CHAPTER FOUR

SHORT MEMORY-ANNUAL FLOW MODELS

Several inherent inadequacies associated with using only a historical sequence in design. Engineers and hydrologists utilize long synthetic sequences which resemble the observed historical record. Even though at present serveral models exist for data generation, Markov and ARMA models have been used widely because of their simplicity. They preserve the low-order moments well but fail to generate events more extreme than those observed in the historical record. Another criticism levelled againt Markov models is their failure to preserve the long-term persistence features observed in the historical sequences. Nevertheless, because of the complexity and excessive computer time involved in operating other models (such as fractional Gaussian noise processes and Broken Line processes, described in the next chapter.), which preserves the long-term persistence effect, Markov and ARMA models are being used extensively for data generation. The aim of this chapter is to compare synthetic sequences obtained from Markov and ARMA models using various generation procedures and some modifications.

4.1 MARKOV MODEL

Markov or autoregressive (AR) models essentially relates the present performance, X_t, of a system to that which occurred at some time, or set of times, in the immediate past and also to a random component ϵ_t. An AR model of order p can be witten as

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \ldots + \Phi_p Z_{t-p} + \epsilon_t \qquad \ldots\ldots\ldots (4.1.1)$$

$$Z_t = (X_t - \bar{X})/s \qquad \ldots\ldots\ldots (4.1.1a)$$

$$\bar{X} = (1/N)\sum_{t=1}^{N} X_t \qquad \text{.........} \quad (4.1.1b)$$

$$s = [1/(N-1)\sum_{t=1}^{N} (X_t-\bar{X})^2]^{1/2} \qquad \text{........} \quad (4.1.1c)$$

Where Φ_i, i=1,2,3,...,p , $\bar{X}$ and s are the autoregressive parameters or weights, mean and standard diviation respectively. The properties of Z_t and ϵ_t are defined by

$$E[Z_t] = E[\epsilon_t] = 0$$
$$var[Z_t] = E[Z_t^2] = \sigma_z^2$$
$$var[\epsilon_t] = E[\epsilon_t^2] = \sigma_e^2 \qquad \text{..........} \quad (4.1.2)$$
$$\rho_k = E[Z_t Z_{t-k}]/\sigma_Z^2 \qquad \text{for } k = 1, 2, 3, ...$$
$$E[\epsilon_t . \epsilon_{t-k}] = E[\epsilon_t . Z_{t-k}] = 0$$

The last equality merely signifies that the current random number is independent of past values of the process. As in previous notation, E denotes the expected or mean value of the term within the parentheses and the variance is abbreviated to var.

Henceforth it is assumed without loss of generality that the variance of the stochastic component is equal to one. This means that, in application, the numbers generated through such equations as Eq.(4.1.1) should be multiplied by the standard deviation of the variable modelled and the mean should then be added to each number. These and other model parameters are estimated from observed sequences.

4.1.1 Estimation of Autoregressive Parameters

If Eq.(4.1.1) is multiplied by Z_{t-1} and expectation are taken,

$$E(Z_t Z_{t-1}) = \Phi_1 E(Z_{t-1}Z_{t-1}) + \Phi_2 E(Z_{t-2}Z_{t-1}) + \Phi_3 E(Z_{t-3}Z_{t-1}) + .. + \Phi_p E(Z_{t-p}Z_{t-1}) + E(\epsilon_t Z_{t-1}) \quad .... \quad (4.1.3)$$

Because $E(\epsilon_t Z_{t-1}) = 0$ and on account of the other properties given above, it follow that

$$\rho_1 = \Phi_1 + \Phi_2\rho_1 + \Phi_3\rho_2 + \ ... \ + \Phi_p\rho_{p-1} \quad .... \quad (4.1.4)$$

Furthermore, if Eq.(4.1.1) is multiplied by $Z_{t-2}, Z_{t-3}, \ldots, Z_{t-p}$ in turn and if expectations are taken after each multiplication, p relationships called the Yule-Walker equations are obtained. These can be represented in matrix form by

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots\cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots\cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdots\cdots & \rho_{p-3} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots\cdots & 1 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \cdot \\ \cdot \\ \cdot \\ \Phi_p \end{bmatrix} \quad (4.1.5)$$

or briefly

$$\rho_p = P_p \Phi_p \quad .... \quad (4.1.5a)$$

A necessary condition for stationarity is that the autocorrelation matrix P_p is positive definite, that is, the determinant and all its principal minors are more than zero. Therefore,

$$\Phi_p = P_p^{-1}\rho_p \quad .... \quad (4.1.6)$$

In order to solve Eq.(4.1.5), autocorrelation coefficients r_1, $r_2, \ldots, r_p$ are substituted in P_p and ρ_p, and hence the Yule-Walker estimates of the autoregressive parameters $\Phi_1, \Phi_2, \ldots, \Phi_p$ are obtained.

By squaring both sides of Eq. (4.1.1) and taking expectations it follows that, because each of the expected values $E(\epsilon_t . Z_{t-1})$, $E(\epsilon_t . Z_{t-2})$, ... , $E(\epsilon_t . Z_{t-p})$ are equal to zero

$$1 = \Phi_1\rho_1 + \Phi_2\rho_2 + \Phi_3\rho_3 + \ldots + \Phi_p\rho_k + \sigma_e^2 \quad \ldots\ldots\ (4.1.7)$$

Hence, σ_e^2 which is the variance of the independent variables ϵ_t is give by

$$\sigma_e^2 = 1 - \Phi_1\rho_1 - \Phi_2\rho_2 - \Phi_3\rho_3 - \ldots - \Phi_p\rho_p \quad \ldots\ldots\ (4.1.8)$$

or

$$\sigma_e^2 = 1 - R^2 \quad \ldots\ldots\ldots\ (4.1.8a)$$

Where

$$R^2 = \Phi_1\rho_1 + \Phi_2\rho_2 + \Phi_3\rho_3 + \ldots + \Phi_p\rho_p \quad \ldots\ldots\ (4.1.8b)$$

is called the coefficient of determination or the square of the multiple correlation coefficient

4.1.1.1 First-Order Model, AR(1)

The first-order autoregressive model AR(1) is known familiarly as a Markov model. It is given by

$$Z_t = \Phi_1 Z_{t-1} + \epsilon_t \quad \ldots\ldots\ldots\ldots\ (4.1.9)$$

if this equation is multiplied throughout by Z_{t-1} and if expectations are taken, because $E(\epsilon_t Z_{t-1}) = 0$ in addition to the assumptions $E(Z_t)^2 = 1$ and $E(Z_t) = 0$, it follows that

$$\Phi_1 = \rho_1 \qquad \text{.................} \qquad (4.1.10)$$

where ρ_1 is the lag-one autocorrelation coefficient and $-1<\rho_1<1$. Also from Eq. (4.1.8) the variance of the independent variables ϵ_t is given by

$$\sigma_e^2 = 1 - \rho_1^2 \qquad \text{................} \qquad (4.1.11)$$

Again, if Eq.(4.1.9) is multiplied in turn by Z_{t-1}, Z_{t-2}, Z_{t-3},... and Z_{t+1}, Z_{t+2}, Z_{t+3}, ...

