

2148-5608A

METHODS FOR MODELLING
PRECIPITATION PERSISTENCE

by

BRENDA ANN WEAK
B.S., Kansas State University, 1970

A MASTER'S REPORT
submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
KANSAS STATE UNIVERSITY
Manhattan, Kansas

1974

Approved by:

A. M. Feyerherm
Major Professor

LD
2668
R4
1974
W43
C.2
Document

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
2.	FIRST-ORDER MARKOV CHAIN.....	2
2.1	Assumptions and Notation.....	2
2.2	Estimation of Parameters.....	4
2.3	Fit of the Model.....	9
2.4	Properties of the Model.....	12
3.	ALTERNATIVE MODELS.....	15
3.1	Higher-Order Markov Chains.....	15
3.2	Modified Log Series.....	17
3.3	Exponential Distribution.....	18
	REFERENCES.....	25

ACKNOWLEDGEMENTS

The author wishes to thank her major professor, Dr. Arlin M. Feyerherm, for assistance in the preparation of this report. The author also wishes to express her appreciation to Dr. Arthur D. Dayton and Dr. George A. Milliken for serving on her committee.

1. INTRODUCTION

Precipitation patterns affect many of man's activities. If man can gain foreknowledge of weather events, he can adjust his activities to make the best use of the weather events. While it is impossible for man to know absolutely what will happen in the future, it is possible for him to use his knowledge of past events to gain insight into what is most likely to happen in the future.

This paper will deal with methods used for modelling persistence in precipitation patterns and using knowledge of this persistence to state probabilities of certain precipitation events. It will deal only with the presence or absence of precipitation. A next step would be to work with amounts of precipitation.

Insight into future precipitation events at a specific location are of interest to the agronomist developing new crops, to the engineer planning an outdoor construction project, to the community planner developing outdoor recreation areas, to the conservationist planning optimum usage of water supplies, and to many others.

Day-to-day precipitation events are not independent. There is persistence. One cannot estimate the probability of a wet day and apply the binomial model to study series of days.

2. FIRST-ORDER MARKOV CHAIN MODEL

2.1 Assumptions and Notation

The most often-used model for studying day-to-day persistence is the first-order, two-state Markov chain defined on a discrete parameter space. This model assumes that if today's precipitation is conditioned on yesterday's precipitation, then further conditioning on past events adds no more information about today's precipitation. A wet day is defined as precipitation exceeding a specified threshold, generally .01 inch. We will be concerned with finding the probability that a day is wet, both marginal and conditional on yesterday's presence or absence of precipitation. From these probabilities other properties of long-term precipitation patterns will be described.

In discussing the first-order Markov chain, the following notation will be used. Define a series of random variables X_t , $t = 1, 2, \dots, 365$, corresponding to the days of the year; $X_t = \text{Dry}$, if the t -th day of the year is dry; $X_t = \text{Wet}$, if the t -th day of the year is wet. To further abbreviate, $X_t = \text{Dry}$ will be written D_t and $X_t = \text{Wet}$ will be written W_t . Six probabilities are of interest for each value of t :

$$P(\text{t-th day of the year is dry}) = P(D_t)$$

$$P(\text{t-th day of the year is wet}) = P(W_t)$$

$$P(\text{t-th day of the year is dry given (t-1)-th day of the year is dry}) = P(D_t/D_{t-1})$$

$P(\text{t-th day of the year is wet given (t-1)-th day of the year is dry}) = P(W_t/D_{t-1})$

$P(\text{t-th day of the year is dry given (t-1)-th day of the year is wet}) = P(D_t/W_{t-1})$

$P(\text{t-th day of the year is wet given (t-1)-th day of the year is wet}) = P(W_t/W_{t-1})$.

It can be shown by basic probability laws that four of these probabilities are functionally related to the other two. If $P(D_t)$ and $P(D_{t-1})$ are given:

$$P(W_t) = 1 - P(D_t)$$

$$P(W_t/D_{t-1}) = 1 - P(D_t/D_{t-1})$$

$$P(D_t/W_{t-1}) = \frac{P(D_t) - P(D_{t-1})P(D_t/D_{t-1})}{1 - P(D_{t-1})}$$

$$P(W_t/W_{t-1}) = 1 - P(D_t/W_{t-1}). \quad (2.1)$$

These are in the form used by Feyerherm and Bark [4].

