

STATISTICS IN THE  
*environmental*  
& EARTH SCIENCES

*edited by*

Andrew T Walden

Department of Mathematics,  
Imperial College of Science, Technology and Medicine,  
London

and

Peter Guttorp

Department of Statistics, University of Washington,  
Seattle

Edward Arnold

A division of Hodder & Stoughton  
LONDON MELBOURNE AUCKLAND

Copublished in the Americas by Halsted Press,  
an imprint of John Wiley & Sons, Inc.  
New York – Toronto

© 1992 A T Walden and P Gutterp

First published in Great Britain 1992  
Copublished in the Americas by Halsted Press, an imprint of  
John Wiley & Sons Inc., 605 Third Avenue, New York, NY 10158

*British Library Cataloguing in Publication Data*

Walden, A.T.

Statistical methods in environmental and earth sciences:  
New developments.

I. Title II. Gutterp, P.  
550.1

ISBN 0-340-54530-5

*Library of Congress Cataloging-in-Publication Data*

Available upon request

ISBN 0 470 21946 7

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronically or mechanically, including photocopying, recording or any information storage or retrieval system, without either prior permission in writing from the publisher or a licence permitting restricted copying. In the United Kingdom such licences are issued by the Copyright Licensing Agency: 90 Tottenham Court Road, London W1P 9HE.

Typeset in Times New Roman 569 by H Charlesworth & Co Ltd, Huddersfield

Printed in Great Britain for Edward Arnold, a division of Hodder and Stoughton Limited, Mill Road, Dunton Green, Sevenoaks, Kent TN13 2YA by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk and bound by Hartnolls Ltd, Bodmin, Cornwall.

# Reviewers

Dr E. Foufoula-Georgiou  
St Anthony Falls Hydraulic Lab.  
University of Minnesota  
USA

Dr J. Haslett  
Dept of Statistics  
Trinity College Dublin  
IRELAND

Prof. D. Oldenburg  
Dept of Geophys. & Astron.  
University of British Columbia  
CANADA

Dr D. B. Percival  
Applied Physics Lab.  
University of Washington  
USA

Dr E. Riskin  
Dept of Electrical Eng.  
University of Washington  
USA

Prof. R. L. Smith  
Dept of Statistics  
University of North Carolina  
Chapel Hill  
USA

Prof. R. E. White  
Dept of Geology  
Birkbeck College  
University of London  
UK

Dr T. Haas  
Dept of Mathematics  
University of Wisconsin-Milwaukee  
USA

Dr Y. Ogata  
Institute of Statistical Mathematics  
Tokyo  
JAPAN

Prof. J. Petkau  
Dept of Statistics  
University of British Columbia  
CANADA

Prof. A. E. Raftery  
Dept of Statistics  
University of Washington  
USA

Prof. N. J. Shackleton  
Godwin Lab. for Quaternary Res.  
Cambridge University  
UK

Prof. D. M. Titterton  
Dept of Mathematics  
University of Glasgow  
Glasgow  
UK

Prof. W. Zucchini  
Dept of Statistics  
University of Cape Town  
SOUTH AFRICA

# Chapter 5 \_\_\_\_\_

## **Modeling daily precipitation – progress and problems**

D. A. Woolhiser

### 5.1 Introduction

Precipitation has a major influence on virtually all human activity. Agricultural operations and many engineering activities are strongly influenced by weather phenomena. Examples of applications where precipitation information is essential include: design of irrigation systems (Palmer *et al.*, 1982), design of agricultural or urban drainage systems, design of earthen covers for landfills or storage sites for low-level nuclear waste (Lane, 1984), selection of farm or construction machinery (Von Barga, 1967), evaluation of erosion hazard and the effects of erosion on agricultural productivity through simulation modeling (Williams and Renard, 1985). New developments in modeling growth and yield of major crops such as corn, wheat, soybeans and cotton will increase the need for precipitation simulation models so that farmers will be able to obtain estimates of the distribution of crop yields as the growing season progresses.

Extensive climatic data, including precipitation, have been collected for many years all over the world with the greatest density of stations in the developed nations. These data were previously available in cumbersome printed form, but are now quite widely available on computer-compatible media. The information content of these data can be summarized by standard statistical analyses and the results presented in tabular or graphical form. For example, extensive analyses were performed by the regional climate committees of State Agricultural Experiment Stations in the United States during the period 1955–67 and the results were presented in a series of reports (Shaw *et al.*, 1960; Dethier and McGuire, 1961; Feyerherm *et al.*, 1965; Gifford *et al.*, 1967).

Individuals from a wide range of disciplines have been interested in the stochastic structure of the precipitation process for a long time. Katz (1984) pointed out that in work published in 1852, Quetelet examined runs of consecutive wet and dry days in Brussels and noted the phenomenon of persistence. To take this dependence into account he developed a special case of what is now called a two-state Markov chain. Many more models have been developed since then, with a major stimulus being provided by the advent of digital computers.

## 5.2 Stochastic models of daily precipitation

Many models have been developed to describe the daily precipitation process. Precipitation is, of course, an intermittent process in continuous time but precipitation data commonly available are the accumulated depths over a period of one day. Although a substantial literature exists on continuous time models of precipitation they are beyond the scope of this paper. At least in the United States, the time of day that the gauges are read (for standard gauges) may vary from station to station and could also change with time. As we shall see, this methodological factor may be quite significant in modeling. The most common approach has been to develop models describing the occurrence (wet–dry) process and to describe the distribution of rainfall amounts on wet days independently.

The seasonal variation of precipitation throughout the year is an important factor in the construction of models. Several approaches have been used to deal with seasonality: assume that parameters vary as step functions for each month, assume that parameters vary as step functions for seasons which may include two or more months, or use polynomials or Fourier series to provide daily variation of parameters. A comparison of two of these approaches will be presented in a later section.

### 5.2.1 Occurrence models

Rainfall occurrence can be viewed as a sequence of random variables  $\{X(t)\}; t = t_1, t_2, \dots, t_T$ , where

$$\begin{aligned} X(t) &= 1 \text{ if rainfall is equal to or greater than a threshold, } d \\ &= 0 \text{ if rainfall is less than } d. \end{aligned} \quad (5.1)$$

The most commonly used model of precipitation occurrence is the two-state Markov chain of first or higher order. Gabriel and Neumann (1962) used a first-order stationary Markov chain to describe rainfall occurrence in Tel Aviv. Non-stationary Markov chains have been used by several investigators (Caskey, 1963; Feyerherm and Bark, 1965; Heerman *et al.*, 1968; Woolhiser and Pegram, 1979; Coe and Stern, 1982, Stern and Coe, 1984). The appropriate order for Markov chain models has been investigated by Chin (1977), Gates and Tong (1976), and Eidsvik (1980). The results varied with the climatic characteristics of the precipitation stations investigated, with the statistical tests used and with the length of record. A first-order chain is adequate for many locations but second or higher order may be required at other locations or during some times of the year.