$$\rho_k = \rho_1^{|k|}, \quad k = 0, \pm1, \pm2, \pm3 \qquad \text{............} \qquad (4.1.12)$$

4.1.1.2 Second-Order Model, AR(2)

Second order is usually the highest lag necessary in representing hydrologic time series. The model takes the form

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \epsilon_t \qquad \text{.............} \qquad (4.1.13)$$

The stationary constraints on Φ_1 and Φ_2 are determined as follows:

$$\begin{aligned} \Phi_2 + \Phi_1 &< 1 \\ \Phi_2 - \Phi_1 &< 1 \qquad \text{..............} \qquad (4.1.14) \\ -1 < \Phi_2 &< 1 \end{aligned}$$

Fig.(4.1) shows the triangular parameter space defined by Eq. (4.1.14). The Yule-Walker equations for the AR(2) model are

$$\begin{aligned} \rho_1 &= \Phi_1 + \Phi_2\rho_1 \\ \rho_2 &= \Phi_1\rho_1 + \Phi_2 \qquad \text{..............} \qquad (4.1.15) \end{aligned}$$

which when solved simultaneously yield

$$\Phi_1 = \rho_1(1-\rho_2)/(1-\rho_1^2)$$
$$\Phi_2 = (\rho_2-\rho_1^2)/(1-\rho_1^2) \qquad \text{............} \quad (4.1.16)$$

or inversly $\rho_1 = \Phi_1/(1-\Phi_2)$

$$\rho_2 = \Phi_2 + \Phi_1^2/(1-\Phi_2) \qquad \text{............} \quad (4.1.17)$$

Fig.(4.2) gives the solution of Eq. (4.1.16) for various values of correlation coefficients ρ_1 and ρ_2. In parctice, sample estimates could be used for the correlation in order to obtain parameter values.

The parameter limits given in Eq. (4.1.14) and the relations between correlations and parameter, given in Eq. (4.1.17) define a region of valid correlation coefficients for the lag-two autoregressive model

$$-1 < \rho_1 < 1$$
$$-1 < \rho_2 < 1 \qquad \text{................} \quad (4.1.18)$$
$$\rho_1^2 < 1/2(\rho_2+1)$$

The admissible regions of parameters and correlations are shown in Fig.(4.1). Notice that this figure could be used as a first -cut criteria for the possible use of an AR(2) with a given set of data.

4.1.2 Application of Markov Model

4.1.2.1 Lag One Markov Model

The lag one Markov model is defined as

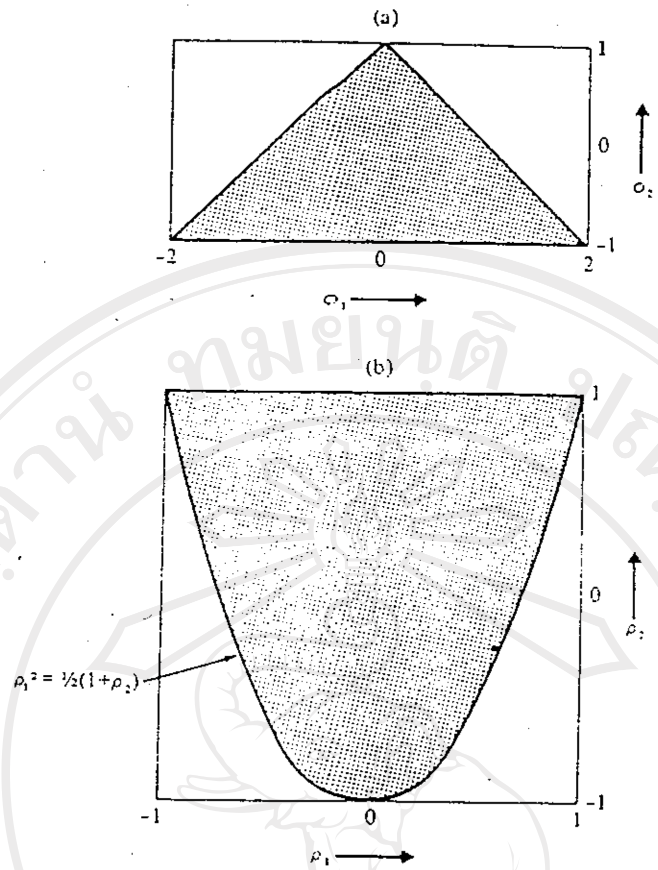

Figure 4.1 Valid regions of the parameters and correlations of a stationary AR(2) process (from Box and Jenkins, 1976)

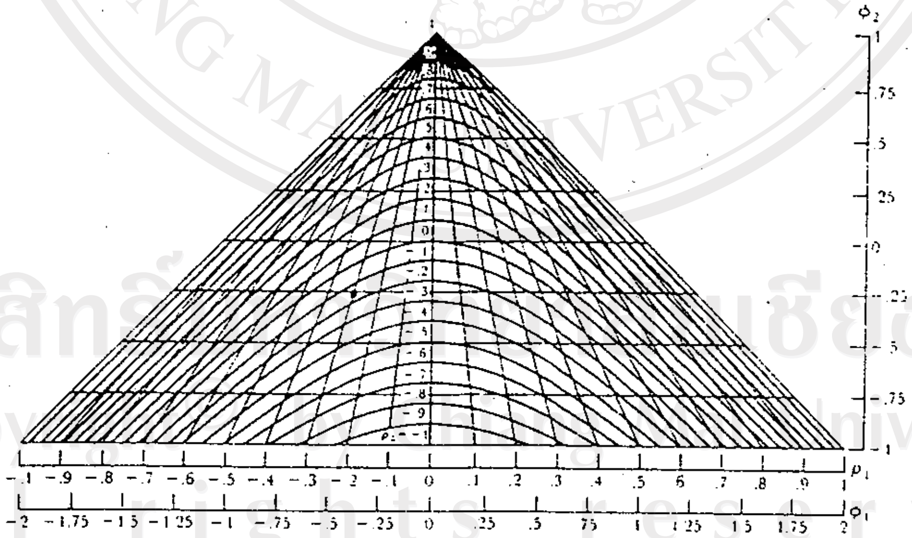

Figure 4.2 Relation between correlations and parameters of an AR(2) model. Diagram may be used for parameter estimation using the method of moments. (From Box and Jenkins, 1976.)

$$Z_t = \rho Z_{t-1} + (1-\rho^2)^{1/2}\epsilon_t \qquad \text{.............. (4.1.19)}$$

where Z_t = flow at time t standardised to have zero mean and unit variance

ρ = lag one autocorrelation coefficient, and

ϵ_t = independent and identifically distributed random numbers having zero mean and unit variance.

If one is interested in generating normally distributed flows, Eq.(4.1.19) can be used with ϵ_t sampled from a normal distribution.