Some writers assume that each of the six probabilities is constant over a given time period such as one month or one season. This assumption will be called the stationary assumption and it will be discussed in further detail later. When this assumption is made, the subscript t will be dropped. Under the stationary model, Gabriel and Neumann [6] defined four of the probabilities in terms of $P(W/W)$ and $P(W/D)$:

$$P(D/W) = 1 - P(W/W)$$

$$P(D/D) = 1 - P(W/D)$$

$$P(W) = \frac{P(W/D)}{1 - P(W/W) + P(W/D)}$$

$$P(D) = 1 - P(W). \quad (2.2)$$

Equations (2.1) and (2.2) are algebraically equivalent if one assumes $P(D_t) = P(D_{t-1}) = P(D)$, $P(D_t/D_{t-1}) = P(D/D)$, etc.

2.2 Estimation of Parameters

Suppose k years of data on a particular sequence of n days during the year are available. Assuming the k years yield independent observations, results by Anderson and Goodman [1] can be applied to derive maximum likelihood estimators for both the stationary and non-stationary first-order Markov models. These estimates are simply the relative frequencies. For the non-stationary model, for each value of t :

$$P(\hat{D}_t) = \frac{\text{total number of years in which the } t\text{-th day of the year is dry}}{\text{total number of years in the sample}}$$

$$P(D_t/\hat{D}_{t-1}) = \frac{\text{total number of years that the } (t-1)\text{-th and } t\text{-th days of the year are dry}}{\text{total number of years that the } (t-1)\text{-th day of the year is dry}}$$

Estimates of the other four probabilities are calculated in terms of these as described in equations (2.2).

Klotz [16] has examined other properties of the stationary estimators given above. He has shown these estimators to be sufficient and to be asymptotically efficient. He shows that,

for k independent samples each of size n , $(nk)^{1/2}(\hat{P}(\hat{D}/D) - P(D/D), \hat{P}(\hat{D}) - P(D))$ has a limiting bivariate normal distribution $N((0,0), \Lambda)$ where:

$$\Lambda = \begin{pmatrix} \frac{P(D/D)P(W/D)}{P(D)} & P(W)P(D/D) \\ P(W)P(D/D) & \frac{P(D)P(W)(1 - 2P(D) + P(D/D))}{P(W/D)} \end{pmatrix}$$

Thus, for large nk the estimators are unbiased and one can derive the following properties:

$$\text{Var}(\hat{P}(\hat{D})) = \frac{P(D)P(W)(1 - 2P(D) + P(D/D))}{nkP(W/D)}$$

$$\text{Var}(\hat{P}(\hat{D}/D)) = \frac{P(D/D)P(W/D)}{nkP(D)}$$

$$\text{Cov}(\hat{P}(\hat{D}), \hat{P}(\hat{D}/D)) = \frac{P(W)P(D/D)}{nk}$$

From the above, the following desirable property of the estimators is noted. As $P(D) \rightarrow P(D/D)$, $\text{Var}(\hat{P}(\hat{D})) \rightarrow \frac{P(D)P(W)}{nk}$, the variance of a binomial distribution with no day-to-day dependence

Some further modifications of these estimates have been suggested. In fitting the stationary model to several sets of data, Weiss [20] noted that there appeared to be a seasonal variation in the probabilities and this should somehow be accounted for. Feyerherm and Bark [4] assumed that the daily probabilities should approximate a continuous curve with at least a one-year period and possibly also periods of shorter length. They expressed the seasonal variation in daily

probabilities by use of partial sums of a Fourier series.

They let:

$$P(D_t) = A_0 + \sum_{h=1}^m (A_h \cos\left(\frac{2\pi h t}{365}\right) + B_h \sin\left(\frac{2\pi h t}{365}\right)) \quad (2.3)$$

$A_0, A_1, \dots, A_m, B_1, B_2, \dots, B_m$ are constants. Relative frequency estimates were obtained for $P(D_t)$ for $t = 1, 2, \dots, 365$. These were written in equations of the form of (2.3) and estimates of $A_0, A_1, \dots, A_m, B_1, B_2, \dots, B_m$ were obtained by the method of least squares. These estimates were then placed in equations (2.3) and new estimates of $P(D_t)$ could be written. Hartley [14] had earlier given a procedure based on an F-test to test which harmonics in such an equation are significantly different from zero and this was applied. Weber and Bayne [17] use a slightly different procedure to test significance of the harmonics. Seasonal variation, in the conditional probabilities, $P(D_t | D_{t-1})$ was also expressed in terms of partial sums of the Fourier series in an equation like (2.3).

