RESEARCH ARTICLES

J. Sci. Soc. Thailand, **6** (1980) 69-80

SEDIMENT VOLUME ACCUMULATED IN A RESERVOIR HAVING CORRELATED INFLOWS

HUYNH NGOC PHIEN and ANAT ARBHABHIRAMA

Asian Institute of Technology, Bangkok, Thailand

PRATHET SUTABUTR

National Energy Administration of Thailand, Bangkok, Thailand

(Received 8 February 1980)

Summary

A relationship between sediment discharge and streamflow was used in this paper either as a scheme for generating sediment data from streamflow record or as a transformation of variable. Assuring that the generated sediment sequence or the streamflow sequence follows some widely used models for correlated variable, an attempt was made to derive the formulas for the expected value and variance of the sediment volume accumulated over a time span in a reservoir having them as inflows.

Introduction

In a previous paper, Phien and Arbhabhirama[1] have presented a statistical approach to estimating the accumulated volume of sediment in reservoirs over their design lifetime. The technique presented in that work was based on the relationship between the total sediment discharge, S, and the streamflow, Q, of the form:

$$S = aQ^b \qquad (1)$$

where a and b are two constants. Since streamflow normally has longer length of record, (1) can be used in two different ways: (1) to generate a sequence of sediment discharge from the streamflow; (2) as a transformation of variable, thus, the distribution of the sediment discharge is deduced from that of the streamflow. The analysis can be then based on the generated sediment sequence or on the transformation of the streamflow distribution.

To continue that work which has dealt with the case of independent variables, the present paper aims to derive the exact formulas for the expected value (or mean) and variance of the accumulated volume of sediment in a reservoir, assuming that these two sequences follow some typical models for correlated data.

General Consideration

Let ε be the trap efficiency of the reservoir under consideration, then ε is normally considered to be a constant in the interval (0, 1). The volume of sediment accumulated in the reservior over a given time span of N years is

$$V_N = \sum_{i=1}^{N} \varepsilon S_i = \varepsilon \sum_{i=1}^{N} S_i \tag{2}$$

where S_i is the annual sediment volume which flows into the reservoir in year i, i = 1, ..., N. It is clear from (2) that the distribution of V_N is readily obtained from the distribution of the sum $\sum_{i=1}^{N} S_i$, which is the value of V_N corresponding to the case $\varepsilon = 1.0$, where all sediment inflows are trapped in the reservoir. Therefore, in the following analysis, it is sufficient to consider this case, and (2) becomes

$$V_N = \sum_{i=1}^{N} S_i \tag{3}$$

Since the annual values of sediment discharge volume are used, it is appropriate to assume that the sequence of the S_i, or the S–sequence, is a stationary process, with mean μ_S and variance Q_S^2:

$$E\,(S_i) = \mu_S$$
$$Var\,(S_i) = Q_S^2$$

where E and Var stand for expected value (or mean) and variance, respectively.

It follows immediately from (3) that

$$E\,(V_N) = N\,\mu_S \tag{4}$$

which is always valid irrespective of the fact that the S–sequence is either independent or dependent.

The variance of V_N is obtained from (3) as

$$Var\,(V_N) = \sum_{i=1}^{N} Var\,(S_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Cov\,(S_i, S_{i+k}) \tag{5}$$

where Cov denotes the covariance of S_i and S_{i+k}. After some algebra, it becomes

$$\text{Var } (V_N) = \sigma_S^2 \, [N+2 \sum_{k=1}^{N-1} (N-k) \, R_k] \tag{6}$$

Where R_k is the autocorrelation at lag k of the S–sequence. If $R_k = 0$, (6) reduces to the expression for the case of independent variables:

$$\text{Var } (V_N) = N \, \sigma_S^2$$

Analysis Based on the Generated S–sequence

In modelling of stationary processes, the appropriate models are the autoregressive process, moving average process and mixed autoregressive-moving average process. For the detailed treatment of these models, it is referred to the work of Box and Jenkins[2]. It should be noted that in this part, both the mean and variance of S can be estimated from the generated sample and thus they are treated as known parameters. Only the variance of the accumulated volume V_N needs to be derived according to the different models employed while its expected value remains unchanged.