4.1.2.2 Modifications to Account for the Skewness

Skewed flows can be generated by modifying ϵ_t in Eq.(4.1.19) There are several ways to do this as discussed in the following sections.

a) Wilson-Hilferty Transformation

To account for skewness, Thomas and Fiering (1962) replaced the random component ϵ_t by using the Wilson - Hilferty transformation as follows:

$$\xi_t = 2/\gamma(\xi)[1+\gamma(\xi)\epsilon_t/6-\gamma(\xi)^2/36]^3-2/\gamma(\xi) \qquad \text{.......... (4.1.20)}$$

where the skewness of ξ_t denoted by $\gamma(\xi)$ is related to the skewness of X denoted as γ_x by

$$\gamma(\xi) = (1-\rho^3)/(1-\rho^2)^{3/2}\gamma_x \qquad \text{.............. (4.1.21)}$$

If ϵ_t is assumed to be normally distributed with zero mean and unit variance, then ξ_t is approximately distributed as Gamma with zero mean, unit variance and skewness $\gamma(\xi)$. With ξ_t as the random component, the lag one Markov process may be used to generate synthetic events that will resemble the historic events in terms of the first three moment and lag one autocorrelation coefficient.

b) Kirby's Modification

McMahon and Miller (1971) showed this transformation breaks down for skewness larger than 2. However, Kirby (1972) provided a modified Wilson-Hilferty transformation which theoretically remains satisfactory over the whole rang of hydrologic interest. Kirby's modification is as follow:

$$\xi_t^m = A\{\max[E, 1+G\epsilon_t/6-(G/6)^2]^3-B\} \quad \ldots\ldots\ldots\ldots \quad (4.1.22)$$

where

$$E = (B-2/(\gamma(\xi)A))^{1/3} \quad \ldots\ldots\ldots\ldots \quad (4.1.23)$$

and A, B and G are given by Kirby (1972) in terms of skewness.

c) Logarithmic Transformation

Synthetic flows which conform to a three parameter log-nomal distribution and which resemble historical flow in terms of lower order moments and serial correlation may be generated as follows (Matalas, 1967b).

If A is assumed to be the lower bound of the variate X_t where (X_t-A) is log normally distributed, then $Y_t = \ln(X_t-A)$ is normally distributed. The mean μ_x, variance σ_x^2, skewness γ_x, and the lag one autocorrelation ρ_x, of the historical data (X_t) are related to the lower bound A, mean μ_y, variance σ_y^2 and lag one autocorrelation coefficient, ρ_y of Y_t by

$$\mu_x = A + \exp(\mu_y + \sigma_y^2/2) \quad .... \quad (4.1.24)$$

$$\sigma_x^2 = \exp\{2(\mu_y + \sigma_y^2)\} - \exp\{2\mu_y + \sigma_y^2\} \quad .... \quad (4.1.25)$$

$$\gamma_x = \{\exp(3\sigma_y^2) - 3\exp(\sigma_y^2) + 2\}/\{\exp(\sigma_y^2) - 1\}^{3/2} \quad (4.1.26)$$

$$\rho_x = (\exp(\sigma_y^2 \rho_y) - 1)/(\exp(\sigma_y^2) - 1) \quad .... \quad (4.1.27)$$

To solve for A, μ_y, σ_y and ρ_y, one begins with Eq. (4.1.26) solves for σ_y. Since this is not explicit in σ_y, an iterative solution such as Newton-Raphson method is required. Once σ_y is computed, μ_y, ρ_y and A can be obtained from Eq. (4.1.24), (4.1.25) and (4.1.27)

An alternative method for determining the parameters is developed by Fiering and Jackson (1971). The method propose a slowly converging iterative solution for the y parameters of the system of Eq. (4.1.24), (4.1.25), (4.1.26) and (4.1.27), but a direct solution is possible by making the substitution

$$\Phi = \exp[\sigma_y^2] \quad .... \quad (4.1.28)$$

Eq. (4.1.26) becomes

$$\gamma_x = (\Phi^3 - 3\Phi + 2)/(\Phi - 1)^{3/2} = (\Phi - 1)^{1/2}(\Phi + 2) \quad .... \quad (4.1.29)$$

When $\Phi = 1$. This last equation shows why the three-parameter log normal model is applicable only to distributions with positive coefficient of skewness: since by definition, Φ is always greater than or equal to 1, the right-hand side is always greater than zero. (The case of $\Phi = 1$ is excluded, since it implies that $\gamma_x = 0$, and the solution is symmetrical about the mean.) After squaring $\gamma_x = (\Phi-1)^{1/2}(\Phi+2)$, one can obtain

$$\Phi^3 + 3\Phi^2 - (4 + \gamma_x^2) = 0 \quad .... \quad (4.1.30)$$

For $\gamma_x > 0$, Eq.(4.1.30) has only one real root, given by Abramowitzand and Stegun (1972, Eq. 3.8.2) as

$$\Phi = [(1+\gamma_x^{2}/2)+(\gamma_x^{2}+\gamma_x^{2}/4)^{1/2}]^{1/3} + [(1+\gamma_x^{2}/2)-(\gamma_x^{2}+\gamma_x^{2}/4)^{1/2}]^{1/3} - 1 \quad \ldots\ldots\ldots \quad (4.1.31)$$

The y statics are easily found to be

$$\sigma_y^{2} = \log \Phi \quad \ldots\ldots\ldots\ldots\ldots \quad (4.1.32)$$

$$\mu_y = (1/2)\log [\sigma_x^{2}/(\Phi^{2}-\Phi)] \quad \ldots\ldots\ldots\ldots\ldots \quad (4.1.33)$$

$$A = \mu_x - [\sigma_x^{2}/(\Phi-1)]^{1/2} \quad \ldots\ldots\ldots\ldots\ldots \quad (4.1.34)$$

$$\rho_y = \{\log[\rho_x(\Phi-1)+1]\}/\log \Phi \quad \ldots\ldots\ldots\ldots\ldots \quad (4.1.35)$$

In Eq.(4.1.35), if $\rho_x < 0$ and $\Phi > (\rho_x-1)/\rho_x$, the argument of the numberator is less than zero, and ρ_y is undefined. If this condition did occur, a circumstance which is unlikely in practice, one would have to turn to use of the gamma distribution for reproducing the coefficient of skewness. From Eq.(4.1.34), one finds that a necessary condition for $A \neq 0$ is

$$(\Phi-1)^{1/2} \neq \sigma_x/\mu_x \quad \ldots\ldots\ldots\ldots \quad (4.1.36)$$

Whether A is greater or less than zero depends on the relative magnitudes of the coefficient of variation, σ_x/μ_x, and the coefficient of skewness, γ_x.