For a set of six locations, Feyerherm and Bark [4] found that the first harmonic was always significant, the second harmonic was always significant for $P(D_t)$ when a wet day was defined as receiving at least .01 inch of precipitation and higher harmonics appeared sometimes. The variance of the estimate obtained in this way is $(\frac{2j+1}{365})\text{Var}(\hat{P}(D_t))$, $j = 1, 2, \dots, 12$, where j is the number of harmonics used in the equation

and $P(\hat{D}_t)$ is the relative frequency estimate of $P(D_t)$. Since j is relatively small, this procedure yields an estimate with a much smaller variance than the separate relative frequency estimates based on a single day or pair of days of the year.

Gringorten [12] noted that obtaining relative frequency estimates in the non-stationary model for $P(D_t)$ and $P(D_{t-1})$ requires a large amount of tedious work. He suggests obtaining relative frequency estimates for $P(W_t)$ and also obtaining an estimate for ρ_0 , the day-to-day correlation coefficient between the standard normal transformed variables of amounts of precipitation. ρ_0 is considered constant throughout the year. This does not mean the conditional probabilities will be constant. In order to use this procedure, one must have records not only of whether a day was wet or dry but also of the actual amounts of precipitation. For a sufficiently large sample of two-day sequences, the amounts of daily precipitation received, y_i , must be transformed to a standard normal distribution. Gringorten suggests doing this graphically. The

estimate of ρ_0 is obtained by computing $r_0 = \frac{\text{Cov}(z_{i-1}, z_i)}{s_{z_{i-1}} s_{z_i}}$

where z_i is the standard normal transformed value for y_i . For each value of $P(\hat{W}_t)$ one must also find z_t where Z is a standard normal random variable and z_t is the constant such that $P(\hat{W}_t) = P(Z > z_t)$. The model proposed by Gringorten is:

$$z_{t/t-1} = z_t - \lambda \exp(-\mu z_t)$$

where $z_{t/t-1}$ is the constant such that $P(Z > z_{t/t-1}) = P(W_t/W_{t-1})$. In this model, λ and μ are parameters numerically defined by $P(W_{t-1})$ and ρ_0 . Gringorten used the method of least squares to draw charts giving values for λ and for μ for given values of ρ_0 and z_{t-1} . Gringorten also gave a more general equation for finding $P(W_t/W_{t-n})$:

$$z_{t/t-n} = z_t - \lambda \exp(-\mu z_t).$$

λ and μ are defined as before, but the correlation coefficient one must enter the charts with is the correlation coefficient, ρ , between the standard normal variates corresponding to amounts of precipitation on two days n days apart. Under the assumption of a Markov chain, $\rho = \rho_0^n$.

Gringorten [12] later suggested a similar model that did not include parameters λ and μ . This model requires an actual amount of precipitation rather than a simplification of "wet" or "dry" for the previous day. This model is:

$$z_{t/t-1} = \frac{z_t - \rho_0 z_0}{(1 - \rho_0^2)^{1/2}}.$$

$z_{t/t-1}$ and z_t are defined as before. z_0 is the constant number such that $P(Z \geq z_0) = P(Y \geq y_0)$ where y_0 is the amount of precipitation received on the previous day. So, if one knows the exact amount of precipitation received on the previous day, this equation can be used.

A simplified method for estimating $P(W_t/W_{t-n})$ is the main advantage of Gringorten's method. From $P(W_t/\hat{W}_{t-n})$, one could of course obtain $P(D_t/\hat{W}_{t-n})$, $P(W_t/\hat{D}_{t-n})$, and $P(D_t/\hat{D}_{t-n})$ using the relationships defined earlier. To test his model, Gringorten compared the estimated conditional probabilities with relative frequency estimates of conditional probabilities. Data used was 18 years of records for Boston, Mass. The rms difference between the estimates of conditional probabilities by the two methods was $\pm .05$. The mean bias of the estimate was .008.

2.3 Fit of the Model

The fit of a first-order Markov chain in modelling conditional precipitation probabilities has been tested with the chi-square goodness-of-fit test. Gabriel and Neumann [6] were the first to suggest the use of this model. They used the stationary model on daily precipitation records for 27 years of rainy seasons in Tel Aviv, defining a wet day as a day receiving at least .1 mm of rain and found the model to fit. Caskey [2] found that 11 years of Denver, Colorado, data fit the Markov assumption where constant probabilities were used for each of the four seasons of the year defining a wet day as a trace or more of precipitation. Weiss [20] reported that the Markov model fit data in France (49 years data), San Francisco (20 years data), Harpendon, England (nine years data), Montreal (76 years data), and Moneton, Canada (49 years data).

