Nordic Hydrology 3, 1972, 214–238
Published by Munksgaard, Copenhagen, Denmark

SINGLE AND MULTI-SITE OPERATIONAL HYDROLOGY

T. A. McMAHON, G. P. CODNER and C. PHILIPS

Monash University, Clayton, Australia

This paper reviews several different single- and multi-site generation models to complement those papers in which the models are developed but not applied. Initially the historical time series is looked at using various time series techniques, including correlograms and spectral analyses, to determine whether a lag one Markov model will satisfactorily represent the historical data.
The single site models which are examined include an empirical model using the historical probability distribution of the random component, the Thomas and Fiering model using both logarithms of the data and Matalas' log-normal transformation equations, and finally Moran's Gamma distribution. Matalas' Residual method with several probability distributions and Moran's Multivariate Gamma technique are used in the multi-site analysis. All models are applied to the streams of the Melbourne metropolitan water supply system.

With increasing utilization of water resources we are beginning to move from an era of *ad hoc* planning to one based more on systems analysis and simulation.

Simulation programs require streamflow as input data. To ensure that representative streamflows are used and to indicate the order of sampling errors involved, several hydrologic traces may be used rather than the single historical record. For this purpose streamflow generation models have been developed which provide replicates of the historical streamflow.

Since the publication by Maass et al. (1962) of the Harvard Water Program activities, considerable use has been made of simulation analysis in water resources projects. Most techniques are based on the synthesis of hydrologic sequences by data generation models. Many streamflow models have been developed for this purpose, but few details of results of application have been published.

The purpose of this paper is to review both single- and multi-site data generation models using monthly data. The models are applied to streams supplying water to the Melbourne metropolitan area. A more detailed description of the development of the models and their application is given by the authors in a research report (McMahon, Codner & Philips 1972, available on request).

It should be noted that we do not consider the implications of the "Noah and Joseph" effects as outlined by Mandelbrot & Wallis (1968). This would require a separate study involving "Fractional Gaussian Noises" (Mandelbrot & Wallis 1969).

TIME SERIES STRUCTURE OF STREAMFLOW

A time series is said to be strongly stationary or stationary in the strict sense if all its moments are time invariant. If only the first two moments are time invariant, the series is said to be weakly stationary or stationary in the wide sense. A Gaussian process that is shown to be weakly stationary is sufficient verification of strict stationarity since all higher-order moments can be expressed in terms of the first two moments.

Hydrologic time series rarely satisfy the condition of strict stationarity. It is well known that the mean and variance of monthly streamflows vary with time. The construction of water resources structures, alterations in land use, and the presence of trends within the data, all tend to produce nonstationary effects. These effects are generally quite small and if necessary can be taken into account. Significant trends can be identified and removed.

The analytical techniques used here require the data to be at least weakly stationary. Nonstationary data require special techniques (Bendat & Piersol 1966) which are beyond the scope of this paper. To obtain the maximum information, the time series should be strongly stationary. However, a large amount of useful information can be obtained from series which are only weakly stationary. Unless otherwise specified, the term stationary will henceforth be used to denote weak stationarity.

A first order stationary time series is obtained by subtracting the monthly mean flow from each monthly flow,

$$X_t' \equiv X_t - M_\tau \qquad \tau = 1, \ldots 12, 1 \ldots 12, \ldots \quad t \equiv 1, \ldots N \tag{1}$$

where N is the length of record in months.

Second order stationarity is obtained by dividing each value of X_t' by the appropriate monthly standard deviation,

$$X_t'' \equiv \frac{X_t'}{\sigma_\tau} \qquad \tau = 1, \ldots 12, 1 \ldots 12, \ldots \quad t \equiv 1, \ldots N \tag{2}$$

Transformations (1) and (2) ensure that the data is weakly stationary.

Monthly streamflow data can be regarded as consisting of four components (Kottegoda 1970): a seasonal component (S_t), a trend component (T_t), a correlated component due to serial correlation (K_t), and a random component (ε_t). Hence streamflow at time t is given by

$$X_t \equiv S_t + T_t + K_t + \varepsilon_t \tag{3}$$

In obtaining a suitable stochastic model, it is necessary to isolate and examine each component. Correlation analysis and power spectral analysis are two powerful techniques for analysing time series.

By studying the underlying structure of historical time series it is possible to postulate a generating model which preserves the essential characteristics of the historical data. One approach is to use an autoregressive model.

Let X_t denote the values at equally spaced time t, t–1, t–2, ... of a stationary time series. The linear autoregressive model is defined as

$$X_t = \sum_{i=1}^{m} A_i X_{t-i} + \varepsilon_t \tag{4}$$

where m denotes the order of the process and the coefficient A_i is a measure of the persistence of past values with the present value. The random uncorrelated component, ε_t, is often referred to as "white noise". Thus,

$$X_t \equiv A_1 X_{t-1} + \varepsilon_t \tag{5}$$

is an autoregressive model (or Markov model) of order one. In this equation, the coefficient is given by $A_1 = \varrho_1$ where ϱ_1 is the lag one serial correlation coefficient. Previous unpublished studies by the authors and the work of Burges (1970) suggest that a first order model normally is adequate.

Correlograms and power spectral density functions are powerful analytical tools for distinguishing among the various generating models. These techniques are now discussed.

A plot of the serial correlation coefficient ϱ_k against the lag k is called a correlogram. The serial correlation coefficient of lag k is defined by

$$\varrho_k \equiv \frac{\mathrm{COV}[X_t, X_{t+k}]}{\left\{\mathrm{VAR}[X_t]\ \mathrm{VAR}[X_{t+k}]\right\}^{1/2}} \tag{6}$$

and is a measure of the degree of linear dependence between events X_t and X_{t+k}. Substituting for the covariance [COV] and variance [VAR], Eq. (6) becomes

$$\varrho_k = \frac{\frac{1}{N-k}\sum_{t=1}^{N-k} X_t X_{t+k} - \frac{1}{(N-K)^2}\sum_{t=1}^{N-k} X_t \sum_{t=1}^{N-k} X_{t+k}}{\left[\frac{1}{N-k}\sum_{t=1}^{N-k} X_t^2 - \frac{1}{(N-K)^2}\left(\sum_{t=1}^{N-k} X_t\right)^2\right]^{1/2}\left[\frac{1}{N-k}\sum_{t=1}^{N-k} X_{t+k}^2 - \frac{1}{(N-K)^2}\left(\sum_{t=1}^{N-k} X_{t+k}\right)^2\right]^{1/2}} \tag{7}$$

A random time series possesses no internal correlation, that is, $\varrho_k = 0$ for all values of $k \geqq 1$. Because of sampling errors within the data it is necessary to test if the serial correlation coefficients are significantly different from zero. Anderson (1941) has developed a satisfactory test for this.

The correlogram is useful in detecting deterministic components which might be mixed with random noise. The correlogram of the noise component will tend towards zero for large time lags, whereas the correlogram of the deterministic component will persist over all lags.

