

Volume 17

# **HYDROLOGY and WATER RESOURCES in ARIZONA and the SOUTHWEST**

**PROCEEDINGS OF THE 1987 MEETINGS  
OF THE  
ARIZONA SECTION --  
AMERICAN WATER RESOURCES ASSOCIATION,  
HYDROLOGY SECTION --  
ARIZONA-NEVADA ACADEMY OF SCIENCE  
AND THE  
ARIZONA HYDROLOGICAL SOCIETY**

**APRIL 18, 1987, NORTHERN ARIZONA UNIVERSITY  
FLAGSTAFF, ARIZONA**

# **ADAPTABILITY OF A DAILY RAINFALL DISAGGREGATION MODEL TO THE MIDWESTERN UNITED STATES**

**Thomas W. Econopouly, D. R. Davis, and D. A. Woolhiser**

## **INTRODUCTION**

Daily disaggregation of rainfall is a technique used to separate a daily rainfall depth into smaller showers. The showers can then be further disaggregated into intensity patterns, which may be used as input for time varying infiltration models (Woolhiser and Osborn, 1985). A stochastic model for the disaggregation of daily summer rainfall in southeastern Arizona was developed by Hershenhorn (1984). Hershenhorn used data collected at a gage on the Walnut Gulch Watershed. Hershenhorn and Woolhiser (1987) found that this model was applicable for locations up to 75 miles away from the original gage. In this paper, we discuss the applicability of the model to two midwestern locations.

## **DATA SET**

The data used in the study were collected on the Agricultural Research Service's Experimental Watersheds at Hastings, Nebraska and McCredie (now Kingdom City) Missouri, and were obtained on magnetic tape from the U.S.D.A.'s Water Data Laboratory. The period of record for Hastings was 1938 through 1967, and for McCredie was 1941 through 1974. Two periods of the year were investigated - May and June, and July and August.

The climate for both of these locations is considered continental, with an annual precipitation cycle that exhibits a winter minimum and a summer maximum. The maximum amount precipitation occurs during June, and over half of the total annual precipitation occurs during the months of May through September. The early summer maximum is a result of the continental influence resulting in increasing temperatures and advection of moisture from the Gulf of Mexico, coupled with still fairly active spring storms (Trewartha, 1981). Approximately 80 percent of the summer precipitation in the region has been classified as frontal in nature (Rudd, 1961).

## **DAILY DISAGGREGATION MODEL**

The mathematical description of the disaggregation model is described in detail by Hershenhorn (1984) and Hershenhorn and Woolhiser (1987). The model's structure requires a daily precipitation amount as an input. Given daily amounts the model can be used to describe the number of showers within a day, the starting time, the depth and the duration for each shower.

### Complete vs. Partial Showers

Two types of showers are defined - a complete shower is one that begins and ends within the same day, and a partial shower is one that begins on one day and ends on another. Table 1 lists the number of two consecutive days with precipitation and the percentage of those containing a partial event. The rank sum test (Hoel, 1971) was used to determine if the sums or products of the daily depth amounts for consecutive wet days with partials and those without partials were different. At the five percent level, there were no significant differences between the consecutive wet days with partials and those without partials.

Table 1. Number of Two Consecutive Wet Days

Location	Period	No. of Two Consecutive Wet Days	Percentage <sup>1/</sup> with Partial
Hastings	May and June	188	34
Hastings	July and August	107	34
McCredie	May and June	272	24
McCredie	July and August	164	21
Walnut Gulch	July and August	217	17

<sup>1/</sup> Number of consecutive wet days when the shower continued through midnight divided by the total number of consecutive wet days.

