WATER RESOURCES RESEARCH, VOL. 32, NO. 9, PAGES 2803–2823, SEPTEMBER 1996

A nonparametric wet/dry spell model for resampling daily precipitation

Upmanu Lall

Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan

Balaji Rajagopalan

Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York

David G. Tarboton

Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan

Abstract. A nonparametric wet/dry spell model is developed for resampling daily precipitation at a site. The model considers alternating sequences of wet and dry days in a given season of the year. All marginal, joint, and conditional probability densities of interest (e.g., dry spell length, wet spell length, precipitation amount, and wet spell length given prior to dry spell length) are estimated nonparametrically using at-site data and kernel probability density estimators. Procedures for the disaggregation of wet spell precipitation into daily precipitation and for the generation of synthetic sequences are proffered. An application of the model for generating synthetic precipitation traces at a site in Utah is presented.

1. Introduction

Synthetically generated sequences of daily precipitation are often used for investigating likely scenarios for agricultural water requirements, reservoir operation for analyses of antecedent moisture conditions, and runoff generation in a watershed. Preserving the characteristics of multiday wet and dry spells is often important in these applications. This paper presents a stochastic model for resampling daily precipitation where the probability distributions functions (pdf's) of alternating wet and dry spell lengths and of rainfall amount are estimated nonparametrically using kernel density estimators. This procedure is equivalent to a bootstrap or sampling with replacement of the observed data sequence of spell lengths and precipitation amounts. It differs from the classical bootstrap in that smoothed rather than empirical distribution functions are used for resampling, and sequential attributes of spells may be preserved. Necessary calibration parameters are chosen automatically from the data set using measures aimed at providing a good fit to the unknown underlying pdf.

Our particular interest was in developing a scheme for synthetic simulation of daily precipitation in mountainous regions in the western United States. Precipitation in these areas is in the form of snow in the winter with orographic and frontal mechanisms dominant. Convective rainfall processes occur in other seasons. Marked differences in the storm tracks and moisture sources over the seasons are observed. A mixture of markedly different mechanisms (some related to the El Nino-Southern Oscillation) leads to the precipitation process in the western United States [*Webb and Bettencourt*, 1992; *Cayan and Riddle*, 1992]. Recognition of such heterogeneities has led to efforts at regime identification and modeling of rainfall conditional on weather types [e.g., *Katz and Parlange*, 1993; *Wilson and Lettenmeier*, 1993; *Bogardi et al*., 1993]. While this is an attractive and necessary approach, deconvolution of mixtures is not always easy from a finite data set and the weather type designations used can be subjective. Traditionally, parametric probability models (e.g., exponential distribution), whose functional form is completely specified by a small set of parameters are used to fit the relevant frequency distributions. Selecting the best such model is tenuous [see *Vogel and McMartin*, 1991] even where mixtures are not of concern.

The work presented here was motivated by the following questions:

1. Is it possible to resample the data while preserving the relative frequencies and conditional relative frequencies of wet and dry spells and precipitation amounts without prior assumptions as to the parametric forms of the underlying probability models?
2. What is a good way to empirically model the relevant pdf for resampling and to guide development of statistical models?
3. Can such a data-based assessment of probabilities or relative frequencies be used to judge the adequacy of conceptual and statistical models posed for daily rainfall?

The first question is relevant not only from a conceptual standpoint but also because organizations (e.g., U.S. Forest Service, U.S. Department of Agriculture) specify a uniform procedure for applications from site to site, where parametric distributions or procedures are used, "models" that work well in some regions/sites fail at others. In our view it is unlikely that a robust parametric framework for model specification and selection can be devised for uniform application given the likely heterogeneity in precipitation generation mechanisms.

Paper number 96WR00565.
0043-1397/96/96WR-00565$09.00

Here we sidestep such issues by using a resampling strategy that honors at-site data directly.

The second question is addressed in paper by B. Rajagopalan et al. Evaluation of kernel density estimation methods for daily precipitation resampling, submitted to *Stochastic Hydrology and Hydraulics*, 1995, hereinafter referred to as submitted manuscript, 1995) where we document our investigations into developing appropriate kernel density estimators for resampling continuous (e.g., precipitation amount) and discrete (e.g., spell length in days) random variables.

With regards to the third question, we argue that the answer is likely to be yes, given that the relevant probability densities can be estimated reliably from the data. However, this is an area that we expect to research formally in the future and discuss only generally here.

We begin with a brief review of available models for simulating daily precipitation and an introduction to the central ideas in kernel density estimation. The nonparametric, alternating wet/dry spell model is presented next, and the resampling/simulation strategy is indicated. Results from an application to a Utah data set follow. The performance of the nonparametric scheme is compared with a simple, parametric alternative. A discussion of applicability, limitations of the approach, and musings on pointers to related work in progress concludes the presentation.

2. Background

Reviews of stochastic precipitation models are offered by *Waymire and Gupta* [1981a, b, c], *Georgakakos and Kavvas* [1987], and *Foufoula-Georgiou and Georgakokas* [1988]. The reader is referred to these papers for an appreciation of the literature and the central issues perceived in the field. While we are aware of the need to look at the concurrent representation of the precipitation process at different timescales, our focus here will only be on daily precipitation. Precipitation models have two components: (1) a model for precipitation occurrence, usually formulated as a Markov process, and (2) a model for precipitation amount, once a precipitation event has been generated. In the latter case, typically a parsimonious member of the exponential family that best fits a given data set is used. A firm basis for such a choice has yet to emerge, and typical tests for selecting between parametric distributions, such as the chi-square test, often lack the power to discriminate between different candidate distributions, since most of the mass of the pdf is concentrated near the origin. This practice is also questionable given our earlier comments that a mix of generating processes likely governs precipitation. A brief discussion of the attributes of some models for daily precipitation occurrence follows.

2.1. Markov Chain Models

The most popular approach is to consider the precipitation occurrence process to be described by a finite state (typically 2, a day is wet W or dry D) Markov Chain (MC) of finite order (typically 1), with seasonally (or time varying) transition probabilities. The basic assumption is that the present state (wet or dry) depends only on the immediate past. The transition probabilities for transitions (i.e., WW, WD, DW, DD) between the two states (W or D) are estimated directly from the data through a counting process. Fourier series methods [*Feyerharm and Bark*, 1965; *Woolhiser et al*., 1988] may be used to parameterize seasonal variations in the transition probabilities. The degree of dependence in time is limited by the order of the MC. *Feyerharm and Bark* [1967] and *Chin* [1977] suggest that the order may need to be seasonally variable as well. Lack of parsimony is a drawback of MC models as the order is increased. A number of researchers [*Hopkins and Robillard*, 1964; *Haan et al*., 1976; *Srikanthan and McMahon*, 1983; *Guzman and Torrez*, 1985] have also stressed the need for multistate MC models that consider the dependence between transition probabilities and rainfall amount.

Chang et al. [1984] and *Foufoula-Georgiou and Georgakakos* [1988] argue that Markov Chain models do not reproduce long term persistence and event clustering very readily. Markov Chain models can be attractive because of their largely nonparametric nature, ease of application and interpretability, and well-developed literature. *Wilson and Lettenmeier* [1993] pursue a hierarchical MC model to describe the daily precipitation process given the heterogeneous generating mechanisms prevalent in the western United States. While this approach addresses the heterogeneity issue, the relative lack of parsimony and shortcomings of the MC model identified above detract from the formulation.

2.2. Wet-Dry Spell Approach

In probabilistic terminology, this approach is also called the alternating renewal model (ARM). The term renewal stems from the implied independence between the dry and wet period length, while the term alternating refers to the fact that wet and dry states alternate. No transition to the same state is possible. An advantage of this representation is that it allows direct consideration of a composite precipitation event, rather than its discontinuous truncation into arbitrary daily segments.

A geometric or a negative binomial distribution [*Roldan and Woolhiser*, 1982] may be used as a model for spell length, where a daily time step is of interest. A probability distribution for wet spell precipitation amount also needs to be developed, as does a procedure for the disaggregation of wet spell precipitation to daily precipitation, for wet spells that are longer than one day.

The primary difficulties cited with the wet/dry spell approach for daily rainfall modeling are (1) the need for disaggregation of wet spell precipitation into daily or event precipitation (this is not an issue if independence in daily precipitation amounts is assumed, since that is typically assumed by Markov Chain models), (2) the justification of the independence between the dry and wet spell lengths at short timescales, and (3) the effective reduction in the sample size by considering spells rather than days. We also find the usual parametric specifications for probability distributions and assumptions of independence of spells in such models objectionable in light of the likely heterogeneous nature of the data of interest to us. However, we do find this structure plausible and address some of the difficulties cited here in our development.

3. Model Formulation

For the nonparametric, seasonal wet/dry spell (NSS) model presented here, the random variables of interest are the wet spell length w (days), dry spell length d (days), daily precipitation amount p (inches), and the wet spell precipitation amount p_w (inches). Note that throughout the paper, wet day precipitation is referred to as daily precipitation. Variables w and d are defined through the set of integers greater than 1 (and less than season length), and p and p_w are defined as

continuous, positive (actually greater than a measurement threshold, e.g., 0.01 inches rather than 0) random variables. A mixed set of discrete and continuous random variables is thus considered. Appropriate season definitions are prescribed by the model user, and model definitions that follow pertain to a given season of the year. The natural sequence of seasons is maintained, and spells in progress at the end of a season are allowed to terminate in the succeeding season.

The general structure of the model is similar to that of a wet/dry spell model. Our model differs from the traditional wet/dry spell model in a number of ways, as illustrated in Figure 1. The dry and wet spell lengths in a season may be dependent. The data are allowed to indicate whether such an assumption is necessary. Rather than fitting parametric probability densities to the data, we consider kernel estimators of the probability mass/density function (pmf/pdf) of wet spell length $f(w)$, dry spell length $f(d)$, wet day precipitation amount $f(p)$, wet spell precipitation amount $f(p_w)$, the joint pmf of wet and dry spell length $f(w, d)$, the joint pdf of wet spell length and wet spell precipitation $f(w, p_w)$, and the conditional pdf of wet spell length given dry spell length $f(w|d)$, dry spell length given wet spell length $f(d|w)$, and wet spell precipitation given wet spell length $f(p_w|w)$.

First, the significance of the dependence between successive wet and dry spell lengths is assessed by computing their sample correlation for each season. The precipitation occurrence process in a given season is described through the conditional pmf's $f(w|d)$ and $f(d|w)$ if the correlation is significant and the marginal pmf's $f(w)$ and $f(d)$ otherwise. The latter with parametrically specified pmf corresponds to the traditional alternating renewal model. The former is a more general dependence structure. Next, one estimates for each season the autocorrelation function for precipitation amounts p_i, $i = 1, \cdots, w$ for each spell length. If these correlations are not significant, it is assumed that there is no "statistical structure" in the within spell precipitation, at least for daily precipitation amounts. In this case, daily precipitation is modeled directly through an estimate of the pdf $f(p)$. If there is evidence for structure in wet spell precipitation, wet spell precipitation p_w becomes the primary variable, and a disaggregation approach that preserves the within spell structure is used to disaggregate p_w to daily precipitation amounts. In most applications using traditional wet/dry spell models or the one presented here, the disaggregation approach is eschewed in favor of treating daily precipitation as an independent random variable.

The decisions on model structure as well as the relevant pdf for each variable for each season are different and are independently estimated. To save on notation, we have chosen not to index any of our variables by season. In summary, the primary differences with the traditional wet/dry spell model are the following: (1) the relevant probability functions are estimated without recourse to prior assumptions as to the parametric form of the model, and (2) a more general conditional dependence structure is admitted.

We stress that while we are ultimately interested in developing a nonparametric model for generating daily precipitation sequences, the nonparametric density estimates generated en route are interesting since they reveal tendencies or structure in the precipitation process. We now describe how the pdf and pmf are estimated. The univariate cases are discussed first followed by the bivariate/conditional cases. The disaggregation approach is finally presented.