A power spectral density function, which is the Fourier transform of the autocovariance function, depicts the total variance of the time series in terms of its frequency components. The advantages of this type of representation are that any cycles within the data are revealed as sharp peaks at the appropriate frequency; trends are revealed as peaks at the zero frequency level, since a trend may be regarded as a cyclic component with an infinitely large period, and random components (white noise) are depicted as horizontal lines since no one frequency contributes any more variance than any other frequency level. The total area under the spectrum is equal to the variance of the series (Julian 1961). Correlograms examine the structure of the time series in terms of the time domain, whereas the spectral density function is concerned with the frequency domain. Theoretically, both correlograms and spectral graphs reveal the same amount of information about a time series, but because of sampling errors some of the characteristics of the data are more clearly discernible with one method than the other. Hence, it is advisable to use both methods together.

For a discrete, finite time series, the "smooth" estimate of the historical spectral density function is given by

$$S'(f) = \frac{1}{m}\left[D_0C_0 + 2\sum_{k=1}^{m-1} D_kC_k \cos\frac{\pi kj}{m} + (-1)^j D_mC_m\right] \tag{8}$$

where C_k is the covariance function of lag k,

f is the frequency, given by $\frac{j}{2hm}$,
h is the sampling interval,
m is the maximum lag,
j is the harmonic number, $j = 0,1, \ldots m$, and
D_k is a generalized lag weighting function.

In this study, the Parzen lag weighting function has been chosen to avoid the problem of negative estimates of the spectrum (Parzen 1967, pp. 142–143).

Model Identification

The objective is to determine the most suitable generating model that describes the historical time series. As a first step, the theoretical and historical correlograms are compared.

The correlogram of a first order Markov process (Eq. 5) is defined by

$$\varrho_k = \varrho_1{}^k; \; k = 0,1 \ldots \tag{9}$$

When ϱ_1 is positive, ϱ_k decreases monotonically from $\varrho_0 = 1$ to $\varrho_\infty = 0$. If ϱ_1 is negative, the correlogram will oscillate about zero with a decreasing amplitude.

Due to the presence of sampling errors, the historical correlogram will exhibit less damping than the corresponding theoretical correlogram. It is therefore necessary to test the goodness of fit between the two correlograms. A large sample test of the goodness of fit for autoregressive schemes has been developed by Quenouille (1947).

It should be noted that the above test does not select the most general autoregressive process. It is possible that a higher order process may provide a better fit even though a lower order process may satisfy the test.

The spectral density function of a first order autoregressive process is given by

$$S'(f) = \frac{(1-\varrho_1{}^2)\ \text{VAR}\ [X(t)]}{1 + \varrho_1{}^2 - 2\varrho_1 \cos 2\pi f} \tag{10}$$

When ϱ_1 is positive, S′ (f) decreases monotonically from a maximum at $f = 0$ to a minimum at $f = 0.5$ cycles/time interval. If ϱ_1 is negative, S′ (f) increases monotonically from a minimum at $f = 0$ to a maximum at $f = 0.5$ cycles/time interval.

Because of sampling errors within the data, confidence limits must be used when comparing the historical spectral density function with the spectral density function of the proposed generating model. The method used in this study is based on a Chi-Square test outlined by Blackman & Tukey (1958).

SINGLE SITE DATA GENERATION

Empirical Model

The first step in applying the above theory to the generation of hydrologic data at a single site is to test for the presence of trends. Unless there is an *a priori* reason for knowing the type of trend, it is better to use a non-parametric test. The test used in this study is based on Kendall's rank correlation coefficient (Kendall & Stuart 1968, p. 419).

At the 95 % confidence level, the test showed that the O'Shannassy monthly flows possess no trend. The flows were then standardised according to Eqs. (1) and (2) and were assumed to be stationary.

Referring to Fig. 1, it can be seen that the correlogram reveals the strong seasonal behaviour of the unstandardised monthly flows. The same information

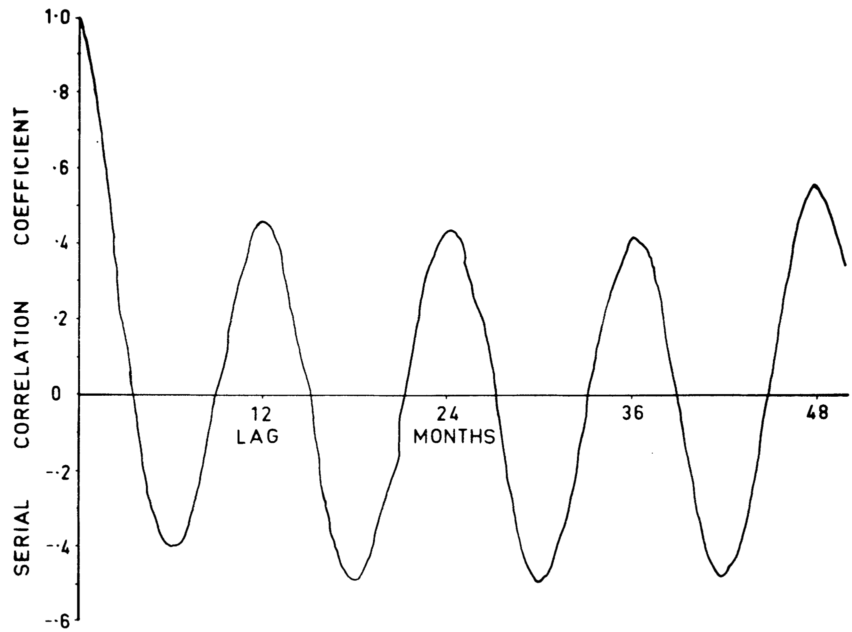

Fig. 1.
Correlogram of historical unstandardized flows for O'Shannassy River.

is revealed in the spectral density graph (Fig. 2) where a peak occurs at f = 0.08 cycles/month corresponding to the twelve-month annual cycle.

Fig. 3 shows the spectral density graph after the seasonal component has been removed. Note the absence of any significant peaks. The standardisation procedure has effectively removed the twelve-month cycle and all of its significant harmonics. Superimposed upon the historical spectral density graph are the confidence limits (Blackman & Tukey 1958) for a first order Markov process (Eq. 5). Clearly, a first order Markov model is an adequate representation of the historical spectral density function.

The correlogram of the standardised flows (Fig. 4) does not specifically identify the order of the Markov process. However, application of Quenouille's test suggests that, at the 95 % level of confidence, a first order Markov model is acceptable.