A. Case of Normal Variables.

(1) Autoregressive Processes. As applied to the modelling of the S–sequence, the autoregressive procese of order m, denoted by AR (m), is written as

$$z_i = \phi_1 \, z_{i-1} + \ldots + \phi_m \, z_{i-m} + i_i \tag{7}$$

where $z_i = S_i - \mu_s$. The random shock t_i is independent of $z_{i-t}, \ldots, z_{i-m}$, and is identically distributed as independent normal variables of zero mean and variance σ_t^2. The stationarity condition is that the roots of the characteristic equation

$$\phi \, (B) = 1 - \phi_1 \, B - \ldots - \phi_m B^n$$

must lie outside the unit circle. In (7), $\phi_1, \ldots, \phi_m$ are the model parameters.

For this process, the autocorrelation at lag k is given by

$$R_k = \phi_1 \, R_{k-1} + \ldots + \phi_m \, R_{k-m}, \; k > 0 \tag{8}$$

and the varirance of z_i is

$$\sigma_z^2 = \sigma_t^2 \; / \; (1 \;\, R_1 \, \phi_1 - \ldots - R_m \, \phi_m)$$

In this case, $\sigma_z = \sigma_S$ ia known, thus the variance of t can be expressed in terms of the variance of S and the parameters of the model as:

$$\sigma_t^2 = \sigma_S^2 \; (1 - R_1 \, \phi_1 - \ldots - R_m \, \phi_m)$$

For the estimation of the model parameters, it is referred to Box and Jenkins[2].

Substuting (8) into (6) yields the expression for the variance of V_N:

$$\text{Var } (V_N) = \sigma_S^2 \, [N + 2 \sum_{k-1}^{N-1} \sum_{i=1}^{m} (N-k) \, \phi_i \, R_{k-i}] \tag{9}$$

The most frequently encountered case is the AR (1) process, for which,

$$R_k = R^k$$

Where R is used in place of R_1. For this model, the variance of V_N is obtained, after some simple manipulations, as

$$\mathrm{Var}\ (V_N) = N\ \sigma_S^2 \left[\frac{1+R}{1-R} - \frac{2R\ (1-R^N)}{N\ (1-R)^2}\right. \tag{10}$$

(2) Moving Average Processes. The moving average model of order n, denoted by MA (n), is as follows

$$z_i = t_i - \theta_1\ t_{i-1} - \ldots - \theta_n\ t_{i-n} \tag{11}$$

where $z_i = S_i - \mu_S$, and t_i are identically distributed as independent normal variables with zera mean and variance σ_t^2. No restrictions are needed on the parameters $\theta_1, \ldots, \theta_n$ of the model to ensure stationarity.

The variance of the process is

$$\sigma_z^2 = (1 + \theta_1^2 + \ldots + \theta_n^2)\ \sigma_t^2$$

which allows σ_t to be determined by $\sigma_z = \sigma_S$ and the model parameters. The autocorrelation at lag k is given by

$$R_k = \begin{cases} (-\theta_k + \sum_{i=1}^{n-k} \theta_i\ \theta_{i+k})\ /\ (1 + \theta_1^2 + \ldots + \theta_n^2) & k = 1, 2, \ldots, n \\ 0 & k > n \end{cases} \tag{12}$$

In practice, the order n of the process is much less than N, therefore, a substitution of (12) into (6) yields

$$\mathrm{Var}\ (V_N) = \sigma_S^2\ [N + 2\ (1 + \theta_1^2 + \ldots + \theta_n^2)^{-1} \sum_{k=1}^{n} (N-k)\ (-\theta_k + \sum_{i=1}^{n-k} \theta_i\ \theta_{i+k})\ ] \tag{13}$$

For the first-order model, $z_i = t_i - \theta_1\ t_{i-1}$,

$$R_1 = -\ \theta_1\ /\ (1 + \theta_1^2)$$

$$R_k = 0,\ k \geq 2$$

and (13) gives:

$$\mathrm{Var}\ (V_N) = \sigma_S^2\ [N - 2\ (N - 1)\ \theta_1\ /\ (1 + \theta_1^2)\ ] \tag{14}$$