The alternating renewal process (ARP) for wet and dry sequences has been investigated by Green (1964), Buishand (1977) and Roldan and Woolhiser (1982). Buishand (1977) found that an alternating renewal model with a truncated negative binomial distribution (TNBD) for the length of wet and dry sequences provided an excellent fit for rainfall data from the Netherlands. He assumed that the parameters of the TNBD were constant within four seasons and noted that in dry climates the sample size for wet

sequences may be very small and the parameters may be unreliable. Roldan and Woolhiser (1982) compared the ARP with truncated geometric distribution of wet sequences and TNBD dry sequences with a first-order Markov chain. For five US stations studied, the first-order Markov chain was superior to the alternating renewal process according to the Akaike information criterion (Akaike, 1974). Roldan and Woolhiser (1982) used Fourier series to describe the seasonal variation of parameters with the parameters for a specific wet or dry sequence determined by the starting day of the sequence. One of the disadvantages of the ARP is that seasonality is difficult to handle.

Foufoula-Georgiou and Lettenmaier (1987) developed a Markov renewal model for rainfall occurrences in which the times between daily rainfall occurrences were sampled from two different geometric probability distributions. The transition from one inter-arrival type to the other was governed by a Markov chain. They derived maximum likelihood and method of moments estimators for the four parameters and assumed that the parameters were constants within each of five seasons.

Smith (1987) introduced a family of models termed Markov-Bernoulli processes that might be used for rainfall occurrence. The process consists of a sequence of Bernoulli trials with randomized success probabilities described by a first-order two-state Markov chain. At one extreme the model is a Bernoulli process, at the other a Markov chain. This process has some very interesting properties including invariance of the process under random thinning, but parameter identification is quite difficult.

The two-state discrete autoregressive moving average (DARMA) model was first used as a model for rainfall occurrence by Buishand (1977) and more recently by Chang *et al.* (1984) and Delleur *et al.* (1989). Buishand found that an alternating renewal process was superior to the DARMA model for data from the Netherlands but that the DARMA model looked more promising in tropical and monsoon areas. Chang *et al.* (1984) and Delleur *et al.* (1989) used four seasons for two stations in Indiana and found that the first-order autoregressive (Markov chain) or the second-order moving-average model were appropriate for different seasons. Buishand (1977) pointed out an important factor, namely that properties of the rainfall in New Delhi cannot be preserved by a model with constant parameters – stochastic parameters are required. This observation may be generally valid for regions with monsoon climates.

### 5.2.2 Distribution of amounts

Let  $Y(t)$  denote the amount of precipitation on day  $t$  and let  $U(t) = Y(t) - d$ , where  $d$  is the smallest amount of rainfall observed or some other threshold for distinguishing between a dry or wet day. It has been observed that the distribution of  $U(t)$  is highly skewed and that it exhibits seasonal variability. In a thorough analysis of data from the Netherlands, Buishand (1977) showed that there was evidence for different distributions of rainfall depending on whether it is a solitary rainy day (type I), a rainy day bounded by a rainy day on one side (type II) or a rainy day bounded on both sides by rainy

days (type III). Buishand (1977) also showed that there was a small but significant correlation between precipitation amounts on successive wet days. He observed that the discrimination between different types of rainfall amounts has little effect on the distribution of 30-day total rainfall.

Introducing dependencies of the type observed by Buishand (1977) into a nonstationary model leads to difficulties in parameter identification and models with a large number of parameters. In efforts to obtain parsimonious models, investigators have frequently assumed that the daily rainfall,  $Y(t)$ , is independent of the occurrence process,  $X(t)$ , and is also independent of rain on previous days  $Y(t-1)$ ,  $Y(t-2)$ , ... An intermediate approach is to assume serial independence of  $Y(t)$  but to allow  $Y(t)$  to depend on  $X(t-1)$ ,  $X(t-2)$ , ... – the chain-dependent process as described by Katz (1977a, b). Stern and Coe (1984) describe techniques for estimating parameters for chain-dependent shifted gamma distributions. Woolhiser and Roldan (1982) found that according to the Akaike Information Criterion (AIC), (Akaike, 1974) the independent mixed exponential distribution was superior to chain dependent gamma distributions for five US stations.

The distribution of precipitation depths on wet days is highly skewed and a shifted gamma distribution has been most commonly used (Ison *et al.*, 1971; Buishand, 1977; Katz, 1977b; Richardson and Wright, 1984; Stern and Coe, 1984).

The mixed exponential has been utilized for the distribution of  $U(t)$  by Smith and Schreiber (1974), Woolhiser and Pegram (1979), and Woolhiser and Roldan (1982). Richardson (1982) compared the exponential, mixed exponential and gamma distributions for several US stations and found the mixed exponential to be superior to the other two.

Stidd (1953) and Nicks (1974) used normalizing transforms of the type often used to transform gamma-like distributions to normal. Other distributions which have been used include the kappa (Mielke, 1973) and the Weibull (Zucchini and Adamson, 1984). It appears unlikely that a single distribution will provide a good fit to daily rainfall data for all climatic regions.

### 5.3 Parameter identification and model selection

One striking feature of the precipitation process in most parts of the world is its seasonal variability. This is demonstrated in Figure 5.1 where the mean number of wet days per 14-day period is plotted for four US stations with quite different climates. If we reduced the number of days in each period to five days or to one day we would find that the pattern would become more erratic due to sampling variability, although the seasonal variations would still be evident.

Parameter identification methods are inextricably related to the particular model used and to the techniques used to represent seasonal variation of parameters. If it is assumed that parameters vary as step functions defined

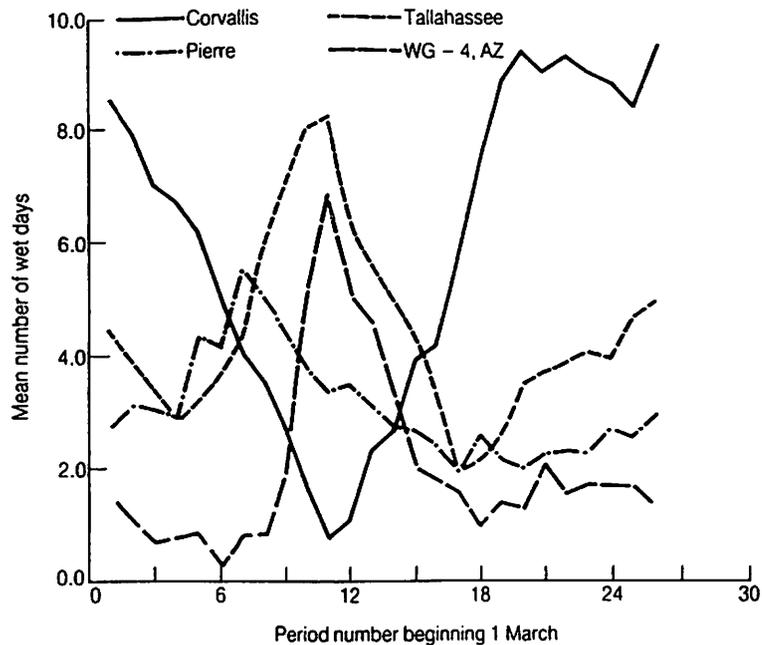


Figure 5.1 Mean number of wet days for 14-day periods for four US stations.