A further check on the validity of assuming a first order Markov model is to test the randomness of the component ε_t. The correlogram and the power

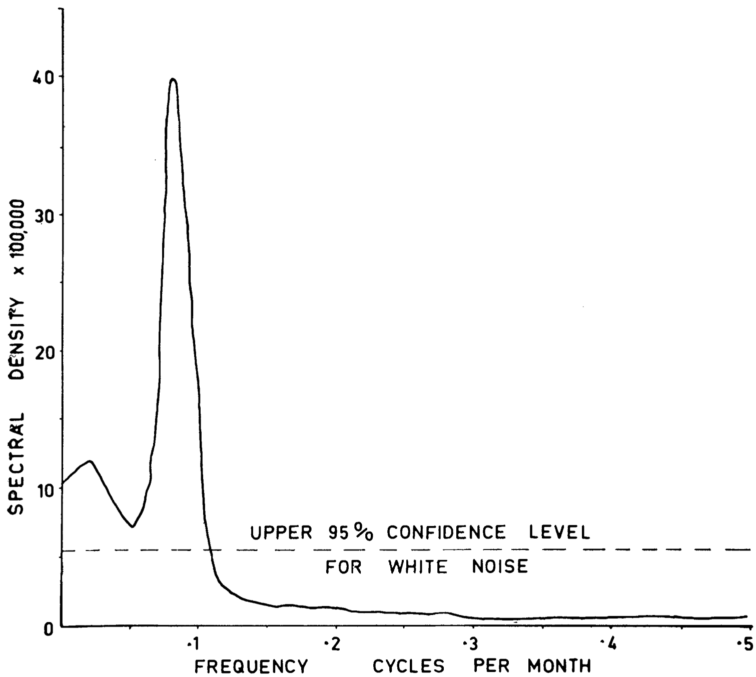

Fig. 2.

Power spectral density function of historical unstandardized flows for O'Shannassy River.

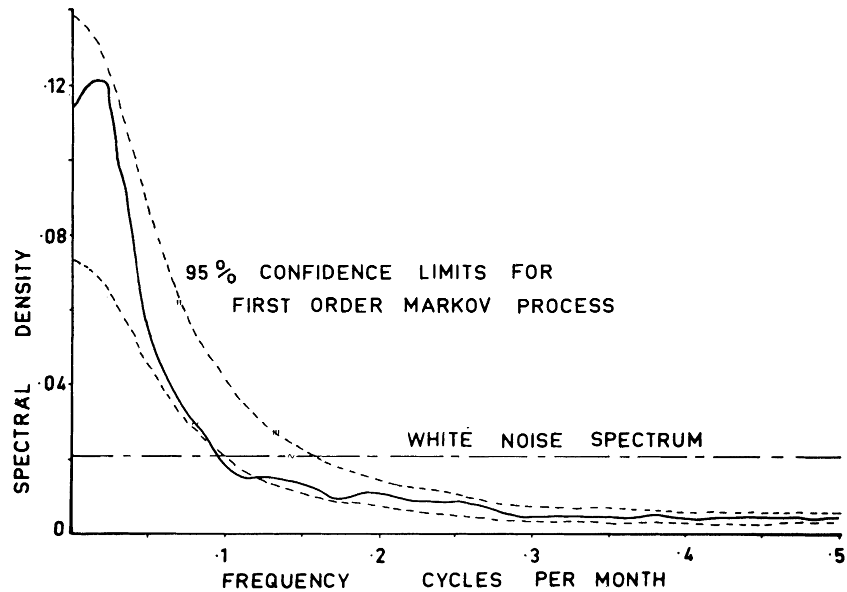

Fig. 3.
Power spectral density function of historical standardized flows for O'Shannassy River.

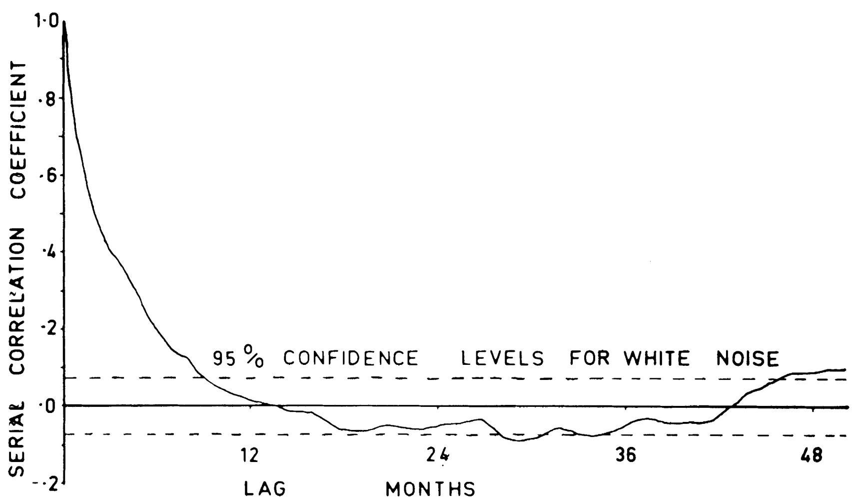

Fig. 4.
Correlogram of historical standardized flows of O'Shannassy River.

spectral density function (not presented in this paper) reveal that, at the 95 % level of significance, ε_t is a truly random uncorrelated component.

Figs. 5 and 6 show that, using the historical distribution of ε_t, the first order Markov model preserves the correlogram and the power spectral density function of the historical flows. Application of the Kolmogorov-Smirnov test (Massey 1951) showed that at the 95 % level of confidence there is no significant difference between the population of generated monthly flows and the population of historical monthly flows.

In contrast to the empirical model, the following single site models assume a lag one Markov effect without testing the validity of the assumption. In addition, it is assumed that there is no trend component.

Log-normal Distribution

One approach to generate data at a single site is to use the Thomas and Fiering model (Thomas & Fiering 1962) which is defined as follows:

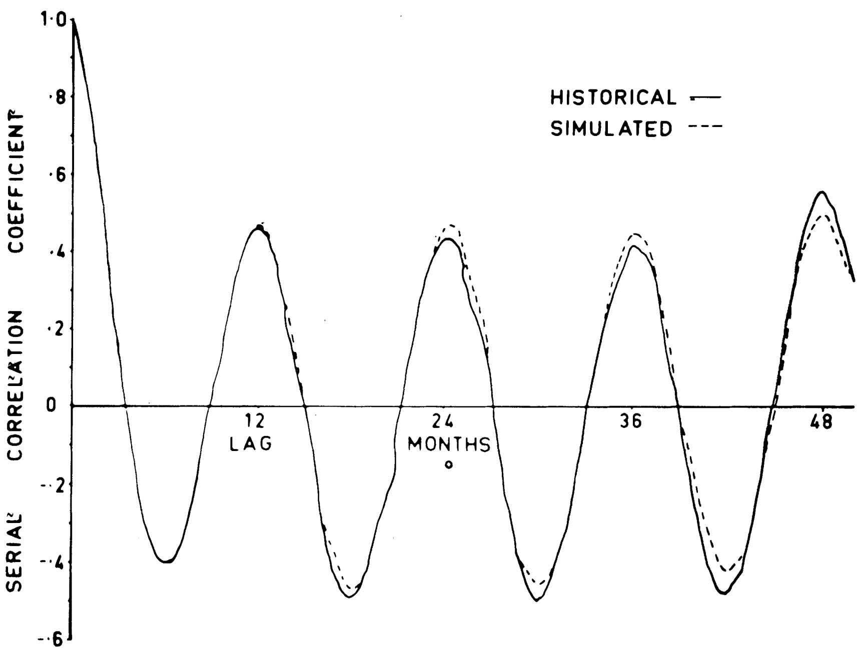

Fig. 5.
Comparison of correlograms based on historical and generated data for O'Shannassy River.

