RANGE AND DEFICIT PROPERTIES OF STATIONARY RANDOM VARIABLES

by

*FELINO P. LANSIGAN**

1. *Introduction*

Within the past three decades, the methods for planning, design and operation of water resource systems have been changing from the use of "rules of thumb" and "engineering judgment" to the more formal type of analysis based on mathematical models. Several methods have been proposed to determine the storage capacities of reservoirs to regulate flows with a given level of assurance. These approaches can be classified as (a) empirical, based on equations or graphical procedures developed from practical experiences, (b) experimental, based on procedures developed using data generation techniques, and (c) analytical, based on the theory of statistics, probability and stochastic processes. Application of the theory of stochastic processes in the design and operation of reservoirs has emerged in recent years as one of the dynamic subjects in statistical hydrology. It has attracted and has appealed to the engineers and statisticians alike because of the inherent stochastic nature of hydrologic phenomena.

Two analytical approaches in determining the required storage capacity of a reservoir are the range analysis and the deficit analysis of partial sums of random variables. Range analysis which assumes an infinite reservoir i.e. topless and bottomless reservoir, is used in the

*Assistant Professor, Institute of Mathematical Sciences and Physics, University of the Philippines at Los Baños, College, Laguna.

design of storage capacities for full regualtion of river flows. On the other hand, deficit analysis which assumes a semi-infinite reservoir, i.e., bottomless reservoir, is used in the case of partial regulation.

Exact and asymptotic expressions for the mean and the variance of the range and deficit properties of partial sums of random variables of stationary process for only small values of the sample size are available in the literature. Approximate expression for the mean range of linearly dependent normal variables has also been proposed elsewhere (Salas, 1972). For higher values of the sample size, however, the mathematical derivation for the mean and the variance of range and deficit becomes extremely cumbersome.

Both the range and the deficit relate to the same stochastic variable of the cumulated partial sums. But the exact form of their relationships is not available. However, the relation of at least the expected values of range and deficit may be investigated for particular cases of practical interest. In this paper, the possibility of estimating the mean and variance of deficit as function of mean range and sample size, respectively, is explored in cases where the stationary input series into a reservoir is either an independent, first-order Markov, or ARMA (1.1) process. The theoretical framework in range and deficit analyses is given in Section 2. Some analytical results and the experimental approach used in the study are presented in Section 3. The results of the simualtion studies are given in Section 4. The paper ends with a summary and conclusions.

2. *Range and Deficit Analysis of Water Storage*

Let $[X_t]$ be a sequence of random variables, and $S_i = \sum_{t=1}^{i} X_t$

where i = 1, 2, . . . , n. The random variable S_i is called the cumulative or partial sum. The maximum partial sum M_n, the minimum partial sum m_n, the range of partial sums R_n, and the maximum accumulated deficit of partial sums D_n for the sumple size n are defined as

$$M_n = \max\,[0, S_1, S_2, \ldots, S_n] \tag{1}$$

$$m_n = \min [0, S_1, S_2, \ldots, S_n]$$

$$R_n = M_n - m_n$$

$$D_n = \max [0, M_1 - S_1, M_2 - S_2, \ldots, M_n - S_n].$$

In storage analysis, the partial sum S_i represents the amount of water in the reservoir. The required storage capacity is usually taken as either the range R_n or the deficit D_n, depending on whether it is full regulation or partial regulation; the study of which is either called range analysis or deficit analysis.

Range analysis is sometimes referred to as the infinite reservoir theory. Although the object of analysis is simply the properties of the partial sums of random variables, one may conceive the existence of a reservoir that can store any water surplus and can supply any water depletion. In deficit analysis, on the other hand, the sequence $[S_i]$ is interpreted to represent the storage level in a semi-infinite reservoir which has a top but no bottom.