Feyerherm and Bark [4] found the model with non-stationary probabilities fit data at six different weather stations in the midwest. Weber and Bayne [19] successfully fit the non-stationary model for the entire state of North Carolina with probabilities varying throughout the state. Hopkins [15] fit the model for three stations in the Canadian prairie provinces. In general, the model has been found to fit data from many different areas with differing climates. Lowry and Guthrie [17] noted that the first-order Markov chain model would not be appropriate for areas having very diverse climates within a season. These are areas where persistence is more important than reflected in a first-order Markov chain. This occurs where there are more long sequences of wet days, more short sequences of wet days, and fewer intermediate-length sequences of wet days than predicted by the first-order Markov chain model. These areas do not seem to occur frequently.

The amount of precipitation needed to define a day as "wet" varies among writers. Lowry and Guthrie [17] examined several sets of data and fit the Markov chain model repeatedly defining a wet day as one receiving more than .01 inch, .05 inch, .10 inch, .20 inch, .30 inch, and .50 inch. He found that, for a given station, the first-order Markov chain fits better when the amount of precipitation defining a wet day is a smaller amount.

As mentioned above, both the stationary and non-stationary

models have been found to fit data. Gabriel and Neumann [6], after testing their stationary model, stated: "If even this simplified model fits the data, there is all the more reason to assume that the more realistic model with continuous seasonal variation of the probabilities must fit." It would be most exact to assume that the marginal and conditional probabilities of precipitation or no precipitation are changing every instant of the day. To assume them constant throughout a given day is a simplification. To assume them constant throughout a given month or season is a further simplification. Thus, one would expect to get a better fit of the model as probabilities are calculated for smaller units of time and the non-stationary model would give a better fit than the stationary model. But the stationary model has been shown to fit satisfactorily by use of the chi-square tests, so this is a possible approximate model. The stationary model has the advantage of requiring less calculations as there are fewer probabilities to be calculated and then fewer probabilities to work with in deriving further properties of the precipitation patterns.

The problem most generally worked with classifies days as wet or dry. It is conceivable that one might be interested in more than two states. Lowry and Guthrie [17] mentioned this problem and noted that it would simply require working with an n-state Markov process rather than a two-state process.

One reason this is not too practical is that it would require more years of data to accurately estimate parameters. There would be more parameters to deal with. And, for practical purposes, a two-state process often provides enough information. No examples of the model being fit to actual data defined in terms of more than two states were found. It is possible that data would not fit the n-state Markov chain, but would fit the two-state Markov chain.

In fitting the first-order Markov chain model, year-to-year independence is assumed. Weiss [20] tested this assumption by dividing 50 years of data into two groups of 25 years of data. Probabilities were enough the same for the two groups to cause Weiss to conclude there was no secular trend.

2.4 Properties of the Model

One of the main reasons for fitting a first-order Markov chain to precipitation patterns is that one can easily obtain many useful properties. The probability that the sequence $x_t, x_{t+1}, \dots, x_{t+n}$ occurs for a given sequence of $n+1$ days is simplified by Feyerherm and Bark [4] to:

$$P(x_t, x_{t+1}, \dots, x_{t+n}) = P(x_t)P(x_{t+1}/x_t)P(x_{t+2}/x_{t+1})\dots \\ P(x_{t+n}/x_{t+n-1}).$$

Gabriel and Neumann [6] stated several properties of the stationary model. The probability of a wet spell of length k (a

dry day, followed by k wet days, followed by a dry day, occurring is: $P(D/W)(P(W/W))^{k-1}$. The probability of a dry spell of length k occurring is: $P(W/D)(P(D/D))^{k-1}$. The distribution of spells by length is geometric. The probability of exactly s wet days among n days following a wet day is:

$$P_1 = (P(W/W))^s (P(D/D))^{n-s} \sum_{c=1}^{c_1} \binom{s}{a} \binom{n-s-1}{b-1} \left(\frac{P(D/W)}{P(D/D)} \right)^b \left(\frac{P(W/W)}{P(D/W)} \right)^a$$

where $c_1 = \begin{cases} n + 1/2 - |2s - n + 1/2| & \text{if } s < n \\ 0 & \text{(and the sum involves only one term) if } s = n \end{cases}$