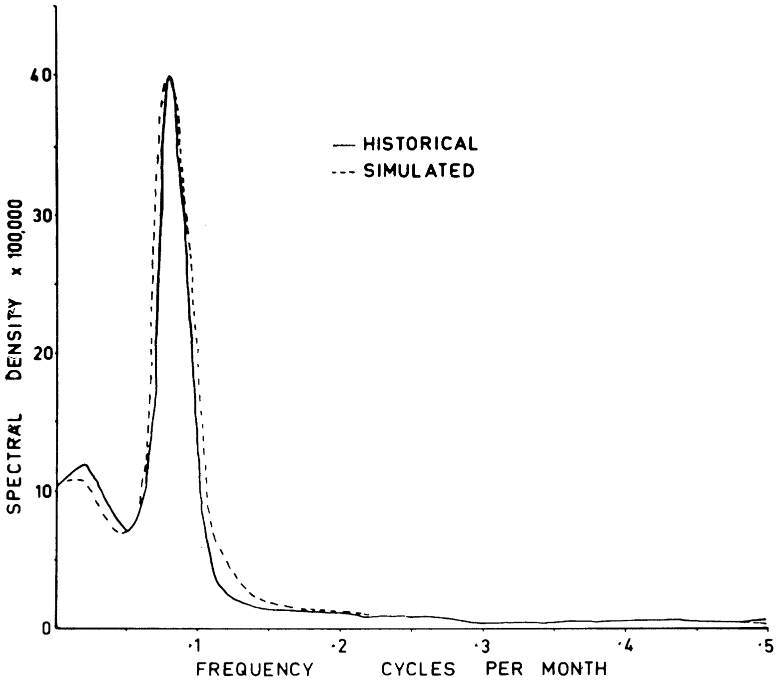

Fig. 6.
Comparison of power spectral density functions based on historical flows and generated data using the empirical model for O'Shannassy River.

$$q_{i+1} = \bar{q}_{j+1} + b_j (q_i - \bar{q}_j) + t_i s_{j+1} (1 - r_j^2)^{1/2} \qquad (11)$$

where q_{i+1}, q_i are flows during $(i+1)^{th}$ and i^{th} seasons, $\bar{q}_{j+1}$, $\bar{q}_j$ are mean flows during $(j+1)^{th}$ seasons within a repetitive annual cycle of seasons, b_j is least squares regression coefficient and r_j is correlation coefficient between flows in j^{th} season and $(j+1)^{th}$ season, s_{j+1} is standard deviation of flows during $(j+1)^{th}$ season and t_i is random normal variate with zero mean and unit variance.

This model preserves the first and second moments of the flows, seasonality, and the first order Markov effects. If the streamflows are not normally distributed, t_i may be replaced by a transformation to provide skewed data. One such transformation is given by Fiering (1967). For streams with high monthly serial correlations and large skews, this model is unsatisfactory. The nature of this difficulty is considered by McMahon & Miller (1971) and is partially overcome by taking a logarithmic transformation of the original data.

One drawback with the log-transformed data model is that the serial correlations of the untransformed flows are not necessarily preserved. Matalas (1967) suggested that this difficulty would be overcome by using moment transformations to obtain mathematically correct parameters. Parameters for a 3-parameter log normal model using Matalas' approach are determined using Eqs. (7), (8), (9) and (12) in Matalas (1967).

The 3-parameter model is defined as follows:

$$q_{i+1} \equiv A_{j+1} + \exp(y_{i+1}) \qquad (12)$$

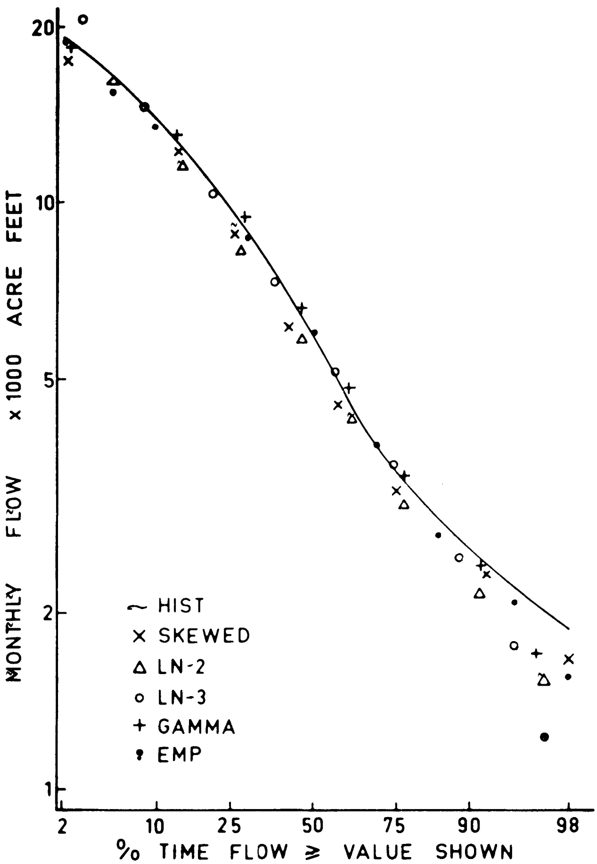

Fig. 7.

Flow duration curves based on historical flows and single site generated flows for O'Shannassy River.

where q_{i+1} is the monthly generated flow, A_{j+1} is the lower bound value such that

$$y_{i+1} \equiv \log (q_{i+1} - A_{j+1}) \tag{13}$$

and,

$$y_{i+1} = \overline{Q}_{j+1} + B_j (y_i - \overline{Q}_j) + t_i S_{j+1} (1 - R_j^2)^{1/2} \tag{14}$$

in which $\overline{Q}_j$, S_j, R_j are the log normal moment estimates of the monthly means, standard deviations and serial correlations, and

$$B_j \equiv \frac{S_{j+1}}{S_j} R_j \tag{15}$$

In the 2-parameter model, the lower bound value A_j is set to zero.