on a monthly or seasonal basis, the techniques are quite straightforward. Maximum likelihood (ML) techniques are typically used to estimate the transition probabilities of Markov chains and ML techniques or method of moments (MOM) are used to estimate parameters for the distribution of daily precipitation and for the distribution of wet and dry periods for alternating renewal models. Buishand (1977) recognized that the small precipitation amounts may be quite unreliable and used ML estimators which included the number of observations below a threshold but not the amount of each observation in this class. Chang *et al.* (1984) used the autocorrelation function to guess the form of a DARMA occurrence model and to obtain initial trial values of the parameters. The final values of the parameters were estimated by an iterative least-squares procedure using the estimated and theoretical autocorrelation functions.

The step function representation of parameters is not intuitively satisfying because we would not expect the precipitation characteristics on adjacent days to vary substantially. The seasonal variation of parameters has been fitted with polynomials (Stern, 1980) but the Fourier series representation has been the most popular (Feyerherm and Bark, 1965; Fitzpatrick and Krishnan, 1967; Ison *et al.*, 1971; Buishand, 1977; Woolhiser and Pegram, 1979; Stern and Coe, 1984; and Zucchini and Adamson, 1984).

Early investigators used ordinary least-squares methods to estimate the Fourier coefficients to describe the seasonal variation of model parameters. These methods are undesirable because the data points represent statistical estimates of parameters and these estimates have unequal variances because

of varying sample size. Furthermore, there is no statistically sound procedure to test the significance of individual harmonics. Woolhiser and Pegram (1979) wrote the likelihood equations for a first-order Markov chain and a mixed exponential distribution of amounts as functions of Fourier coefficients and used numerical methods to estimate the coefficients that maximized the likelihoods. They used likelihood ratio tests to determine if harmonics were significant. Stern and Coe (1984) improved upon these techniques in two ways. First they used the logit transform

$$g(t) = \log \{p(t)/[1 - p(t)]\} \quad (5.2)$$

to transform transition probabilities,  $p(t)$ , which are bounded by 0 and 1 into  $g(t)$  which is bounded by plus and minus infinity. They then describe  $g(t)$  by a Fourier series and write the likelihood function as the sum of the logs of the likelihoods of binomial distributions. Since the binomial distribution is a member of the exponential family of distributions, the model is a generalized linear model (Nelder and Wedderburn, 1972) and the package GLIM (Baker and Nelder, 1978) can be used for fitting the model. Stern and Coe (1984) used a shifted gamma distribution for the distribution of rainfall amounts and fitted a Fourier series to the log of the mean of the distribution

$$g(t) = \log [\mu(t)]. \quad (5.3)$$

If  $g(t)$  is linear in the parameters then this is also an example of a generalized linear model and the GLIM package can again be used for fitting.

Zucchini and Adamson (1984) also used the logit transform for the transition probabilities, but used a Newton–Raphson iteration procedure to estimate the parameters. The procedures for parameter estimation using the logit transform are faster than those used by Woolhiser and Pegram (1979) and the parameters of the harmonics are estimated simultaneously. The logit transform leads to superior log likelihoods in climates where a period of very little rain is followed rather quickly by a wet season. Woolhiser and Pegram (1979) had to introduce a penalty function to provide a constraint on the transition probabilities. In Arizona I have found that this constraint is activated and the optimization scheme cannot adequately fit the dry–dry transition probabilities. This problem is illustrated in Figure 5.2. Note the rapid change in  $p_{00}$  from a value near to 1.0 in June to less than 0.6 in July. The Woolhiser and Pegram (1979) method could not accommodate this rapid variation as the penalty function became active and higher harmonics were rejected. With the logit transform the variation was fitted quite closely and substantially higher log likelihoods were obtained. This advantage has not been noted in comparisons made for several stations in more humid areas of the United States. In fact the algorithm described by Woolhiser and Roldan (1986) led to higher likelihoods in a majority of the cases examined, but did require substantially more computer time.

The problem of model selection has been addressed by Gates and Tong (1976), Chin (1977), Eidsvik (1980), Katz (1981), and Zucchini and Adamson (1984). Gates and Tong (1976) used the Akaike information criterion (AIC), Akaike (1974) to determine the appropriate order of Markov chains. Chin

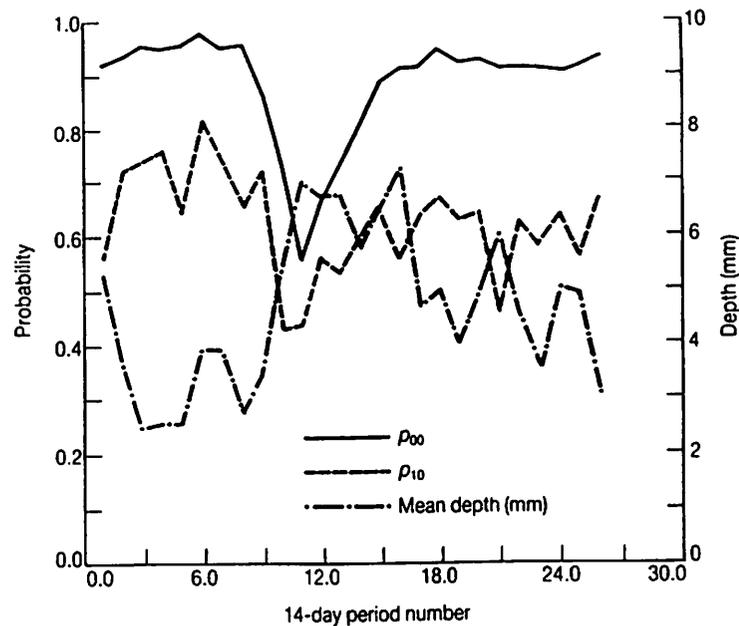


Figure 5.2 Markov chain-mixed exponential model parameters for Walnut Gulch 4, Arizona. 14-day periods beginning 1 March.

(1977) used the AIC technique to identify the appropriate order for Markov chains for more than 100 stations in the United States and found that higher than first-order chains were sometimes required. Eidsvik (1980) used a similar approach for several rainfall stations in Norway. He found that the required order of the chain increased with record length and that with large sample size the AIC minima are not well defined, suggesting significant uncertainty. Katz (1981) showed that the AIC has a substantial probability of overestimating chain order. He then used the Bayesian information criterion (BIC) proposed by Schwarz (1978) and the AIC to estimate the correct order using simulated data. He found that the BIC had a tendency to underfit and proposed a modification to the BIC to correct this problem.