In the case of 2 parameter transformation $y_t = \ln(x_t)$, Eq. (4.1.26) is omitted from the set of transformation equations. In the case μ_y, σ_y and ρ_y can be solved explicitly from Eq. (4.1.24), (4.1.25) and (4.1.27). For this distribution, the skewness, C_s, is related to the coefficient of variation, C_v, in the following maner (Chow, 1964).

$$C_s = 3C_v + C_v^3 \qquad \text{..............} \quad (4.1.37)$$

d) **Beard's Procedure**

Beard (1972) introduced another procedure for generating streamflows:

Step 1. Compute the logarithm of each streamflow quantity. If one or more streamflow items are zero, a small increment, such as 0.1 percent of the mean annual flow is added to each quantity before taking the logarithm.

Step 2. Compute the mean($\bar{X}$), standard deviation (s) and coefficient of skewness (g) of the log values and stardandize them to have zero mean and unit variance

$$t_i = (x_i - \bar{X})/s \qquad \text{...............} \quad (4.1.38)$$

Step 3. Transform these standardized values to normal distribution using the inverse Wilson-Hilferty transformation as follows

$$K_i = 6/g\{(g/2t_i+1)^{1/3}-1\}+g/6 \qquad \text{..........} \quad (4.1.39)$$

where K_i is the normalized variate.

Step 4. Compute the lag one autocorrelation coefficient (r) of these normalized values and generate the standardized variates using

$$K_{i+1} = rK_i + (1-r^2)^{1/2}\epsilon_{i+1} \qquad \text{...........} \quad (4.1.40)$$

Step 5. Transform each generated variate by first inputing the skewness and then the mean and standard deviation

$$t_i = \{[g/6(K_i-g/6)+1]^3-1\}2/g \qquad \text{.........} \quad (4.1.41)$$

$$x_1 = \bar{x} + t_1 s \qquad \ldots\ldots\ldots\ldots \quad (4.1.42)$$

Step 6. Exponentiate the values obtained in step 5 and subtract the small increment added in step 1. If a negative value results, set it to zero.

4.2 AUTOREGRESSIVE-MOVING-AVERAGE MODELS, (ARMA)

Autoregressive and moving-average models can be combined to model processes that otherwise would be operationally impossible to represent with single finite AR or MA models. An ARMA(p,q) model takes the form

$$(1-\Phi_1 B-\Phi_2 B^2- \ldots -\Phi_p B^p)Z_t = (1-\Theta_1 B-\Theta_2 B^2- \ldots -\Theta_q B^q)\epsilon_t$$
$$\Phi(B)Z_t = \Theta(B)\epsilon_t \qquad \ldots\ldots\ldots\ldots \quad (4.2.1)$$

where B is the backward shift operator (i.e. $BZ_t = Z_{t-1}$)

The stationarity and invertibility conditions of the ARMA(p,q) model correspond to those of the component MA and AR models. For stationarity and invertibility, the roots of $\Phi(B)$ and of $\Theta(B)$ must lie outside the unit circle.

The autocovariance function is found by multiplying $Z_t = \Phi_1 Z_{t-1}+ \ldots + \Phi_p Z_{t-p}+\epsilon_t-\Theta_1\epsilon_{t-1}- \ldots -\Theta_q\epsilon_{t-q}$ by Z_{t-k} and finding expected values,

$$\gamma_k = E[Z_t Z_{t-k}] = \Phi_1\gamma_{k-1} + \ldots +\Phi_p\gamma_{k-p}+\gamma_{ze}(k)$$
$$-\Theta_1\gamma_{ze}(k-1)- \ldots -\Theta_q\gamma_{ze}(k-q), \quad \ldots\ldots \quad (4.2.2)$$

where

$$\gamma_{ze}(k) = E[Z_{t-k}\epsilon_t] \qquad \ldots\ldots\ldots \quad (4.2.3)$$
$$\gamma_{ze}(k-1) = E[Z_{t-k}\epsilon_{t-1}] \qquad \ldots\ldots\ldots \quad (4.2.4)$$

The value for $\gamma_{ze}(k)$ will be zero as long as $k > 0$, since it is not correlated for values of Z before t. The value for $\gamma_{ze}(k)$ will not be zero for $k \leqslant 0$.

With the above in mind, it should be clear that for $k > q$, the autocovariance (and autocorrelation) in Eq. (4.2.2) reduces to that of an AR(p) model:

$$\gamma_k = \Phi_1\gamma_{k-1} + \dots + \Phi_p\gamma_{k-p} \quad \text{for } k > q \quad \dots\dots \quad (4.2.4a)$$

$$\rho_k = \Phi_1\rho_{k-1} + \dots + \Phi_p\rho_{k-p} \quad \text{for } k > q \quad \dots\dots \quad (4.2.4b)$$

For values of k less than or equal to q, the autocovariance will be a function of the moving-average terms and will depend on all coefficients $\Phi_1,\dots,\Phi_p$, $\theta_1,\dots,\theta_q$, and the variance $\sigma_e^{\,2}$. The ARMA(p,q) model then has the convenient property that its first q autocorrelations depend on moving-average terms as well as autoregressive terms. After q lags, autoregressive behavior takes over from the last correlation value.

The variance of the process is given by Eq.(4.2.2) for $k = 0$. Evaluation of the variance requires the solution of $\gamma_1,\dots,\gamma_p$.

4.2.1 Estimation of ARMA(1,1) Parameters

The estimation of the parameters of the ARMA(p,q) model is not a straightforward procedure. An algorithm can be formulated for the purpose by following the method used for the basic ARMA(1,1) model in this section. With regard to the number of parameters, parsimony has been suggested. This means that p+q ought to be a minimum. For example, an ARMA(1,1) model is preferable to an AR(3) model if it is found that both types fit an observed sequence.

A popular, and useful, model in hydrology is

$$Z_t - \Phi_1 Z_{t-1} = \epsilon_t - \theta_1\epsilon_{t-1} \quad \dots\dots\dots\dots \quad (4.2.5)$$

$$(1-\Phi_1 B)Z_t = (1-\theta_1 B)\epsilon_t \quad \dots\dots\dots\dots \quad (4.2.6)$$

Stationarity and invertibility conditions correspond to the individual AR(1) and MA(1) models and so imply that the parameter region is

$$-1 < \Phi_1 < 1 \qquad \text{.........} \quad (4.2.7)$$

$$-1 < \theta_1 < 1 \qquad \text{.........} \quad (4.2.8)$$

Fig.(4.3) shows this admissible parameter space.