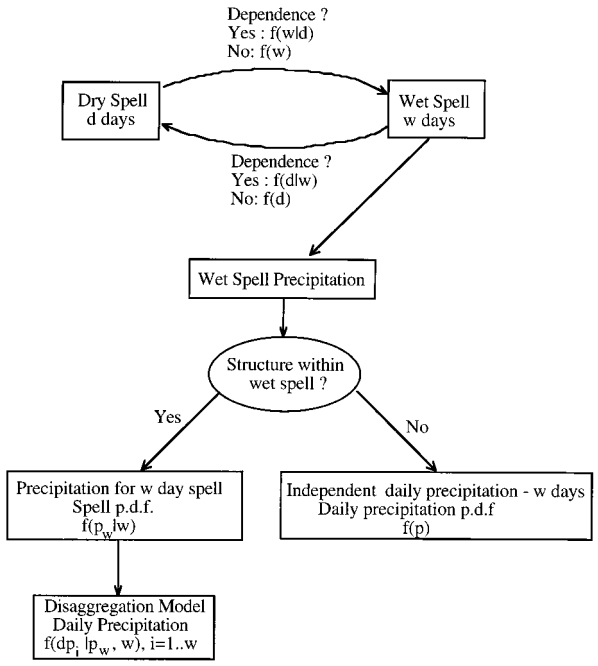

Figure 1. Structure of the wet/dry spell precipitation model.

3.1. Nonparametric Kernel Function Estimation

Nonparametric estimation of probability and regression functions is an emerging area in stochastic hydrology. A review of recent applications is offered by *Lall* [1995]. A function approximation method is considered nonparametric if (1) it is capable of approximating a large number of target functions, (2) it is "local" in that estimates of the target function at a point use only observations located within some small neighborhood of the point, and (3) no prior assumptions are made as to the overall functional form of the target function. A histogram is a familiar example of such a method. Such methods do have parameters (e.g., the bin width of the histogram) that influence the estimate at a point. However, they are different from "parametric" methods where the entire function is indexed by a finite set of parameters (e.g., mean and standard deviation), and a prescribed functional form.

Kernel density estimation is a nonparametric method of estimating a pdf from data that is related to the histogram. Recent expository monographs that develop these ideas include [*Silverman*, 1986; *Scott*, 1992; *Härdle*, 1991]. Given a set of observations $x_1, \cdots, x_n$ (in general x may be a scalar or a vector), the kernel density estimate (kde) is defined as

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where $K(\ \ )$ is a weight or kernel function and h is a bandwidth.

The idea is illustrated through Figure 2. Consider the definition of probability as a relative frequency of event occurrence. Now an estimate of the probability density at a point x (refer to points x_1 and x_2 in Figure 1) may be obtained if we consider a box or window of width $2h$ centered at x and count the number of observations that fall in such a box. The estimate $\hat{f}(x)$ is then (number of x_i that lie within $[x - h, x + h]$)/$(2hn)$). In this example, we have used a rectangular kernel ($K(t) = 1/2$ for $|t| < 1$, 0 else; $t = (x - x_i)/h$) for

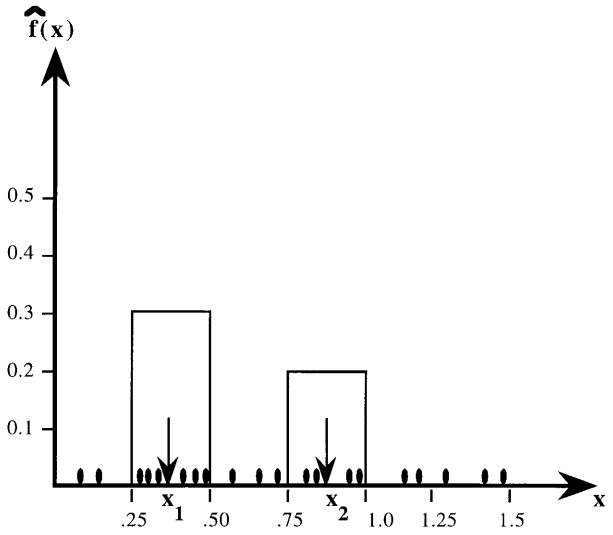

Figure 2. Example of kernel density estimate using 20 points with an histogram, $h = 0.125$. The variables x_1 and x_2 are points of estimate.

the estimate in the locale of x. As the sample size n grows, one could shrink the bandwidth h such that asymptotically the underlying pdf is well approximated. Note that for a finite sample this is much like describing a histogram, except that the "bins" are being centered at each observation or at each point of estimate, as desired. From the point of view of resampling, one can treat each observation (x_i) as being equally likely to occur in the window $x_i + h$ and resample it uniformly in that interval, for this example. Clearly, one is not restricted to rectangular kernels.

The "parameters" of this method are the kernel function or "local density" and the bandwidth h. A valid pdf estimate is obtained for any $K(\ )$ that is itself a valid pdf. Symmetry of $K(\ )$ is assumed for unbounded data to ensure pointwise unbiasedness of the estimate. For bounded data, special boundary kernels that correspond to the interior kernels are used in the boundary region, to assure unbiasedness. Finite variance of $K(\ )$ is assumed to ensure that $\hat{f}(x)$ has finite variance. This still leads to a wide choice of functions for $K(\ )$. It turns out that in terms of the mean square error (MSE) of $\hat{f}(x)$ the choice of $K(\ )$ is not crucial. Different kernels can be made equivalent under rescaling by choosing appropriate bandwidths. A Gaussian kernel with a large bandwidth can give MSE of $\hat{f}(x)$ comparable to that using a rectangular kernel with a smaller bandwidth. Thus, given a kernel function, the focus shifts to appropriate specification or estimation of the bandwidth.

It is important to note that specifying a kernel function does not have the same implications as choosing a parametric model for the whole density because the focus remains on a good pointwise or local approximation of the density rather than on fitting the whole curve directly. Different choices of $K(\ )$ still yield a local approximation of the underlying curve point by point. One can understand this by thinking of a weighted Taylor series approximation to $f(x)$ at a point x. The interplay between the h and $K(\ )$ can be thought of in terms of the interval of approximation and a weight sequence used to localize the approximation. The length of the interval (or bandwidth in this case) is more important in terms of approximation error. However, the tail behavior of the $K(\ )$ is important in the resampling context since it relates to the likely degree of extrapolation of the process. Some typically used kernels are listed in Table 1.

The sense in the statistics literature [e.g., *Silverman*, 1986] is that the choice of kernel is secondary in estimating $f(x)$, and research has focused on choosing an appropriate bandwidth optimally (in a likelihood or MSE sense) from the data. The bandwidth may vary by location (i.e., value of x) being larger where the data are sparser. Bandwidth and kernel selection issues and the success of the kernel scheme for approximating discrete, continuous, and bivariate pdf's are discussed by *Rajagopalan et al*. (submitted manuscript, 1995). Here we present the estimators that we recommend be used for the NSS model.

Table 1. Examples of Kernel Functions

	Continuous Random Variable, Univariate		
Kernel			
Normal			$K(t) = (2\pi)^{-1/2}\, e^{-t^2/2}$
Epanechnikov			$K(t) = 0.75(1 - t^2) \quad \lvert t\rvert \le 1$
Bisquare			$K(t) = 0.9375(1 - t^2)^2 \quad \lvert t\rvert \le 1$
	Discrete Random Variable, Univariate (DK) Estimation*		
Interior region (i.e., $L \ge h + 1$) (quadratic)	$K(t) = at^2 + b$	$\lvert t\rvert \le 1$	$a = -3h/(1 - 4h^2)$, $b = 3h/(1 - 4h^2)$
Left boundary (quadratic)			
for the case $1 < L < h + 1$†	$K(t) = at^2 + b$	$\lvert t\rvert \le 1$	$a = -D/2h(h + L)\,[1/(E/4h^3 - CD/12h^3\ 9h + L))]$, $b = [1 - aC/6h^2]\ 1/(h + L)$
for the case $L = 1$‡	$K(t) = at^2 + b$	$\lvert t\rvert \le 1$	$a = -d/2h^2\,[1/(E/4h^3 - CD/12h^4)]$, $b = [1 - aC/6h^2]\ 1/h$

Note that $t = (x - x_i)/h$.

*Note that $t = (L - j)/h$, and L is the point at which density is estimated.

†Where $C = h(h - 1)(2h - 1) + (L - 2)(i - 1)(2L - 3)$; $D = -h(h - 1) + (L - 2)(L - 1)$; and $E = -[h(h - 1)]^2 + [L - 2)(L - 1)]^2$.

‡Where $C = h(h - 1)(2h - 1)$; $D = -h(h - 1)$; and $E = -[h(h - 1)]^2$.

3.2. Kernel Estimation of Continuous, Univariate pdf's

The continuous, univariate pdf's of interest to us are $f(p)$, the pdf of daily precipitation, and $f(p_w)$, the pdf of wet spell precipitation for a season. The data set in the first case is composed of n_p days of daily precipitation values, p_i, for all days with measurable precipitation, in season s for the y year record. For p_w the data set is composed of n_w wet spells with total precipitation $p_{w,j}$ for each spell of length w, in season s for the y year record.

A logarithmic transform of the precipitation data prior to density estimation is often considered. Such a transformation is also attractive in the kde context. It can provide an automatic degree of adaptability of the bandwidth (in real space), thus alleviating the need to choose variable bandwidths with heavily skewed data, and also alleviating problems that the kde has with pdf estimates near the boundary (e.g., the origin) of the sample space. The resulting kde can be written as

$$\hat{f}(r) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{hr}K\left(\frac{\log(r) - \log(r_i)}{h}\right) \tag{2}$$

where h is the bandwidth of the log-transformed data, and r is p or p_w, and n is correspondingly n_p or n_w.

The bandwidth h is chosen using a recursive method of *Sheather and Jones* [1991] that minimizes the average mean integrated square error (MISE) of $\hat{f}[\log(r)]$. Figures 3a and 3b provide an illustration of the kernel estimated pdf and the underlying true pdf for two situations described in Table 2.

3.3. Kernel Estimation of Discrete Univariate pmf

In this section we present procedures for the estimation of the discrete, univariate probability mass functions $f(d)$ and $f(w)$ for each season s. This corresponds to the assumption of independence between w and d in a traditional alternating renewal model. We adopt the discrete kernel estimator (DKE) developed by *Rajagopalan and Lall* [1995a] for pmf estimation. The DK estimator for the pmf $\hat{f}(L)$, where L is either w or d, and n is the corresponding sample size is given as

$$\hat{f}(L) = \sum_{j=1}^{L_{\max}} K_d\left(\frac{L-j}{h}\right)\tilde{p}_j \tag{3}$$

where $\tilde{p}_j$ is the sample relative frequency (n_j/n) of spell length j, n_j is the number of spells of length j, $L_{\max}$ is the maximum observed spell length (note that $\sum_{j=1}^{L_{\max}}\tilde{p}_j = 1$), $K_d(\ \ )$ is a discrete kernel function, and L, j, and h are positive integers. The kernel function $K_d(\ \ )$ is given as

$$K_d(t) = at_j^2 + b \qquad |t| \le 1 \tag{4}$$

The expressions for a and b for the interior of the domain, $L > h + 1$, and the boundary region, $L < h$, are given in Table 1.

The bandwidth h is estimated by minimizing a least squares cross validation (LSCV) function given as

$$\mathrm{LSCV}(h) = \sum_{j=1}^{L_{\max}}[\hat{f}(j)]^2 - 2\sum_{j=1}^{L_{\max}}\hat{f}_{-j}(j)\tilde{p}_j \tag{5}$$

where, $\hat{f}_{-j}(j)$ is the estimate of the pmf of spell length j, formed by dropping all the spells of length j from the data. This method has been shown by *Hall and Titterington* [1987] to automatically adapt the estimator to an extreme range of sparseness types. Monte Carlo results showing the effectiveness of the DKE with bandwidth selected by LSCV are presented by *Rajagopalan and Lall* [1995a]. Figures 3c and 3d show examples of the DKE for two situations described in Table 2.

3.4. Kernel Estimation of Bivariate and Conditional pdf

The bivariate pdf of interest to us are $f(w, d)$ and $f(w, p_w)$. The conditional pdf of interest are $f(w|d)$, $f(d|w)$, and $f(p_w|w)$. Recall that the conditional pdf $f(y|x)$ of a random variable y given x is given as $f(x, y)/f(x)$, where $f(x, y)$ is the joint pdf of x and y, and $f(x)$ is the unconditional pdf of x. Since we have discussed univariate kernel density estimation, the key step is to show how the bivariate density may be evaluated.