I have calculated the log likelihood functions of the Markov chain and the mixed exponential model as a function of the number of parameters for the Fourier series representation and for two-step function representations (14- and 28-day periods). I analyzed data for the same stations studied by Roldan and Woolhiser (1982). The results are presented graphically for one station in Figures 5.3 and 5.4. The likelihood functions, of course, increase as the number of parameters increase, but after the first 8–10 parameters the rate of increase is slow. The AIC shows a minimum with the Fourier series representation. The results for the other stations were consistent with Figures 5.3 and 5.4.

The representation of seasonality is relevant in selecting the order of Markov chains, for Diggle (1984) has pointed out that stochastic dependence

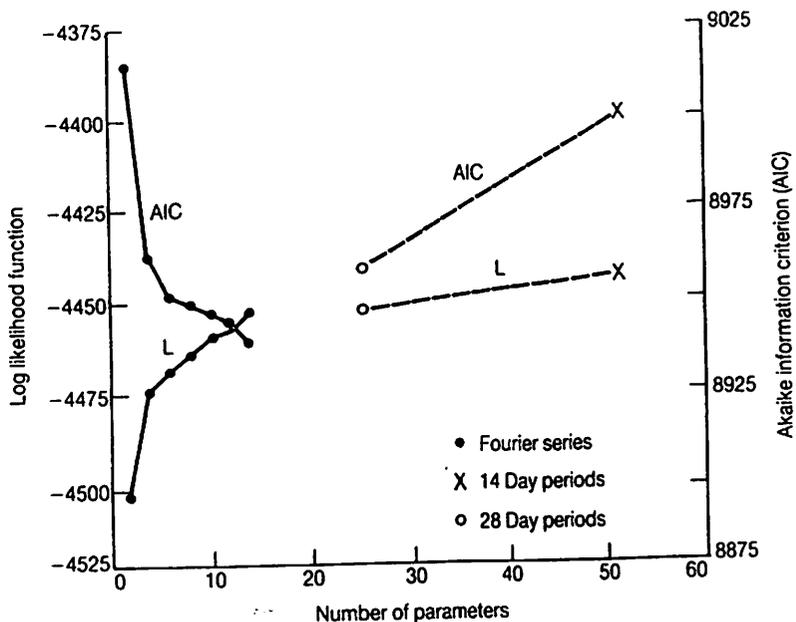


Figure 5.3 Log likelihood and AIC versus number of parameters – Markov chain.

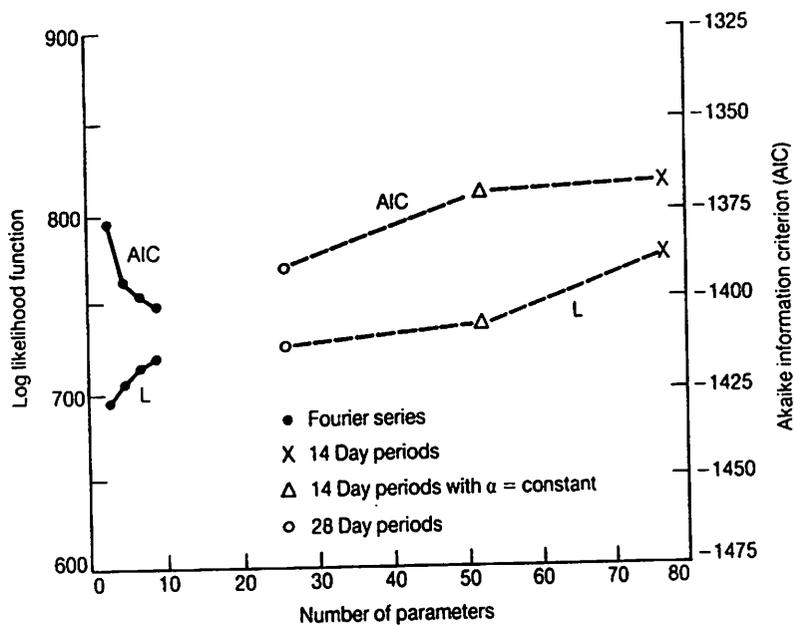


Figure 5.4 Log likelihood and AIC versus number of parameters – mixed exponential distribution.

and deterministic trends may be indistinguishable in practice. It should also be noted that the choice of model should depend to a great extent on the proposed use of the model. A model with a relatively simple structure that can be used in practice (agricultural decision making for example) can be superior to a more complicated model that is unlikely to be used. The important question then becomes: 'How sensitive are the decisions to be made to the precipitation model structure?' Pickering *et al.* (1989) examined this question by comparing three stochastic weather models with historical data as input to a nutrient loss simulation model.

## 5.4 Some problems with precipitation data

Persons who have collected precipitation data in the field and have some training in meteorology and statistics are usually quite aware of problems that may arise that affect the quality of the data. Unfortunately the collectors of field data often do not have training in meteorology and statistics.

Systematic errors and nonhomogeneities may be introduced into rainfall data by changing the type or location of a rain gauge or by construction of buildings nearby or the growth of trees. These factors should be noted in the station history and the user should always be aware of this potential problem.

Frequently there is substantial diurnal variability in the occurrence of rainfall as shown in Figure 5.5. In the United States precipitation data are gathered by a variety of organizations. The National Weather Service (NWS) has a relatively sparse network of first-order stations where an observer is always present. These data are very reliable and midnight is used as the day delimiter. The cooperative observer network is much denser but the observation time varies, usually being around 8.00 a.m. or 5.30–6.00 p.m. The Agricultural Research Service (ARS), the US Geological Survey (USGS) and others often operate rain gauges in conjunction with hydrological research. These data are obtained in analog or digital form by weighing recording gauges or in digital form with tipping-bucket gauges.

It has been well documented that tipping-bucket gauges under register during very intense rainstorms. The weighing-recording gauges with analog output on a chart may underestimate the number of days with small amounts of rainfall if the chart drum revolves daily or more frequently because the thickness of the pen trace may hide small rises. Unless rain gauges are shielded, all will show an underestimate of precipitation when it is accompanied by wind.

Woolhiser and Roldan (1986) found that the time of observation was a significant factor introducing 'noise' into both the occurrence process as defined by a first-order Markov chain and in the distribution of rainfall depth. They attributed this to diurnal variation of rainfall and to evaporation of small rainfall amounts for gauges read in the afternoon.