Using Eq. (4.2.2), the autocovariance function is

$$\gamma_0 = \Phi_1\gamma_1 + \sigma_e^2 - \theta_1\gamma_{ze}(-1) \qquad \text{........} \quad (4.2.9)$$

$$\gamma_1 = \Phi_1\gamma_0 - \theta_1\sigma_e^2 \qquad \text{........} \quad (4.2.10)$$

$$\gamma_k = \Phi_1\gamma_{k-1} \qquad k \geqslant 2. \qquad \text{........} \quad (4.2.11)$$

To obtain $\gamma_{ze}(-1)$ Eq. (4.2.5) is multiplied by ϵ_{t-1} and expectations are taken:

$$\gamma_{ze}(-1) = E[Z_t\epsilon_{t-1}] = (\Phi_1-\theta_1)\sigma_e^2 \qquad \text{..........} \quad (4.2.12)$$

Using Eq.(4.2.12) in Eq.(4.2.9) the autocovariance function of the process is obtained as

$$\gamma_0 = (1-\theta_1^2-2\Phi_1\theta_1)/(1-\Phi_1^2)\sigma_e^2 \qquad \text{...........} \quad (4.2.13)$$

$$\gamma_1 = (1-\Phi_1\theta_1)(\Phi_1-\theta_1)/(1-\Phi_1^2)\sigma_e^2 \qquad \text{...........} \quad (4.2.14)$$

$$\gamma_k = \Phi_1\gamma_{k-1} \qquad k \geqslant 2 \qquad \text{...........} \quad (4.2.15)$$

Note that the autocovariance will decay exponentially from a starting value γ_1, which is dependent on θ_1. The sign of γ_1 (and ρ_1) is defined by $\Phi_1-\theta_1$. The sign of Φ_1 determines if the correlation decay is smooth or alternates in sign.

The correlation function is given by

$$\rho_1 = (1-\Phi_1\theta_1)(\Phi_1-\theta_1)/(1+\theta_1^2-2\Phi_1\theta_1) \qquad \text{...........} \quad (4.2.16)$$

$$\rho_k = \Phi_1\rho_{k-1} \qquad k \geqslant 2 \qquad \text{...........} \quad (4.2.17)$$

The relationships shown in Eq. (4.2.16) and (4.1.17) and the invertibility -stationarity parameter space define an admissible region for the first two correlations.

$$|\rho_2| < |\rho_1| \qquad \ldots\ldots\ldots\ldots\ldots \quad (4.2.17a)$$

$$\rho_2 > \rho_1(2\rho_1+1) \qquad \rho_1 < 0 \qquad \ldots\ldots\ldots\ldots\ldots \quad (4.2.17b)$$

$$\rho_2 > \rho_1(2\rho_1-1) \qquad \rho_1 > 0 \qquad \ldots\ldots\ldots\ldots\ldots \quad (4.2.17c)$$

Fig.(4.3) illustrates the above region; correlation outside that space indicate that the ARMA(1,1) is not a good model. Fig.(4.4) diagrams the solution of parameter Φ_1 and θ_1 in terms of ρ_1 and ρ_2 as given by Eq.(4.2.16) and (4.2.17). Fig.(4.5) gives typical forms of the autocorrelation expected for various regions of the parameter space.

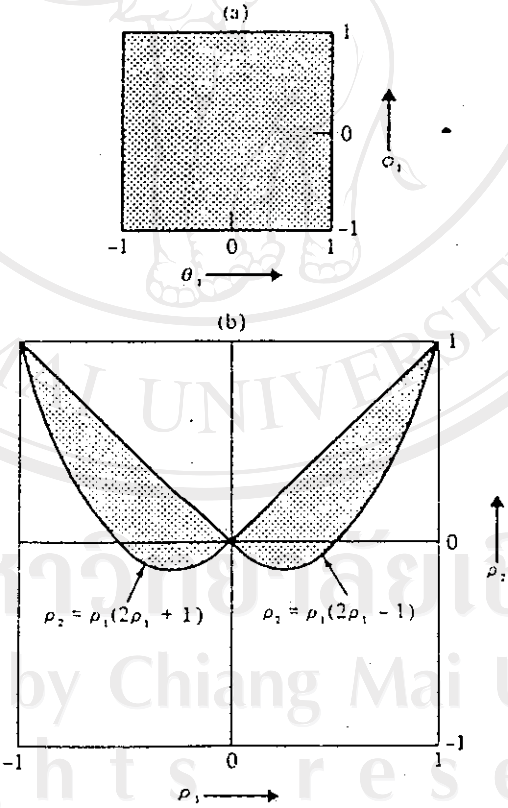

Figure 4.3 Valid regions for the parameters and correlations of a stationary and invertible ARMA(1,1) process(from Box and Jenkins, 1976).

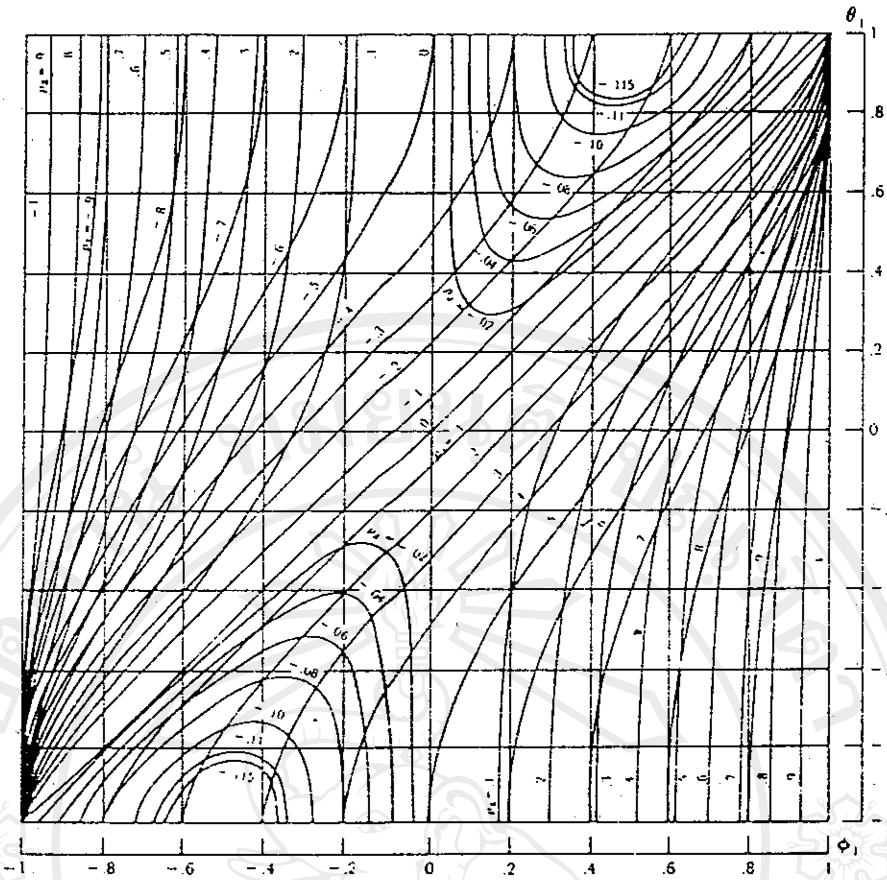

Figure 4.4 Relation between correlations and parameters for a stationary and invertible ARMA(1,1). Diagram may be used for parameter estimation using the method of moments. (From Box and Jenkins, 1976).