Bivariate kernel density estimators may be constructed in much the same manner as their univariate counterparts, that is, through the convolution of appropriate kernel functions. Two types of bivariate kernel functions, radially symmetric and product kernels, are popular. *Wand and Jones* [1992] argue that for typical generalizations of the univariate kernels, there is little to choose between these representations. They point out that it is more important to choose bandwidths in each direction appropriately. We chose to use a product of univariate kernels for the bivariate kernel to allow a natural extension of the univariate kde presented here to discrete, bivariate, or mixed (continuous and discrete) bivariate situations. The joint pdf are estimated as follows:

$$\hat{f}(w, d) = \frac{1}{n_{sp}}\sum_{i=1}^{n_{sp}}K_d\left(\frac{w-w_i}{h_w}\right)K_d\left(\frac{d-d_i}{h_d}\right) \tag{6}$$

$$\hat{f}(p_w, w) = \frac{1}{n_w p_w h_{p_w}}\sum_{i=1}^{n_w}K\left(\frac{\log(p_w) - \log(p_{w_i})}{h_{p_w}}\right) \cdot K_d\left(\frac{w-w_i}{h_w}\right) \tag{7}$$

where n_{sp} is the number of consecutive wet and dry spells on record for season s, over the y year record, n_w is the number of wet spells.

The conditional pdf are given by

$$\hat{f}(w|d) = \sum_{i=1}^{n_{sp}}K_d\left(\frac{w-w_i}{h_w}\right)K_d\left(\frac{d-d_i}{h_d}\right)\Big/\sum_{i=1}^{n_{sp}}K_d\left(\frac{d-d_i}{h_d}\right) \tag{8}$$

$$\hat{f}(d|w) = \sum_{i=1}^{n_{sp}}K_d\left(\frac{w-w_i}{h_w}\right)K_d\left(\frac{d-d_i}{h_d}\right)\Big/\sum_{i=1}^{n_{sp}}K_d\left(\frac{w-w_i}{h_w}\right) \tag{9}$$

$$\hat{f}(p_w|w) = \frac{1}{p_w h_{p_w}}\sum_{i=1}^{n_w}K\left(\frac{\log(p_w) - \log(p_{w_i})}{h_{p_w}}\right) \cdot K_d\left(\frac{w-w_i}{h_w}\right)\Big/\sum_{i=1}^{n_w}K_d\left(\frac{w-w_i}{h_w}\right) \tag{10}$$

We see from (8) to (10) that the kde of the conditional pdf represents a weighted average of the relative frequency of values of the dependent variable that correspond to a "weighted" neighborhood of the conditioning point. It will be seen in section 3.7 that for simulation it is not necessary to compute

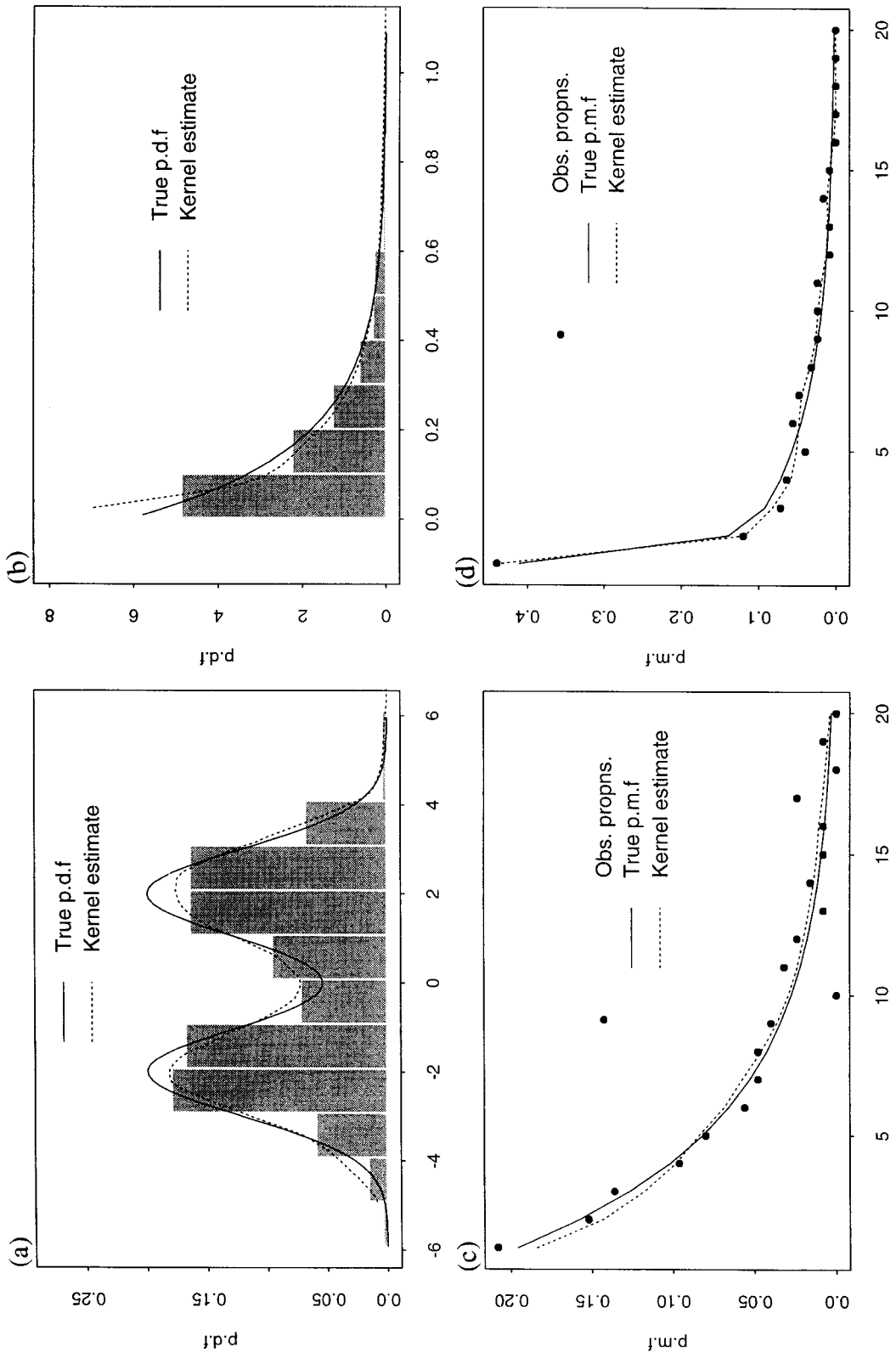

Figure 3. True probability density functions (pdf) and Kernel estimated pdf of data generated (a) from $0.5[N(-2, 1) + N(2, 1)]$, along with the histogram of the data; and (b) from Exp(0.15), along with the histogram of the data; (c) true probability mass functions (pmf) and Kernel estimated pmf of data generated from Geom(0.2), along with the observed proportions; and (d) from 0.3Geom(0.9) + 0.7Geom(0.2), along with the observed proportions.

Table 2. Statistics of Known Distributions From Which a Sample of Size 250 was Taken to Test k.d.e. Methods

Figure	Parent	Method	Sample Mean	Sample Standard Deviation	Kernel Bandwidth
3a	$\{0.5N(-2, 1) + 0.5N(2, 1)\}$	Epanechnikov kernel, SJ bandwidth	−0.00	2.26	1.22
3b	Exp (0.15)	Log transform, Epanechnikov kernel, SJ bandwidth	0.16	0.18	0.94
3c	Geom (0.2)	Quadratic kernel, DK estimator, LSCV bandwidth	5.11	4.19	6
3d	{0.3 Geom (0.9) + 0.7 Geom (0.2)}	Quadratic kernel, DK estimator, LSCV bandwidth	3.92	4.02	2

the joint and conditional pdf, estimation of the bandwidths alone is sufficient.

McLachlan [1992, pp. 306–308] discusses the simultaneous selection of bandwidths in each coordinate versus the use of the optimal univariate bandwidths in each direction. It is not clear that the additional effort of simultaneous selection of the two bandwidths is justified. Consequently, we choose the bandwidths h_w, h_d, and h_{p_w} by the methods described for the univariate case.

As an illustration, a sample of size 250 is generated from a bivariate geometric distribution Geom (0.6, 0.2) were used to test this procedure. The surface of the observed proportions is plotted in Figure 4a, the true density surface is shown in Figure 4b, the kernel estimated density surface is on Figure 4c and the difference between the true and kernel estimates are plotted in Figure 4d. The bandwidth was 3 in the x direction and 6 in the y direction. To illustrate the conditional kde, a slice is taken from the joint density in Figure 4c and presented in Figure 4e.

In the precipitation data sets we have investigated thus far, the correlation between w and d is generally weak, and the serial correlation between daily precipitation for fixed spell length w is also weak. Thus, in most cases the univariate pdf suffice. However, for the sake of completeness we describe a nonparametric, kernel-based disaggregation strategy for disaggregating a w day precipitation p_w into w daily precipitation amounts p_i.

3.5. Wet Spell Precipitation Disaggregation

We follow the approach of *Aitchison and Lauder* [1985] for analyzing compositional data. A basic requirement for the disaggregation process is that $\sum_{i=1}^{w} p_i = p_w$. Consider the rescaling $x_i = p_i/p_w$, so that $0 < x_i < 1$, and $\sum x_i = 1$. Recognizing that the effective degrees of freedom are $(w - 1)$, we can write $x_w = 1 - \sum_{i=1}^{w-1} x_i$. *Aitchison and Lauder* [1985] now apply the transform

$$y_i = \log (x_i/x_w) \qquad i = 1, \cdots, w - 1 \tag{11}$$

The multivariate pdf $f(\mathbf{x})$, where $\mathbf{x}$ is a vector of length $(w - 1)$ representing the first $(w - 1)$ proportions, is then estimated using the kernel method with a logistic normal kernel and n_w wet spells of length w as

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n_w} \frac{1}{n_w} L(\mathbf{x}, \mathbf{x}_i, \mathbf{y}, \mathbf{y}_i, h)$$

$$= \sum_{i=1}^{n_w} \frac{\exp[-0.5(\mathbf{y} - \mathbf{y}_i)^T \mathbf{S}_\mathbf{y}^{-1} (\mathbf{y} - \mathbf{y}_i)/h^2]}{n_w (2\pi)^{(w-1)/2} h^{(w-1)} \det (\mathbf{S}_\mathbf{y})^{1/2} \prod_{j=1}^{w} x_{ji}} \tag{12}$$

where i is a spell index, $\mathbf{y}$ is a vector of length $(w - 1)$ as defined in (12), x_{ji} represents the value of the jth component of $\mathbf{x}$ for the ith spell, $L(\mathbf{x}, \mathbf{x}_i, \mathbf{y}, \mathbf{y}_i, h)$ is the logistic normal kernel, h is a bandwidth, and $\mathbf{S}_\mathbf{y}$ is the sample covariance matrix of $\mathbf{y}$, estimated using a robust method [see *Huber*, 1981]. The bandwidth h is selected using maximum likelihood cross validation, that is, choosing h to maximize $\Pi_{i=1}^{n_w} f_{-i}(\mathbf{x}_i)$, where $f_{-i}(\mathbf{x}_i)$ is the estimate of $f(\mathbf{x})$ at $\mathbf{x}_i$ obtained by dropping the ith point. *Aitchison and Lauder* [1985] demonstrated that performance of this algorithm is comparable to parametric alternatives with sample sizes ranging from 23 to 95 for 2–10 components.

The use of the sample covariance matrix $\mathbf{S}_\mathbf{y}$ of $\mathbf{y}$ as the covariance matrix for the kernel function for $\mathbf{y}$ leads to some degree of preservation of the covariance structure of the components of $\mathbf{y}$ and hence of the disaggregated daily precipitation amounts p_i. It also mitigates the effect of choosing x_w, rather than say x_1 as the normalizing variable in the transformation of (12).

