OBJECTIVE AND SUBJECTIVE ANALYSIS OF TRANSITION PROBABILITIES OF MONTHLY FLOW ON AN EPHEMERAL STREAM

William Dvoranchik[a]

Lucien Duckstein[a]

Chester C. Kisiel[b]

INTRODUCTION

In order to generate synthetic sequences of streamflows on aridland channels for subsequent evaluation of proposed water resource projects, it is necessary to study the structure of historical sequences of flow. Given the 58-record of monthly flows on Rillito Creek at Tucson, Arizona, the objective of the research was to determine its stochastic properties so as to develop a model of the historic flows. The stochastic properties of interest included (a) statistical moments (arithmetic mean, variance and skewness) over each of the twelve months over the period of record, (b) frequency of dryness (number of months with zero flows), (c) crosscorrelation between flows over adjacent months, and (d) state transition probabilities for flows in a specified range in one month given that flows were in a specified range in a previous month. This study was intent on a critical evaluation of the value of the lag-one autoregressive model (Fiering, 1967) for modeling of ephemeral flow. The findings of this previously unreported study led to consideration of event-based models (Kisiel, et al., 1971; Denny, et al., 1971).

[a]Department of Systems and Industrial Engineering, University of Arizona

[b]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona.

The complete time series of observed flows on any stream has many properties of interest both to scientists, engineers and managers. Our particular focus here is geared to ephemeral flow of semi-arid regions. Rillito flows have been studied in our research project for the purpose of developing methodology for engineering purposes and for scientific description. The strategy involves an interplay of data and models. Previous studies on the Rillito have focused on the annual flood maxima (Davis and Dvoranchik, 1971; Davis, et al., 1972) and on the individual flow events as an alternating wet-dry sequence (Baran, et al., 1971; Denny, et.al, 1971). In this paper, we focus on the monthly flow sequence.

STOCHASTIC PROPERTIES OF THE MONTHLY FLOWS

The utility of the methods of time series analyses of streamflow has been amply demonstrated for channels with perennial flows (Fiering, 1967). However, when the time series contains zero flow over 53% of the months in the record (as is the case for Rillito Creek - see Table 1) serious questions arise about the interpretation of autocorrelation plots (correlograms) and variance spectra for such sequences. Crosscorrelations computed for streamflows in adjacent months (as shown in Table 2) do not reflect the storage capabilities of the basin because there is no base flow. The only explanation for such persistence in flows would be in the persistence of regional meteorological conditions. Even though this factor dominates flows in more temperate climates, it seems more natural to model aridland streamflows in terms of alternating sequences of wet and dry periods and to include in the model a provision for meteorological transitions.

Tables 1 and 2 summarize the results of a statistical study of total monthly streamflow on Rillito Creek for the period 1909 to 1966. A bimodal population of flow and no-flow periods is clearly indicated in the results. But the 53% of the

TABLE 1.--STATISTICS OF TOTAL MONTHLY STREAMFLOW[a]

Month	Mean[b]	Variance	Standard Deviation	Coefficient of Variation	Skewness	Zero Flows	Percent Zero Flows
Oct.	2.54	7.32	2.71	1.07	7.54	38	0.66
Nov.	8.20	188.3	13.72	1.67	7.53	41	0.71
Dec.	115.80	63,800.	252.60	2.18	7.83	40	0.69
Jan.	54.32	17,080.	130.70	2.41	7.54	31	0.53
Feb.	54.26	9,080.	98.03	1.81	7.67	31	0.53
Mar.	31.62	1,910.	43.71	1.38	7.54	25	0.43
Apr.	4.46	36.91	6.08	1.36	7.52	44	0.76
May	4.16	65.53	8.10	1.95	7.51	54	0.93
June	3.11	12.46	3.53	1.14	7.51	47	0.81
July	36.80	8,072.	89.84	2.44	7.54	6	0.10
Aug.	45.25	2,573.	50.73	1.12	7.53	4	0.07
Sept.	22.29	2,519.	50.19	2.25	7.56	12	0.21

[a]Excluding zero flows.

[b]In acre-feet (AF)

[c]There were 373 months with zero flows over a total of 696.0 data points. The percentage of zero flows over this record is 0.53.

TABLE 2.--MONTHLY FLOWS ON RILLITO CREEK (1909-1966) CROSS CORRELATION ON MONTHS[a]

	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.
Oct.	1.00	.25	.39	.27	.33	.28	.37	-.06	-.13	-.11	-.24	-.14
Nov.	.25	1.00	.31	.20	.56	.30	.47	-.04	-.09	-.05	-.17	-.11
Dec.	.39	.31	1.00	.40	.88	.55	.62	-.03	-.08	-.08	-.14	.04
Jan.	.27	.20	.40	1.00	.45	.40	.37	-.04	-.08	-.08	.00	-.05
Feb.	.33	.56	.88	.45	1.00	.53	.69	-.05	-.10	-.09	-.10	-.01
Mar.	.28	.30	.55	.40	.53	1.00	.46	.40	.16	-.08	-.12	-.10
Apr.	.37	.47	.62	.37	.69	.46	1.00	-.04	-.10	.21	-.12	-.09
May	-.06	-.04	-.03	-.04	-.05	.40	-.04	1.00	.52	-.05	-.11	-.05
June	-.13	-.09	-.08	-.08	-.10	.16	-.10	.52	1.00	.01	-.02	-.05
July	-.11	-.05	-.08	-.08	-.09	-.08	.21	-.05	.01	1.00	.50	-.00
Aug.	-.24	-.17	-.14	-.10	-.10	-.12	-.12	-.11	-.02	.50	1.00	.26
Sept.	-.14	-.11	-.04	-.05	-.01	-.10	-.19	-.05	-.05	-.00	.26	1.00

[a]Confidence limits on the cross correlation coefficients were based on a two-tailed test at the five percent significance level:

Lower limit of confidence band is approximately -0.28
Upper limit of confidence band is approximately +0.28

months in which no flow occurred is deceptive because in many of the months the flows were recorded only a few days within a month. For example, in the month of December over the period of record, the mean flow was 26 cfs but 70% of the variance of the monthly flows came from one average value of 976 cfs. This condition repeats itself in many other streamflow records and provides a strong argument for stochastic models that preserve the timing of flows within the month.