Data from cooperative stations usually show a reduced number of wet days and greater mean daily amounts as compared to data from nearby first-order stations. Part of this is due to observation time, but part is due to the

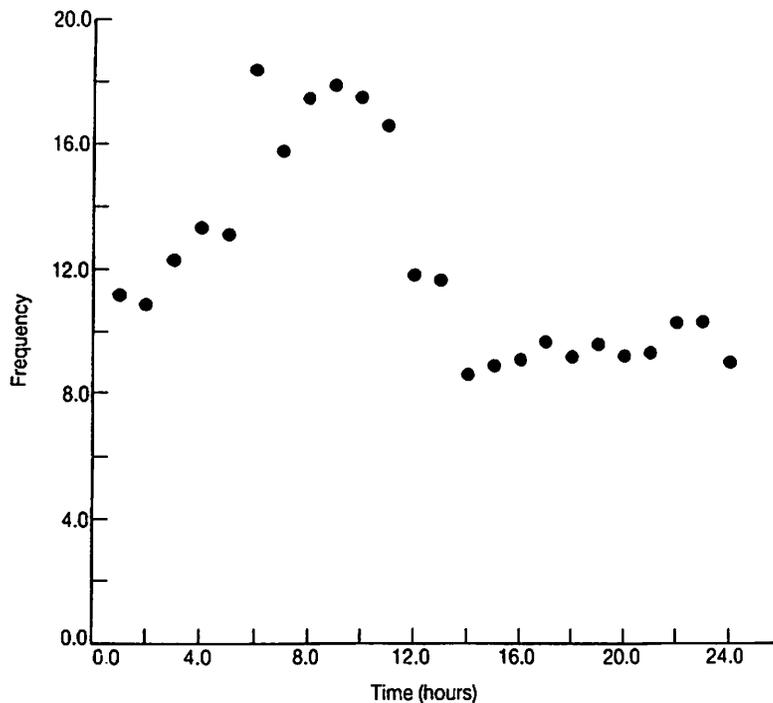


Figure 5.5 Hourly frequencies of precipitation, Hastings, Nebraska.

fact that the observer has missed making an observation on one day and the total accumulation is recorded on the following day. The effect of these methodological factors and errors is primarily reflected in the occurrence process and in the smaller amounts of rainfall. Histograms of rainfall depth are usually erratic, reflecting the rather small sample size, and in many cases it is noted that observers favor certain depths. For example some observers will record an excess of rainfall at 0.05 inches (1.25 mm) and a deficit at 0.04 (1.00 mm) and 0.06 inches (1.50 mm).

These practical problems have a bearing on model selection and on the testing criteria. For example, it may well be a waste of time to develop a complex model to fit the distributions of runs of wet and dry days precisely, because these statistics are very sensitive to the observation time and show wide variations among adjacent stations.

## 5.5 Application to practical problems

Agricultural applications of daily precipitation models fall into two categories: (1) short-run problems such as irrigation management, pest management and planning field operations and (2) long-range problems, including irrigation system design, drainage design, farm planning and examination of the national impact of erosion and flood-control policy. For the short-run

problems, the precipitation models can be used in a simulation mode along with the models of irrigation scheduling (Hubbard and Wilhite, 1987), pest monitoring (Welch, 1984) and crop yield (Jones and Kiniry, 1986). With initial conditions known, several equally likely precipitation sequences can be simulated and distributions of various functionals such as time to next irrigation, time to pesticide application or crop yield can be obtained. The consequences of various management decisions can then be evaluated in a probabilistic sense.

Because these models require more than just precipitation as input, multivariate models including such variables as temperature and solar radiation must be used (see Richardson and Wright, 1984).

Stern and Coe (1984) demonstrate the use of rainfall models to obtain distributions of soil water content as a function of time and the distributions of dry spells. Zucchini and Adamson (1984) identified parameters for 2550 climatic stations in South Africa and used simulation procedures to examine the spatial characteristics of climatic indices, optimal planting time for maize and wheat and characteristics of drought. Because the parameters of rainfall models provide a very concise description of precipitation climatology, a combination of these parameters along with temperature and soils information should prove useful in determining the adaptability of plant species.

Some of the best examples of long-run or policy problems that can be examined using precipitation models include the evaluation of potential nonpoint pollution from agriculture through the use of the CREAMS model (Knisel, 1980) and the evaluation of the effects of erosion on crop productivity (Williams and Renard, 1985).

Many engineering activities are weather dependent and daily rainfall models can be useful in estimating the probability that a project may be disrupted by rain. Weather records or rainfall models are also useful in developing design criteria for drainage works, small dams, urban storm drainage structures and in the evaluation of the trafficability of unpaved areas.

## 5.6 Some new approaches

### 5.6.1 Conditioning models on monthly amounts

Wilks (1989) has developed an interesting precipitation model in which he estimated parameters for a Markov chain and gamma distributions of daily amounts separately for months in the lower 30%, middle 40% and upper 30% of the climatological distributions of total monthly precipitation. These classes correspond to those used by the Climate Analysis Center (CAC) of the National Oceanic and Atmospheric Administration for 30–90-day forecasts. Simulations of daily precipitation sequences can then be conditioned on the monthly forecasts.

Long-term simulations can be carried out by using probability mixtures of the conditional parameter sets which reproduce the observed probabilities of transitions among dry, near-normal, and wet months. These transition

probabilities are represented by a three-state, first-order Markov chain. Wilks (1989) used generalized likelihood ratio tests to show that the increase from four to ten parameters per month is justified by the data. Distributions of monthly rainfall totals simulated using the conditionally derived suites of parameters exhibited upper and lower tails that were closer to the observed distributions than those obtained by unconditional parameters. It would be interesting to try this approach for stations with a monsoon climate.

### 5.6.2 Southern oscillation index

Simulation studies using daily precipitation models with annually periodic parameters typically result in underestimation of the variance of monthly and annual precipitation (Buishand, 1977; Zucchini and Adamson, 1984; Woolhiser *et al.*, 1988). This may be due to changes in data-collection techniques during the period of record, real long-term trends, or the assumption of annual periodicity may be incorrect because of large-scale meteorological circulation patterns that do not exhibit annual periodicities. One such phenomenon that has attracted recent scientific interest is the Southern Oscillation (SO). The Southern Oscillation is ‘a coherent variation of barometric pressures at interannual intervals that is related to weather phenomena on a global scale, particularly in the tropics and subtropics’ (Enfield, 1989). The SO is the atmospheric counterpart to El Niño, the warm, southward-flowing current along the coast of southern Ecuador and northern Peru that appears at irregular intervals. The SO is quantitatively described by the Southern Oscillation Index (SOI), the time series of the anomalies of atmospheric pressure differences between Papeete (Tahiti) and Darwin (Australia).

Several recent studies have documented statistical relationships between the El Niño–Southern Oscillation, commonly referred to as ENSO, and weather patterns and precipitation in the western United States and South America (Caviedes, 1975, 1984; Redmond and Koch, 1991). These studies have typically involved regression analyses between ENSO and annual or seasonal precipitation for groups of stations. These results prompt the question: ‘Is it possible to incorporate the information in the SOI into a stochastic daily precipitation model?’ In the remainder of this section I will describe one possible approach to answering this question.