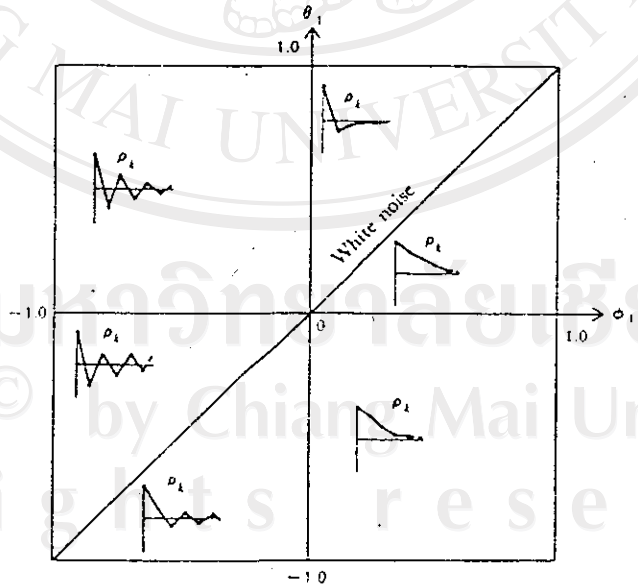

Figure 4.5 Autocorrelation function for various ARMA(1,1) models (from Box and Jenkins, 1976).

4.2.2 Application of ARMA(1,1) Model

Normally distributed streamflows X_t can be generated by using

$$X_t - \mu = \Phi(X_{t-1}-\mu) + \sigma\sigma_e(\epsilon_t - \Theta\epsilon_{t-1}) \qquad \ldots\ldots\ldots\ldots \quad (4.2.18)$$

where μ = mean of the flow series

σ = standard deviation of the flow series

ϵ_t = normally distributed independent random variable with zero mean and unit variance

and

$$\sigma_e^2 = (1-\Phi^2)/(1+\Theta^2-2\Phi\Theta) \qquad \ldots\ldots\ldots\ldots \quad (4.2.19)$$

4.2.3 Modifications to Account for the Skewness

a) Wilson-Hilferty Transformation

If the flow series X_t is required to have a skewness γ_x, then the terms ϵ_t in Eq. (4.2.18) must be replaced by a skewed random variable ξ_t given by

$$\xi_t = \{[1+\gamma_e\epsilon_t/6-\gamma_e^2/36]^3-1\}2/\gamma_e \qquad \ldots\ldots\ldots\ldots \quad (4.2.20)$$

where the skewness γ_e of the random variable ϵ_t is given by

$$\gamma_e = \gamma_x\{(1+\Phi^2-2\Theta\Phi)/(1-\Phi^2)\}/ \\ \{(1-\Theta^3+3\Phi\Theta^2-3\Phi^2\Theta)/(1-\Phi^3)\} \qquad \ldots\ldots\ldots\ldots \quad (4.2.21)$$

If the skewness have a value larger than 3 then use Kirby's modification.

b) Logarithmic Transformation

Another method to deal with skewed data is to use a log normal transformation. The mean, standard deviation , lag one autocorrelation coefficient in the log domain and the location parameter are obtained from Eq. (4.1.24) - (4.1.27).

In the case of three-parameter log normally distributed process with an ARMA(1,1) autocorrelation function in the normal domain the autocorrelation in the X domain is given by

$$\rho_x(1) = \{\exp(\rho_y(1)\sigma_y^{2})-1\}/\{\exp(\sigma_y^{2})-1\} \qquad \ldots\ldots\ldots\ldots \quad (4.2.22)$$

$$\rho_x(k) = \{\exp(\rho_y(1)\Phi_y^{|k|-1}\sigma_y^{2})-1\}/\{\exp(\sigma_y^{2}-1)\} \quad \ldots\ldots\ldots \quad (4.2.23)$$

where

$$\rho_y(1) = \{(1-\Phi_y\theta_y)(\Phi_y-\theta_y)\}/\{1+\theta_y^{2}-2\Phi_y\theta_y\} \qquad \ldots\ldots\ldots \quad (4.2.24)$$

As is true in the lag one Markov case, the nonlinear transformation results in an autocorrelation function in the X domain, which is distorted from the theoretical form for an ARMA(1,1) process. Since use of the ARMA(1,1) process to approximate self-similar hydrologic time series requires knowledge of the values of Φ_x and θ_x (the equivalent ARMA(1,1) parameters in the X domain) some equivalence must be established between the parameters in the X and Y domains. This equivalence will obviously be only approximate, since the three-parameter log normally distribution generated sequences will not be truly ARMA(1,1). The most straightforward approach is to establish the equivalence by equating the correlation coefficients of a (theoretical) log normal ARMA(1,1) process and those of the approximate sequence derived by transforming a normally distributed ARMA(1,1) sequence. If we assume that Φ_x and θ_x are (known) desired values, with the corresponding Φ_y and θ_y required for a given, this approach leads to

$$\rho_y(1) = \ln[1+\rho_x(1)(\exp(\sigma_y^{2})-1)]/\sigma_y^{2} \qquad \ldots\ldots\ldots \quad (4.2.25)$$

$$\rho_y(2) = \Phi_y \rho(1) = \ln[1+\rho_x(2)(\exp(\sigma_y^2)-1)]/\sigma_y^2 \quad \ldots\ldots \quad (4.2.26)$$

from which

$$\Phi_y = \{\ln[1+\rho_x(2)(\exp(\sigma_y^2)-1)]\}/\{\ln[1+\rho_x(1)(\exp(\sigma_y^2)-1)]\} \quad \ldots\ldots\ldots \quad (4.2.27)$$

To solve for θ_y, let $\rho_y(1) = C = C(\Phi_x,\theta_x,\sigma_y)$; than

$$\{(1-\Phi_y\theta_y)(\Phi_y-\theta_y)\}/(1+\theta_y^2-2\Phi_y\theta_y) = C \quad \ldots\ldots\ldots \quad (4.2.28)$$

where C is constant for a given ρ_x and σ_y . This may be written as

$$\theta_y^2 + A\theta_y + 1 = 0 \quad \ldots\ldots\ldots \quad (4.2.29)$$

where

$$A = \{\Phi_y^2+1-2\Phi_y C\}/\{C-\Phi_y\} \quad \ldots\ldots\ldots \quad (4.2.30)$$

Of the two roots of (4.2.29), only one satisfies the stationarity condition (Box and Jenkins, 1970). In addition, the requirement $A^2 \geqslant 4$ must be met to yield real solutions for θ_y. The effect of this requirment is to establish a constraint $\theta_x \leqslant P(\Phi_x,\sigma_y)$. This functional relationship is shown by Burges and Lettenmaier (1975); however, in practice this requirment will be important only for $\Phi_x \leqslant \theta_x$, which corresponds to negative values of the correlation coefficients of lags one and two in the X domain, not a condition of practical hydrologic interest.

c) **Beard's Procedure**

Another method to generate skewed flows is to use Beard's Procedure with an ARMA(1,1) model.