and a and b are the least integers not smaller than $1/2(c-1)$ and $1/2c$ respectively. The probability of exactly s wet days among n days following a dry day is:

$$P_2 = (P(W/W))^s (P(D/D))^{n-s} \sum_{c=1}^{c_0} \binom{s-1}{b-1} \binom{n-s}{a} \left(\frac{P(D/W)}{P(D/D)} \right)^a \left(\frac{P(W/D)}{P(W/W)} \right)^b$$

where $c_0 = \begin{cases} n + 1/2 - |2s - n - 1/2| & \text{if } s > 0 \\ 0 & \text{(and the sum involves only one term) if } s = 0 \end{cases}$

and a and b are defined as above. The probability of s wet days among any n days is then:

$$P(W)P_1 + P(D)P_2.$$

For large n , the distribution of the number of wet days in n days approaches normal with a mean of $nP(W)$ and a variance of $nP(W)P(D)\frac{1 + P(W/W) - P(W/D)}{1 - P(W/W) + P(W/D)}$. Caskey [2] stated the following:

The probability of at least one wet day in n days is

$$1 - P(D)(P(D/D))^{n-1}.$$

From Parzen [18], pages 242-243, the mean first passage and mean recurrent times can be obtained. Given that $X_t = \text{Dry}$, the mean number of days until a wet day next occurs is $\frac{1}{P(W/D)}$. Given that $X_t = \text{Dry}$, the mean number of days until a dry day next occurs is $\frac{1}{P(D)}$. Similarly, mean first passage time from wet to dry is $\frac{1}{P(D/W)}$ days and the mean recurrence time of a wet day is $\frac{1}{P(W)}$. From Parzen [18], page 197 the following relationships between precipitation today and precipitation n days ago can be obtained:

$$P(D_t/D_{t-n}) = \frac{P(D/W) + P(W/D)(P(D/D) + P(W/W) - 1)^n}{2 - P(D/D) - P(W/W)}$$

$$P(W_t/D_{t-n}) = 1 - P(D_t/D_{t-n})$$

$$P(W_t/W_{t-n}) = \frac{P(W/D) + P(D/W)(P(D/D) + P(W/W) - 1)^n}{2 - P(D/D) - P(W/W)}$$

$$P(D_t/W_{t-n}) = 1 - P(W_t/W_{t-n})$$

3. ALTERNATIVE MODELS

3.1 Higher-order Markov Chains

Use of a first-order Markov chain model to describe precipitation persistence assumes that if today's weather is conditioned on yesterday's weather, then no further conditioning on past events is necessary. Use of a second-order Markov chain model assumes that if today's weather is conditioned on yesterday's weather and weather two days ago, then no further conditioning on past events is necessary. Thus, a second-order chain contains all the information of a first-order chain plus some additional information. In describing today's weather, a third-order Markov chain would also condition on weather three days ago, a fourth-order Markov chain would also condition on weather four days ago, etc. Thus, if the first-order model does not adequately fit data for a particular area, one might investigate whether a higher-order Markov chain would be an adequate model. The first-order chain has been shown to fit data for many areas, and if the first-order chain fits, it would be preferable over higher-order chains because it contains fewer parameters and is simpler to work with.

Lowry and Guthrie [17] fit models of zero order (binomial), first-order, second-order, and third-order Markov chains to several different sets of data in the Southwest using stationary probabilities throughout a month. They found that in

areas showing too many long sequences of wet days, too many short sequences of wet days, and too few intermediate-length sequences of wet days for a first-order model, that a higher-order Markov chain gave a substantially better fit. Feyerherm and Bark [4] investigated the use of higher-order Markov chains for precipitation data in the midwest. They found for most areas and most times of the year only minor improvement could be gained by using more than a first-order model although there were some cases where the fit may be improved.

Wiser [22] has proposed four modifications to the Markov chain for cases in which the first-order Markov chain does not fit. These modified models fit the data the same as higher-order Markov chains, but specify certain relationships between the parameters which reduce the number of parameters to be estimated and provide models which are easier to work with.

Feyerherm and Bark [5] described a modification of the second-order Markov chain that was found to fit data in April in the midwest. They obtained separate estimates for $P(W_t/W_{t-1}, W_{t-2})$ and $P(W_t/W_{t-1}, D_{t-2})$ as one would for a second-order Markov chain, but found one estimate for $P(W_t/D_{t-1}, D_{t-2}) = P(W_t/D_{t-1}, W_{t-2}) = P(W_t/D_{t-1})$ as one would for a first-order Markov chain. This resulted in a model with three parameters--simpler than the second-order Markov chain and more complex than the first-order Markov chain. The model was specifically

designed to fit data for this particular location and this particular time of year.