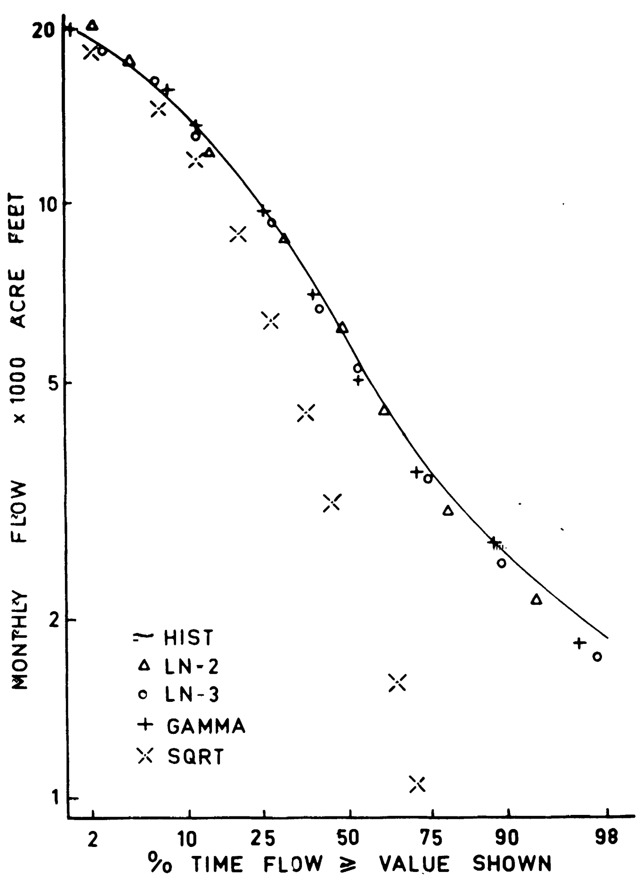

Fig. 8.
Flow duration curves based on historical flows and multi-site generated flows for O'Shannassy River.

Gamma Distribution

An alternative to the log normal distribution discussed above is the gamma distribution for which Matalas has produced transformation equations so long as the historical skews are less than $2\sqrt{2}$. As this is not the case for O'Shannassy River, Moran's Multivariate Gamma Distribution technique (Moran 1970) was used for the limiting case of one site. This method is discussed in a later section.

MULTI-SITE GENERATION

In multi-site data generation not only is it necessary to preserve the distribution and serial correlation of flows at each site but also their spatial correlations. Three procedures have been proposed to do this. The key station method (Hufschmidt & Fiering 1966) and the principal component approach (Fiering 1964) both have drawbacks (Fiering 1964, Matalas 1967).

The third procedure is the residual method approach suggested by Matalas (1967). Matalas uses mathematically correct transformation equations to preserve the historical parameters through the log normal transformation. This is in constrast to the method used by Young & Pisano (Jennings 1969, Young & Pisano 1968) in which logarithms of the data are taken before calculating parameters for use in the generating process.

Matalas Residual Approach

Several probability distributions may be chosen to represent the historical data. One which tries to take account of the skewness of the data is the 3-parameter log normal distribution with lower bound A_j^p such that

$$y_i^p = \log_e (x_i^p - A_j^p) \qquad (16)$$

where y_i^p is normally distributed log values for i^{th} interval and p^{th} station, x_i^p is historical or generated streamflow, and j is month.

Having assumed a probability distribution the appropriate transformation equations between the parameters of $x^{(p)}$ and $y^{(p)}$ may be determined. For the 3-parameter log normal distribution the appropriate moment equations and lag zero and lag one cross correlations equations are given as Eqs. (7), (8), (9), (12), (24) in Matalas (1967).

Having obtained the transformed log parameters it is possible to establish the matrices in the generating equation. The multivariate log synthetic sequences, y_i^p may be generated by using a multivariate weakly stationary generating process that is defined for m sites as

$$[y_{i+1}] = [A][y_i] + [B][\varepsilon_{i+1}] \tag{17}$$

In Eq. (17) $[y_i]$, $[y_{i+1}]$ are $(m \times 1)$ matrices whose p^{th} elements are

$$\frac{y_i^p - \bar{Q}_j^p}{S_j^p} \text{ and } \frac{y_{i+1}^p - \bar{Q}_{j+1}^p}{S_{j+1}^p}, \text{ respectively.}$$

These are the residuals for the i^{th} and $(i+1)^{th}$ consecutive periods, and $[\varepsilon_{i+1}]$ is an $(m \times 1)$ matrix whose elements are random normal deviates. $\bar{Q}_j^p$ and S_j^p are the monthly log normal moment estimates of the means and standard deviations for season j at site p. [A] and [B] are both $(m \times m)$ matrices whose elements consist of a combination of the lag zero and lag one cross correlations between sites. Thus, each flow is related to the flows at the neighbouring sites in both the present and preceding time interval.

The elements of [A] and [B] must be defined in such a way that the generated multivariate sequences resemble the historical multivariate sequences in terms of mean, variance, skew, serial correlation, lag zero and lag one cross correlations for all sites. A detailed derivation of the equations for determining [A] and [B] is given by Matalas (1967), but only the end result is noted here.

The elements of [A] and [B] are defined by

$$[A] = [M_{-1}][M_0]^{-1} \tag{18}$$

$$[B][B]^T = [M_0]-[M_{-1}][M_0]^{-1}[M_{-1}]^T \tag{19}$$

where $[M_0]$ is an $(m \times m)$ matrix with diagonal elements of unity and off diagonal elements representing the lag zero cross correlations, and $[M_{-1}]$ is an $(m \times m)$ matrix with diagonal elements of the serial correlation at each site and off diagonals composed of the lag one cross correlations between sites. The cross correlations used in matrices $[M_0]$ and $[M_{-1}]$ are all non-seasonal values.

After the residuals have been generated by Eq. (17), two inverse transformations have to be applied. The first inverse transformation replaces the cyclic seasonal component using the transformed log parameters and is defined by $(y_{i+1}^p \cdot S_{j+1}^p + \bar{Q}_{j+1}^p)$. This yields cyclic log values to which the inverse transform equation

$$x_{i+1}^p = \exp\,(y_{i+1}^p) + A_{j+1}^p \tag{20}$$

is applied to give multivariate synthetic sequences.

The gamma distribution is another distribution commonly used. As set out in the single site section, Matalas' gamma transformation equations should not be applied where skews are greater than $2\sqrt{2}$. Because of this, Moran's Multivariate technique was used.

Moran's Multivariate Gamma Technique

The assumption is made that the multiple sites have a multivariate gamma distribution such that the flows at each site are gamma distributed. The shape and scale factors are independently determined for each site on a monthly basis using the maximum likelihood equations given in Thom (1958). These parameters are assumed to be estimates of the multivariate gamma distribution parameters. This approach is necessary because of the complexity of the maximum likelihood equations for anything greater than the bivariate gamma cases (Moran 1969).

Using Moran's method (Moran 1970) the historical flows are then transformed to corresponding normal variates with zero mean and unit variance by means of Eqs. (21) and (22).

$$U_{kj}{}^{p} \equiv ((\beta_j{}^{p})^{\gamma_j{}^{p}} \Gamma(\gamma_j{}^{p}))^{-1} \int_0^{X_{kj}{}^{p}} e^{-x\beta_j{}^{p}} \; x^{(\gamma_j{}^{p}-1)} \, dx \tag{21}$$

$$U_{kj}{}^{p} \equiv (2\pi)^{-1/2} \int_{-\infty}^{y_i{}^{p}} e^{-1/2x^2} \, dx \tag{22}$$

where $U_{kj}{}^{p}$ are marginally distributed random variables for the p^{th} site, k^{th} year and j^{th} month,

$X_{kj}{}^{p}$ is streamflow for p^{th} site, k^{th} year and j^{th} month, and

$y_i{}^{p}$ is normally distributed random variables for p^{th} site and i^{th} interval.

