 and 52.2 mm for the 2 extreme situations of $w = 0$ and $w = 1$, respectively. These limiting standard deviations are somewhat smaller than the corresponding observed conditional standard deviations of monthly total precipitation (values listed previously). As for the case of a single unconditional chain-dependent process, the individual conditional chain-dependent processes exhibit some degree of overdispersion (see also Katz & Parlange 1993).

Table 3 gives more-detailed statistics of January daily amount and monthly total precipitation for 2 new hypothetical values of the index state probability w. The odds of above normal SLP [i.e. $w/(1 - w)$] are either doubled or halved, corresponding to
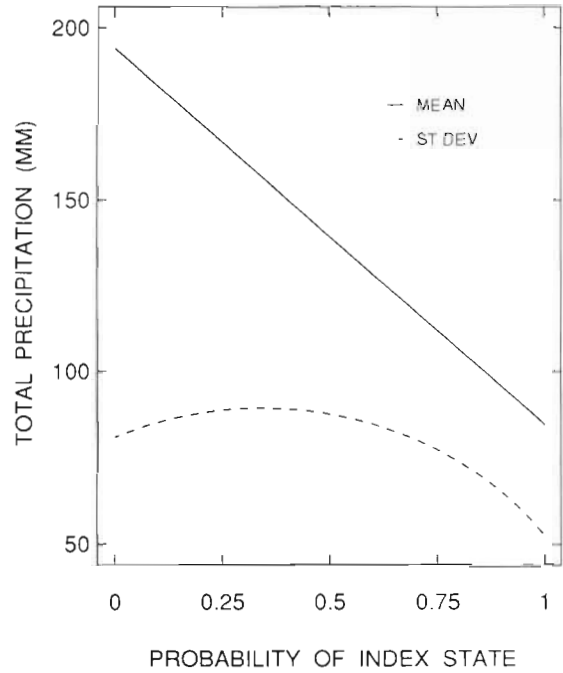

Fig. 2. Hypothetical changes in mean and standard deviation of January total precipitation at Chico as function of probability that January mean sea level pressure is above normal (i.e. index $I = 1$)

new values of $w = 0.689$ and 0.357, respectively. For sake of comparison, some of the induced statistics for the present climate listed in Table 2 are included in Table 3 as well. Both the standard deviation of daily precipitation amount and of monthly total precipitation vary along with the changes in the respective means. Whether for daily amount or for monthly total, the coefficient of variation (i.e. standard deviation divided by mean) is greater for the new drier climate (i.e. $w = 0.689$) than for the present climate and less for the new wetter climate (i.e. $w = 0.357$) than for the present climate. Given that regions drier on the average do tend

Table 3. Daily and monthly precipitation statistics for January at Chico associated with 2 hypothetical new values of circulation state probabilities (induced statistics for present climate also included)

Precipitation statistic		Probability of above normal SLP w		
		0.526[a]	0.689	0.357
Daily amount (mm)	Mean	4.40	3.82	5.00
	SD	10.51	9.76	11.20
Autocorrelation daily amount	lag 1	0.144	0.144	0.140
Monthly total (mm)	Mean	136.4	118.4	154.9
	SD	86.8	80.6	89.1
[a]Present climate				

to have larger coefficients of variation (i.e. greater relative variability), this result is reasonable. Finally, little effect on the autocorrelation function of daily precipitation amount is evident, despite changes being theoretically possible (see Eq. 17).

6. SUMMARY AND CONCLUSIONS

Statistical downscaling is approached as a problem involving mixtures of stochastic processes. The focus is on the model induced through a combination of the individual conditional models that sometimes arise in downscaling applications. Such conditional models are typically taken as identical in form to those traditionally fit unconditionally. In this case, the properties of statistical mixtures imply that the induced model will be more complex than the corresponding unconditional model. As an example of downscaling, a time series of daily precipitation amount at Chico is modeled conditional on a monthly index of large- (or regional) scale atmospheric circulation. A stochastic model fitted unconditionally to daily precipitation amount badly underestimates the observed variance of monthly total precipitation, whereas the induced model virtually eliminates this overdispersion phenomenon. Varying the circulation state probabilities to create climate change scenarios through downscaling produces a linear response in the mean of daily amount or monthly total precipitation, but a nonlinear response in the corresponding variance.

The nature of the statistical downscaling application that was considered could be generalized in several respects. For instance, the index on which daily precipitation amount is conditioned could be varied from day to day as in Bardossy & Plate (1991), rather than held constant over an entire month. Instead of just daily precipitation amount, other daily weather variables like minimum and maximum temperature (e.g. as specified in WGEN, Richardson & Wright 1984) could be conditioned on atmospheric circulation (e.g. as in Hughes et al. 1993). The more complex situation of conditional modeling of the same weather variable at several locations simultaneously could also be considered (Hughes et al. 1993, Hughes & Guttorp 1994). A challenging statistical problem concerns the situation in which the circulation index is not observed, but only hypothesized to exist. The technique of 'hidden Markov modeling' could be applied to this problem (Zucchini & Guttorp 1991, Hughes & Guttorp 1994).