We will assume that the Markov chain–mixed exponential model (MCME) (Woolhiser and Pegram, 1979) is appropriate for the stations investigated. In this model precipitation occurrence is described by a first-order Markov chain and the mixed exponential distribution is used for the distribution of daily rainfall, given that rain occurs.

Let

$$\begin{aligned} X(t) &= 0 && \text{if day } t \text{ is dry, } t = t_1, \dots, t_T \\ &= 1 && \text{if day } t \text{ has rain over a threshold, } d. \end{aligned} \quad (5.4)$$

We assume that  $\{X(t)\}$  is a first-order Markov chain with transition probabilities

$$P_{ij}(t) = P[X(t) = j | X(t-1) = i], \quad i, j = 0, 1, \quad (5.5)$$

Let  $Y(t)$  be the amount of precipitation on day  $t$  when  $X(t) = 1$ . We assume that  $Y(t)$  is serially independent and is independent of  $X(t-1)$ . Let the random variable  $U(t) = Y(t) - d$  be distributed as a mixed exponential (ME)

$$f_i(u) = \alpha(t)/\beta(t) \exp(-u/\beta(t)) + (1 - \alpha(t))/\delta(t) \exp(-u/\delta(t)) \quad (5.6)$$

where  $0 < u < \infty$ ,  $d$  is a threshold (in the United States normally 0.01 inch (0.25 mm)),  $0 < \alpha(t) < 1$ ,  $0 < \beta(t) < \delta(t)$ . The mean,  $\mu(t)$ , is given by

$$\mu(t) = \alpha(t)\beta(t) + [1 - \alpha(t)]\delta(t). \quad (5.7)$$

The model is nonhomogeneous so the parameters  $p_{00}(t)$ ,  $p_{10}(t)$ ,  $\alpha(t)$ ,  $\beta(t)$ , and  $\mu(t)$  are written in the polar form of a finite Fourier series

$$G_i(t) = G_{i0} + \sum_{k=1}^{m_i} [C_{ik} \sin(2\pi tk/365 + \phi_{ik})] \quad (5.8)$$

where  $i = 1, 2, \dots, 5$ ,  $G_i(t)$  is the value of the  $i$ th parameter on day  $t$ ,  $m_i$  is the maximum number of harmonics,  $G_{i0}$  is the mean,  $C_{ik}$  = amplitude of the  $k$ th harmonic and  $\phi_{ik}$  = phase angle of the  $k$ th harmonic for the  $i$ th parameter. Instead of using the Markov chain transition parameters directly we use the logit transform as demonstrated by Stern and Coe (1984) and Zucchini and Adamson (1984):

$$g_{ij}(t) = \log\{p_{ij}(t)/[1 - p_{ij}(t)]\}. \quad (5.9)$$

The Fourier series are fit to the logits and the transition probabilities are obtained by the inverse transform

$$p_{ij}(t) = \exp[g_{ij}(t)]/[1 + \exp[g_{ij}(t)]]. \quad (5.10)$$

To incorporate the effect of the SOI let us suppose that the periodic parameters are perturbed by a lagged linear function of the SOI

$$G'_i(t) = G_i(t) + b_i S(t - \tau_i) \quad (5.11)$$

where  $b_i$  and  $\tau_i$  are parameters to be estimated from the data and  $S(t)$  is the SOI on day  $t$ . Both of the parameters of the Markov chain and the mean,  $\mu(t)$ , of the mixed exponential distribution were assumed to be affected by the SOI.

The Fourier coefficients for the logits were estimated by maximum likelihood techniques as described by Zucchini and Adamson (1984) and coefficients for the parameters of the mixed exponential distribution were estimated by numerical maximum likelihood as described by Woolhiser and Roldan (1986). Then the characteristics of the likelihood response surface for the Markov chain and the mixed exponential were investigated by varying the parameters  $b_i$  and  $\tau_i$ . A monthly SOI series was used, so  $S(t)$  is represented as a step function.

Data from stations in Arizona, Idaho and Oregon were analyzed in this preliminary study. Some characteristics of the data are shown in Table 5.1. Note that there is a large range in the mean annual precipitation and in the mean number of wet days per year.

Results of perturbing the periodic logits of the transition probabilities are shown in Table 5.2. The increase in log likelihood resulted in a minimum

Table 5.1 Precipitation stations analyzed

Station	Years of record	Annual precipitation (mm)	Mean number wet days
<b>Arizona</b>			
Phoenix	1949–81	177	34.39
Prescott	1953–81	478	68.97
Tucson	1948–81	277	49.88
Walnut Gulch 4	1955–87	305	53.61
<b>Idaho</b>			
Boise	1940–89	304	91.38
Grangeville	1940–84	600	121.82
Reynolds Cr. 116	1962–81	469	105.45
Reynolds Cr. 163	1962–87	1128	132.15
<b>Oregon</b>			
Corvallis	1948–87	1078	156.68
Crater Lake	1947–87	1732	142.02
Bonneville Dam	1950–87	1945	171.87

Table 5.2 Effect of SOI on log likelihood functions for occurrence of rainfall – first-order Markov chain

Station	Unperturbed $L$	Perturbed $L$	$b_i$	Lag (days)
<b>Arizona</b>				
Phoenix	-3306.560	-3300.482*	+0.120	95
Prescott	-4240.921	-4236.580*	+0.100	95
Tucson	-4277.916	-4269.211*	+0.142	95
Walnut Gulch 4	-4140.479	-4129.321*	+0.147	93
<b>Idaho</b>				
Boise	-3696.153*	-3695.075	+0.04	330
Grangeville	-9576.560	-9573.180*	-0.04	104
Reynolds Cr. 116	-3767.502*	-3766.384	+0.053	150
Reynolds Cr. 163	-5199.465	-5196.788*	+0.060	0
<b>Oregon</b>				
Corvallis	-7320.965*	-7320.580	-0.02	90
Crater Lake	-7487.811*	-7487.811	0.00	0
Bonneville Dam	-7126.041*	-7124.312	-0.06	30

\* Minimum AIC

AIC for six stations, with the Arizona stations being most strongly affected. The signs of the coefficients are fairly consistent with previous studies, with a negative SOI leading to more rainfall in the Southwest and the opposite effect in the Pacific Northwest.