The coefficients of variation noted in Table 1 are quite typical of aridland conditions. The skewness reflects the wide scatter in observed flows in each month. For these reasons, no attempt was made to fit probability distributions to each of the 12 sets of monthly flows.

Table 2 gives the crosscorrelation matrix for each month with respect to the remaining 11 months. Of significance in this table is the substantial positive crosscorrelation between the set of months from October to May. These strong seasonal linkages and the extreme variability of monthly flow totals are important features to be preserved in any proposed stochastic model.

Upon exclusion of the zero flows (373 months), the remaining set of months with flows (323) were taken as a group, irrespective of seasonal pattern to determine if a negative exponential probability distribution were a suitable hypothesis to describe the average monthly flows. By means of the chi-squared test it was inferred that the exponential distribution could be not accepted because of the large number of extreme events in the tails of the empirical frequency distribution. Thus, the designation of a month as a suitable time basis for massaging of hydrologic data is arbitrary and can lead to misleading results at least in ephemeral streams. As in many cases, the use to which the results are to be put will determine whether averages over large time intervals (month or year) are to be preferred to measures of variability and temporal structure over shorter time intervals.

Confidence bands on mean and variance of monthly flows

Because statistics, like the mean $\bar{x}$ and variance s^2, based on record length n, are estimates of their population values, μ and σ^2, respectively, it is desirable to localize or circumscribe the variability of these estimates. The variance of $\bar{x}$, (Var x), and of s^2, (Var s^2), must be determined by assuming an underlying population from which samples have been collected. On repetitive sampling of this population, other sets of 58 monthly flows could be obtained and new values of $\bar{x}$ and s^2 found. Because such sampling in nature is not possible, a theoretical or a computer simulation approach is necessary to find the sampling distributions of $\bar{x}$ and s^2. In what follows we assume a normal distribution $N(\mu,\sigma^2)$ for the population even though monthly flows are usually skewed and assume independence of a month's flow from year to year. The latter assumption is easy to justify in terms of zero serial correlation between monthly flows, whereas normality is difficult to justify except for reasons of theoretical convenience. At least, the resulting limits of confidence in estimates of $\bar{x}$ and s^2 from "short" records give a crude measure of parameter variability and serve as a cautionary warning when making economic analyses based on the single estimates from the historical record.

The two-sided confidence region for the mean μ may be constructed from (Lindgren, 1968):

$$\bar{x} - \frac{\sigma}{n^{1/2}} z_{1-\frac{\alpha}{2}} < \mu < \bar{x} + \frac{\sigma}{n^{1/2}} z_{1-\frac{\alpha}{2}} \qquad (1)$$

in which z is the standard normal variate, N(0,1), with zero mean and unit variance at a significance level of α (usually taken as 0.05). The two-sided confidence region for the variance σ^2 may be constructed from (Lindgren, 1968):

$$\frac{n\,s^2}{\chi^2_\alpha} > \sigma^2 > \frac{n\,s^2}{\chi^2_{1-\alpha}} \qquad (2)$$

In inequality (1) no assumption of normality is necessary and because of the central limit theorem the Var $\bar{x}$ = Var x/n for mutually independent random variables. But in inequality (2) for sample variances the "central limit theorem effect" does not hold. In both inequalities (1) and (2), the confidence regions should get tighter (smaller) around the true mean μ and variance σ^2, respectively, as n becomes large provided the assumptions hold approximately. The rate of convergence is slower, however, for the bounds on σ^2.

Concerning confidence limits for the skewness shown in Table 1, Kendall (1947) illustrates how the sampling variance of a parameter (moment in this case) depends on a moment that has twice the order of moment of interest. Hence, calculations of sampling variance of mean, variance, and skew depend, respectively, on the 2nd, 4th and 6th order moments. Because estimates of 6th order moments are notoriously weak for available sample sizes in hydrology, no effort was made to construct confidence regions for the skew. The available confidence regions for the skew are based on $N(\mu,\sigma^2)$ populations; for gamma (Pearson Type III), log-gamma, and other skew populations computer simulation experiments would be necessary to get the form of the sampling distribution for the sample moments.

For Rillito Creek monthly flows Table 3 gives the separate confidence regions for monthly means and variances. For example, in the month of December, the 95% confidence region on the mean is (-2.572 < 35.950 < 74.471) and on the variance it is (15,782 < 21, 949 < 33,298). The region is not tight and so not as much confidence can be placed in $\bar{x}$ and s^2 as would be desirable. It is well known that a future 58-year record (in which any design or plan would have to function) would

Table 3. Separate Confidence Regions on Monthly Means and Variances on Rillito Creek (Period of 1909-1966)

			Confidence Limits (95%) - may be rounded off to 2-3 digits.			
			On μ Alone		On σ^2 Alone	
MONTH	MEAN(cfs)	VARIANCE	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND
October	.887	3.9	.362	1.392	2.8	6.0
November	2.403	67.0	.275	4.533	48.2	101.7
December	35.950	21,949.9	-2.572	74.471	15,982.1	33,298.2
January	25.286	8,538.4	1.238	49.331	6,139.1	12,952.8
February	25.256	5,128.9	6.637	43.876	3,687.7	7,780.5
March	17.993	1,322.9	8.540	27.446	950.5	2,005.5
April	1.076	12.1	.170	1.981	8.7	18.4
May	.287	4.6	-.270	.843	3.3	7.0
June	.590	3.7	.090	1.091	2.7	5.6
July	32.997	7,349.9	10.706	55.287	5,277.5	11,149.8
August	42.127	2,526.5	29,059	55.186	1,816.6	3,832.9
September	17.681	2,071.5	5,849	29.514	1,489.4	3,142.5

not have the same statistical properties as the past record. This fact is an important justification for the emergence of stochastic hydrology. Note that the negative value of the lower bound on the mean is a consequence of skewness in the original data and emphasizes an inherent weakness about reasoning in terms of variances of skew populations.