The lag zero and lag one cross correlations of the normally distributed $y_i{}^{p}$ values are then found. To generate normally distributed values, Moran suggests that canonical analysis be used. This procedure was not adopted here; instead the cross correlations were used to develop [A] and [B]. These matrices were then used in conjuction with Matalas' generating equation (Eq. 17).

It is unlikely that the cross correlations of the normally distributed data will be the same as those of the historical flows so that the matrices used to generate the residuals will be in error. The cross correlations of the residuals and not the historical flows will be the quantity preserved.

Table 1.
Streamflow characteristics of the Melbourne water supply system

River	Gauging site	Gauge number	Area (km^2)	Non-seasonal monthly				Max. mean	Min. mean	Annual		
				Mean	Coeff. of var.	Skew	Ser. corr.	Av. mean	Av. mean	Coeff. of var.	Skew	Ser. corr.
Yarra	Doctor's Creek	229100	332	14.9	0.98	1.6	0.64	1.9	0.22	0.41	1.0	0.12
O'Shannassy	O'Shannassy	229103	127	8.8	0.65	1.2	0.76	1.7	0.42	0.29	0.8	0.04
Watts	Maroondah	229105	104	6.7	0.69	1.5	0.67	1.6	0.51	0.32	0.7	0.16
Plenty	Toorourrong Reservoir	229110	93	2.0	0.66	1.8	0.67	1.5	0.58	0.38	0.9	0.37
Graceburn Creek	Graceburn	229108	23	1.5	0.68	1.4	0.68	1.5	0.43	0.34	0.5	0.24
Coranderrk Creek	Coranderrk	229104	18	1.6	0.67	1.3	0.64	1.5	0.39	0.28	0.5	0.02
Thomson	Aberfeldy	–	347	13.9	0.86	1.8	0.52	1.9	0.35	0.37	1.4	0.16

Mean is in units of millions of cubic metres.

Table 2.
Non-seasonal and annual parameters based on historical and single site generated flows for O'Shannassy River

Model	Monthly non-seasonal				Annual			
	Mean	Std. dev.	Skew	Ser. corr.	Mean	Std. dev.	Skew	Ser. corr.
HIST	8.8	5.7	1.2	0.76	106.0	31.0	0.7	0.04
SKEWED	8.8	5.8	1.6	0.78	106.0	31.0	1.1	0.13
	(0.4)	(0.6)	(0.7)	(0.03)	(4.9)	(6.4)	(0.9)	(0.13)
LN-3	9.1	7.3	2.3	0.74	109.0	42.0	0.9	0.08
	(0.5)	(0.7)	(0.5)	(0.02)	(5.9)	(6.0)	(0.5)	(0.15)
LN-2	8.9	5.9	1.5	0.78	106.0	32.0	1.0	0.06
	(0.5)	(0.5)	(0.3)	(0.02)	(5.3)	(4.2)	(0.4)	(0.15)
GAMMA	8.5	5.6	1.2	0.77	102.0	28.0	0.5	0.07
	(0.2)	(0.2)	(0.4)	(0.02)	(3.2)	(3.0)	(0.4)	(0.16)
EMP	8.8	5.7	1.2	0.75	105.0	28.0	0.5	0.07
	(0.4)	(0.4)	(0.2)	(0.01)	(4.8)	(2.8)	(0.3)	(0.14)

Gamma parameters are median values of ten replicates of 50 years' length. Other generated parameters are means of fifty replicates of 50 years' length. Means and standard deviations are in units of millions of cubic metres. Values in brackets are standard deviations of appropriate generated parameters.

Once the residuals have been generated, they are transformed back to alternative historical flows by applying first the inverse of Eq. (22) then the inverse of Eq. (21).

The main disadvantage with this procedure is the time required to set up the twelve monthly tables for U_j^p and X_{kj}^p and then to search through the tables during the inverse transformation.

RESULTS

The five single site models – empirical, skewed log data, two- and three-parameter log normal, and gamma* – were applied to the O'Shannassy River data,

* Abbreviations of model names used in tables are empirical EMP, two-parameter log normal LN-2, three-parameter log normal LN-3, skewed log data SKEWED, gamma GAMMA, square root SQRT.

Table 3.
Monthly parameters based on historical and single site generated flows for O'Shannassy River.

Statistic	Model	J	F	M	A	M	J	J	A	S	O	N	D
Mean	HIST	5.4	3.8	3.1	4.2	6.5	8.8	12.1	14.9	14.3	13.0	10.6	8.3
	SKEWED	5.4	3.8	3.7	4.2	6.7	8.8	12.2	14.9	14.3	13.0	10.5	8.1
	LN-3	5.7	4.4	3.8	4.2	6.4	8.5	13.3	16.9	14.3	13.3	10.7	7.5
	LN-2	5.4	3.9	3.8	4.2	6.4	8.5	12.5	15.3	14.7	13.1	10.6	8.0
	GAMMA	5.3	3.7	3.6	4.1	6.3	8.0	12.1	14.7	14.3	13.0	10.9	8.1
	EMP	5.4	3.8	3.7	4.1	6.4	8.5	11.8	14.8	14.2	13.0	10.7	8.4
Standard deviation	HIST	2.0	1.1	1.2	1.9	3.7	5.1	5.2	5.7	5.2	4.9	4.7	4.8
	SKEWED	2.0	1.0	1.2	1.9	4.3	5.3	5.6	6.0	5.3	4.8	4.6	4.2
	LN-3	2.7	3.3	1.7	1.6	3.6	4.6	9.3	13.4	5.3	6.0	4.6	2.7
	LN-2	2.3	1.5	1.4	1.7	3.3	4.6	6.3	6.8	6.2	5.1	4.3	3.9
	GAMMA	2.0	1.1	1.2	1.6	3.1	4.2	5.3	5.3	4.8	5.1	4.7	3.9
	EMP	2.0	1.1	1.2	1.7	3.6	4.8	9.3	5.6	5.2	4.9	4.8	4.8
Skew	HIST	1.2	0.5	0.8	1.6	1.7	2.0	0.9	0.4	1.0	0.7	0.9	3.4
	SKEWED	1.3	0.5	0.8	1.4	2.4	1.9	1.3	1.0	1.1	0.5	1.0	1.9
	LN-3	1.3	1.2	0.9	1.0	1.3	1.4	1.3	1.1	0.9	0.7	0.8	1.4
	LN-2	1.2	1.0	0.8	1.1	1.3	1.4	1.4	1.2	1.2	0.9	1.0	1.3
	GAMMA	0.8	0.5	0.5	1.0	0.8	0.9	0.8	0.5	0.4	0.8	0.8	1.3
	EMP	0.8	0.8	0.7	0.8	0.8	0.9	1.0	0.8	0.9	0.9	0.9	0.9
Serial correlation	HIST	0.86	0.64	0.56	0.44	0.64	0.82	0.68	0.73	0.61	0.76	0.66	0.70
	SKEWED	0.86	0.69	0.59	0.39	0.77	0.87	0.67	0.74	0.77	0.77	0.73	0.90
	LN-3	0.98	0.81	0.42	0.40	0.61	0.94	0.94	0.73	0.75	0.74	0.33	0.85
	LN-2	0.92	0.68	0.51	0.33	0.62	0.87	0.79	0.80	0.66	0.75	0.50	0.80
	GAMMA	0.73	0.76	0.66	0.69	0.74	0.74	0.70	0.71	0.69	0.74	0.69	0.74
	EMP	0.66	0.69	0.66	0.68	0.67	0.66	0.69	0.70	0.66	0.68	0.70	0.67

Gamma parameters are median values of ten replicates of 50 years' length. Other generated parameters are mean values of 50 replicates of fifty years' length. Means and standard deviations are in units of millions of cubic metres.