Using (13), one can evaluate the pdf of the first $(w - 1)$ ratios x_i of daily precipitation to wet spell precipitation. A stochastic realization of these ratios can then be generated. The last ratio x_w is obtained by noting that all the ratios have to sum to one. Daily precipitation values are then obtained by multiplying x_i by p_w. This disaggregation procedure generalizes the logistic normal based disaggregation procedure through the use of the kernel method and admits multimodality and heterogeneity in the pdf of daily rainfall in a wet spell. A problem with any wet/dry spell model is that as w increases, n_w typically decreases. Consequently, this disaggregation scheme may not be practical for large w unless long records are available. Also, it fails to "borrow" information from spells of length other than the one generated. However, that can be a problem even for the usual parametric schemes.

3.6. Generation of Synthetic Sequences

Since our goal here is to generate random samples that are similar to the observed sequence, a "raw" bootstrap or resampling of the data with replacement from the observed data sequence could be considered as an alternative to sampling from the kde. Such a strategy would be equivalent to sampling from the empirical distribution function of the data. The kde can be thought of as a smoothed (moving average) estimate of the derivative of the empirical distribution function. Sampling from the kde can lead to a reduced variance of the Monte Carlo design [*Silverman*, 1986, p. 145]. It also avoids the problem with the bootstrap where a number of the historical values are repeated in a generated sample and provides an ability to fill in and extrapolate to a limited extent beyond the observed values.

Synthetic precipitation sequences at a site are generated continuously from season to season. A dry spell is first generated using $\hat{f}(d)$. Following the strategy indicated in Figure 1, a wet spell is generated using $\hat{f}(w)$ or $\hat{f}(w|d)$. Precipitation for each of w days is then generated using $\hat{f}(p)$ or $\hat{f}(p_w|w)$ followed by $\hat{f}(p_i|p_w)$. A dry spell is then generated using $\hat{f}(d)$ or

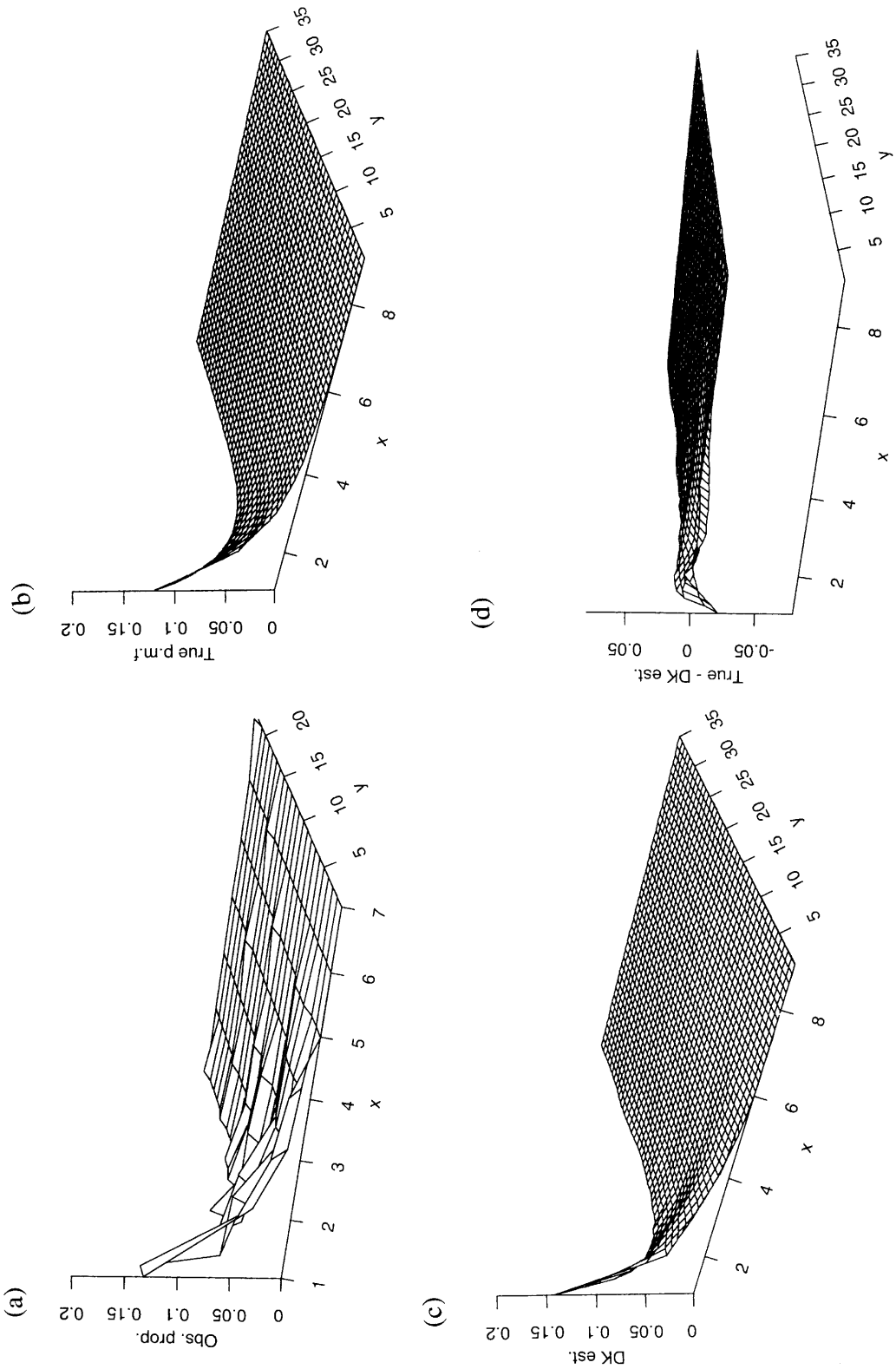

Figure 4. Surface plot of (a) observed proportions of data generated from Geom(0.6, 0.2), (b) true pmf of Geom(0.6, 0.2), (c) Kernel estimated pmf of data generated from Geom(0.6, 0.2), and (d) difference between Kernel estimated pmf and the true pmf of data generated from Geom(0.6, 0.2), and (e) A conditional slice from Figure 4b and Figure 4c, conditioned at $y = 5$, along with the observed proportions at $y = 5$.

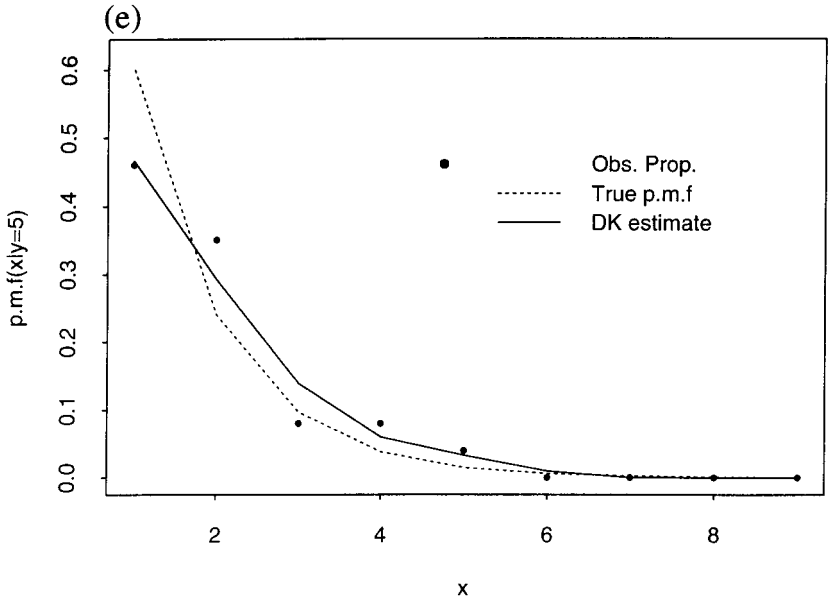

Figure 4. (continued)

$\hat{f}(d|w)$, and the process repeats. If a season boundary is crossed, the pdf used switch to those for the new season.

For the univariate continuous case ($\hat{f}(r)$), the random variate (r) of interest can be generated readily from the kernel density following a two-step procedure [*Devroye*, 1986, p. 765]. Consider the original sample (r_i, $i = 1, \cdots, n$) from which the kernel density (that depends on r, r_i, and h) was constructed using a Kernel function $K(\ \ )$. To generate a random number r that follows the estimated distribution, first sample a random integer j uniformly between 1 and n, that is, identify the historical data point to perturb. Now generate a random variate U from the probability density corresponding to the kernel function $K(\ \ )$, (e.g., $K(u) = 3/4(1 - u^2)$ for the Epanechnikov kernel). The random variate r is then given by $(r_j + Uh)$. This reinforces the notion that the kernel density estimator is formed as a convolution of local densities centered at each observation and that the generated sequence will constitute a smoothed bootstrap of the data. Any of a number of standard procedures (e.g., based on order statistics or rejection) for sampling from a density may be used to generate U from the density $K(\ \ )$. *Devroye* [1986, p. 765] provides examples for the Epanechnikov kernel. The discrete random variables (w and d) are generated directly from the estimated cumulative mass function.

A similar strategy is possible for sampling from the conditional pdf as well. Consider two continuous variables x and y. The conditional kernel density $\hat{f}(y|x)$ is given by

$$\hat{f}(y|x) = \frac{1}{h_y}\sum_{i=1}^{n} K\left(\frac{y - y_i}{h_y}\right) K\left(\frac{x - x_i}{h_x}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_x}\right)$$

$$= \frac{1}{h_y}\sum_{i=1}^{n} wt_i K\left(\frac{y - y_i}{h_y}\right) \tag{13}$$

where $wt_i = K(x - x_i/h_x)/\Sigma_{i=1}^{n} K(x - x_i/h_x)$. Now note that $\Sigma_{i=1}^{n} wt_i$ is equal to 1, and hence we can view the wt_i values as providing the probability metric with which the ith point should be selected. Define F as the set of probabilities wt_i. Sample an integer $j \in [1, n]$ using F. Now sample a variate U from the density corresponding to the kernel function for y. The variate of interest is then $y = Uh + y_j$. The discrete variate case follows as before.

4. Model Application

The model described was applied to daily rainfall data from the Silver Lake station in Utah. Forty-four years of daily rainfall data were available from 1948 to 1992. For this application we have divided the year into four seasons: season 1 (January–March), season 2 (April–June), season 3 (July–September), season 4 (October–December). Alternate season definitions as well as variable season lengths could be used. The demarcation of precipitation seasons can be based on the kernel smoothing procedures described by *Rajagopalan and Lall* [1995b]. Silver Lake is one of the higher-elevation stations in Utah, situated at 40°36′N latitude, 111°35′W longitude, and at an elevation of 8740 feet (2664 m). Most of the precipitation comes in the form of winter snow and season 4 rainfall. We see from Table 3 that season 4 (fall) has the highest mean wet day precipitation and maximum wet day precipitation, while season 1 (winter) has the highest percentage of yearly precipitation. Season 1 (winter) has the highest average wet spell length and the longest wet spell length. For the dry spells, season 3 (summer) has the highest average dry spell length and the longest dry spell length.

The successive wet-dry spell and dry-wet spell length correlations for the data from Silver Lake, Utah, were all near zero for each season. We present a representative scatterplot of the length of successive wet and dry spells for season 1 in Figure 5; the line in Figure 5 is the locally weighted regression (LOESS) smooth [*Cleveland*, 1979]. There is little evidence of even nonlinear structure in the relationship. The correlations between daily precipitation amount on successive days within a spell were also found to be near 0. Consequently, we simulated the wet and dry spells alternately using the unconditional densities

Table 3. Statistics From the Historical Precipitation Record, Silver Lake, Utah, 1948–1992

Statistic	Season			
	1*	2†	3‡	4§
Average wet spell length, days	2.6	2.2	1.85	2.5
Standard deviation of wet spell length, days	2.2	1.7	1.2	1.9
Fraction of wet days	0.62	0.44	0.36	0.55
Longest wet spell length, days	21	11	10	18
Average dry spell length, days	3.0	5.1	6.0	4.0
Standard deviation of dry spell length, days	2.80	6.0	6.0	4.0
Fraction of dry days	0.38	0.56	0.64	0.45
Longest dry spell length, days	19	42	45	24
Average wet day precipitation, inches, cm	0.37, 0.94	0.33, 0.84	0.26, 0.66	0.40, 1.02
Standard deviation of wet day precipitation, inches, cm	0.37, 0.94	0.33, 0.84	0.30, 0.76	0.42, 1.07
Fraction of yearly precipitation	0.35	0.20	0.12	0.30
Maximum wet day precipitation, inches, cm	3.7, 9.4	3.0, 7.6	1.90, 4.8	3.5, 8.9

*January–March.
†April–June.
‡July–September.
§October–December.