Partial showers were separated into two smaller showers which were assigned to the day on which each occurred. For the remainder of this paper a partial shower will be considered as either parts of the shower that crossed midnight. A log likelihood ratio test was performed on the distributions of partial and complete showers depths at Walnut Gulch (Hershenhorn, 1984), it found that the complete and partial depths could be described by the same distribution. The log likelihood ratio test was also performed using the Hastings and McCredie data and it was found that only the McCredie - July and August, complete and partial depths could be described by one distribution. However, the two sample Kolmogorov-Smirnov (KS) test could not distinguish between any of the midwestern complete or partial shower depth distributions. Thus, to make the description of the rainfall process more tractable, the partial and complete depths were assumed to be samples from the same distribution. This assumption was necessary in order to distribute daily depths into multiple shower depths.

### Distribution of Number of Showers Given Daily Precipitation

A joint distribution was required to describe the number of showers in a day, given a daily precipitation amount. The joint

distribution of the number of showers per day, and the daily amount can be written as the product of the conditional and marginal distributions:

$$H_{N,Z}(n,z) = G_{N|Z}(n|z)F_Z(z) \quad (1)$$

where:

N = the number of showers.

Z' = the daily precipitation depth minus a threshold.

Because the lower limit of the daily observation was 0.01 inch, the threshold was set at 0.009 inch. The threshold was set at this value so all small depth values would remain in the data set.

In this study, the marginal distribution of daily rainfall was developed so that the goodness of fit of the conditional distribution could be tested. When the overall model is used, the daily depth will be obtained from historical data or perhaps a climate generating model. The marginal distribution chosen for test purposes was the Mixed Exponential distribution (Smith and Schreiber, 1974) which density has the form:

$$f_Z(z) = \frac{\alpha}{\theta_1} \exp(-z/\theta_1) + \frac{(1-\alpha)}{\theta_2} \exp(-z/\theta_2) \quad (2)$$

where:

$\alpha$  - is a weighting parameter ( $0 \leq \alpha \leq 1$ )

$\theta_1$  and  $\theta_2$  are parameters

Hershenthorn (1984) found that the Shifted Negative Binomial (SNB) distribution provided a good fit for the conditional distribution of the number of showers given daily rainfall. The probability mass function of the SNB is written as:

$$P(N=n) = \binom{n+r-2}{n-1} p^r (1-p)^{n-1}; n=1,2,\dots \quad (3)$$

Variables p and r were allowed to vary with daily depth. The Walnut Gulch data indicated that the number of showers given Z, asymptotically approached a limiting value. However, the McCredie data suggested that the expected number of showers asymptotically approached a straight line with a positive slope (figure 1), and thus a new functional form was developed for p and r. This new functional form accounts for this factor and includes the Walnut Gulch function as a special case. The functional form for p and r is now:

$$p = \exp(-A1 \times Z) \quad (4)$$

$$r = (E - 1.0) \times p / (1.0 - p) \quad (5)$$

$$E = A2 + A3 \times Z + (1.0 - A2) \times \exp(-A4 \times Z) \quad (6)$$

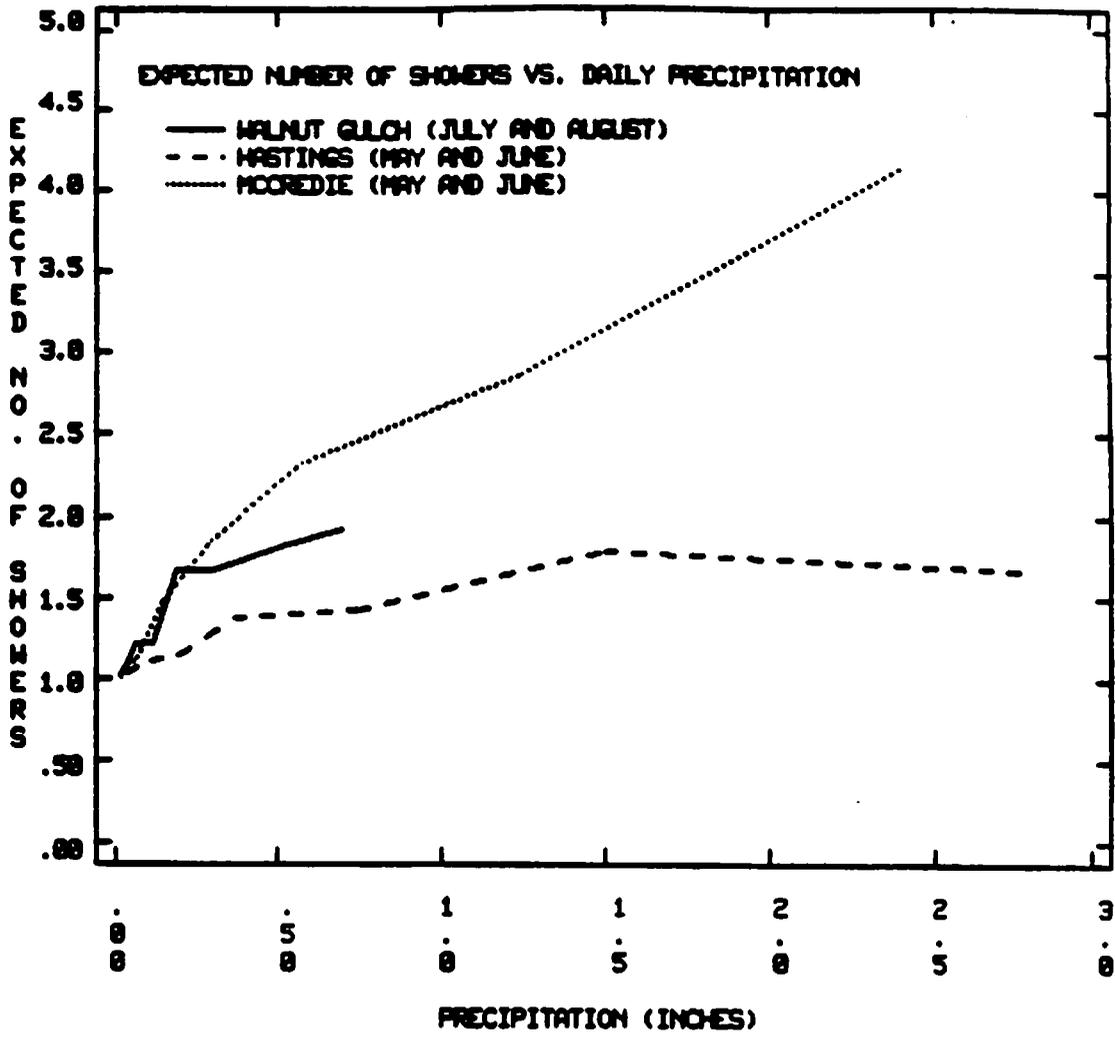


Figure 1. Expected Number of Showers vs. Daily Precipitation.

where:

E - is the expected amount of showers given a daily precipitation amount.

Z - is daily precipitation depth.

A1, A2, A3 and A4 are fitted parameters.

Parameter values were obtained for the midwestern data by numerical maximum likelihood techniques. The null hypothesis that the sample data were taken from the identified bivariate distribution could not be rejected for the McCredie data sets, according to a Chi-Squared goodness of fit test, at the 1 percent level.

### Individual Shower Depths

Given a daily rainfall depth, Z, and the number of showers per day, N, individual shower amounts,  $Y_1, Y_2, \dots, Y_N$  need to be determined. The summation of all the individual depths which occur within a day must equal that day's rainfall depth. The ratio technique developed for the Walnut Gulch data (Hershenthorn, 1984), also represented the midwestern locations. The technique used ratios developed from shower amounts and daily depth. The ratios were then fitted to either a Beta Fourier distribution or a Uniform distribution. The ratio technique was used to describe up to 6 showers within a day at Walnut Gulch. The McCredie data had up to 13 showers per day.