Step 1. Compute the logarithm of streamflow data. In the case where the data contain zero flows, add a small increment such as 0.1 of mean annual flow to each flow before taking the logarithm.

Step 2. Compute the mean ($\bar{x}$), standard deviation (s) and skewness (g) of the log values (x_i).

Step 3. Standardize the log values to have zero mean and unit variance

$$t_i = (x_i-\bar{x})/s \qquad \text{................} \quad (4.2.31)$$

Step 4. If these standardized values exhibit skewness transform this to a normal distribution by using

$$K_i = 6/g\{(gt_i/2+1)^{1/3}-1\} + g/6 \qquad \text{...............} \quad (4.2.32)$$

Step 5. Since an ARMA(1,1) model is assumed in this case, its parameters Φ and Θ are estimated by the method of moment for the tranformed values K_i.

Step 6. gennerate normal variates using

$$K_{i+1} = \Phi K_i+\sigma_e(\epsilon_{i+1}-\Theta\epsilon_i) \qquad \text{.............} \quad (4.2.33)$$

Step 7. Transform each generated value using the inverse transformations of Eq.(4.2.32) and (4.2.31)

$$t_i = \{(1+K_ig/6-g^2/36)^3-1\}2/g \qquad \text{..............} \quad (4.2.34)$$

$$X_i = \bar{X} + st_i \qquad \text{..............} \quad (4.2.35)$$

Step 8. Find the antilogarithm of the values obtained in step 7, and subtract the increment added in step 1. If any negative value results, set it to zero.

4.3 APPLICATION FOR ACTUAL DATA

Using AR(1) and ARMA(1,1) models 1000 years of streamflows were generated for each river. The AR(1) can be applied all streams while the ARMA(1,1) model is limited in certain cases where the parameters are outside a region for the invertibility-stationarity condition. The procedures adopted are AR(1) and ARMA(1,1) and are applied to various rivers (Table 3.1). Abbreviations for the models are as follows.

Hist.	- Historical values
AR1WHT	- AR(1) with Kirby's modified W-H tranformation
AR1LT-2	- AR(1) with 2 parameter log normal distribution
AR1LT-3	- AR(1) with 3 parameter log normal distribution
AR1BP	- AR(1) with Beard's procedure
ARMAWHT	- ARMA(1,1) with Kirby's modified W-H tranformation
ARMALT-2	- ARMA(1,1) with 2 parameter log normal distribution
ARMALT-3	- ARMA(1,1) with 3 parameter log normal distribution
ARMABP	- ARMA(1,1) with Beard's procedure

4.4 DISCUSSION OF RESULTS

The results obtained by applying the models to various rivers are tabulated in Table 4.1 - 4.3. The following describes the results in terms of principal statistics.

4.4.1 Mean, Standard Deviation and Lag One Autocorrelation Coefficient

All the model results slightly overestimated the mean for all of the rivers. All the models except AR1BP, ARMALT-2, ARMALT-3 and ARMABP slightly underestimated the standard deviation for all the rivers, while the model ARMABP overestimated it for all of the rivers. The model AR1BP overestimated it for most of the rivers except Ping River, Wang River and Nam Mae Chaem. The model ARMALT-2 and ARMALT-3 overestimated for Nam Mae Chaem Nam Mae Rim and Ngao River. All the models except AR1BP preserved the lag one autocorrelation coefficient. In the case of Wang River and Nam Mae Taeng, the model AR1BP slightly overestimated the parameters.

4.4.2 Skewness and Hurst Coefficient

All the models preserved the skewness except a few cases. Model AR1LT-2 slightly overestimated the skewness for Yom River, Nan River, Ngao River and Nam Pat. The AR1LT-2 model slightly underestimated skewness and Hurst coefficient for Ping River. The model AR1BP overestimated the skewness for Yom River, Nan River, Nam Mae Taeng, Ngao River and Nam Pat. In the case of Ngao River, the model ARMALT-3 resulted in a small negative skewness where as the historical value is positive. The Hurst Coefficient is found that all the models resulted lower than the corresponding historical values for all of the rivers.

4.4.3 Maximum, Minimum and Percentage of Zero Flows

All the models resulted the maximum values larger than the historical values and the minimum values lower than the historical values for most of the rivers except model AR1WHT and AR1LT-3. The minimum values are larger than the historical values for Ping River. Except for the models AR1LT-2, AR1BP, ARMALT-2 and ARMABP, all the other models resulted in some negative flows being generated.

4.5 SUMMARY

From the results, it was observed that all the models slightly overestimated the mean and the Hurst coefficient values were lower than the historical values for all of the rivers. The others parameters (standard deviation, skewness and lag one autocorrelation) from AR(1) model are comparable with ARMA(1,1) model. When modified, the former model can preserve the parameters better than the latter model.

TABLE 4.1 COMPARISON OF MEAN ,STANDARD DEVIATION AND SKEWNESS FOR HISTORIC DATA AND GENERATED SEQUENCES FROM MODEL

PARAMETER	MODEL	PING	WANG	YOM	NAN	MAE TAENG	MAE CHAEM	MAE RIM	MAE KHAN	NGAO	NAM PAT
Mean	Hist.	2018.061	255.085	2631.718	2655.370	625.856	1145.244	160.193	384.669	96.617	395.580
x 10^6 cu.m	AR1WHT	2080.294	262.017	2711.227	2756.169	641.358	1183.807	166.140	402.560	101.297	406.361
	AR1LT-2	2090.763	262.078	2700.937	2750.774	643.952	1188.463	168.443	404.471	100.791	405.478
	AR1LT-3	2089.614	262.018	2710.305	2759.103	642.433	1188.740	168.306	405.174	101.255	406.089
	AR1BP	2091.055	263.716	2714.111	2766.564	644.812	1185.781	167.952	406.169	101.932	406.952
	ARMAWHT	-	-	2712.917	2751.934	-	1194.758	171.375	-	104.550	404.657
	ARMALT-2	-	-	2705.518	2744.978	-	1204.331	181.365	-	104.252	404.657
	ARMALT-3	-	-	2712.570	2754.137	-	1204.461	181.508	-	104.261	401.635
	ARMABP	-	-	2708.881	2767.493	-	1213.811	171.621	-	103.771	402.462
Standard	Hist.	684.676	113.692	1118.134	1028.907	342.952	362.210	68.406	171.605	40.637	239.624
Deviation	AR1WHT	675.263	111.728	1092.905	1007.469	336.358	357.814	67.435	169.567	39.474	235.247
	AR1LT-2	677.692	111.289	1094.377	1013.794	336.060	360.336	68.293	170.930	40.227	236.069
x10^6 cu.m	AR1LT-3	677.956	111.294	1090.932	1008.801	336.100	360.080	68.262	170.727	39.720	235.778
	AR1BP	665.226	112.884	1128.807	1066.373	375.094	360.031	69.114	176.066	43.253	256.539
	ARMAWHT	-	-	1093.931	1006.030	-	361.635	67.463	-	40.038	234.823
	ARMALT-2	-	-	1097.349	1009.166	-	368.720	75.239	-	42.159	232.178
	ARMALT-3	-	-	1093.003	1006.218	-	368.161	75.942	-	40.730	232.371
	ARMABP	-	-	1125.177	1067.493	-	373.591	71.751	-	44.126	250.481
Skewness	Hist.	1.218	1.466	0.735	0.498	2.195	0.894	1.588	0.979	0.15	1.815
	AR1WHT	1.019	1.344	0.647	0.382	2.097	0.708	1.335	0.783	0.094	1.766
	AR1LT-2	0.81	1.303	1.237	0.972	1.694	0.728	0.985	1.041	1.02	1.988
	AR1LT-3	0.932	1.34	0.655	0.373	2.051	0.681	1.132	0.726	0.056	1.775
	AR1BP	0.773	1.525	1.047	0.961	4.178	0.712	1.474	0.999	0.608	2.908
	ARMAWHT	-	-	0.648	0.381	-	0.664	1.223	-	0.051	1.771
	ARMALT-2	-	-	1.24	0.963	-	0.702	0.922	-	0.978	1.951
	ARMALT-3	-	-	0.657	0.369	-	0.636	1.078	-	-0.002	1.744
	ARMABP	-	-	1.041	0.962	-	0.673	1.446	-	0.588	2.826