($\hat{f}(w)$ and $\hat{f}(d)$), and used $\hat{f}(p)$ to describe the daily precipitation process. We also performed conditional simulations using the densities $\hat{f}(w|d)$ and $\hat{f}(d|w)$ for each season. The results of these simulations were very similar in terms of the performance measures (see section 4.1 below) to those from the unconditional simulations. As is to be expected, the conditional simulations exhibit slightly greater variability. Results for the conditional simulations are not presented here for the sake of brevity. They are available electronically by e-mail from the authors.

We first list some measures of performance that were used to compare the historical record and the model simulated record, and then outline the experimental design. As emphasized earlier in the manuscript our goal is to reproduce the frequency structure (i.e., the underlying pdf). One would then expect that the usual statistics are reproduced.

4.1. Performance Measures

The seven performance measures are as follows: (1) probability distribution function of wet spell length, dry spell length, and wet day precipitation in each season; (2) mean of wet spell length, dry spell length, and wet day precipitation in each season; (3) standard deviation of wet spell length, dry spell length, and wet day precipitation in each season; (4) length of longest wet spell and dry spell in each season; (5) maximum wet day precipitation in each season; (6) percentage of yearly precipitation in each season; and (7) fraction of wet and dry days in each season.

4.2. Experiment Design

The resampling process proceeded as follows:

1. Wet and dry spells for each season are determined from the daily precipitation data. Spells that cross seasonal bound-

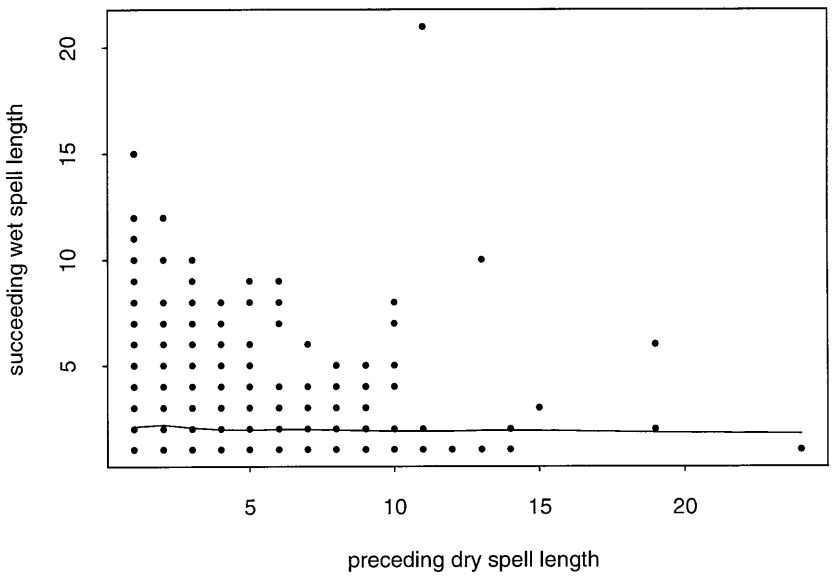

Figure 5. Scatterplot of preceding dry spell length and following wet spell length in season 1, along with the locally weighted regression (LOESS) smooth (solid line).

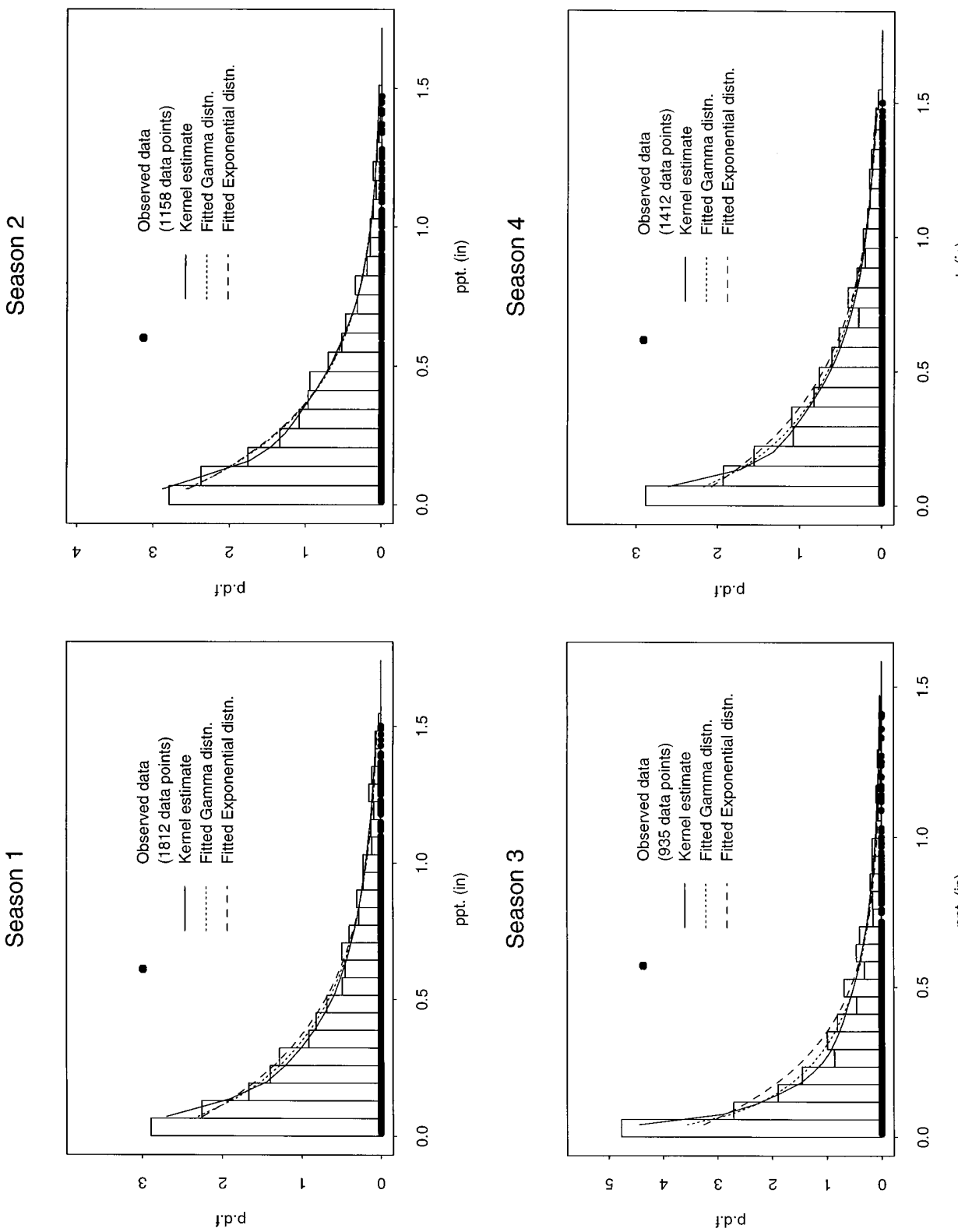

Figure 6. Plots of pdf of wet day precipitation for the four seasons at Silver Lake, Utah, estimated using Sheather Jones Log procedure, the fitted exponential distribution, fitted gamma distribution, and histogram of the observed data.

aries are truncated at the season boundary and included in the appropriate seasons. We recognize that this could have the effect of introducing a small bias in the spell characteristics for a given season. Missing data are skipped, and the spell count is restarted with the next event.

2. Probability density/mass functions are fitted for the wet day precipitation, wet spell lengths, and dry spell lengths for each season using the kernel estimators recommended in section 3.

3. Twenty-five synthetic records of 44 years each (i.e., the historical record length) are simulated.

4. The statistics of interest are computed for each simulated record and for each season and compared to statistics of the historical record using boxplots.

5. Results

In this section we present comparative results (using the performance measures listed in section 4.1) of the NSS model for the Silver Lake data. The statistics (pdfs) of the simulated records are compared with those for the historical record using boxplots. A box in the boxplots (e.g., Figure 7) indicates the interquartile range of the statistic computed from 25 simulations, the line in the middle of the box indicates the median simulated value. The solid lines correspond to the statistic of the historical record. The boxplots show the range of variation in the statistics from the simulations and also show the capability of the simulations to reproduce historical statistics. The plots of the pdf are truncated to show a common range across seasons and to highlight differences near the origin (mode).

5.1. Wet Day Precipitation

Figure 6 shows that the fitted kernel densities for wet day precipitation amount are similar to the histogram of the recorded data in all four seasons. They differ from the fitted exponential and gamma distribution, particularly in seasons 3 (summer) and 4 (fall). The kernel estimated pdf's of the simulated data reproduce the pdf of the historical data quite well, as can be seen in Figure 7. The other statistics are reproduced well by the model, as can be seen from the boxplots in Figure 8.

5.2. Wet Spell Length

Figure 9 shows that the pmf's of wet spell length estimated by DKE and the fitted geometric distribution are very close (except perhaps for season 1 (winter)). In this case one could argue for using the geometric distribution rather than DKE. However, the "loss" in using DKE is small and for uniform application across sites, DKE may still be a better choice. The pmf of wet spell length from the simulations reproduce the historical pdf very well in all the seasons as can be noted from Figure 10, suggesting that the model is performing well in reproducing the underlying frequency structure. Figure 11 shows that the mean, standard deviation, fraction of wet days, and longest wet spell length are all well reproduced by the model.

5.3. Dry Spell Length

Figure 12 shows that the dry spell length pmf estimated by DKE and the fitted geometric distribution are generally similar with the most difference in season 3 (summer), which we noted as being the most "active" with regard to dry spell length extremes. Observationally, we know that there are dry summers with little rainfall activity and other summers with intermittent, stagnating precipitation systems in this area. Thus we would expect a mixture of mechanisms generating dry spells to show up in this season.

The pmf of wet spell length from the simulations reproduce the historical pdf very well in all the seasons as can be noted from Figure 13, suggesting that the model is performing well in reproducing the underlying frequency structure. Figure 14 shows that the statistics of the dry spell length are also well reproduced.

The reader may be tempted to suggest formal tests to check for a mixture of the geometric distributions in this case as an alternative to the kernel density estimate. While this may be a fruitful activity (we did consider it), it gets harder to perform and/or justify as we consider arbitrary, finite component mixtures. An advantage of the DKE employed here is that it readily admits such mixtures without requiring that they be hypothesized or formally identified. We feel that this provides a more direct and parsimonious representation of this sort of structure if present in the data.

6. Summary and Conclusions

A nonparametric methodology for simulating daily precipitation is presented in this paper. The traditional wet/dry spell model is extended to (1) consider heterogeneity in the pdf of precipitation or wet/dry spell length and (2) consider dependence between wet/dry spell length, and between wet spell length and spell precipitation. The latter may or may not be important for rainfall data. All functions of interest are estimated nonparametrically. The primary intended use of the model is as a simulator that is faithful to the historical data sequence. The pdf's evaluated are also likely to be of use for justifying the use of other formal, parametric models of the underlying process.

While a rather flexible framework is provided by the model proposed, it is not without a price. Sample sizes needed for estimating the pdf of interest are likely to be larger than for parametric estimation. However, the nonparametric specification of the pdf leads to robustness with respect to the misspecification of the parametric model which may be valuable if the use of a particular model is to be legislated across a variety of sites and regions with different attributes. Only a crude treatment for seasonal nonstationarity is offered. This is something we expect to address in the future.