Table 4.
Storage estimates for O'Shannassy River

Model	Single site	Multi-site
HIST	8.6 (Based on 59 years)	
SKEWED	+ 1 % 8.1 — 16 %	–
LN-3	–	+ 37 % 9.4 + 3 %
LN-2	+ 11 % 8.4 — 11 %	+ 40 % 9.0 + 1 %
GAMMA	+ 19 % 8.8 — 13 %	+ 31 % 9.8 — 1 %
EMP	+ 20 % 9.9 — 3 %	–

Storage estimates for 50 % draft and 1 % probability of failure are in millions of cubic metres. Generated estimates are the median values of ten replicates of 50 years' length. Percentages represent the variation of eight of ten generated estimates about the historical storage estimate.

one stream in the Melbourne water supply system. (For characteristics, see Table 1.) Using each model, fifty replicates of 50 years' length were generated and appropriate monthly and annual statistics calculated (Tables 2 & 3). In addition, storage estimates were calculated for 10 replicates of 50 years of synthesized data (Table 4). A slightly modified version of Gould's monthly procedure (Joy & McMahon 1972) was used to determine the storage capacity required to yield a constant draft equivalent to 50 % of the mean monthly flow and a 1 % probability of failure. Of the five models examined, the skewed log data and two-parameter log normal models perform better overall than the others, and within the limits of this analysis appear to be satisfactory procedures for generating monthly flows.

The four multi-site procedures – two- and three-parameter log normal,

Table 5.
Non-seasonal and annual parameters based on historical and multi-site generated flows for O'Shannassy River

Model	Monthly non-seasonal				Annual			
	Mean	Std. dev.	Skew	Ser. Corr.	Mean	Std. dev.	Skew	Ser. Corr.
HIST	8.8	5.7	1.2	0.76	106	31	0.7	0.04
LN-3	8.4	5.8	1.3	0.78	103	31	0.7	0.02
	(0.5)	(0.5)	(0.4)	(0.02)	(5.4)	(4.1)	(0.4)	(0.10)
LN-2	8.9	6.0	1.4	0.79	107	35	0.9	–0.02
	(0.6)	(0.7)	(0.2)	(0.01)	(7.5)	(6.2)	(0.4)	(0.13)
GAMMA	8.5	5.3	1.0	0.77	102	25	0.42	0.03
	(0.5)	(0.4)	(0.1)	(0.01)	(6.2)	(4.1)	(0.5)	(0.08)
SQRT	4.9	6.8	2.3	0.65	106	47	1.2	–0.04
	(1.0)	(1.1)	(0.3)	(0.06)	(11.7)	(10.1)	(0.6)	(0.09)

Generated parameters are median values of ten replicates of 50 years' length. Means and standard deviation are in units of millions of cubic metres. Values shown in brackets are standard deviations of appropriate generated parameters.

square root, and gamma models – were applied to the seven streams forming the Melbourne water supply system. Ten replicates, each 50 years long, were generated at each site. However, results for only one site, O'Shannassy, are reported. Monthly non-seasonal and annual statistics are summarized in Table 5 where standard deviations of the generated statistics are also shown. These are based, however, on ten items of data and, therefore, provide only a general indication of the variability in the generated statistics. Monthly variations are given in Table 6. Multi-site models are complicated because of the need to preserve lag zero and lag one cross correlation. Correlations between O'Shannassy and the six other streams are given in Table 7. For the same conditions as adopted previously, storage estimates are set out in Table 4.

Based on the generated parameters, the two parameter log normal model performs more satisfactorily than the other models. The gamma storage values are higher than the storage estimates determined using the other generated sequences.

Table 6.
Monthly parameters based on historical and multi-site generated flows for O'Shannassy River

Statistic	Model	J	F	M	A	M	J	J	A	S	O	N	D
Mean	HIST	5.4	3.8	3.7	4.2	6.5	8.8	12.1	14.9	14.3	13.0	10.6	8.3
	LN-3	5.3	3.6	3.6	4.1	6.3	8.5	12.0	14.8	13.9	12.6	10.4	7.6
	LN-2	5.6	3.8	3.6	4.2	6.5	8.5	11.7	14.8	14.2	13.2	11.0	8.9
	GAMMA	5.4	3.8	3.7	4.2	6.4	8.3	11.5	14.1	13.7	13.2	10.5	8.4
	SQRT	3.2	2.5	2.1	2.1	3.9	4.7	6.0	8.3	7.9	7.9	6.2	5.4
Standard deviation	HIST	2.0	1.1	1.2	1.9	3.7	5.1	5.2	5.7	5.2	4.9	4.7	4.8
	LN-3	2.0	1.0	1.2	1.7	3.6	4.9	5.6	6.3	4.9	4.9	4.7	3.8
	LN-2	2.2	1.1	1.5	2.1	4.4	5.6	5.9	6.2	5.8	5.4	5.2	5.1
	GAMMA	1.9	1.1	1.1	1.7	3.2	4.1	4.7	5.4	4.6	4.9	4.4	3.7
	SQRT	3.3	1.7	2.0	2.7	5.3	7.5	7.4	9.1	9.0	8.3	6.9	6.2
Skew	HIST	1.2	0.5	0.8	1.6	1.7	2.0	0.9	0.4	1.0	0.7	0.9	3.4
	LN-3	1.4	0.5	0.7	1.3	1.3	1.9	0.9	0.5	0.7	0.6	0.8	2.6
	LN-2	1.4	0.8	1.2	1.3	1.7	1.9	1.5	1.1	0.9	0.9	1.2	1.7
	GAMMA	0.5	0.3	0.6	0.8	0.9	0.7	0.7	0.7	0.4	0.5	1.0	0.8
	SQRT	1.4	1.0	1.3	1.7	1.9	2.6	2.0	1.5	1.2	1.6	2.0	2.1
Serial Correlation	HIST	0.86	0.64	0.56	0.44	0.64	0.82	0.68	0.73	0.61	0.76	0.66	0.70
	LN-3	0.76	0.77	0.78	0.79	0.79	0.73	0.77	0.73	0.75	0.79	0.72	0.74
	LN-2	0.81	0.82	0.81	0.80	0.76	0.76	0.80	0.77	0.79	0.80	0.81	0.77
	GAMMA	0.64	0.73	0.70	0.68	0.72	0.63	0.72	0.68	0.75	0.70	0.71	0.69
	SQRT	0.63	0.71	0.61	0.66	0.64	0.66	0.67	0.67	0.71	0.65	0.66	0.52

Generated parameters are median values of ten replicates of 50 years' length.
Means and standard deviations are in units of millions of cubic metres.