Results of perturbing the seasonally varying mean of the ME are shown in Table 5.3. The perturbed mean precipitation resulted in the minimum AIC for all stations and the signs of the coefficients are consistent with expectations except for Boise, ID. Again the Arizona stations exhibit the strongest

**Table 5.3** Effects of perturbing the mean of the ME distribution on the log likelihood function

Station	Unperturbed $L$	Perturbed $L$	$b_i$ (mm)	Lag (days)
Arizona				
Phoenix	876·973	876·180*	-0·43	90
Prescott	950·163	955·856*	-0·60	60
Tucson	1184·797	1187·926*	-0·39	75
Walnut Gulch 4	1092·404	1096·816*	-0·45	89
Idaho				
Boise	2166·453	2170·052*	-0·19	0
Grangeville	4162·459	4164·689*	+0·13	60
Reynolds Cr. 116	1795·836	1803·659*	+0·39	101
Reynolds Cr. 163	705·685	714·536*	+0·40	180
Oregon				
Corvallis	2889·396	2894·771*	+0·20	104
Crater Lake	-1188·772	-1185·422*	+0·254	94
Bonneville Dam	-638·453	-635·349*	+0·254	115

\* Minimum AIC

effects. The most common lag is about 90 days, which raises interesting possibilities of conditioned simulations of rainfall based upon the SOI for the past 90 days. However, this analysis is preliminary and we cannot reach strong conclusions based on the evidence presented. For example, although the Markov chain-mixed exponential appears to fit well for the Arizona and Reynolds Creek data at least a second-order Markov chain may be required for the Oregon stations. A more thorough analysis using more stations and higher-order Markov chains is presently under way.

### 5.6.3 Elevation effects

It is often desirable to have daily precipitation data in mountainous regions but frequently data are only available for valley stations and usually these data do not represent conditions at higher elevations. Therefore it would be useful if relationships could be developed between precipitation model parameters and characteristics of higher elevation sites such as elevation or annual precipitation so that simulation model parameter sets could be estimated. Hanson *et al.* (1989) utilized data from a network of rain gauges in southwest Idaho, United States, to determine if such relationships could be found. Fourier coefficients describing the seasonal variability for the parameters of the MCME model were estimated for each station and regression relationships were obtained between means, amplitudes and phase angles and annual precipitation. Highly significant relationships were found between the logits of the mean  $p_{00}$ ,  $p_{10}$ , the amplitudes of the first harmonic of both  $p_{00}$  and  $p_{10}$  and annual precipitation,  $P_a$ . For the distribution of amounts, the logit of the weighting parameter,  $\alpha$ , showed a linear decrease with  $P_a$ , the mean  $[\alpha\beta + (1 - \alpha)\delta]$  and the amplitude of the first harmonic

increased linearly with  $P_a$ , and the phase angle of the first harmonic exhibited a nonlinear decrease with  $P_a$ . The regression relationships were used to estimate a complete set of parameters for four sites and the statistical properties of 50 years of simulated data were compared with historical data at the same sites. Mean monthly precipitation and number of wet days were closely preserved. There appeared to be a slight tendency for the simulated data to have a lower variance than the historical data. The procedure appears promising but it must be tested over a larger area and information must be provided so that the user can evaluate the potential errors involved.

## 5.7 Discussion

Models to describe the daily precipitation process are well developed and a great deal of progress has been made recently in developing techniques for parameter estimation – particularly when the seasonal variation is described by Fourier series. Various practical problems related to data collection may affect the models chosen and may limit the degree of fit that can be obtained. Increasing use of crop yield models and models simulating runoff, erosion and chemical transport will lead to a greater demand for models for simulating precipitation and other weather variables.

A shortcoming of existing models is an inability to maintain the variance of simulated monthly and annual totals. Conditioning model parameters on monthly amounts (Wilks, 1989) or perturbing periodic parameters with the SOI both result in better agreement between the variance of simulated and observed annual total precipitation. These approaches should be investigated more thoroughly and the statistical problems and the practical impact should be assessed.

## Acknowledgement

I wish to thank T. O. Keefer for his assistance in the SOI analysis.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19 (6), 716–23.
- Baker, R.J. and Nelder, J.A. (1978). *The GLIM System*, NAG, Oxford.
- Buishand, T.A. (1977). Stochastic modeling of daily rainfall sequences. *Meded. Landbouwhoges. Wageningen*, 77 (3), 211 pp.
- Caskey, J.E. (1963). A Markov chain model for the probability occurrence in intervals of various lengths. *Mon. Weather Rev.*, 91, 298–301.
- Caviedes, C.N. (1975). El Niño 1972: Its climatic, ecological and human implications. *Geograph. Rev.*, 65, 493–509.
- Caviedes, C.N. (1984). El Niño 1982–83. *Geograph. Rev.*, 74, 268–90.
- Chang, T.J., Kavvas, M.L., and Delleur, J.W. (1984). Daily precipitation modeling by

- discrete autoregressive moving average processes. *Water Resources Res.*, **20**, 565–80.
- Chin, E.H. (1977). Modeling daily precipitation occurrence process with Markov chain. *Water Resour. Res.*, **13** (6), 949–56.
- Coe, R. and Stern, R.D. (1982). Fitting models to daily rainfall data. *J. Appl. Met.*, **21**, 1024–31.
- Delleur, J.W., Chang, T.J., and Kavvas, M.L. (1989). Simulation models of sequences of dry and wet days. *J. Irrig. and Drainage Engr.*, **115** (3), 344–57.
- Dethier, B.E. and McGuire, J.K. (1961). The climate of the Northeast. Probability of selected weekly precipitation amounts in the Northeast region of the U.S. *Agronomy Mimeo 61-4*. Cornell Univ. Agr. Exp. Stn., Ithaca, NY.
- Diggle, P.J. (1984). Discussion of paper by Stern and Coe. *J. Roy. Statist. Soc. A*, **147**, Part 1, 27–8.
- Eidsvik, K.J. (1980). Identification of models for some time series of atmospheric origin with Akaike's information criterion. *J. Appl. Met.*, **19** (4), 357–69.
- Enfield, D.B. (1989). El Niño, past and present. *Rev. Geophys.*, **21** (1), 159–87.
- Feyerherm, A.M. and Bark, L.D. (1965). Statistical methods for persistent precipitation patterns. *J. Appl. Met.*, **4**, 320–8.
- Feyerherm, A.M., Bark, L.D., and Burrows, W.C. (1965). Probabilities of sequences of wet and dry days in South Dakota. *North Central Region Res. Pub. 161 and Tech. Bull. 139h*. Kansas Agric. Exp. Stn., Manhattan, KS.
- Fitzpatrick, E.A. and Krishnan, A. (1967). A first order Markov model for assessing rainfall discontinuity in Central Australia. *Arch. Met. Geophys. Biokl.*, **B13**, 270–86.
- Foufoula-Georgiou, E. and Lettenmaier, D.P. (1987). Markov renewal model for rainfall occurrence. *Water Resour. Res.*, **23** (5), 875–84.
- Gabriel, K.R. and Neumann, J. (1962). A Markov chain model for daily rainfall in Tel Aviv. *Quart. J. Roy. Met. Soc.*, **88**, 90–5.
- Gates, P. and Tong, H. (1976). On Markov chain modelling to some weather data. *J. Appl. Met.*, **15**, 1145–51.
- Gifford, R.O., Ashcroft, G.L., and Magnuson, M.D. (1967). Probability of selected precipitation amounts in the Western Region of the U.S. *Western Regional Res. Pub. and T-8*. Nevada Agric. Exp. Stn., Reno, NV.
- Green, J.R. (1964). A model for rainfall occurrence. *J. Roy. Statist. Soc., Series B*, **26**, 345–53.
- Hanson, C.L., Osborn, H.B., and Woolhiser, D.A. (1989). Daily precipitation simulation model for mountainous areas. *Trans. ASAE*, **32** (3), 865–73.
- Heerman, D.F., Finkner, M.D., and Hiler, E.A. (1968). Probability of sequences of wet and dry days for eleven Western States and Texas. *Colorado Agric. Exp. Stn. Tech. Bull.*, 117.
- Hubbard, K.G. and Wilhite, D.A. (1987). A demonstration and evaluation of the use of climate information to support irrigation scheduling and other agricultural operations. *CAMaC Prog. Report 87-4*. Center for Agricultural Meteorology and Climatology, Univ. of Nebraska, Lincoln.
- Ison, N.T., Feyerherm, A.M., and Bark, L.D. (1971). Wet period precipitation and the gamma distribution. *J. Appl. Met.*, **10** (4), 658–65.
- Jones, C.A. and Kiniry, J.R. (eds) (1986). *CERES-MAIZE: A Simulation Model of Maize Growth and Development*, Texas A & M University Press, College Station, TX, 194 pp.
- Katz, R.W. (1977a). An application of chain-dependent processes to meteorology. *J. Appl. Prob.*, **14**, 598–603.
- Katz, R.W. (1977b). Precipitation as a chain-dependent process. *J. Appl. Met.*, **16** (7), 671–6.