### Shower Starting Times

A partial shower was defined as one that ends or begins at midnight, so once the durations had been determined, the time of the start of a shower was determined. Hershenthorn (1984) used the Mixed Beta distribution to describe normalized starting times for the complete showers. The starting times were normalized by defining the starting times as time (in hours) from midnight and dividing by 24. The Mixed Beta distribution can be written as:

$$f_{Tc}(t) = w \left[ \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} t^{\alpha_1-1} (1-t)^{\beta_1-1} \right] + (1-w) \left[ \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} t^{\alpha_2-1} (1-t)^{\beta_2-1} \right] \quad (7)$$

where:

$\Gamma$  - is the gamma function

w - is a weighting parameter ( $0 \leq w \leq 1$ )

$\alpha_1, \beta_1, \alpha_2,$  and  $\beta_2$  are parameters

The Mixed Beta distribution has some undesirable features. The distribution is not necessarily periodic. However, the process that we wish to describe is periodic. The distribution can not be explicitly integrated. If numerically integrated, care must be taken at the tails where the Mixed Beta distribution may take on infinite values. To overcome the problems with the Mixed Beta, a Fourier distribution with a mean of one and two harmonics was tested to determine if it could replace the Mixed Beta. The Fourier distribution is periodic and can be explicitly integrated. The Fourier distribution can be written as:

$$f_{Tc}(t) = 1.0 + \alpha \cos(2\pi t + \beta) + \delta \cos(4\pi t + \rho) \quad (8)$$

where:

- $\alpha$  - the first amplitude.
- $\beta$  - the first phase angle.
- $\delta$  - the second amplitude
- $\rho$  - the second phase angle.

Parameter values for the Mixed Beta distribution and the Fourier distribution were obtained by using numerical maximum likelihood techniques for the individual data sets. The data sets were divided into subclasses, which were grouped according to the number of showers per day, up to 6 showers per day. All days with 6 or more showers were grouped together. One-sample KS tests were performed to determine whether the historical cumulative distribution functions (CDF's) were significantly different from the CDF's calculated from the parameter values optimized over the entire data set. At the one percent level, all the historical CDF's for each subclass were not significantly different from the Fourier CDF's. The Hastings, July and August - 1 shower per day data was the only subclass that was significantly different, at the one percent level, from the Mixed Beta CDF's.

The hypothesis of the independence of multiple starting times on a day with more than one shower, and whether they can be described by order statistics was tested for both the Mixed Beta and the Fourier distributions. Let  $F_{Tr}(t)$  represent the CDF of the starting time of the  $r^{\text{th}}$  shower on a day in which  $n$  showers occur (Kendall and Stuart, 1977). The general form is:

$$F_{Tr}(t) = \frac{n!}{(r-1)! (n-r)!} \left[ \sum_{i=0}^{n-r} \frac{\binom{n-r}{i} (-1)^i}{i+r} (F_{Tc})^{i+r} \right] \quad (9)$$

where  $F_{Tc}$  is the CDF of all shower starting times as given by the CDF of the Mixed Beta or Fourier distribution.

The empirical distribution functions were compared with the theoretical distributions given the assumption of independence. The one-sample KS test was used to test the hypothesis that the shower starting

times were independent. When the order statistics were calculated from the Mixed Beta distribution, there were 5 instances out of 36 when the distributions were considered significantly different at the one percent level. Tests with the Fourier distribution revealed 3 instances out of 36 where the distributions were significantly different.

The Mixed Beta and the Fourier distribution appear to fit the data equally well. Therefore, because of its favorable properties previously mentioned, the Fourier distribution was chosen as the distribution to use to describe the starting times of complete showers.

**Starting times for one shower per day.** To reflect the relation between depth and starting time, the shower starting times were conditioned upon depth. However, we will describe only one shower per day in this manner because the statistical description of more than one shower per day becomes intractable. The joint distribution of shower depth and time of day may be written as:

$$f_{X,Tc}(x,t) = f_{X/Tc}(x/t) g_{Tc}(t) \quad (10)$$

where:

$f_{X/Tc}(x/t)$  is the conditional distribution of depth given time of day.