For the second-order model,

$$R_1 = -\ \theta_1\ (1 - \theta_2)\ /\ (1 + \theta_1^2 + \theta_2^2)$$

$$R_2 = -\ \theta_2\ /\ (1 + \theta_1^2 + \theta_2^2)$$

$$R_k = 0,\ k \geq 3$$

and (13) becomes:

$$\mathrm{Var}\ (V_N) = \sigma_S^2\ \{N - 2\ (1 + \theta_1^2 + \theta_2^2)^{-1}\ [\ (N - 1)\ \theta_1\ (1 - \theta_2) + (N - 2)\ \theta_2]\} \tag{15}$$

(3) Mixed Autoregressive–Moving Average Processes. The model is

$$z_i = \theta_1 z_{i-1} + \ldots + \phi_m z_{i-m} + t_i - \theta_1 t_{i-1} - \ldots - \theta_n t_{i-n} \quad (16)$$

which is often denoted by ARMA (m, n), where m and n are the orders of the model.

An ARMA Process of considerable importance in practice is the first-order autoregressive–first-order moving average, ARMA (1, 1), process. It is commonly written as

$$z_i - \phi z_{i-1} = t_i - \phi t_{i-1} \quad (17)$$

and is also referred to as the ARIMA (1, 0, 1) process (See O'Connell[3] and Sen[4]). For this model,

$$R_k = \phi^{k-1} (\phi - \theta) (1 - \phi\theta) (1 + \theta^2 - 2\phi\theta), \; k \geq 1 \quad (18)$$

In view of (18), the variance of V_N in (6) can be written as

$$\mathrm{Var}(V_N) = \sigma_S^2 [N + 2 (\phi - \theta) (1 - \phi\theta) (1 + \theta^2 - \phi\theta)^{-1} \sum_{k=1}^{N-1} (N - k) \phi^{k-1}] \quad (19)$$

It can be easily shown that

$$\sum_{k=1}^{N-1} (N - k) \phi^{k-1} = \frac{N(1 - \phi) - (1 - \phi^N)}{(1 - \phi)^2}$$

whence (19) becomes:

$$\mathrm{Var}(V_N) = \sigma_S^2 \left\{N + \frac{2(\phi - \theta) (1 - \phi\theta)}{(1 + \theta^2 - \phi\theta) (1 - \phi)^2} [N(1 - \phi) - (1 - \phi^N)\right\} \quad (20)$$

B. Case of Lognormal Variables.

When the S–sequence is found to have a skewness coefficient which is significantly different from zero, the above models can no longer be used. In this case, lognormal distribution should be attempted. By extending the definition of Vincens *et al.*[5], a sequence of lognormal variables is said to follow a model if the sequence of the logarithms follows that model.

The expected value of V_N is still given by (4). Further derivation is needed for the variance.

Let r_k denotes the autocorrelation at lag k of the s–sequence, where $s_i = \ln S_i$, then it follows from the Appendix that

$$R_k = [(1 + \eta_S^2)^{r_k} - 1] / \eta_S^2 \quad (21)$$

in which $n_S = \sigma_S / \mu_S$ is the variation coefficient of the S–sequence. Substituting (21) into (6) yields:

$$\mathrm{Var}(V_N) = \sigma_S^2 \{N + 2 \eta_n^{-2} \sum_{k=1}^{n-1} (n - k) [1 + \eta_S^2)^{r_k} - 1]\} \quad (22)$$

(1) Autoregressive Processes. The model of (7) now applies to the variables $z_i = s_i - \mu_s$, where μ_s is the expected value of the s–sequence. This expected value can be expressed in terms of μ_s and σ_s [See Yevjevich[6]] as follows:

$$\mu_s = (1/2) \ln [\mu_S^4 / (\mu_S^2 + \sigma_S^2]$$

The autocorrelation r_k is given by

$$r_k = \sum_{i=1}^{m} \phi_i \, r_{k-i}$$

Where ϕ_i are the model parameters which can be estimated from the logarithms of the S–sequence by the method described by Box and Jenkins[2].