Simultaneous confidence regions may be constructed by assuming independence of $\bar{x}$ and s^2. This is justified when the underlying population is $N(\mu,\sigma^2)$. In this case the joint probability distribution of $\bar{x}$ and s^2 is described by a normal-chi square distribution (in reality the normal-gamma distribution NG because the chi-square belongs to the gamma family). The NG distribution may give tighter bounds than those shown in Table 3 because of the additional prior information introduced on the joint distribution (provided the assumptions hold). Table 4 gives the simultaneous confidence regions for μ based on the lower and upper values of σ^2 given in Table 3. The mean discharge given in the second column of Table 4 may be inserted between the lower and upper bounds in each of the three cases to evaluate the spread for both independent and simultaneous confidence regions. The significance of the spread may be judged in terms of the cost of one acre-foot of water for agricultural purposes ($10/AF say) or for municipal purposes ($100/AF). Such figures may be used to assess the costs of errors of overestimation and underestimation of mean monthly flows.

STATE TRANSITION PROBABILITIES

Modeling the transition from one flow range to another flow range has been in terms of (a) use of the serial correlation or crosscorrelation of flows at adjacent time periods, or (b) use of the probability distribution that describes the transition from one state of flow to another state.

Table 4. Simultaneous Confidence Regions for Monthly Means Based on Variance Bounds shown in Table 3

				95% Confidence bounds on μ and σ^2 are taken together; numbers may be rounded off to 2-3 digits.			
		95% Confidence bounds on μ alone		Based on lower bound on σ^2		Based on upper bound on σ^2	
Month	Mean* (cfs)	Lower**	Upper	Lower**	Upper	Lower**	Upper
October	.887	.362	1.392	.243	1.314	.440	1.511
November	2.403	.275	4.533	-.217	4.204	.602	5.023
December	35.950	-2.572	74.471	-11.470	68.580	3.320	83.370
January	25.286	1.238	49.331	-4.254	45.646	4.926	54.826
February	25.256	6.637	43.876	2.376	41.016	9.496	48.136
March	17.993	8.540	27.446	6.373	16.003	9.983	19.613
April	1.076	.170	1.981	-.027	1.843	.309	2.179
May	.287	-.270	.843	-.398	.759	-.185	.972
June	.590	.090	1.091	-.111	1.014	.166	1.291
July	32.997	10.706	55.287	5.537	51.267	14.727	60.457
August	42.127	29.059	55.196	26.037	53.147	31.107	58.217
September	17.681	5.847	29.514	3.121	27.661	7.701	32.241

* The only justification for the number of significant figures in the mean monthly discharge rate is to permit more accurate conversion to acre-foot of total flow within the month.

**Negative lower confidence bounds arise because of the normal assumption made on the distribution of the means. In a sense, negative values suggest a major weakness of the notion of confidence bounds.

The lag-one autoregressive model (Fiering, 1967) :

$$x_i - \overline{x} = r_1 (x_{i-1} - \overline{x}) + \varepsilon_i \quad (3)$$

exemplifies the first approach; in eq.(3) x_i and x_{i-1} are the flows in adjacent time periods i and i-1, r_1 is the lag-one serial correlation coefficient, and ε_{i-1} is a random component that allows for preservation of the variance of x_i or of both the variance and skewness (g_1). Then, either a N(0,1) distribution or a gamma distribution, G(0,1,g) is sampled to obtain E_i values at each time i. Eq. (3) is not proposed as a model for ephemeral flows but is presented here and later only to illustrate the conceptual relationship between deterministic and stochastic concepts.

The second approach is more general in that one is willing to assume more about the overall properties of the stochastic process. The probability distribution of a state transition may be postulated outright on the basis of prior knowledge or may be determined from empirical analysis of actual data. The latter direction requires large sample sizes as will be shown shortly for monthly flows on Rillito Creek. In the former case, it is common in hydrology to assume a bivariate normal distribution for the flow x_i or for some transform like the logarithm (log x_i); the use of logarithms is, however, fraught with difficulties when the moment properties of the original x_i's are compared with those moment properties derived from the inverse transforms of log x_i obtained from Eq. (3). This arises because the logarithmic transform does not induce a normal distribution for most hydrologic variables (except perhaps for annual flood extremes, mean annual flow, or precipitation extremes in some cases). It is for these reasons and others that we chose to pursue our analysis in the "natural" space of the flows rather than in the logarithmic space.

The transition from one state (flow) at time t, S(t), to another state at time (t+1), S(t+1), may be generally written as

$$S(t+1) = F\,[S(t),\, X(t)\,] \tag{4}$$

wherein F is a state transition function that may be deterministic or stochastic and X(t) is an input disturbance at time t. For example, in eq.(3), if $\varepsilon_i = 0$, then F is deterministic and the transition from x_{i-1} to x_i is given by the linear regression equation:

$$x_i - \bar{x} = r_1(x_{i-1} - \bar{x}) \tag{5}$$

wherein r_1 is simply the least squares estimate of the slope of a straight line. Eq.(5) is the discrete form of the classical hydrograph recession equation:

$$x(t) = x(o)\,\exp\,(-Kt) \tag{6}$$

in continuous time wherein x(o) is the initial flow rate for the recession and K is the recession or decay constant. Dividing eq.(6) by $x(t-1) = x(o)\,\exp\,(-K(t-1))$ we obtain:

$$x(t) = x(t-1)e^{-K} \tag{7}$$

from which we find $r_1 = e^{-K}$ or $\ln(1/r_1) = K$, thus permitting us to compute easily the recession constant K from the serial correlation coefficient r_1. In this manner, we see an interconnection between deterministic and stochastic concepts; the transition from x(t-1) to x(t) is deterministic at any one time instant t but the overall pattern is stochastic because x(t) varies stochastically.