$g_{Tc}(t)$  the marginal distribution of starting times - the Fourier distribution.

The following marginal distribution for depth was obtained by integrating the joint distribution over time:

$$g_X(x) = \int g_{Tc}(u) f_{X/Tc}(x/u) du \quad (11)$$

From Bayes' theorem the conditional distribution of time given depth:

$$f_{Tc/X}(t,x) = [g_{Tc}(t) f_{X/Tc}(X/Tc)] / g_X(x) \quad (12)$$

The cumulative probability of the conditional time given depth is:

$$P(T < t / x) = \int_0^t f_{Tc/X}(u/x) du \quad (13)$$

Preliminary investigations revealed that the Exponential distribution may be used to describe the distribution of shower depths within two hour intervals. Therefore, an Exponential distribution with a time varying parameter was chosen for the conditional distribution of depth of a shower given time of day. The conditional distribution may be written as:

$$f_{X/Tc}(x/t) = 1.0/l(t) \left[ \exp(-x/l(t)) \right] \quad (14)$$

where the mean  $l(t)$ , is represented by a fourier series with two terms:

$$l(t) = A + B \cos(2\pi x + C) + D \cos(4\pi x + E) \quad (15)$$

The parameters A, B, C, D and E are obtained by numerical optimization using maximum likelihood techniques. To improve the fit of this distribution to the data set and to reduce the number of showers to be simulated, only showers greater than 0.10 of an inch were included.

The Chi-Square goodness of fit test was used to determine whether the theoretical joint distribution was significantly different from the historical distribution of shower depths and starting times. At the one percent significance level, the only data set that was considered significantly different is the July and August data set from Hastings. Inspection of the cumulative plot of starting times (figure 2) for the one per day showers reveals a high frequency of points at 6:00 am. The occurrence of this many points at one time probably is an artifact due to the techniques used to convert analog rainfall data into a digital form. The number of points at 6:00 am made it difficult to fit a curve through these data.

#### Joint Distribution of Shower Duration and Amount

Shower durations were determined from the corresponding shower depth. The joint distribution of shower duration and amount may be written as:

$$h_{y, D}(y, d) = g_{D/Y}(d/y) f_Y(y) \quad (16)$$

Where d is the shower duration and y is the shower depth minus a threshold amount. The shower depth is obtained from the storm ratio technique.

Hershenthorn (1984) used a bivariate log-normal distribution for  $h_{y, D}(y, d)$ . Different functional relationships between depth and duration were tested for linear dependency for the complete and partial showers, individually. The best linear relationship obtained for the complete showers, with a threshold equal to 0.009, for three of the four midwestern data sets, was when the duration was transformed to its natural log and the depth was raised to the one-third power. However, the hypothesis that the residuals came from a normal distribution with a mean equal to zero and a standard deviation equal to the standard error of estimate could not be accepted by the Chi-Squared test at the one percent significance level, for half of the data sets.

The threshold for the complete showers was increased to 0.099 inches, to reduce the number of data points, and to facilitate the fit of a regression line. In future simulations, showers of less than 0.10 inch will be treated in a simpler fashion than showers greater than or equal to 0.10 inch. The highest coefficient of determination for two of the midwestern data sets resulted when the durations were transformed to their natural logs and the depths were not transformed. The remaining data sets highest coefficient of determination resulted when the duration was transformed to its natural log and the depth was transformed to its square root.