Another research topic relates to the overdispersion phenomenon of monthly or seasonal total precipitation. Some climate researchers have utilized this difference between modeled and observed variances as a measure of 'potential predictability' (e.g. Shea et al.

1995). Should the increase in induced variance of monthly total precipitation achieved through conditioning be viewed as representing actual predictability as opposed to a reduction in potential predictability? One way to attack this question is to fit unconditional models more complex than an ordinary chain-dependent process. In a preliminary analysis of the same Chico data set, Katz & Parlange (1996) found that the estimated variance of monthly total precipitation does increase substantially for more complex models, but remains an underestimate of the observed value (in particular, still somewhat lower than the induced value). In other words, the overdispersion phenomenon is not solely attributable to an inadequate model for high frequency variations, but to the existence of low frequency variations as well.

Appendix 1. Conditional moments

For an accessible treatment of the geometric approach to conditional expectation based on Hilbert space theory see Brockwell & Davis (1991: Chap. 2). The mean of the conditional distribution of the random variable X given the index I, $E(X \mid I)$, is itself a random variable. That is,

$$E(X \mid I) = E(X \mid I = i) \text{ with probability } \Pr\{I = i\}$$

$$\text{where } i = 0, 1, . \qquad (A1)$$

Now X has an orthogonal decomposition (Bhattacharya & Waymire 1990, p 642)

$$X = E(X \mid I) + [X - E(X \mid I)] \qquad (A2)$$

Here 'orthogonality' means that the 2 random variables appearing in the decomposition (Eq. A2) are uncorrelated, that is

$$\text{corr}[E(X \mid I), X - E(X \mid I)] = 0 \qquad (A3)$$

The definitions of conditional variance and conditional covariance follow directly from the basic concept of conditional expectation (Eq. A1):

$$\text{var}(X \mid I) = E\{[X - E(X \mid I)]^2 \mid I\} \qquad (A4)$$

$$\text{cov}(X, Y \mid I) = E\{[X - E(X \mid I)][Y - E(Y \mid I)] \mid I\}$$

The decomposition (Eq. A2) can be applied to X to derive Eq. (3) for the variance and to both X and Y to derive Eq. (4) for the covariance

LITERATURE CITED

Bardossy A, Plate EJ (1991) Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. J Hydrol 122:33–47

Bhattacharya RN, Waymire EC (1990) Stochastic processes with applications. Wiley, New York

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control, revised edn. Holden-Day, San Francisco

Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer-Verlag, New York

Cayan DR, Peterson DH (1989) The influence of North Pacific atmospheric circulation on streamflow in the west. In: Peterson DH (ed) Aspects of climate variability in the Pacific and the western Americas. Geophysical Monograph 55, American Geophysical Union, Washington, DC, p 375–397

Cox DR (1983) Some remarks on overdispersion. Biometrika 70:269–274

Everett BS, Hand DJ (1981) Finite mixture distributions. Chapman & Hall, London

Gabriel KR (1959) The distribution of the number of successes in a sequence of dependent trials. Biometrika 46: 454–460

Gregory JM, Wigley TML, Jones PD (1993) Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. Clim Dyn 8: 299–310

Hughes JP, Guttorp P (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 30:1535–1546

Hughes JP, Lettenmaier DP, Guttorp P (1993) A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation. Water Resour Res 29:3303–3315

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 (1993) Effects of an index of atmos-

pheric circulation on stochastic properties of precipitation. Water Resour Res 29:2335–2344

Katz RW, Parlange MB (1995) Generalizations of chain-dependent processes: application to hourly precipitation. Water Resour Res 31:1331–1341

Katz RW, Parlange MB (1996) Overdispersion phenomenon in stochastic modeling of daily precipitation. Preprints, Thirteenth Conference on Probability and Statistics in Atmospheric Sciences. American Meteorological Society, Boston, p 291–294

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

Lloyd EH (1974) What is, and what is not, a Markov chain? J Hydrol 22:1–28

Nitsche G, Wallace JM, Kooperberg C (1994) Is there evidence of multiple equilibria in planetary wave amplitude statistics? J Atmos Sci 51:314–322

Richardson CW, Wright DA (1984) WGEN: a model for generating daily weather variables. U.S. Dept of Agriculture, ARS-8. NTIS, Springfield

Shea DJ, Sontakke NA, Madden RA, Katz RW (1995) The potential for long-range prediction over India for the southwest monsoon season: an analysis of variance approach. Preprints, Sixth International Meeting on Statistical Climatology. University College, Galway, p 474–477

Titterington DM, Smith AFM, Makov UE (1985) Statistical analysis of finite mixture distributions. Wiley, Chichester

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

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

Wilks DS (1989) Conditioning stochastic daily precipitation models on total monthly precipitation. Water Resour Res 25:1429–1439

Zucchini W, Guttorp P (1991) A hidden Markov model for space-time precipitation. Water Resour Res 27:1917–1923