An example of a probabilistic state transition F is given by a Markov chain:

$$P\,(\,x_i \mid x_{i-1}) = P\,(\,x_i \mid x_{i-1},\, x_{i-2},\, \ldots) \tag{8}$$

in which the probability $P(x_i|x_{i-1})$ tells us that the probability of being in state x_i at time i depends only on the state at time i-1 and not on the infinite past. The generality of eq.(8) is not well understood because too often it is misrepresented as eq.(3). In fact, eq.(8) encompasses nonlinear state transitions in contrast to that given by eqs.(3) and (5). The Markov condition encompasses both linear and nonlinear differential equations that might be used as models for F [S(t), X(t)]. A further discussion of these issues is given by Denny, et al., (1971) and also by Kisiel and Duckstein (1971) who show how nonlinear behavior of the hydrograph affects the predictions given by eq.(3).

If prior information justifies use of a bivariate normal distribution, then eq.(8) may be written as a joint density function:

$$p(x_i, x_{i-1}) = \frac{1}{2\pi\, \sigma_{x_i}\, \sigma_{x_{i-1}} (1-r_1{}^2)^{1/2}} \exp\left(- \frac{z_{x_i}{}^2 - 2r_1\, z_{x_i}\, z_{x_{i-1}} + z_{x_{i-1}}{}^2}{2(1-r_1{}^2)}\right) \tag{9}$$

in which

$$z_{x_i} = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}} \tag{10}$$

σ_{x_i} and $\sigma_{x_{i-1}}$ are the standard deviations of x_i and x_{i-1}, respectively, and μ_{x_i} and $\mu_{x_{i-1}}$ are the means of x_i and x_{i-1}, respectively. If the stochastic process $\{x_i;\ i=1,2,\ldots\}$ is stationary, then $\mu_{x_i} = \mu_{x_{i-1}}$ and $\sigma_{x_i} = \sigma_{x_{i-1}}$, and r_1 is constant over any pair of adjacent time intervals. Thus, we have given the necessary assumptions that justify computation of $\bar{x}$ and s as estimates μ_x and s_x and of r_1 over a sample of size n. If eq.(9) is valid as a model for probabilistic state transitions, then either tables of correlated normal variates or a computer program may be used to simulate a process like that given by eq.(3).

Empirical evaluation of state transition probabilities

State transition probabilities for monthly flows on the Rillito were evaluated for transitions from a state of zero flow in one month to zero flow in the following month or to one of ten other bounded flow intervals. This was done by counting the number of times a monthly flow satisfied these conditions. The process was repeated for transitions from finite flow to zero or finite flow in the next month. The procedure was repeated for the 12 pairs of adjacent months and are given in Table 5.

From Table 5 we note that many of the cells in the 12 x 12 state transition matrix are empty. This sparsity is a function of record length n and size of class intervals (eleven in this case) used to capture the fineness of state transitions. Each cell gives the absolute frequency f_{ij} with which the particular transition occurred in the record; thus, there were 31 transitions from zero flow to zero flow for the October-November pair. To convert this to an estimate of probability of transition from state i to state j, P_{ij}, use

$$\frac{f_{ij}}{\sum_{j=1}^{12} f_{ij}} = \frac{f_{ij}}{f_i} \tag{11}$$

in which f_i is the total number of state transitions in row i. Note that $\sum_{j=1}^{12} P_{ij} = 1$ by definition because it is certain that a state transition has occurred. Stable estimates of P_{ij} are possible only if sufficient data were available. However, inspection of the state transition tables suggests that many of the cells should have a finite value rather than zero. In the next section, we put forward a subjective procedure for filling in these cells.

Subjective estimation of state transition probabilities

Herein, a procedure is proposed that allows for prior knowledge of the investigator (including his biases). There is no claim that consistent results are possible from one person to another. Nonetheless, serious consideration of the approach seems warranted for two reasons: (a) inadequacy of "objective" methods as outlined earlier in this paper, and (b) common use of subjective guessing in actual hydraulic practice. Pending resolution of problems in the stochastic modeling of ephemeral flows, a structured subjective approach to data analysis offers the chance of consistent results from person to person. Each person's

Table 5. Frequency of specified state transitions from month i to month i+1 for monthly flows (in cfs) on Rillito Creek (for period of 1909-1966).

(1) October-November state transitions

OCT State i \ NOV State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	31	6	-	-	1	-	-	-	-	-	-	-
0-10	10	7	1	1	-	1	-	-	-	-	-	-
11-20	-	-	-	-	-	-	-	-	-	-	-	-
21-30	-	-	-	-	-	-	-	-	-	-	-	-
31-40	-	-	-	-	-	-	-	-	-	-	-	-
41-50	-	-	-	-	-	-	-	-	-	-	-	-
51-60	-	-	-	-	-	-	-	-	-	-	-	-
61-80	-	-	-	-	-	-	-	-	-	-	-	-
81-100	-	-	-	-	-	-	-	-	-	-	-	-
101-150	-	-	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	-	-	-	-	-	-	-	-	-
251-largest value	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. continued

(2) November-December state transitions

NOV State i \ DEC State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	34	4	-	1	-	-	-	-	1	-	-	1
0-10	6	3	1	-	-	1	-	-	1	-	1	-
11-20	-	1	-	-	-	-	-	-	-	-	-	-
21-30	-	-	-	-	-	-	-	-	-	-	-	1
31-40	-	-	-	1	-	1	-	-	-	-	-	-
41-50	-	-	-	-	-	-	-	-	-	-	-	-