3.2 Modified Log Series

Green [10] proposed a model that combines the first-order Markov chain model with a log series model. Before Gabriel and Neumann suggested the first-order Markov chain for modelling precipitation persistence, Williams [21] had proposed a model where the probabilities of runs of 1, 2, ..., r, ... dry (wet) days are estimated by:

$$aq, \frac{aq^2}{2}, \dots, \frac{aq^r}{r}, \dots$$

where a and q are parameters to be estimated. The Markov chain assumes probabilities of runs of 1, 2, ..., r, ... dry (wet) days are of the form:

$$aq, aq^2, \dots, aq^r, \dots$$

Green's model assumed these same probabilities to be of the form:

$$\frac{aq}{1+b}, \frac{aq^2}{2+b}, \dots, \frac{aq^r}{r+b}, \dots$$

where $0 < b < \infty$. This model has added another parameter, b. When $b = 0$, the model is exactly the same as Williams' log series model. As b becomes larger, the model approaches the Markov chain model. Green tested this model for several sets

of data and found that it fit 29 out of 33 sets of data, including some that fit the Markov model, some that fit the log series model and some that fit neither. Thus, this model might be used, but it would be suggested only when the first-order Markov chain model does not fit because: (1) This model contains three parameters and the Markov chain model only two; (2) This model does not easily provide a lot of additional probabilities as the Markov chain model did; and, (3) Estimates of the parameters for this model are more difficult to obtain. Green could not find maximum likelihood estimators for these parameters. Instead, he used an asymptotically equivalent method, the method of minimum chi-squared. This is an iterative procedure done by computer and requiring much more time than the Markov chain estimators.

3.3 Exponential Distribution

Another model suggested by Green [8] begins by assuming that dry spells and wet spells form an alternating renewal process and the lengths of dry spells form an exponential distribution with density function $\alpha e^{-\alpha t}$, while lengths of wet spells form an exponential distribution with density function $\beta e^{-\beta t}$. These spells are studied in continuous time rather than in terms of days as the Markov chain model. A wet spell is a series of all wet "instants". This model is a continuous parameter Markov process. The Markovian dependence is instant-

to-instant, not day-to-day. From this model can be derived properties of wet and dry days where a wet day is defined to be a day that contains at least one wet instant and a dry day is defined to be a day containing no wet instants. It was found that this model led to the conclusion that states of the day approximately form a Markov chain of order one.

All equations of interest are written in terms of α and β :

$$P(d) = P(\text{dry instant}) = \frac{\beta}{\alpha + \beta}$$

$$P(w) = P(\text{wet instant}) = \frac{\alpha}{\alpha + \beta}$$

$$P(D) = P(d)P(\text{a day is dry given that it starts dry}) =$$

$$\frac{\beta}{\alpha + \beta} \int_1^{\infty} \alpha e^{-\alpha t} dt = \frac{\beta e^{-\alpha}}{\alpha + \beta}$$

$$P(W) = 1 - P(D) = \frac{\alpha + \beta - \beta e^{-\alpha}}{\alpha + \beta}.$$

α and β are considered constant over a given time period resulting in $P(D)$ and $P(D/D)$ constant over a given time period. So, this corresponds to the stationary Markov chain model. Green gives some useful equations for probabilities related to wet and dry days:

$$\begin{aligned} Q_n &= P(\text{wet day given a preceding run of } n \text{ wet days}) \\ &= P(W_t / W_{t-1}, W_{t-2}, \dots, W_{t-n}) \\ &= 1 - b + \frac{b-a}{Q_{n-1}}, \quad n=1, 2, \dots \end{aligned} \quad (3.1)$$

$$\begin{aligned}
P_n &= P(\text{wet day given a preceding run of } n \text{ wet days} \\
&\quad \text{preceded by a dry day}) \\
&= P(W_t/W_{t-1}, W_{t-2}, \dots, W_{t-n}, D_{t-n-1}) \\
&= 1 - b + \frac{b-a}{P_{n-1}}, \quad n=1,2,\dots
\end{aligned} \tag{3.2}$$

In these equations $a = \frac{\beta e^{-\alpha} - \beta e^{-2\alpha-\beta}}{\alpha+\beta}$ and $b = e^{-\alpha} - e^{-\alpha-\beta}$.