Table 7.
Spatial correlations between O'Shannassy and other rivers shown

Model	Yarra		O'Shannassy		Watts		Plenty		Graceburn		Coranderrk		Thomson	
	Lag 0	Lag 1	Lag 0	Lag 1	Lag 0	Lag 1	Lag 0	Lag 1	Lag 0	Lag 1	Lag 0	Lag 1	Lag 0	Lag 1
HIST	0.94	0.67	1.0	0.76	0.91	0.64	0.74	0.61	0.84	0.66	0.87	0.61	0.84	0.59
LN-3	0.96	0.75	1.0	0.78	0.93	0.72	0.77	0.64	0.89	0.71	0.90	0.70	0.86	0.65
LN-2	0.96	0.76	1.0	0.79	0.93	0.74	0.78	0.66	0.89	0.71	0.91	0.72	0.88	0.70
GAMMA	0.93	0.70	1.0	0.77	0.92	0.69	0.75	0.62	0.84	0.66	0.89	0.66	0.84	0.59
SQRT	0.79	0.49	1.0	0.65	0.82	0.54	0.63	0.46	0.71	0.48	0.77	0.51	0.73	0.46

Correlations are median values of ten replicates of 50 years' length.

CONCLUSIONS

The following conclusions are drawn from this study of one streamflow system. For the single-site analysis only one river of this system has been used. Although some models have been found to be unsatisfactory, we are not necessarily implying that they will be inadequate in other generation studies.

Single Site

1. All models except the three-parameter log normal generate sequences that closely resemble the historical sequences in terms of monthly and non-seasonal means and standard deviations.

2. No model adequately simulates skewness.

3. The skewed log data model generates monthly serial correlation estimates closer to the historical values than those generated by the other models. The gamma and empirical models produce near constant values.

4. Storage estimates based on 50 years of generated data yield median values ranging for the models from –6 % to +14 % of the historical estimates.

Multi-site

1. The square root model is unacceptable and is excluded from the remaining conclusions.

2. All models preserve monthly means and standard deviations.

3. The two- and three-parameter log normal models reproduce the seasonal variation of skew. A feature of the gamma model is the low value of generated monthly skews.

4. No model preserves the seasonal variation of the serial correlation.

5. All models preserve the historical lag zero and lag one cross correlations between sites.

6. A feature of the multi-site models is that they all produce storage estimates which are high relative to the historical estimate.

ACKNOWLEDGEMENT

Part of this project was sponsored by the Water Research Foundation of Australia. Streamflow data was provided by the Melbourne Metropolitan Board of Works.

REFERENCES

Anderson, R. L. (1941) Distribution of the serial correlation coefficient. *Ann. Math. Stat. 13,* 1–13.

Bendat, J. S. & Piersol, A. G. (1966) *Measurement and analysis of random data.* J. Wiley & Sons, N.Y.

Blackman, R. B. & Tukey, J. W. (1958) *The measurement of power spectra from the point of view of communications engineering.* Dover Publications Inc., N.Y.

Burges, S. J. (1970) *Use of stochastic hydrology to determine storage requirements for reservoirs – A critical analysis.* Report EEP-34, Stanford University.

Fiering, M. B. (1964) Multivariate technique for synthetic hydrology. *A.S.C.E. J. Hydraul. Div. 90* (5), 43–60.

Fiering, M. B. (1967) *Streamflow synthesis.* Macmillan, 139 pp.

Hufschmidt, M. M. & Fiering, M. B. (1966) *Simulation techniques for design of water-resource systems.* Harvard University Press, Cambridge, Massachusets.

Jennings, M. E. (1969) Discussion on operational hydrology using residuals by Young and Pisano. *A.S.C.E. J. Hydraul. Div. 95* (4), 1468–1469.

Joy, C. S. & McMahon, T. A. (1972) Reservoir-yield estimation procedures. *I.E.Aust. Civil Eng. T.* (in press).

Julian, P. R. (1961) *A study of the statistical predictability of stream-runoff in the Upper Colorado River Basin.* University of Colorado.

Kendall, M. G. & Stuart, A. (1968) *The advanced theory of statistics.* Vol. 3, 2nd Edition, Charles Griffin and Co. Ltd., London.

Kottegoda, N. T. (1970) Statistical methods of river flow synthesis for water resources assessment. *I.C.E.,* Suppl. 18, Paper 7339 S.

Maass, A. et al. (1962) *Design of water resources systems.* Harvard University Press, 620 pp.

Mandelbrot, B. B. & Wallis, J. R. (1968) Noah, Joseph and operational hydrology. *Water Resources Research* (5), 909–918.

Mandelbrot, B. B. & Wallis, J. R. (1969) Computer experiments with fractional Gaussian noises. *Water Resources Research 5 (1),* 260–267.

Massey, F. J. Jr. (1951) The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Stat. Assoc. 46,* 68.

Matalas, N. (1967) Mathematical assessment of synthetic hydrology. *Water Resources Research 3 (4),* 937–945.

McMahon, T. A., Codner, G. P. & Philips, C. (1972) *Stochastic generation of streamflow.* Civil Engineering Research Report No. 2, Monash University.

McMahon, T. A. & Miller, A. J. (1971) Application of Thomas and Fiering model to skewed data. *Water Resources Research 7* (5), 1338–1340.

Moran, P. A. P. (1970) Simulation and evaluation of complex water systems operation. *Water Resources Research 6* (6), 1737–1742.

Moran, P. A. P. (1969) Statistical inference with bivariate gamma distribution. *Biometrika 56,* 627–634.

Parzen, E. (1967) *Time series analysis papers. Holden-Day,* 142–143.

Quenouille, H. H. (1947) A large-sample test for the goodness of fit of autoregressive schemes. *J. Roy. Stat. Soc. 110,* 123–129.

Thomas, H. A. & Fiering, M. B. (1962) Mathematical synthesis of streamflow sequences for the analysis of river basins simulation. In: Maass et al.: *Design of water resources systems.* Harvard University Press, Chapter 12.
Thom, H. G. S. (1958) A note on the gamma distribution. *Monthly Weather Rev. 86 (4),* 117–122.
Young, G. K. & Pisano, W. C. (1968) Operational hydrology using residuals. *A.S.C.E. J. Hydraul. Div. 94* (4), 909–923.

Received 1 April 1972.

Address:
Dr. T. A. McMahon,
Department of Civil Engineering,
Monash University,
Clayton, Victoria, 3168,
Australia.