A number of issues of interest to stochastic precipitation modelers were not discussed here. The foremost is the behavior of the proposed model at different timescales. We view our developments as "operational" and relevant to the timescale of the data, which was daily. Spell definitions are tenuous at best at finer timescales and sample sizes drop rapidly as longer timescales (e.g., monthly or annual) are considered. Thus, while the scaling issue is of theoretical and practical interest, it is difficult to formally assess how such a model may fit in. It is an issue we expect to explore in due course. A second issue is the need to incorporate climatic or precipitation "types" [e.g., *Bogardi et al*., 1993; *Wilson and Lattenmaier*, 1993] into the daily precipitation model. We feel that implicit consideration of some of these factors is provided by our model by admitting an arbitrary mixture of generating mechanisms. Transitions between generating mechanisms are not explicitly modeled. However, their relative frequencies ought to be reproduced. Given limited data sets and the potentially large number of generating mechanisms this may be all that is reliably feasible

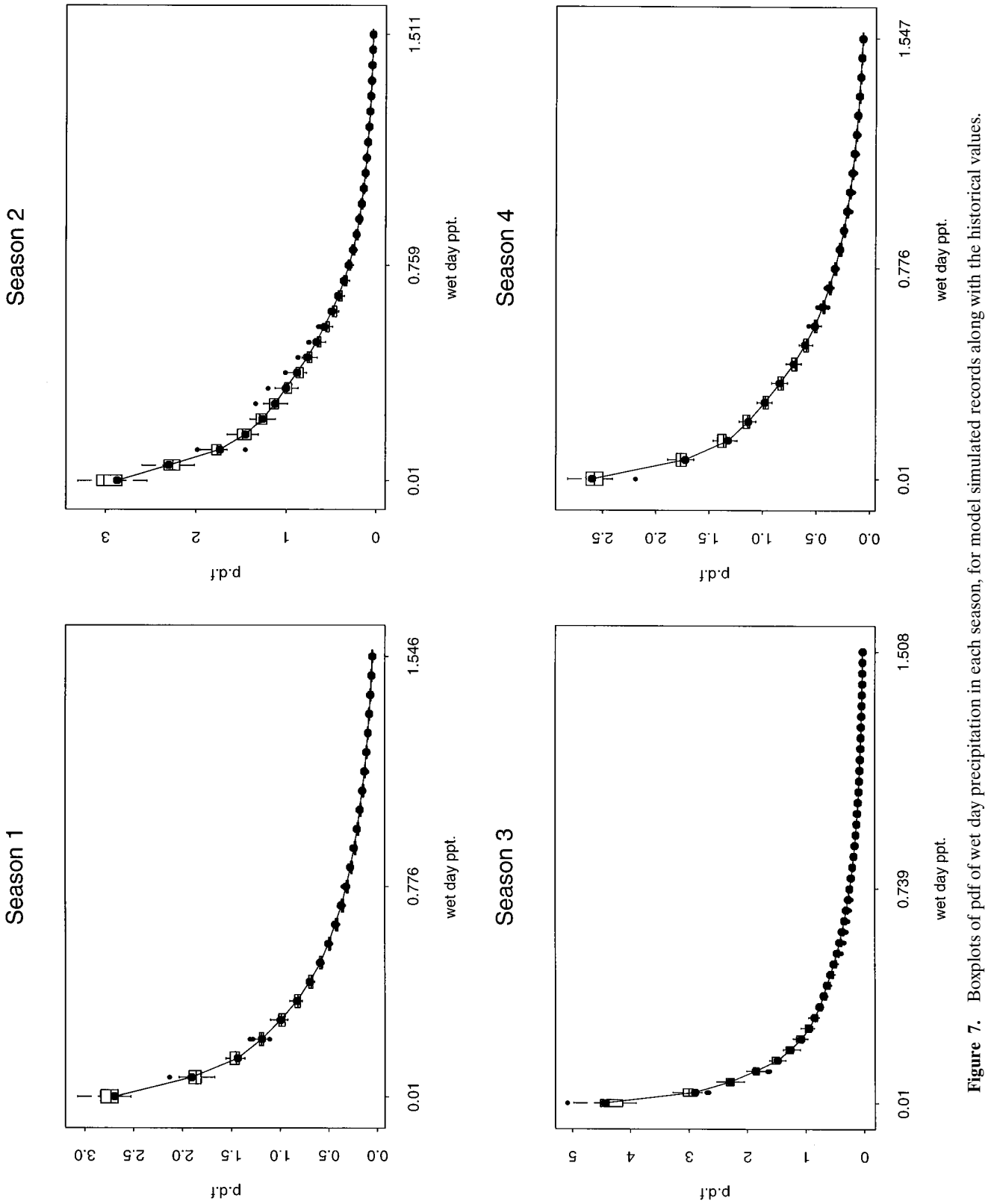

Figure 7. Boxplots of pdf of wet day precipitation in each season, for model simulated records along with the historical values.

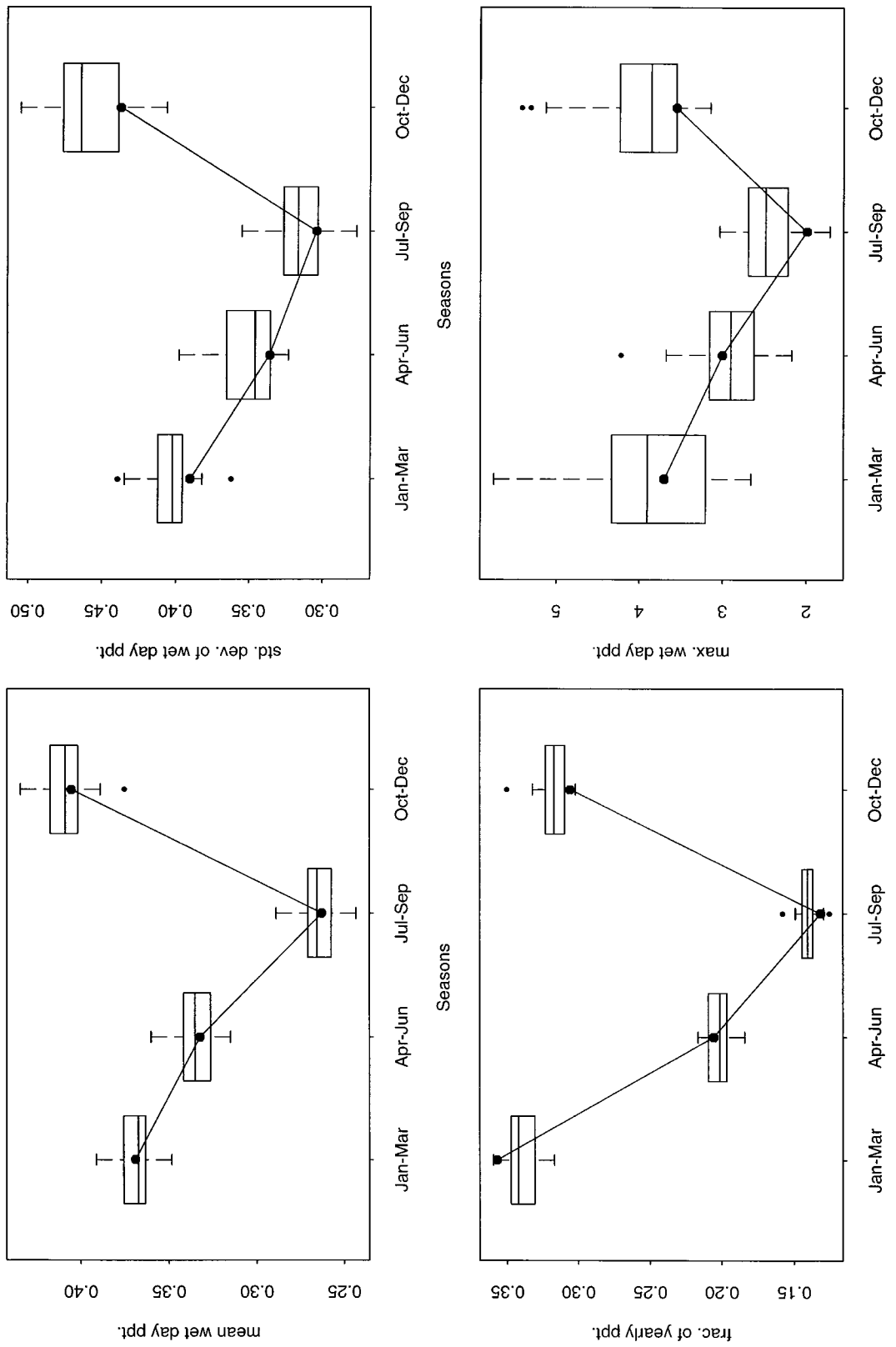

Figure 8. Boxplots of mean, standard deviation, fraction of yearly precipitation, and maximum precipitation of wet day precipitation in each season for model simulations along with the historical values.

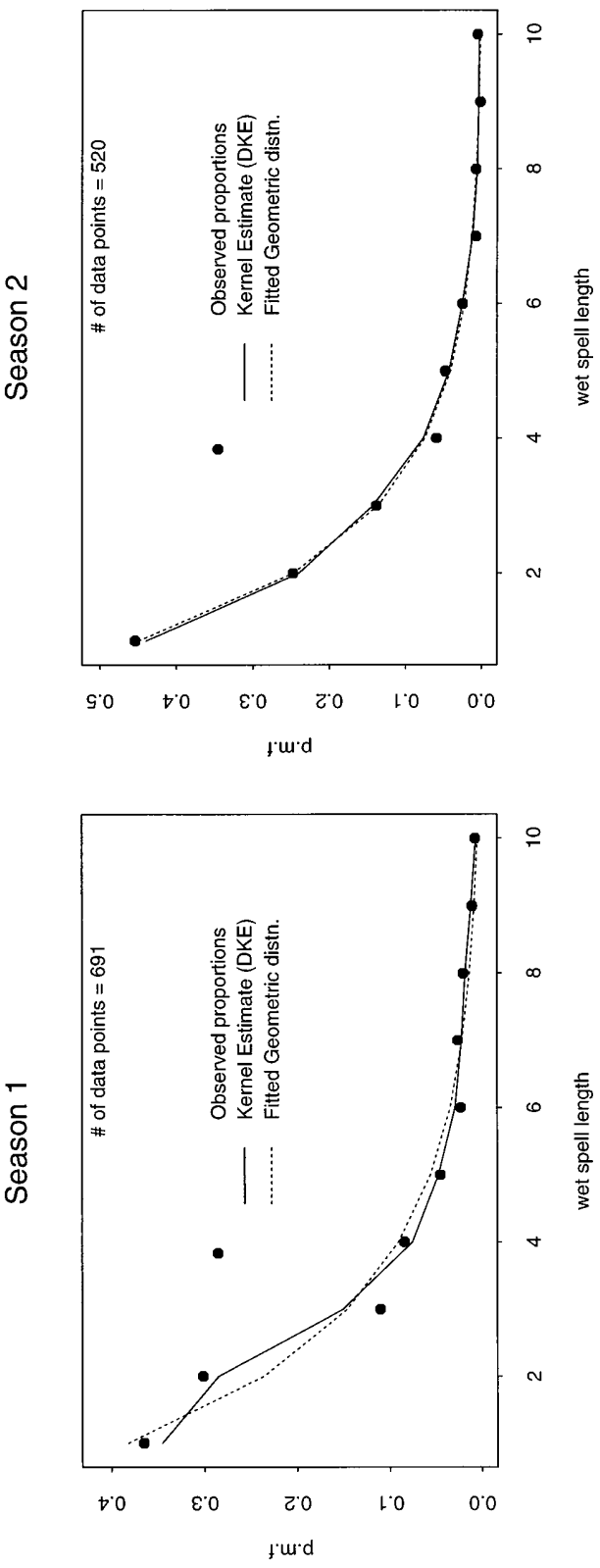

Figure 9. Plots of pmf of wet spell length for the four seasons at Silver Lake, Utah, estimated using discrete kernel estimator (DKE). Along with the fitted geometric distribution and observed proportions.