The hypothesis of linearity of regression was tested by using a correlation ratio test (Kendall and Stuart, 1977). When all the four

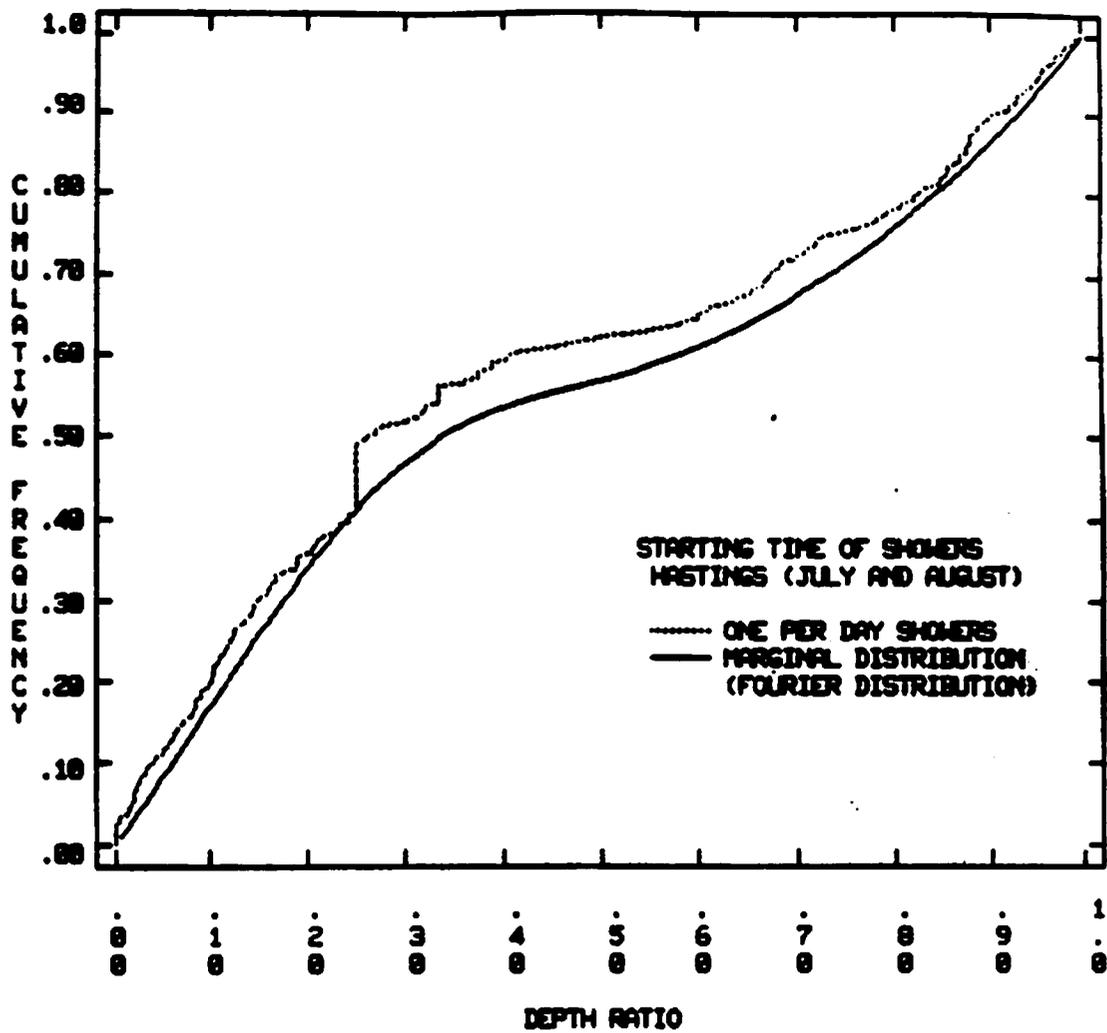


Figure 2. Starting Time of Showers for Hastings (July and August).

data sets' durations were transformed to their natural logs and the depths were allowed to remain the same, all four transformed data sets passed the linearity test at the five percent level. When the data sets' durations were transformed to their natural logs and the depths were transformed to their square, three of the data sets passed the test at the five percent level, and the Hastings, May and June data passed at the one percent level.