Equations (3.1) and (3.2) are recursive and Q_0 and P_0 are defined:

$$Q_0 = P(W) = \frac{\alpha+\beta-\beta e^{-\alpha}}{\alpha+\beta} \tag{3.3}$$

$$P_0 = P(W/D) = 1 - \beta e^{-3\alpha-\beta} - e^{-\alpha} + e^{-\alpha-\beta} \tag{3.4}$$

P_n and Q_n are related:

$$P_n = \frac{a}{1-Q_n} - b$$

Green also shows that where c is the positive root of the equation $c = 1 - b + \frac{b-a}{c}$, for $n \geq 0$, both P_n and Q_n alternate about c . They are greater than c for odd n and less than c for even n , and approach c as a limit as n increases. For any n , Q_n is closer to c than P_n .

Using the relations between conditional probabilities in a first-order Markov chain discussed previously, it can be shown:

$$P(D/W) = \frac{(1 - \beta e^{-3\alpha-\beta} - e^{-\alpha} + e^{-\alpha-\beta}) \beta e^{-\alpha}}{\alpha + \beta - \beta e^{-\alpha}}. \tag{3.5}$$

This is an approximate expression since states of days form an approximate Markov chain. In addition one could apply previously discussed properties of sequences of wet and dry days in the Markov chain model to this model, substituting in appropriate probabilities as defined in equations (3.3), (3.4), and (3.5) and get approximate results.

Some additional information can be obtained from this model that cannot be obtained from the Markov Chain model. The mean lengths of dry and wet spells are respectively $\frac{1}{\alpha}$ and $\frac{1}{\beta}$ days and the probability that the last instant of a wet day is wet is $\frac{\alpha}{\alpha+\beta-\beta e^{-\alpha}}$ [8]. Parzen [18] states the following results. If it is wet at time $t = 0$ then the probability of a wet instant at time t is:

$$\frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

and the probability of a dry instant at time t is:

$$\frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

If it is dry at time $t = 0$ then the probabilities of wet and dry instants at time t are, respectively:

$$\frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

and:

$$\frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

Another possible use of this model described by Crovelli [3] is determining expected total amount of precipitation during a given time interval. In order to do this one must define and estimate a precipitation rate in inches per unit time for the wet state. Crovelli also suggests some methods for dealing with a process of more than two states.

Crovelli gave an example where parameters α and β were estimated by a "careful investigation" of five years of hourly precipitation records. Green stated that maximum likelihood estimators for α and β would be difficult to obtain. He instead obtained the relative frequency estimates for $P(W/D)$ and $P(W/W)$, set them equal to the expressions obtained from equations (3.4) and (3.5) and solved for α and β . This gives approximate estimates since states of the day approximately form a first-order Markov chain.

Crovelli [3] noted that Grayman and Eagleson [7] reported that this model had been found to fit five years of Boston rainfall data by use of a curve-fitting procedure. Green fit the model to the Tel Aviv data used by Gabriel and Neumann and found that the significance level of the chi-square test using the exponential model is slightly higher than using the Markov chain model, indicating that Green's model gives a better fit. Green [9] later fit the model to some of the data

Weiss had used. The exponential model gave a satisfactory fit in some cases where the Markov chain model did not.

The exponential model requires $P(W/W) > P(W/D)$. Green did some further investigation and found that the two models in general were more alike when $P(W/W)$ and $P(W/D)$ were nearly equal.

Whether or not the exponential model is preferable to the Markov chain model would depend partly on one's purposes in studying weather patterns. If one is interested only in presence or absence of precipitation on a daily basis and one is not interested in precipitation changes within shorter units of time, the Markov chain model would be adequate. The Markov chain model requires somewhat less calculations. The exponential model allows one to estimate day-to-day probabilities only with the stationary assumption which is somewhat of a simplification and may not give as accurate results as the non-stationary assumption. If the first-order Markov chain model is not found to give a satisfactory fit for a certain area or time of the year, the exponential model may in some cases provide an adequate fit. If a first-order Markov chain does not fit, the exponential model may be preferable to a higher-order Markov chains if both these models fit because the exponential model requires only two parameters.

In conclusion, one can gain much information about precipitation patterns by taking into consideration day-to-day

persistence of precipitation. The first-order two-state Markov chain is a very simple model and has been found to fit data in many areas. Some modifications to this model and some alternatives to the model have also been shown to be of interest.