TABLE 4.2 COMPARISON OF LAG ONE AUTOCORRELATION, HURST COEFF., AND MAX. FOR HISTORIC DATA AND GENERATED SEQUENCES FROM MODEL

PARAMETER	MODEL	PING	WANG	YOM	NAN	MAE TAENG	MAE CHAEM	MAE RIM	MAE KHAN	NGAO	NAM PAT
Lag one auto-correlation coefficient	Hist.	0.297	-0.061	-0.068	0.21	-0.108	0.377	0.403	0.378	0.293	-0.22
	AR1WHT	0.308	-0.058	-0.058	0.228	-0.113	0.393	0.411	0.392	0.312	-0.221
	AR1LT-2	0.312	-0.059	-0.066	0.223	-0.112	0.393	0.418	0.392	0.307	-0.228
	AR1LT-3	0.311	-0.06	-0.059	0.228	-0.115	0.394	0.417	0.394	0.315	-0.227
	AR1BP	0.315	-0.008	-0.062	0.208	-0.044	0.309	0.37	0.379	0.307	-0.284
	ARMAWHT	-	-	-0.059	0.228	-	0.403	0.41	-	0.328	-0.223
	ARMALT-2	-	-	-0.067	0.225	-	0.397	0.428	-	0.309	-0.229
	ARMALT-3	-	-	-0.06	0.229	-	0.398	0.423	-	0.341	-0.228
	ARMABP	-	-	-0.062	0.208	-	0.318	0.362	-	0.314	-0.286
Hurst coefficient	Hist.	0.714	0.729	0.694	0.734	0.682	0.815	0.842	0.863	0.781	0.662
	AR1WHT	0.648	0.591	0.607	0.655	0.574	0.669	0.66	0.666	0.675	0.56
	AR1LT-2	0.644	0.593	0.593	0.646	0.584	0.68	0.68	0.674	0.66	0.565
	AR1LT-3	0.662	0.593	0.605	0.659	0.581	0.681	0.677	0.684	0.678	0.567
	AR1BP	0.665	0.599	0.594	0.644	0.589	0.666	0.663	0.672	0.667	0.567
	ARMAWHT	-	-	0.61	0.649	-	0.705	0.769	-	0.737	0.535
	ARMALT-2	-	-	0.597	0.64	-	0.714	0.779	-	0.721	0.532
	ARMALT-3	-	-	0.609	0.652	-	0.715	0.776	-	0.745	0.535
	ARMABP	-	-	0.586	0.645	-	0.726	0.697	-	0.704	0.539
Maximum	Hist.	4255.187	547.347	5318.144	4738.287	1925.64	2239.832	393.304	867.603	196.355	1163.758
	AR1WHT	5646.94	919.334	7697.145	6815.787	3080.534	2826.036	565.22	1207.95	233.916	1976.711
	AR1LT-2	5335.139	935.81	9157.771	8156.787	2896.037	2768.45	494.705	1259.411	312.261	1967.777
	AR1LT-3	5514.76	944.632	7786.712	6803.509	3114.75	2722.272	513.284	1152.679	232.836	1881.872
	AR1BP	5213.045	1022.817	8797.134	8412.359	5017.23	2829.681	593.284	1273.553	275.817	2577.938
	ARMAWHT	-	-	7670.518	6818.306	-	2762.924	552.833	-	232.818	1996.745
	ARMALT-2	-	-	9068.504	8144.777	-	2746.74	537.814	-	312.812	1940.477
	ARMALT-3	-	-	7733.46	6798.65	-	2700.428	565.129	-	232.438	1873.243
	ARMABP	-	-	8793.792	8413.703	-	2826.013	595.193	-	270.49	2270.1

TABLE 4.3 COMPARISON OF MINIMUM AND % OF ZERO FLOWS FOR HISTORIC DATA AND GENERATED SEQUENCES FROM MODEL

PARAMETER	MODEL	PING	WANG	YOM	NAN	MAE TAENG	MAE CHAEM	MAE RIM	MAE KHAN	NGAO	NAM PAT
Minimum	Hist.	689.726	547.347	914.18	1292.405	325.894	593.739	70.529	43.288	39.784	122.47
	AR1WHT	797.532	93.794	355.145	83.914	144.602	361.673	39.869	30.198	0	0
	AR1LT-2	656.024	69.101	755.939	772.75	125.449	388.984	36.911	84.094	24.068	66.381
	AR1LT-3	739.624	71.513	277.917	117.73	174.825	360.467	46.647	26.028	0	45.191
	AR1BP	660.14	78.886	637.527	699.66	263.918	389.713	54.903	12.448	12.448	93.16
	ARMAWHT	-	-	339.257	99.286	-	378.598	41.866	-	0	0
	ARMALT-2	-	-	750.598	775.086	-	395.14	36.178	-	24.377	69.716
	ARMALT-3	-	-	268.708	118.532	-	364.829	46.021	-	0	48.89
	ARMABP	-	-	643.893	698.804	-	392.67	55.563	-	12.244	96.812
Percentage of zero flows	Hist.	0	0	0	0	0	0	0	0	0	0
	AR1WHT	0	0	0	0	0	0	0	0	0.6	0.4
	AR1LT-2	0	0	0	0	0	0	0	0	0	0
	AR1LT-3	0	0	0	0	0	0	0	0	0.6	0
	AR1BP	0	0	0	0	0	0	0	0	0	0
	ARMAWHT	-	-	0	0	-	0	0	-	0.5	0.4
	ARMALT-2	-	-	0	0	-	0	0	-	0	0
	ARMALT-3	-	-	0	0	-	0	0	-	0.6	0
	ARMABP	-	-	0	0	-	0	0	-	0	0