- Katz, R.W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics*, **23** (3), 243–9.
- Katz, R.W. (1984). Discussion of the paper by Dr. Stern and Mr. Coe. *J. Roy. Statist. Soc. A*, **147**, Part 1, 29.
- Knisel, W.G. (ed.) (1980). *CREAMS: A Field Scale Model for Chemicals, Runoff and Erosion from Agricultural Management Systems*, U.S. Dept. of Agric., Conservation Research Report No. 26, 640 pp.
- Lane, L.J. (1984). Surface water management: a user's guide to calculate a water balance using the CREAMS model. *LA-10177-M. Manual UC-70B*, Los Alamos National Laboratory, Los Alamos, N.M.
- Mielke, P.W. (1973). Another family of distributions for describing and analyzing precipitation data. *J. Appl. Met.*, **10** (2), 275–80.
- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J. Roy. Statist. Soc. A*, **135**, 370–84.
- Nicks, A.D. (1974). Stochastic generation of the occurrence, pattern, and location of maximum amount of daily rainfall. *Proc. Symp. Statistical Hydrology*, Misc. Publ. No. 1275, U.S. Dept. of Agriculture, ARS. 154–71.
- Palmer, W.L., Barfield, B.J., and Hann, C.T. (1982). Sizing farm reservoirs for supplemental irrigation of corn. Part I: Modeling reservoir size yield relationships. *Trans. ASAE*, **25**, 372–6.
- Pickering, N.B., Stedinger, J.R., and Haith, D.A. (1989). Weather Input for Non-Point source Pollution Models. *J. Irrig. and Drainage Engr.*, ASCE, **114** (4), 674–90.
- Redmond, K.T. and Koch, R.W. (1991). Surface climate and streamflow variability in the Western United States and their relationship to large scale circulation indices. *Water Resour. Res.*, **27** (9), 2381–99.
- Richardson, C.W. (1982). A comparison of three distributions for the generation of daily rainfall amounts. In Singh, V.P. (ed.), *Statistical Analysis of Rainfall and Runoff*, pp. 67–78, Proc. Int. Symp. on Rainfall-Runoff Modeling, Water Resour. Publ., 700 pp.
- Richardson, C.W. and Wright, D.A. (1984). *WGEN: A Model for Generating Daily Weather Variables*. U.S. Dept. of Agric., Agric. Res. Svc., ARS-8, 83 pp.
- Roldan, J. and Woolhiser, D.A. (1982). Stochastic daily precipitation models I. A comparison of occurrence models. *Water Resour. Res.*, **18** (5), 1451–9.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–4.
- Shaw, R.H., Barger, G.L., and Dale, R.F. (1960). *Precipitation Probabilities in the North Central States*, Bull. 753, Missouri Agric. Exp. Stn., Columbia, MO.
- Smith, J.A. (1987). Statistical modeling of daily rainfall occurrences. *Water Resour. Res.*, **23** (5), 885–93.
- Smith, R.E. and Schreiber, H.A. (1974). Point processes of seasonal thunderstorm rainfall, 2. Rainfall depth probabilities. *Water Resour. Res.*, **10** (3), 418–23.
- Stern, R.D. (1980). Analysis of daily rainfall at Samaru, Nigeria, using a simple two-part model. *Arch. Met. Geophys. Biokl., Ser. B*, **28**, 123–35.
- Stern, R.D. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *J. Roy. Statist. Soc. A*, **147**, Part 1, 1–34.
- Stidd, C.K. (1953). Cube root normal precipitation distributions. *Amer. Geophys. Union Trans.*, **34** (1), 31–5.
- Von Barga, K. (1967). A systems approach to harvesting alfalfa hay, *Trans. ASAE*, **10** (3), 318–19.
- Welch, S.M. (1984). Developments in computer-based IPM extension delivery systems. *Ann. Rev. Ent.*, **29**, 359–81.
- Wilks, D.S. (1989). Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.*, **23** (6), 1429–39.

- Williams, J.R. and Renard, K.G. (1985). Assessments of soil erosion and crop productivity with process models (EPIC). In *Soil Erosion and Crop Productivity*, Follett, R.F. and Stewart, B.A. (eds), American Society of Agronomy and Crop Science Society of America, Madison, WI, pp. 67–103.
- Woolhiser, D.A. and Pegram, G.G.S. (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variation of parameters in stochastic daily precipitation models. *J. Appl. Met.*, **18** (1), 34–42.
- Woolhiser, D.A. and Roldan, J. (1982). Stochastic daily precipitation models 2. A comparison of distribution of amounts. *Water Resour. Res.*, **18** (5), 1461–8.
- Woolhiser, D.A. and Roldan, J. (1986). Seasonal and regional variability of parameters for stochastic daily precipitation models: South Dakota, U.S.A. *Water. Resour. Res.*, **22** (6), 965–78.
- Woolhiser, D.A., Hanson, C.L., and Richardson, C.W. (1988). *Microcomputer Program for Daily Weather Simulation*, U.S. Dept. of Agric., Agric. Res. Svc., ARS-75, 49 pp.
- Zucchini, W. and Adamson, P.T. (1984). *The Occurrence and Severity of Droughts in South Africa*, Dept. of Civil Engr., Univ. of Stellenbosch and Dept. of Water Affairs, WRC Report No. 91/1/84, 198 pp.