REFERENCES

- [1] Anderson, T.W. and Goodman, L.A. 1957. Statistical inference about Markov chains. Annals of Mathematical Statistics 28: 89-110.
- [2] Caskey, J.E. 1963. A Markov chain model for the probability of precipitation occurrence in intervals of various lengths. Monthly Weather Review 91: 298-301.
- [3] Crovelli, R.A. 1973. Stochastic models for precipitation. International Symposium on Uncertainties in Hydrologic and Water Resource Systems. American Meteorological Society.
- [4] Feyerherm, A.M. and Bark, L.D. 1965. Statistical methods for persistent precipitation patterns. Journal of Applied Meteorology 5: 606-624.
- [5] Feyerherm, A.M. and Bark, L.D. 1967. Goodness of fit of a Markov chain model for sequences of wet and dry days. Journal of Applied Meteorology 6: 770-773.
- [6] Gabriel, K.R. and Neumann, J. 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society 88: 90-95.
- [7] Grayman, W.M. and Eagleson, P.S. 1969. Streamflow record length for modelling catchment dynamics. Massachusetts Institute of Technology Department of Civil Engineering Hydrodynamics Laboratory Report No. 114.
- [8] Green, J.R. 1964. A model for rainfall occurrence. Journal of the Royal Statistical Society Series B 26: 345-353.
- [9] Green, J.R. 1965. Two probability models for sequences of wet or dry days. Monthly Weather Review 93: 798-801.
- [10] Green, J.R. 1969. A generalized probability model for sequences of wet and dry days. Monthly Weather Review 98: 238-241.
- [11] Gringorten, I.I. 1966. A stochastic model of the frequency and duration of weather events. Journal of Applied Meteorology 5: 606-624.
- [12] Gringorten, I.I. 1970. Modelling conditional probability. Journal of Applied Meteorology 10: 646-657.

- [13] Gringorten, I.I. 1972. Conditional probability for an exact noncategorized initial condition. Monthly Weather Review 100: 796-798.
- [14] Hartley, H.O. 1949. Tests of significance in harmonic analysis. Biometrika 36: 194-201.
- [15] Hopkins, J.W. and Robillard, P. 1964. Some statistics of daily rainfall occurrence for the Canadian prairie provinces. Journal of Applied Meteorology 3: 600-602.
- [16] Klotz, J. 1973. Statistical inference in Bernoulli trials with dependence. Annals of Statistics 1: 373-379.
- [17] Lowry, W.P. and Guthrie, D. 1968. Markov chains of order greater than one. Monthly Weather Review 96: 798-801.
- [18] Parzen, E. 1962. Stochastic Processes. Holden-Day, San Francisco.
- [19] Saucier, W.J., Weber, A.H., and Bayne, C.K. 1973. Precipitation variability over North Carolina. Water Resources Research Institute of the University of North Carolina Report No. 84.
- [20] Weiss, L.L. 1964. Sequences of wet or dry days described by a Markov chain probability model. Monthly Weather Review 92: 169-176.
- [21] Williams, C.B. 1951. Sequences of wet and of dry days considered in relation to the logarithmic series. Quarterly Journal of the Royal Meteorological Society 78: 91-96.
- [22] Wiser, E.H. 1965. Modified Markov probability models of sequences of precipitation events. Monthly Weather Review 93: 511-516.

METHODS FOR MODELLING
PRECIPITATION PERSISTENCE

by

BRENDA ANN WEAK
B.S., Kansas State University, 1970

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1974

Presence or absence of precipitation at one point in time is not independent of the presence or absence of precipitation at earlier times. In developing a statistical model of rainfall occurrence, this persistence should be considered.

The first-order Markov chain model considered on a discrete day-to-day parameter space assumes that if today's precipitation is conditioned on yesterday's precipitation then further conditioning on past events adds no more information. This model has been shown, by means of a chi-square goodness-of-fit test to fit data from many different climates. The maximum likelihood estimators of the probabilities involved in the model have been shown to be simple relative frequencies. Some properties of the estimators have been described and some alternative estimation procedures have been discussed.

For areas or times of the year where there is more persistence in precipitation patterns, a first-order Markov chain will sometimes not be an adequate model, but a higher-order Markov chain will add enough information to provide an adequate model. Another model that has been proposed is based on the log series and this model is a more general model than the first-order Markov chain.

One can also consider not day-to-day changes, but instant-to-instant changes. This is a continuous time Markov process where lengths of dry spells and lengths of wet spells each have exponential distributions.