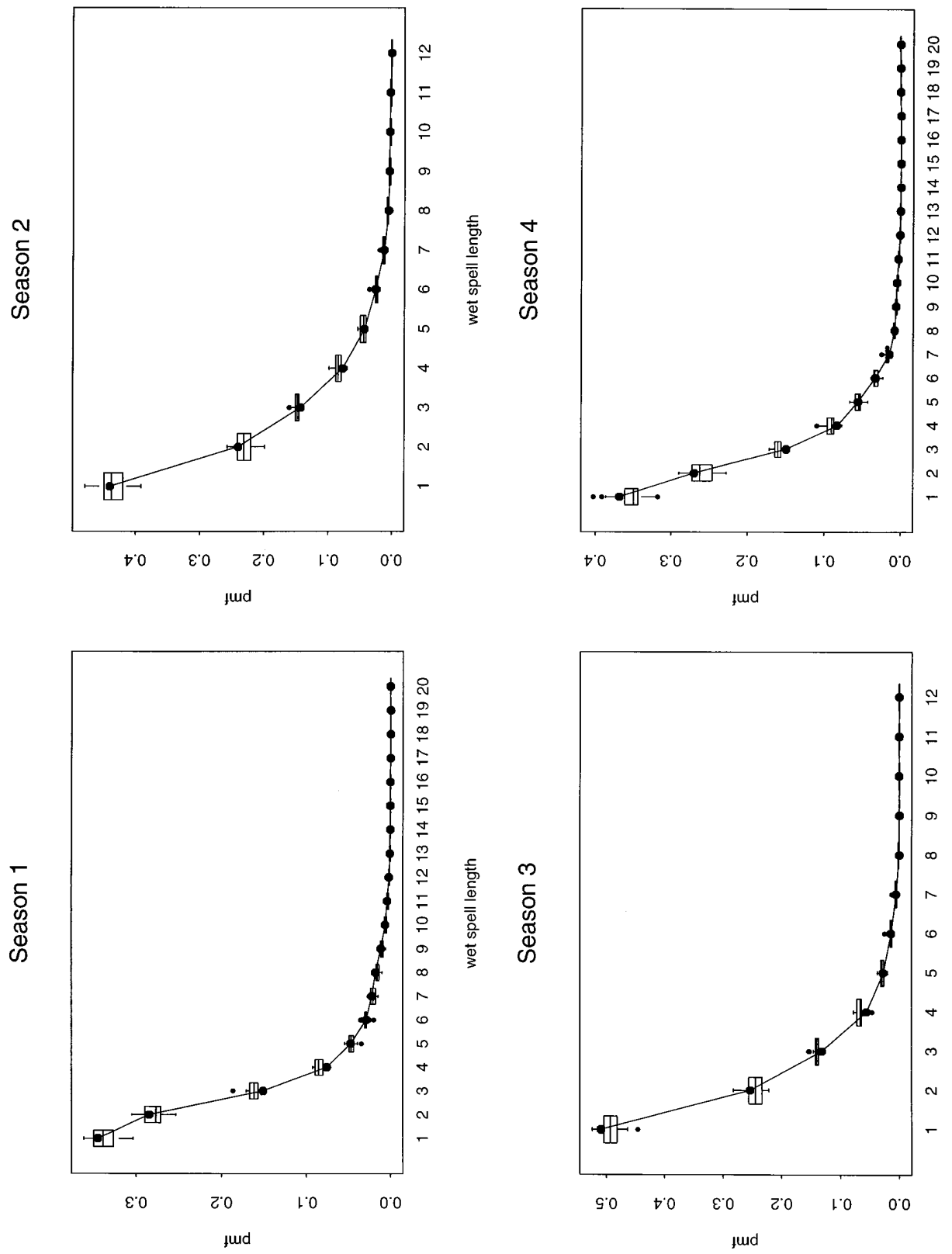

Figure 10. Boxplots of pmf of wet spell length in each season for model simulated records along with the historical values.

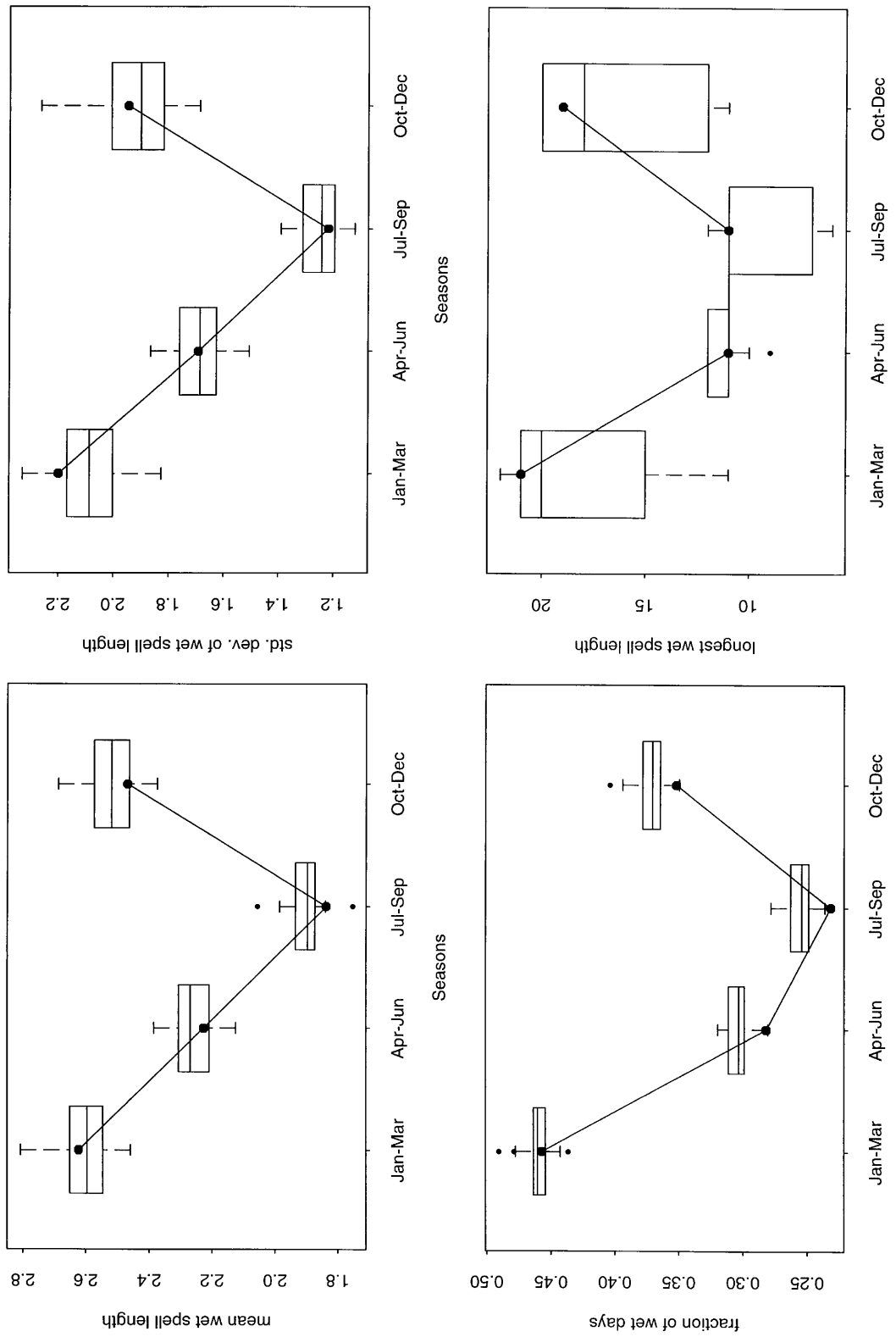

Figure 11. Boxplots of mean, standard deviation, fraction of wet days, and longest wet spell length in each season for model simulations made along with the historical values.

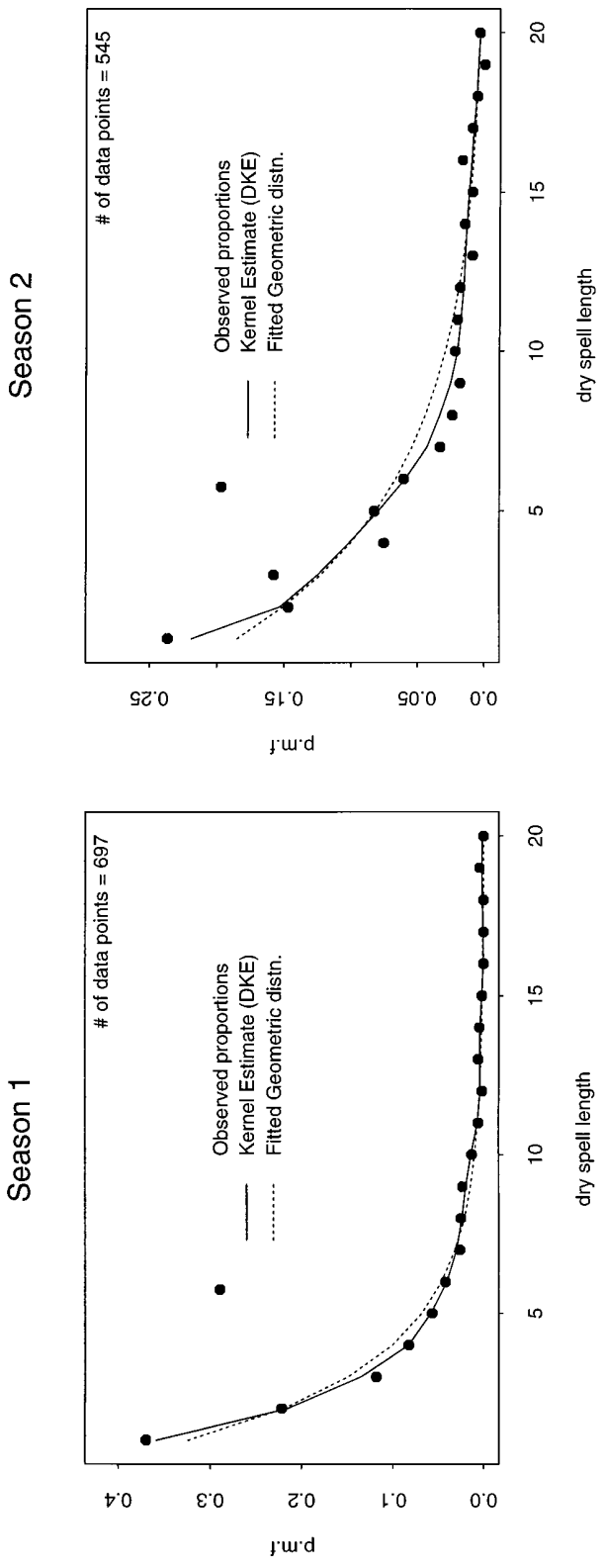

Figure 12. Plots of pmf of dry spell length for the four seasons at Silver Lake, Utah, estimated using DKE. Along with the fitted geometric distribution and observed proportions.

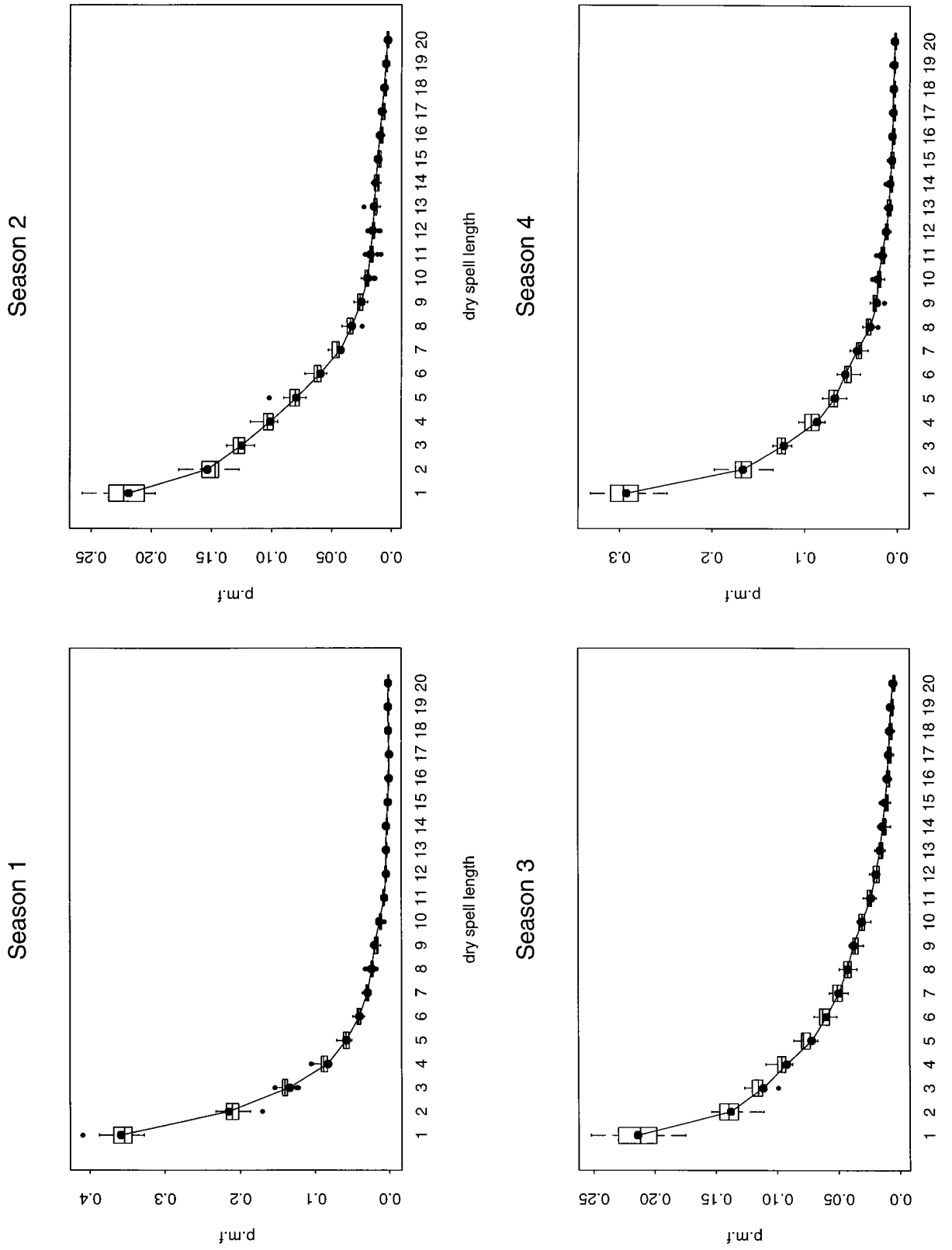

Figure 13. Boxplots of pmf of dry spell length in each season for model simulated records along with the historical values.