For the first-order autoregressive model, $r_k = r^k$, where r is used instead of r_1. Equation (22) then becomes:

$$\text{Var}\ (V_N) = \sigma_S^2 \{N + 2\ \eta_S^{-2} \sum_{k=1}^{N-1} (N - k)\ [(1 + \eta_S^2)^{r^k} - 1]\} \quad (23)$$

(2) Moving Average Procresses. For this model, the expression for r_k is

$$r_k = \begin{cases} (-\theta_k + \sum_{i=1}^{n-k} \theta_i\, \theta_{i+k}) / (1 + \theta_1^2 + \ldots + \theta_n^2) & k = 1, \ldots, n \\ 0 & k > n \end{cases}$$

wher $\theta_1, \ldots, \theta_n$ are the model parameters as applied to the sequence of $s_i - \mu_s$. For the case of the first-order model, MA (1),

$$r_1 = -\theta_1 / (1 + \theta_2^1)$$
$$r_k = 0,\ k \geq 2$$

and (22) gives

$$\text{Var}\ (V_N) = \sigma_S^2 \{N + 2\ \eta_S^{-2}\ (N - 1)\ [(1 + \eta_S^2)^{-\theta_1 / (1 + \theta_1^2)} - 1]\} \quad (24)$$

(3) Mixed Autoregressive–Moving Average Processes. For the case of the ARMA (1, 1) or ARIMA (1, 0, 1) process,

$$r_k = \phi^{k-1}\ (\phi - \theta)\ (1 - \phi\theta)\ /\ (1 + \theta^2 - 2\ \theta\phi),\ k > O.$$

Substituting this expression into (22) results in

$$\text{Var}\ (V_N) = \sigma_S^2 \{N + 2\ \eta_S^{-2} \sum_{k=1}^{N-1} [(1 + \eta_S^{-2})^{\phi^{k-1} (\phi - \theta)(1 - \phi\theta) / (1 + \theta^2 - 2\theta\phi)} - 1]$$

Analysis Based on a Transformation of the Streamflow Distribution

The case of normal variables must be excluded from the analysis to ensure that Q^b is defined. Thus, in the following, the annual streamflows are assumed to follow the lognormal distribution.

If Q is a lognormal variable, then (1) shows that S is also a lognormal variables. The expected value and variance of S have been shown (See Phien and Arbhabhirama[1]) to be expressed in terms of the expected value and variance of Q, denoted by μ_Q and σ_Q^2, respectively, as follows:

$$\mu_S = a\ \mu_Q^{\ b(2-b)}\ (\mu_Q^2 + \sigma_Q^2)^{(b-1)/2}$$

$$\sigma_S^2 = a^2\ \mu_Q^{2b}\ \ (1 + \eta_Q^2)^{b\ (b-1)}\ [\ (\ + \eta_Q^2)^{b^2} - 1] \qquad (26)$$

where $\quad n_Q = \sigma_Q / \mu_Q$ is the variation coefficient of Q.

The expected value of V_N is therefore given by

$$E(V_N) = N\ \mu_Q^{\ b(2-b)}\ (\mu_Q^2 + \sigma_Q^2)^{b(b-1)/2} \qquad (27)$$

which is independent of the model employed.

In order to obtain the variance of V_N, one can substitute (26) into (6). However, the autocorrelation R_k in (6) is obtained from S, but not from Q, it is necessary to establish the relationship between the autocorrelations at the same lag of S and Q.

Let $q = \ln Q$, and $s = \ln S$, then the basic relationship between S and Q in (1) gives:

$$s = \ln a + bq$$

Consequently the standardized variables of s and q are equal:

$$(s - \mu_s) / \sigma_s = (q - \mu_q) / \sigma_q$$

with obvious notation. This relationship shows that if q_1, q_2, ... have the same distribution as q, and s_1, s_2, ... have the same distribution as s, then

$$\mathrm{Cov}\ (s_i, s_{i+k}) / \sigma_S^2 = \mathrm{Cov}\ (q_i, q_{i+k}) / \sigma_q^2$$

In other words, the autocorrelations of s and q are equal at any lag. The autocorrelation at lag k of q can thus be denoted by r_k, the autocorrelation at lag k of s as employed before. Since both S and Q are lognormal variables, eq. (A. 3) in the Appendix can be applied to both the two sequences, resulting the following important conclusion: *The autocorrelations of S and Q at the same lag are equal.* This result permits the use of R_k to denote the autocorrelation at lag k of Q as well. With this in mind, a substitution of (26) into (6) yields:

$$\mathrm{Var}\ (V_N) = f\ (a, b, Q)\ [N + 2 \sum_{k=1}^{N-1} (N - k)\ R_k]$$

Where $f(a, b, Q) = \sigma_S^2$ as given in (26)