(3) December-January state transitions

DEC State i \ JAN State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest value
0	29	7	1	-	1	1	-	-	-	-	-	1
0-10	2	3	-	2	1	-	-	-	-	-	-	-
11-20	-	1	-	-	-	-	-	-	-	-	-	-
21-30	-	1	-	1	-	-	-	-	-	-	-	-
31-40	-	-	-	-	-	-	-	-	-	-	-	-
41-50	-	-	-	-	-	-	-	1	-	-	1	-
51-60	-	-	-	-	-	-	-	-	-	-	-	-
61-80	-	-	-	-	-	-	-	-	-	-	-	-
81-100	-	2	-	-	-	-	-	-	-	-	-	-
101-150	-	-	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	-	-	1	-	-	-	-	-	-
251-largest value	-	1	-	-	-	-	-	-	-	-	-	1

Table 5. continued

(4) January-February state transitions

JAN State i \ FEB State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	23	5	-	-	-	-	2	-	1	-	-	-
0-10	5	4	3	1	-	-	-	-	1	-	-	-
11-20	-	1	-	-	-	-	-	-	-	-	1	-
21-30	-	2	-	-	-	-	-	1	-	-	-	-
31-40	2	-	-	-	-	-	-	-	-	-	-	-
41-50	-	1	-	-	-	-	-	1	-	-	-	-
51-60	-	-	-	-	-	-	-	-	-	-	-	-
61-80	-	-	-	-	-	-	-	-	-	-	1	-
81-100	-	-	-	-	-	-	-	-	-	-	-	-
101-150	-	-	-	-	-	-	-	-	-	-	-	-
151-250	1	-	-	-	-	-	-	-	-	-	-	-
251-largest value	-	-	-	-	1	-	-	-	-	-	-	1

Table 5. continued

(5) February-March state transitions

FEB State i \ MAR State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	21	5	-	-	2	-	1	1	-	1	-	-
0-10	1	9	1	-	-	-	-	-	-	2	-	-
11-20	2	1	-	-	-	-	-	-	-	-	-	-
21-30	-	-	-	1	-	-	-	-	-	-	-	-
31-40	-	-	-	-	-	-	1	-	-	-	-	-
41-50	-	-	-	-	-	-	-	-	-	-	-	-
51-60	1	1	-	-	-	-	-	-	-	-	-	-
61-80	-	-	1	-	-	-	-	-	-	1	-	-
81-100	-	1	-	1	-	-	-	-	-	-	-	-
101-150	-	-	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	-	2	-	-	-	-	-	-	-
251-largest value	-	-	-	-	-	-	-	-	-	-	1	-

Table 5. continued

(6) March-April state transitions

MAR State i \ APR State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	23	2	-	-	-	-	-	-	-	-	-	-
0-10	11	5	1	-	-	-	-	-	-	-	-	-
11-20	2	-	-	-	-	-	-	-	-	-	-	-
21-30	2	-	-	-	-	-	-	-	-	-	-	-
31-40	2	2	-	-	-	-	-	-	-	-	-	-
41-50	-	-	-	-	-	-	-	-	-	-	-	-
51-60	2	-	-	-	-	-	-	-	-	-	-	-
61-80	-	1	-	-	-	-	-	-	-	-	-	-
81-100	-	-	-	-	-	-	-	-	-	-	-	-
101-150	3	1	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	1	-	-	-	-	-	-	-	-
251-largest value	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. continued

(7) April-May state transitions

APR State i \ MAY State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	41	2	1	-	-	-	-	-	-	-	-	-
0-10	11	1	-	-	-	-	-	-	-	-	-	-
11-20	1	-	-	-	-	-	-	-	-	-	-	-
21-30	1	-	-	-	-	-	-	-	-	-	-	-
31-40	-	-	-	-	-	-	-	-	-	-	-	-

(8) May-June state transitions

MAY State i \ JUN State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	44	10	-	-	-	-	-	-	-	-	-	-
0-10	3	-	-	-	-	-	-	-	-	-	-	-
11-20	-	1	-	-	-	-	-	-	-	-	-	-
21-30	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. continued

(9) June-July state transitions

JUNE State i \ JULY State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	6	21	6	4	3	1	1	1	2	-	-	2
0-10	-	5	1	-	1	1	2	-	-	1	-	-
11-20	-	-	-	-	-	-	-	-	-	-	-	-

(10) July-August state transitions

JUL State i \ AUG State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	1	2	-	1	2	-	-	-	-	-	-	-
0-10	2	12	3	-	3	2	-	1	-	3	-	-
11-20	-	-	2	2	1	-	1	-	-	1	-	-
21-30	1	-	-	-	1	-	-	2	1	-	-	-
31-40	-	-	-	-	1	-	-	-	-	1	-	-
41-50	-	-	-	-	-	1	-	1	1	0	-	-
51-60	-	1	-	1	-	-	1	-	-	-	-	-
61-80	-	-	-	-	-	-	-	-	-	1	-	-
81-100	-	-	-	-	-	1	-	-	-	-	1	-
101-150	-	1	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	-	-	-	-	-	-	-	-	-
251-largest value	-	-	-	-	-	-	-	1	-	-	-	1

