NOTES AND CORRESPONDENCE

A Comparison of Conditional and Unconditional Probabilities for Wet- and Dry-Day Sequences¹

DAVID M. HERSHFIELD

USDA Hydrograph Laboratory, Beltsville, Md.

17 April 1970 and 30 June 1970

1. Introduction

The unconditional probability of a day with precipitation equal to or greater than a selected threshold, or the complementary probability of a day with less than a particular value, can be computed from climatic tables, i.e., from the aggregated or total occurrence (proportion) data. However, the conditional probabilities of a precipitation day following a precipitation day (or a dry day following a dry day) are rarely available and can only be obtained by direct enumeration of the daily rainfall series or from frequency distributions displaying the length of wet (or dry) sequences.

The utility of these conditional probabilities has been demonstrated in determining the frequency of wet- and dry-day sequences in a number of papers (Feyerherm and Bark, 1965, 1967; Green, 1970; Hershfield, 1970). The conditional probability is the one parameter required for the geometric distribution in the first-order Markov chain probability model of such a sequence (Gabriel and Neumann, 1962). For dry sequences, the geometric distribution is of the form

$$Pr(k) = p^{k-1}(1-p), \quad k=1,2,3,\dots, \quad 0 < p < 1, \quad (1)$$

where k is the length of the sequence in days and p is the conditional probability of a dry day following a dry day.

For wet sequences, the geometric distribution is of the form

$$Pr(m) = q^{m-1}(1-q), \quad m=1,2,3,\dots, \quad 0 < q < 1, \quad (2)$$

where m is the length of the wet-day sequences and q

is the associated conditional probability. Eqs. (1) and (2) are consistent with those of Gabriel and Neumann presented for the wet and dry conditional probabilities.

The purpose of this note, which is an outgrowth of work with the climatological problem of estimating the frequency of wet- and dry-day sequences for agricultural problems, is to provide some relationships, based on observed data, between the conditional and unconditional probabilities.

2. Analysis

A wet sequence is defined as a sequence of one or more days where each day has an amount equal to or above a certain threshold value. Similarly, a dry day has a value less than the threshold. In terms of the relative frequency definition of probability, the unconditional

FIG. 1. Station location map and percentage conditional probabilities $P(W|W)$ for 0.10-inch threshold.

¹ Contribution of the Soil and Water Conservation Research Division, Agricultural Research Service, U. S. Department of Agriculture.

FIG. 2. Relationships between conditional probabilities P_c and unconditional probabilities P_u for four thresholds of wet- and dry-day sequences.

FIG. 3. Nomogram for estimating $P(D|D)$ and $P(W|D)$ from the unconditional probability P_u of a dry day.

probabilities (sometimes referred to as historical or climatological) of the events wet (W) and dry (D) are given by

$$P(W) = \frac{N_W}{N}, \quad (3)$$

$$P(D) = \frac{N_D}{N}, \quad (4)$$

where N is the total size of the sample, N_W the total number of precipitation days, and N_D the number of dry days. On the other hand, the conditional probability of the event W , given that the preceding event was W , is given by

$$P(W|W) = \frac{N_{WW}}{N_W} = p, \quad (5)$$

where N_{WW} denotes the number of times in which a rain day was preceded by a rain day and N_W the total number of rain days. Similarly, the conditional probability for dry days is

$$P(D|D) = \frac{N_{DD}}{N_D} = q. \quad (6)$$

Fig. 1 shows the location of the 33 stations used in the

analysis. The data are for the 10-year period 1951–60. The values of N_W , N_D , N_{WW} and N_{DD} were determined by direct enumeration from which the associated unconditional and conditional probabilities were computed. Fig. 2 displays the relationships between the two types of probabilities for each of four thresholds for wet and dry sequences. A least-squares line and associated correlation coefficient (r) and standard error of estimate (SEE) were computed for each set of data. Of interest is that the range of probabilities in both directions decreases with increasing threshold for both the wet and dry sequences. For the dry sequences, the unconditional probabilities can be used to provide an excellent estimate for the conditional probabilities and this relationship improves with increasing threshold. In fact, at the 0.50-inch threshold, both probabilities may be considered equal in magnitude for all practical purposes.

For the wet sequences, the strength of the relationship between the two types of probabilities decreases with increasing threshold. Even though there are statistically significant relationships (1% level) for all four thresholds, the SEE is so large that an estimate of q from (2) could result in a large error in $Pr(m)$ even for small values of m . Thus, the wet-day aggregative outcome data are not amenable to areal generalization. Reference might then be made to the functional relationships presented by Feyerherm and Bark (1965) to estimate p . Except for Fresno, Calif., the geographical distribution of the conditional probabilities, $P(W|W)$ for the 0.10-inch threshold (Fig. 1), exhibits a pattern quite similar to that for the number of precipitation days above the same threshold. The Fresno value is a function of the seasonal nature of precipitation on the Pacific coast.

The nomogram of Fig. 3 was constructed from the dry-day relationships of Fig. 2. If the samples used in the analysis are considered adequate, then the nomogram can provide a convenient short-cut method for estimating $P(D|D)$ and its complement $P(W|D)$ from readily available published data.

REFERENCES

- Feyerherm, A. M., and Deam Bark, 1965: Statistical methods for precipitation patterns. *J. Appl. Meteor.*, **4**, 320–328.
- , and —, 1967: Goodness of fit of a Markov chain model for sequences of wet and dry days. *J. Appl. Meteor.*, **6**, 770–773.
- Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95.
- Green, J. R., 1970: A generalized probability model for sequences of wet and dry days. *Mon. Wea. Rev.*, **98**, 238–241.
- Hershfield, D. M., 1970: Generalizing dry day frequency data. *J. Amer. Waterworks Assn.*, **62**, 51–54.