In view of (A.3) in the Appendix, this can be rewritten as:

$$\mathrm{Var}\ (V_N) = f\ (a, b, Q) \left\{N + 2\ \eta_Q^{-2} \sum_{k=1}^{N-1} (N - k)\ [\ (1 + \eta_Q^2)^{r_k} - 1]\right\} \qquad (28)$$

(1) *Autoregressive Processes.* The model of (7) now applies to the variables $z_i = q_i - \mu_q$. The autocorrelation r_k is given by

$$r_k = \sum_{i=1}^{m} \phi_i \, r_{k-i}$$

where the parameters ϕ_i are determined from the logarithms of the streamflows.

For the AR (1) model, $r_k = r^k$, r being employed in place of r_1, and (28) becomes:

$$\text{Var}(V_N) = f(a, b, Q) \left\{N + 2\,\eta_Q^{-2} \sum_{k=1}^{N-1} \left[(1 + \eta_Q^2)^{r^k} - 1\right]\right\} \quad (29)$$

(2) *Moving Average Processes.* The model of (11) applies to the variables $z_i = q_i - \mu_q$, by which the autocorrelation r_k is given by

$$r_k = \begin{cases} \left(-\theta_k + \sum_{i=1}^{n-k} \theta_i \, \theta_{i+k}\right) / \left(1 + \theta_1^2 + \ldots + \theta_n^2\right) & k = 1, \ldots, n \\ 0 & k > n \end{cases}$$

For the MA (1) model,

$$r_1 = -\theta_1 / (1 + \theta_1^2)$$

$$r_k = 0, \; k \geq 2$$

Thus, (28) gives

$$\text{Var}(V_N) = f(a, b, Q) \left\{N + 2(N-1)\,\eta_Q^{-2}\left[(1 + \eta_Q^2)^{-\theta_1/(1+\theta_1^2)} - 1\right]\right\} \quad (30)$$

(3) *Mixed Autoregressive—Moving Average Processse.* For the ARMA (1, 1) or ARIMA (1, 0, 1) process,

$$r_k = \phi^{k-1} (\phi - \theta)(1 - \phi\theta) / (1 + \theta^2 - 2\phi\theta), \; k > 0$$

in which, the model parameters ϕ and θ are of course estimated using the logarithms of the streamflows. Substituting this expression of r_k into (28) gives:

$$\text{Var}(V_N) = f(a, b, Q) \left\{N + 2\,\eta_Q^{-2} \sum_{k=1}^{N-1} \left[(1 + \eta_Q^2)^{\phi^{k-1}(\phi-\theta)(1-\phi\theta)/(1+\theta^2-2\theta\phi)} - 1\right]\right\}$$

Application

From several samples of suspended sediment and streamflow of the Pasak River at Kang Khoi station (Thailand), application of the least squares method yields the following equation which relates the suspended load, G (in tons/day) to the streamflow, Q (in m^3/s):

$$G = 10.60\, Q^{1.05}$$

Assuming the bed load to be 20% of the suspended load and a unit weight of 80 $lb./ft^3$ which are commonly experienced in this country, one obtains a relationship between the total sediment discharge and the streamflow as follows:

$$S = 3.62 \times 10^3 Q^{1.05} \tag{32}$$

where S is in m^3/year and Q is in m^3/s.

Analysis Based on the Generated Sequence

At the considered station, data for Q have been made available for 63 years (1914–1976). Using(32), 63 annual values for sediment discharge are generated. For this generated sequence, only the first-order autocorrelation, being equal to 0.343 is found to be significantly different from zero at 5% probability level (student t-test). In addition, the skewness coefficient is 0.487 which is also significantly different from zero, thus the first-order autoregressive model for lognormal variables may be employed.