Table 5. continued

(11) August-September state transitions

AUG State i \ SEPT State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	3	-	1	-	-	-	-	-	-	-	-	-
0-10	6	5	3	1	-	-	-	1	-	-	-	-
11-20	-	4	-	-	-	-	-	-	-	1	-	-
21-30	-	2	1	-	-	1	-	-	-	-	-	-
31-40	1	5	2	-	-	-	-	-	-	-	-	-
41-50	-	2	1	-	-	1	-	-	-	-	-	-
51-60	-	1	1	-	-	-	-	-	-	-	-	-
61-80	-	4	-	-	-	-	1	-	-	-	-	-
81-100	1	1	-	-	-	-	-	-	-	-	-	-
101-150	1	2	1	-	-	-	-	1	-	-	-	1
151-250	-	-	-	-	-	-	-	-	1	-	-	-
251-largest value	-	1	-	-	-	-	-	-	-	-	-	-

Table 5. continued

(12)September-October state transitions

SEPT State i \ OCT State j	0	0-10	11-20	21-30	31-40	41-50	51-60	61-80	81-100	101-150	151-250	251-largest values
0	7	5	-	-	-	-	-	-	-	-	-	-
0-10	17	10	-	-	-	-	-	-	-	-	-	-
11-20	7	3	-	-	-	-	-	-	-	-	-	-
21-30	-	1	-	-	-	-	-	-	-	-	-	-
31-40	0	-	-	-	-	-	-	-	-	-	-	-
41-50	2	-	-	-	-	-	-	-	-	-	-	-
51-60	1	-	-	-	-	-	-	-	-	-	-	-
61-80	2	-	-	-	-	-	-	-	-	-	-	-
81-100	1	-	-	-	-	-	-	-	-	-	-	-
101-150	1	-	-	-	-	-	-	-	-	-	-	-
151-250	-	-	-	-	-	-	-	-	-	-	-	-
251-largest value	-	1	-	-	-	-	-	-	-	-	-	-

insight can suggest modifications so as to strengthen the consistency of the procedure.

The objective is to construct a square matrix of subjective state transition probabilities so as to generate sequences of monthly flow that possess properties similar to the historical record. Only in the probabilistic (and statistical) sense are the synthetic and historic sequences presumed to be similar.

In the procedure, a freedom of choice exists for (a) size of m in the m x m square matrix, (b) size of flow interval for each cell of the matrix, (c) subjectiveness in filling in the empty cells of the matrix, and (d) generating function that determines variable magnitude of flow.

The following set of rules and guidelines are given for the sake of consistency:

(a) Based on the number of data points, determine the size m of the matrix to obtain meaningful probabilities. (We had 696 data points in 58 years and used 12 x 12 matrices since the range of monthly flows was 0 cfs to 976 cfs with sample means ranging from 0.287 cfs/month to 42.13 cfs/months. In this manner, we hoped to allow freedom to have sufficient variability in the synthetically generated flows.)

(b) Choose the flow intervals for the cells. They do not have to be evenly spaced; in semi-arid and arid regions, a cell should be allocated for a flow of 0 cfs.

(c) Compute the "historic" transition probability matrices between each pair of months by entering the number of flows that fill a cell given that the previous month's flow fell in a certain interval. From these numbers compute the transition probabilities for each cell of the matrix.

(Hundredths of cfs is precise enough for most cases. Note that the hundredths arise only because of averaging; no claim is made that flows are measured with such precision.) Make a note of the mean and maximum flow for each month; confidence regions as given in Tables 3 and 4 may help in the subjective analysis.

(d) Based on the mean, maximum flow, historic probabilities and number of occurrences, compute or fill in the matrices subjectively using the following rules and guidelines:

(1) Summation of transition probabilities across each row should equal one.

(2) Precision to the hundredths.

(3) Rows in the vicinity of the mean (where the majority of flows are expected) are most sensitive from a modeling viewpoint than rows where less flows are expected.

(4) If a row has p_{ij} of 0.50 and 0.50 in two cells, it is probable that this row contained only 2 flows. Whereas a row with p_{ij} of 0.10 in each of 10 cells probably was based on 10 flows. Thus, more subjective weight should be placed on rows where it appears that a large number of flows was used to compute p_{ij}; less weight should be placed in rows where few occurrences arise. Knowledge of f_i, the absolute frequency, for each row would help in this evaluation of the p_{ij}'s.

(5) When a zero p_{ij} is encountered between finite p_{ij}'s, the empty cell is assigned a p_{ij} that reasonably fits between the adjacent p_{ij}'s.

(6) Usually the first 2 or 3 rows of the matrix are of most significance since the majority of flows will probably fit in these intervals (see

Table 5). The simulation results would be more sensitive in this region than in higher-indexed rows.

(7) If p_{ij} is the last probability in a row and $j \neq m$, then assign a finite p to $p_{i(j+1)}$ and zero to $i^{\underline{th}}$ row cells with column index greater than j+1. Thus, if maximum flow in a row is 40 cfs and it lies in cell #5 out of 10, then the last finite probability in that row should be assigned to cell #6 and cells #7 to 10 will have zero p. This rule does not apply if the last cell in the row has finite p. The rule allows generation of flows larger than actually recorded.

(8) A certain consistency should be maintained in the subjective matrices both in the months and between months. For example, the p_{ij} in the extreme cells should be greater than or equal to the p in the previous cell because of possible dependence from month to month (see Table 2) and because a higher flow in a month is more probable if the preceding month's flow was high (see Table 1 for relative values of means for adjacent months).

To illustrate the above rules, consider the simple 6 x 6 matrix in Table 6. Little weight is assigned to the p_{ij} in the lower 4 rows because a small number of occurrences entered the cells.

The last two authors of this paper used the above rules to fill in the cells shown in Table 5. A Student's t-test showed that the historic and simulated mean flows for the months were statistically similar but the Fisher's ratio test on variances showed that their simulated variances were "out of the ball park." The latter result was produced by the tendency to assign too high or a very small probability to transitions to high flows, that is, the senior authors had a tendency

Table 6. Illustration of development of subjective matrix (Maximum flow = 220 cfs; $\overline{x}$ = 25.0 cfs).