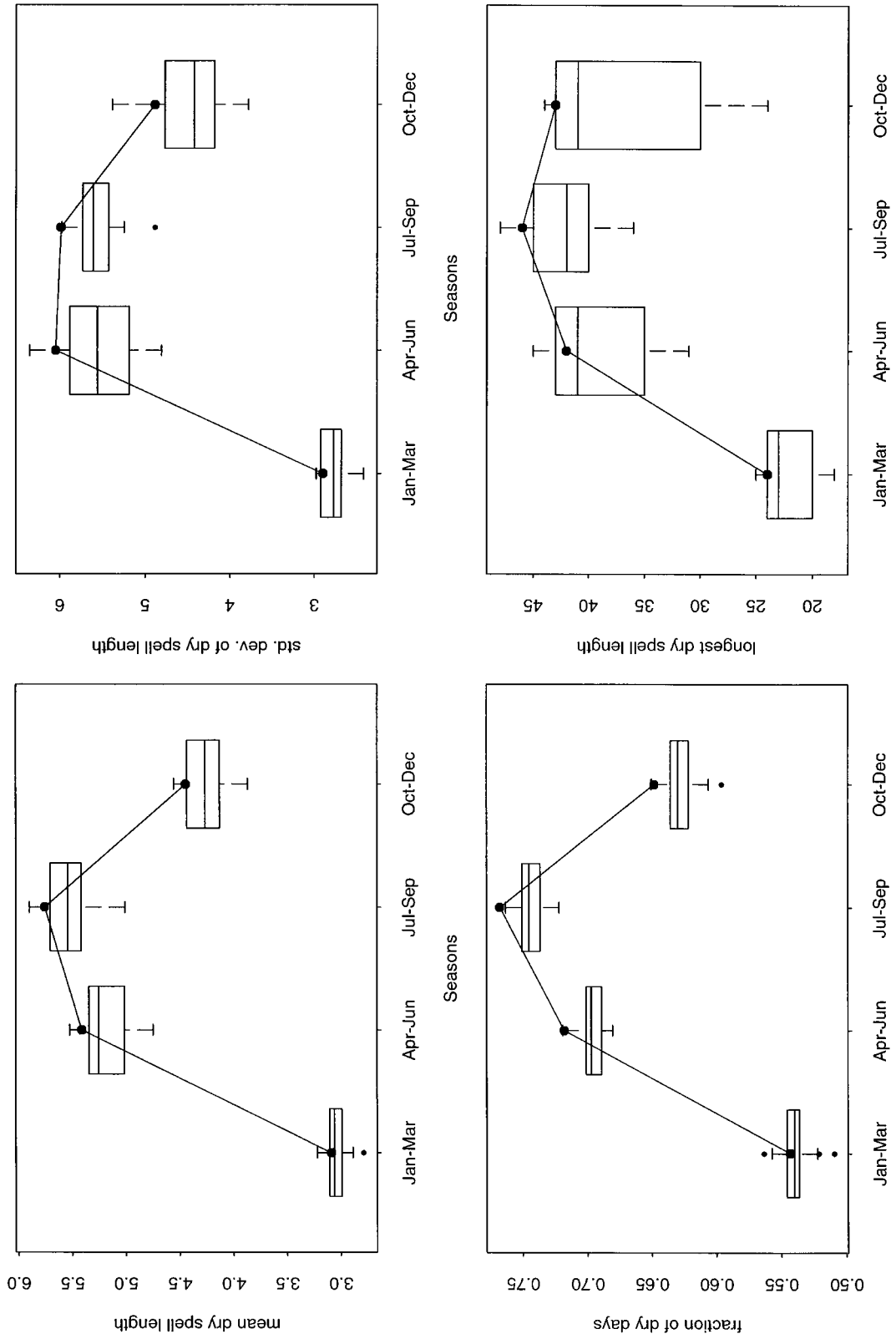

Figure 14. Boxplots of mean, standard deviation, fraction of dry days, and longest dry spell length in each season for model simulations made along with the historical values.

in a number of cases. Finally, there is the question of regionalization and/or portability of the method. The nonparametric approach clearly enjoys broader applicability than its parametric competitors. On the other hand, it may be less amenable to direct regionalization as is sometimes done in terms of the parameters of a parametric distribution. It is meaningless to talk of a regional bandwidth. It may be more fruitful to develop a space-time nonparametric precipitation model with a nonhomogeneous point process structure that is inferred from the data.

Acknowledgments. Partial support of this work by the U.S. Forest Service under contract INT-92660-RJVA, Amend 1 is acknowledged. We are grateful for discussions with D. S. Bowles, the Principal Investigator for this project. Finally, we thank M. C. Jones, H. G. Müller, S. J. Sheather, J. Simonoff, and J. Dong for stimulating discussions on kernel density estimation, review, and provision of relevant manuscripts and codes.

References

Aitchison, J., and I. J. Lauder, Kernel density estimation for compositional data, *Appl. Stat*., *34*(2), 129–137, 1985.

Bogardi, I., I. Matyasovszky, A. Bardossy, and L. Duckstein, Application of space-time stochastic model for daily precipitation using atmospheric circulation patterns, *J. Geophys. Res*., *98*, 16,653–16,667, 1993.

Cayan, D., and L. Riddle, Atmospheric circulation and precipitation in the Sierra Nevada, Managing water resources during global change, paper presented at Conference of American Water Resources Association, Tucson, Ariz., 1992.

Chang, T. J., M. L. Kavvas, and J. W. Delleur, Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resour. Res*., *20*, 565–580, 1984.

Chin, E. H., Modeling daily precipitation occurrence process with Markov Chain, *Water Resour. Res*., *13*, 949–956, 1977.

Cleveland, W. S., Robust locally weighted regression and smoothing scatter plots, *J. Am. Stat. Assoc*., *74*, 829–836.

Devroye, L., *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

Feyerherm, A. M., and L. D. Bark, Statistical methods for persistent precipitation patterns, *J. Appl. Meteorol*., *4*, 320–328, 1965.

Feyerherm, A. M., and L. D. Bark, Goodness of fit of a markov chain model for sequences of wet and dry days, *J. Appl. Meteorol*., *6*, 770–773, 1967.

Foufoula-Georgiou, E., and D. P. Lettenmaier, A markov renewal model for rainfall occurrences, *Water Resour. Res*., *23*, 875–884, 1987.

Foufoula-Georgiou, E., and K. P. Georgakakos, Recent advances in space-time precipitation modeling and forecasting, *Recent Advances in the Modelling of Hydrologic Systems*, NATO ASI Ser. 1988.

Georgakakos, K. P., and M. L. Kavvas, Precipitation analysis, modeling, and prediction in hydrology, *Rev. Geophys*., *25*(2), 163–178, 1987.

Guzman, A. G., and C. W. Torrez, Daily rainfall probabilities: Conditional upon prior occurrence and amount of rain, *J. Clim. Appl. Meteorol*., *24*(10), 1009–1014, 1985.

Haan, C. T., D. M. Allen, and J. O. Street, A markov chain model of daily rainfall, *Water Resour. Res*., *12*, 443–449, 1976.

Hall, P., and D. M. Titterington, On smoothing sparse multinomial data, *Aust. J. Stat*., *29*(1), 19–37, 1987.

Härdle, W., *Smoothing Techniques With Implementation in S*, Springer-Verlag, New York, 1991.

Hopkins, J. W., and P. Robillard, Some statistics of daily rainfall occurrence for the canadian prairie provinces, *J. Appl. Meteorol*., *3*, 600–602, 1964.

Huber, P. J., *Robust Statistics*, New York, John Wiley, 1981.

Katz, R. W., and M. B. Parlange, Effects of an index of atmospheric circulation on stochastic properties of precipitation, *Water Resour. Res*., *29*, 2335–2344, 1993.

Lall, U., Recent advances in nonparametric function estimation, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994*, *Rev. Geophys*., *33*, 1093–1102, 1995.

McLachlan, G. J., *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley, New York, 1992.

Rajagopalan, B., and U. Lall, A kernel estimator for discrete distributions, *J. Nonparametric Stat*., *4*, 409–426, 1995a.

Rajagopalan, B., and U. Lall, Seasonality of precipitation along a meridian in the western U.S., *Geophys. Res. Lett*., *22*(9), 1081–1084, 1995.

Roldan, J., and D. A. Woolhiser, Stochastic daily precipitation models, 1, A comparison of occurrence processes, *Water Resour. Res*., *18*, 1451–1459, 1982.

Scott, D. W., *Multivariate Density Estimation: Theory, Practice and Visualization*, John Wiley, New York, 1992.

Sheather, S. J., and M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc*., *Ser. B*, *53*, 683–690, 1991.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.

Srikanthan, R., and T. A. McMahon, Stochastic simulation of daily rainfall for Australian stations. *Trans. ASAE*, 754–766, 1983.

Vogel, R. M., and D. E. McMartin, Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution, *Water Resour. Res*., *27*, 3149–3158, 1991.

Wand, J. S., and M. C. Jones, Comparison of smoothing parameterizations in bivariate kernel density estimation, *J. Am. Stat. Assoc*., *88*(422), 520–528, 1992.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 1, A review of the stochastic rainfall models, *Water Resour. Res*., *17*(5), 1261–1272, 1981a.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 2, A review of the theory of point processes, *Water Resour. Res*., *17*(5), 1273–1285, 1981b.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 3, Some applications of the point process theory to rainfall processes, *Water Resour. Res*., *17*(5), 1287–1294, 1981c.

Webb, R. H., and J. L. Bettencourt, Climatic variability and flood frequency of the Santa Cruz river, Pima County, Arizona, *U.S. Geol. Surv. Water Supply Pap*., *2379*, 1992.

Wilson, L. L., and D. P. Lettenmaier, A hierarchical stochastic-model of large-scale atmospheric circulation patterns and multiple station daily precipitation, *J. Geophys. Res*., *97*, 2791–2809, 1993.

Woolhiser, D. A., C. L. Hanson, and C. W. Richardson, Microcomputer program for daily weather simulation, *Rep. 75*, 49 pp., *Agric. Research Serv*., *U.S. Dep. of Agric*., Washington, D. C., 1988.

U. Lall and D. G. Tarboton, Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322-8200. (e-mail: ulall@kernel.uwrl.usu.edu;

B. Rajagopalan, Lamont-Doherty Earth Observatory of Columbia University, P.O. Box 1000, Route 9W, Palisades, NY 10964-8000. (e-mail: rbala@rosie.ldgo.columbia.edu)

(Received February 13, 1995; revised February 15, 1996;
accepted February 16, 1996.)