The expected value and standard deviation of S, and the first-order autocorrelation of lnS have the following estimates.

$$\mu_S \doteq 3.48 \times 10^5 \text{ m}^3/\text{year} \qquad \sigma_S = 1.40 \times 10^5 \text{ m}^3/\text{year} \qquad r = 0.270$$

For a reservoir to be designed based upon the data available to Kang Khoi, the expected value and variance of the accumulated sediment volume in N years, N = 10(10)100, can be computed by using (4) and (23). The results for $\varepsilon = 1.0$ are shown in Table 1.

TABLE 1. EXPECTED VALUE AND VARIANCE OF SEDIMENT VOLUME COMPUTED FROM DATA OF PASAK RIVER.

N	Mean (10^6 m^3)		Variance (10^{12} m^6)	
Years	(1)	(2)	(1)	(2)
10	3.484	3.484	0.313	0.316
20	6.969	6.967	0.644	0.652
30	10.453	10.451	0.975	0.987
40	13.938	13.934	1.306	1.322
50	17.422	17.418	1.637	1.658
60	20.907	20.902	1.968	1.993
70	24.391	24.385	2.299	2.328
80	27.875	27.869	2.630	2.663
90	31.360	31.352	2.962	2.999
100	34.844	34.836	3.293	3.334

Note: (1) based on the sediment sequence
(2) based on the streamflow sequence

Analysis Based on the Streamflow Sequence

Only the first-order autocorrelation of the annual streamflow (0.344) for the Pasak River is found to be significantly different from zero at 5% probability level. With a high value of the skewness coefficient (0.435), the first-order autoregressive model for lognormal variables is used. The estimated values of the expected value and standard deviation of Q, and the first-order autocorrelation of lnQ are respectively as

$$\mu_Q \doteq 77.16 \text{ m}^3/\text{s} \qquad \sigma_Q = 29.54 \text{ m}^3/\text{s} \qquad r = 0.270$$

The expected value and variance of V_N are then computed using (27) and (29). The results are also given in Table 1.

Discussion

(1) As seen in the analysis, when Q is lognormally distributed, S is also lognormally distributed, and hence the same model should be used for sediment discharge and streamflow sequences. In the present case, the first-order autoregressive model for lognormal variables is employed, mainly because of its simplicity and popularity.

(2) The results in Table 1 show that the two different approaches provide close values for the expected value and variance of the accumulated sediment volume. This may indicate the appropriateness of the model in use.

Summary and Conclusions

The present paper investigates the distribution of the accumulated volume of sediment in a reservoir over N years through derivation of the expressions for its expected value and variance corresponding to the case where sediment inflows are correlated. All the three models, namely, the autoregressive, moving average, and mixed autoregressive-moving average models are considered and due attention is paid to their most commonly used forms; i.e., the AR(1), MA(1), and ARMA(1, 1) or ARIMA(1, 0, 1) models. All the results are presented in a from which is ready for use. Since there are several sequences involved in the analysis, some repetition is made to specify clearly the sequence and the model under consideration in each case.

The expected value of the accumulated volume of sediment is independent of the model considered, while the variance depends upon the autocorrelation of the inflow sequence. The expressions for the variance corresponding to different models are given. The contribution of the persistence in the inflow sequence, expressed by its autocorrelations, is explicitely indicated in these results.

To obtain the accurate estimates of the expected value and variance of the accumulated sediment volume over a given time span, sediment and streamflow sequences must be correctly modelled. This could be achieved by following the three stages recommended by Box and Jenkins[2] and by incorporating recent experiences and developments in improved Box-Jenkins methods.

List of Symbols

The following symbols have been used in this paper:

a, b	=	constants in the relationship between sediment and streamflow
E	=	expected value (mean)
f (a, b, Q)	=	variance of S in terms of a, b, mean, and variance of Q
G	=	suspended load
ln	=	natural logarithm
m	=	order of the autoregressive process
n	=	order of the moving average process
N	=	time span in years
Q	=	streamflow
q	=	lnQ
R_k	=	autocorrelation at lag k of Q (and S)
r_k	=	autocorrelation at lag k of q (and s)
S	=	sediment discharge
s	=	lns
t_i	=	random shock (independent normal variable)
z_i	=	correlated variables used in a model
V_N	=	volume of sediment accumulated in a reservior over N years
Var	=	variance
ε	=	trap efficiency
η	=	variation coefficient
θ_i	=	parameters involved in moving average model
μ	=	expected value (mean)
σ	=	standard deviation
ϕ_i	=	parameters involved in autoregressive process

Acknowledgements

The authors wish to express their sincere thanks to the Asian Institute of Technology for its valuable support which made this research possible.