"Historic" matrix

Month i \ Month j = i+1	0	1-20	21-50	51-100	101-200	> 200
0	.55	.33	.12	-	-	-
1-20	.61	.29	.08	.02	-	-
21-50	.40	-	.20	-	.20	.20
51-100	-	.50	-	-	-	.50
101-200	-	-	-	-	-	-
> 200	-	-	-	1.0	-	-

Subjectively assigned matrix:

	1	2	3	4	5	6
1	.55	.30	.13	.02	-	-
2	.60	.30	.05	.03	.02	-
3	.50	.30	.10	.05	.03	.02
4	.50	.30	.10	.05	.03	.02
5	.50	.30	.10	.05	.03	.02
6	.50	.30	.10	.05	.03	.02

to fill in all cells. This suggests that an upper limit should be placed on the guessing of probabilities at the extremities. Other results, consistent with the significant difference in variances, were the production of higher extremes than historically observed and decrease in number of zero flows. However, the cross-correlation matrix (Table 2) turned out quite well. The results suggest that the matrix might also be tested for finer breakdown of flow values.

Explanation of computer generation model

The subjective matrices are taken and converted to cumulative probability matrices across each row. An initial flow is taken, say 0, and the next flow is computed by random number generation as follows:

1) Go to row dictated by last flow generation
2) Get a pseudo-random number R
3) Compare that number to cumulative probabilities in the specified row to ascertain location of the cell in that row
4) Get another pseudo-random number R
5) Compute new flow by taking the lower bound of specified cell and adding some function of the generated number in 4) to the lower bound. Return to 1).

The function of the random number (R) used in this routine was R^2. This yields an exponential function. There is only intuition behind the choice of this function, but the model is not highly sensitive to any function so long as it is somewhat between this type of exponential and a uniform distribution. Approximate bounds could be stated as $R^3 \leq f(R) \leq R$.

Flow charts on the computer program are given in Appendix I. This represents the main generation procedure.

SUMMARY AND CONCLUSIONS

A critique of statistical properties of monthly flows on an ephemeral stream has been given. An ad hoc subjective procedure is proposed for sequential generation of the monthly flows. Such a method seems justified for managerial purposes that are not concerned with the variability of flow within the month and with the large number of days within any month for which the flow is zero. It is our considered judgment that ephemeral flows should be modeled by starting with historical daily flows (at least). If this is done, then more meaningful monthly flow models (and then synthetic sequences) could be built. These would serve as a check on alternative monthly flow models (subjective matrix method proposed in this paper, lag-one autoregressive model, harmonic model, bivariate normal model, or bivariate log-normal model). Modern computers permit low cost analysis of hydrologic records at smaller time intervals like a day (or even an hour).

REFERENCES

Baran, N., C.C. Kisiel and L. Duckstein, 1971. A stochastic analysis of flows on Rillito Creek. Proceedings, 1st Annual Meeting Arizona Section - AWRA & Hydrology Section - Ariz. Academy of Science, Tempe, Ariz., April.

Clainos, D., 1972. Input specifications to stochastic decision models, M.S. thesis, Dept. of Systems and Industrial Engineering, University of Arizona, Tucson, Arizona.

Denny, J.L., C.C. Kisiel and S. Yakowitz, 1971. Statistical inference on streamflow processes with Markovian characteristics. Proceedings, Symp. on Statistical Hydrology, Tucson, Ariz. (to be released by U.S. Government Printing Office as a U.S. Dept. of Agriculture publication in 1972).

Fiering, M., 1967. Streamflow Synthesis, Harvard University Press, Cambridge, Mass. p. 139.

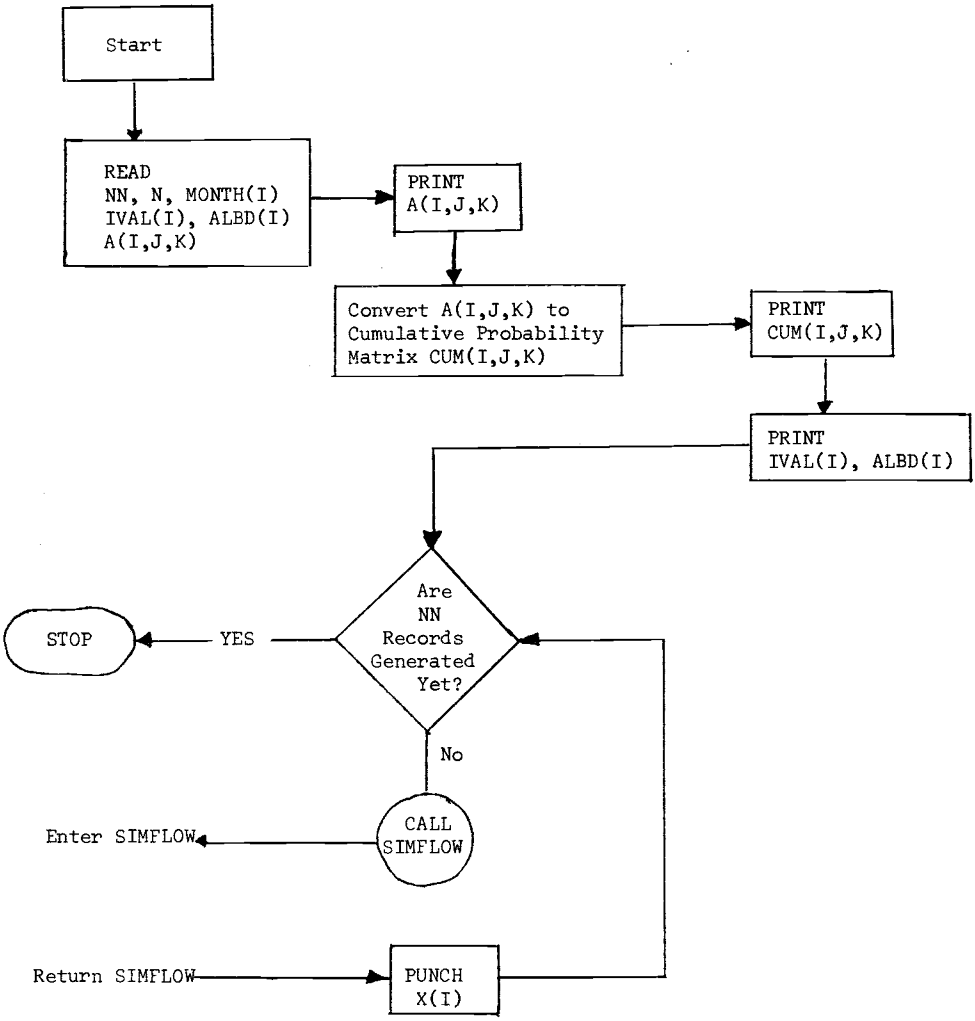
Kendall, M.G., 1947. Advanced Theory of Statistics, Vol. 1, Charles Griffin and Co., Ltd., London, England.