Testing was also performed to determine if the residuals from the regression were from a normal distribution with a mean of zero and a standard deviation equal to the standard error of estimate. To determine if the standard deviation of the residuals was constant the Chi-Square test was performed on four subclasses of each data set. The subclasses were set up so each subclass had the same sample size. Four of the total of sixteen subclasses' residuals could not be considered from a normal distribution, at the one percent level, for the log duration vs. depth relationship. At the one percent level, four of the sixteen subclasses' residuals could not be considered from a normal distribution for the log duration vs square root of depth relationship.

The residuals for all of the four midwestern data were also tested for normality without breaking up the data sets into subclasses. At the one percent level, just the McCredie July and August data set's residuals could not be considered as from a normal distribution for the log duration vs depth relationship. Only the Hastings, May and June residuals did not pass the significance testing when the log duration vs square root of depth relationship was examined.

The log duration vs. depth relationship used for the complete data sets was tested for linearity and normality, with the partial data sets. The threshold was set to 0.009 inch. At the five percent level, the data sets transformed in this manner passed all the tests.

## CONCLUSIONS

The daily disaggregation technique developed by Hershenhorn (1984) from summer precipitation data collected in southeastern, Arizona required slight modification to be used to describe spring and summer precipitation for two midwestern stations.

Two of the modifications which were needed improved the tractability of the model. The Fourier distribution was used to replace the Mixed Beta distribution for the description of the starting times of the showers. The replacement made the description of the starting times more theoretically correct because the Fourier distribution is periodic where the Mixed Beta is not. The Fourier density distribution is also much easier to integrate for its use in simulations. This replacement may be suitable for the Arizona data as well. A change was made in the functional forms of the  $p$  and  $r$  values of the Shifted Negative Binomial distribution, which improved the old functional form by allowing the parameters to take additional functional forms.

The duration vs. depth relationship which resulted in the greatest linear dependency was different at the midwestern locations from that

in Arizona. The residuals of the complete shower depths minus the threshold of 0.009 inch were not normally distributed about the regression line of the greatest linear dependency, thus the threshold for the complete showers had to be increased to 0.099 inch. This change of threshold should not reduce the applicability of the model since precipitation intensities from showers of less than 0.10 inch are rarely needed in models which use time varying infiltration techniques.

The disaggregation technique was improved by conditioning the starting times of the one shower per day on the depth of the shower. This allows the daily disaggregation to partially describe the diurnal fluctuation of depth. This addition may also be used to describe the diurnal fluctuation of depth in southeastern Arizona.

#### REFERENCES CITED

- Hershenhorn, J.S. 1984. Stochastic Modeling of Daily Rainfall Sequences. M.S. Thesis, University of Arizona.
- Hershenhorn, J.S. and D.A. Woolhiser, 1987 (In Press). Disaggregation of Daily Rainfall. Journal of Hydrology.
- Hoel, P.G. 1971. Introduction to Mathematical Statistics. John Wiley and Sons, Inc., New York.
- Kendall, M., and Stuart, A., 1977. The Advanced Theory of Statistics, Vol. 2. Charles Griffin and Company, Ltd., London.
- Rudd, R.D. 1961. Summer Frontal Precipitation in the United States Area of Daf Climate. J. Geophys. Res., 66: 125-130.
- Smith, R.E., and Schreiber, H.A. 1974. Point Processes of Seasonal Thunderstorm Rainfall. 2. Rainfall Depth Probabilities. Water Resources Research, Vol. 10, No. 3, pp. 418-428.
- Trewartha, G.T. 1981. The Earth's Problem Climates, 2nd Edition. The University of Wisconsin Press, Madison, 1981.
- Woolhiser, D.A. and Osborn, H.B. 1985. A Stochastic Model of Dimensionless Thunderstorm Rainfall, Water Resour. Res. 21:511-522.