Appendix

Relationship between the autocorrelations of Y and lnY for lognormal variable.

Let $\{Y\} = \{Y_1, Y_2, ...\}$ denote a stationary process of lognormal variables with mean μ, standard deviation σ. If $\{Y\}$ is the (stationary) stochastic process of $Y_1, Y_2, ...$ where $Y_i = \ln Y_i$, $i = 1, 2, ...$ then for any $k \geq 0$,

$$Y_i \; Y_{i+k} = \exp(Y_i) \; \exp(Y_{i+k}) = \exp \; (Y_i + Y_{i+k})$$

Since Y_i and Y_{i+k} are normal variables, $Y_i + Y_{i+k}$ is a normal variable, and thus $Y_i \; Y_{i+k}$ is a lognormal variable. The expected value of $Y_i \; Y_{i+k}$ can then be expressed in terms of the expected value and the variance of Y as follows: (see Yevjevich[6], p.p. 135–136)

$$E(Y_i \; Y_{i+k}) = \exp\,[E(Y_i + Y_{i+k}) + (1/2)\, \mathrm{Var}\,(Y_i + Y_{i+k})]$$

$$= \exp\,[2E\,(Y) + (1 + r_k)\, \mathrm{Var}\,(Y)] \qquad (A.1)$$

where r_k denotes the autocorrelation at lag k of the Y sequence. Expressing E (Y) and Var (Y) in terms of μ and σ, eq. (A.1) yields:

$$E\,(Y_i \; Y_{i+k}) = \exp\,[\ln\,(\frac{\mu^4}{\mu^2 + \sigma^2}) + (1 + r_k)\,\ln\,(\frac{\mu^2 + \sigma^2}{\mu^2})\,]$$

$$= \exp\,[\ln\,\mu^2 + r_k\,\ln\,(1 + \eta^2)\,]$$

or

$$E\,(Y_1 \; Y_{1+1}) = \mu^2\,(1 + \eta^2)^{r_k} \qquad (A.2)$$

where $\mu = \sigma/\mu$ is the variation coefficent of Y.

The autocorrelation at lag k of Y is given by

$$R_k = \frac{E\,(Y_i \; Y_{i+k}) - E\,(Y_i)\,E\,(Y_{i+k})}{\mathrm{Var}\,(Y)}$$

which, in view of eq. (A.2), becomes

$$R_k = [\mu^2\,(1 + \eta^2)^{r_k} - \mu^2]\,/\,\sigma^2$$

or

$$R_k = [\,(1 + \eta^2)^{r_k} - 1]\,/\,\eta^2 \qquad (A.3)$$

Equation A.3 establishes the relationship between the autocorrelations of the lognormal variables and their respective logarithms at the same lag. For $k = 1$, eq. (A.3) reduces to the result given by Matalas[7].

References

1. Phien, H.N. and Arbhabhirama, A. (1979). A Statistical Analysis of the Sediment Volume Accumulated in Reservoirs, *J. Hydrol.*, **44**, 231–240 (Amsterdam).
2. Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis:* Forecasting and Control, Holden-Day, San Francisco, (2nd edition).
3. O'Connell, P.E. (1971), A Simple Stochastic Modelling of Hurst's Law. *Proc. of the Warsaw Symposium* (IAHS Publication No. 100), Vol. **1**, 169–187.
4. Sen, Z. (1977), The Small Sample Expectation of Population and adjusted Ranges. *Water Resour. Res.* **13**, 975–980.
5. Vincens, G.J., Rodriquez-Iturbe, I. and Schaake, J.C. (1974), *A Bayesian Approach to Hydrologic Time Series Modelling*, MIT Report, No. 181.
6. Yevjevich, V. (1972), *Probability and Statistics in Hydrology*, Water Resources Publications, Fort Collins, Colorado.
7. Matalas, H.C. (1967), Mathematical Assessment of Synthetic Hydrology, *Water Resour. Res.* **3(4)**, 937–945.