Kisiel, C.C. and L. Duckstein, 1971. Time and frequency domain analysis of a causal bivariate stochastic process. Proceedings, Int'l Symposium on Mathematical Models in Hydrology, Warsaw, Poland (July).

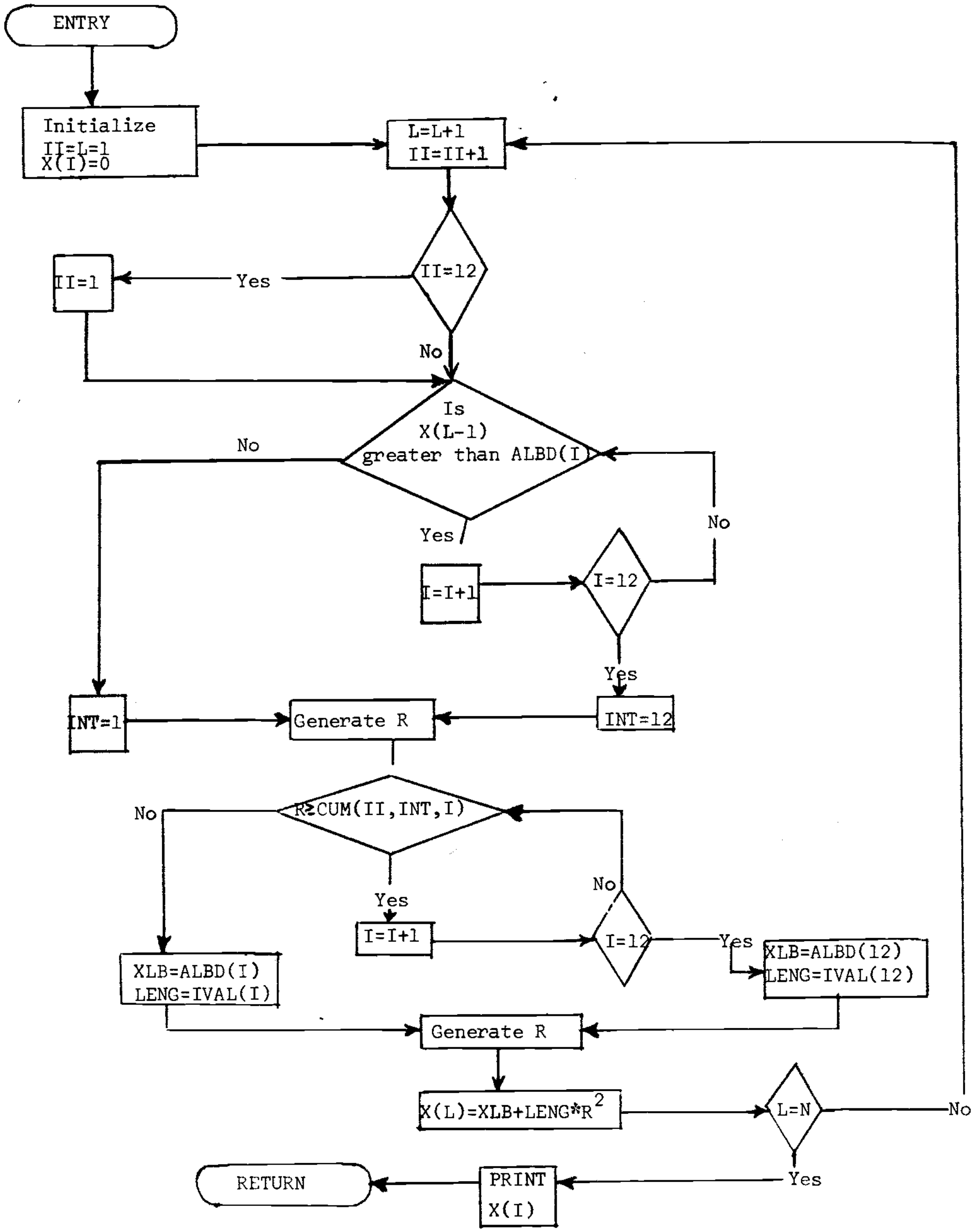
Lindgren, B.W., 1968. Statistical Theory, The MacMillan Co., New York.

APPENDIX I

Flowchart for Streamflow Model



Subroutine SIMFLOW



Appendix I (continued)

Definitions of Variables Used in Flowchart

NN	number of records to be generated
N	number of flows per record
MONTH (I)	array of abbreviations for the months
IVAL (I)	interval width
ALBD (I)	lower bound of the interval
A (I,J,K)	subjective matrix for month i, row j, column k
CUM (I,J,K)	cumulative subjective matrix for month i, row j, column k
X (I)	generated streamflows
II	month counter
L	flow counter
INT	row number of preceding flow
XLB	lower bound for next flow
LENG	width of interval