The exact expected range $E(R_n)$ of random variables with a general covariance structure for the sample size $n \leqslant 3$ was first derived by Salas (1972). He also gave approximate expressions for the mean and the variance of the range of periodic-stochastic processes. Earlier, Yevjevich (1967) conjectured that the expected range of linearly dependent normal variables can be represented by

$$E(R_n) = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} i^{-1} \, [\mathrm{Var}(S_i)]^{1/2} \tag{2}$$

where $\mathrm{Var}(S_i) = i\sigma^2$ with σ^2 the population variance. In the case of AR (1) sequence, this expression reduces to

$$E(R_n) = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} i^{-1/2} \, . \, [1 + \frac{2}{i} \sum_{k=1}^{i-1} (i-k)\rho_k]^{1/2} \tag{3}$$

Simulations carried out by Salas (1972) showed that the variance of the range is linearly related to the sample size n. Meanwhile, the lack of theoretical work on defict analysis has been recognized by Gomide

(1975) who first presented the analytical treatment of the distribution of the deficit. He showed that the analytical approach to deal with range and defict analyses for stationary random variables follows directly from the theory of Markov chains. Troutman (1976) later derived the asymptotic distribution of the deficit D_n for the general case of dependent, periodic net inputs and the special cases of dependent, non-periodic net inputs to the reservoir.

For the special case of the sample size n=1, the exact expected values $E(R_n)$ and $E(D_n)$ for independent standard normal variables are given by Anis and Lloyd (1953) and Gomide (1975), respectively, as

$$E(R_1) = \sqrt{\frac{2}{\pi}} \simeq 0.7978 \tag{4}$$

and

$$E(D_1) = \frac{1}{\sqrt{2\pi}} \simeq 0.3989 \tag{5}$$

These expressions are also valid even for dependent normal variables since only one random variable is considered. Therefore, all curves of $E(D_n)$ versus $E(R_n)$ must pass through the point (0.3989, 0.7978).

On the other hand, the asymptotic expressions for $E(R_n)$ and $E(D_n)$ are given by (Gomide, 1975 and Troutman, 1976)

$$E(R_n) = \sqrt{\frac{8n}{\pi}} \cdot \sigma \tag{6}$$

and

$$E(D_n) = \sqrt{\frac{\pi n}{2}} \cdot \sigma \tag{7}$$

Combining the above equations yields the asymptotic relation between $E(D_n)$ and $E(R_n)$ as

$$E(D_n) = \frac{\pi}{4} E(R_n) \simeq 0.785398 \, E(R_n) \tag{8}$$

Asymptotically, the mean deficit as function of the mean range can be determined from the line with slope equal to $\pi/4$ and passing through the origin.

3. *Experimental Approach in the Analysis of Range and Deficit*

For practical applications, a simple relation between $E(D_n)$ and $E(R_n)$ for moderate sample size n is very useful. In this study, the data generation method is used to determine the means and variances of R_n and D_n for sample sizes $n = 2, 3, 4, 5, 10, 15, 20, 25, 30, 40,$ 50, 60, 100, 250 based on a sample of 100,000 normal random numbers. The sequences of net inputs into the reservoir generated are either independent normal variables, first-order Markov dependent or AR(1) variables, and autoregressive moving average (ARMA(1, 1)) dependent for various values of the parameter set.

Based from preliminary experimental results, a high linear correlation exists between $E(D_n)$ and $E(R_n)$. Moreover, experimental curves obtained by data generation suggest that the variance of deficit is linearly related to the sample size n. Thus, the empirical approach used in this study is to relate $E(D_n)$ to $E(R_n)$ and the $\text{Var}(D_n)$ as a function of the sample size n by the least squares fitting techniques. Specifically, $E(D_n)$ and $\text{Var}(D_n)$ are expressed as

$$E(D_n) = A + B \,.\, E(R_n) \tag{9}$$

$$\text{Var}(D_n) = \sigma^2 \,(A + B \,.\, n) \tag{10}$$

where the A's and B's are parameters to be estimated and σ^2 the variance of the sequence.

4. *Results and Discussion*

Using the regression parameters A and B estimated via simulation and substituting the expression for $E(R_n)$ given by Eq. (3) into Eq. (9), the $E(D_n)$ of the independent standard normal variables is estimated. Figure 1 shows the comparison of the expected deficit obtained from simulated samples and those computed by Eq. (9) for the independent standard normal variables. The relative differences are less than 2.0 percent for sample size n greater than 3.

The exact variances of the deficit and the range for finite values of n are not known even for i.i.d normal variables. Thus, an approximate equation is obtained using the data generation method. The experimental curve for VAR(D_n) is derived by simulation. The plot of the values of VAR(D_n) against n suggests that a straight line fit is a good approximation. For i.i.d standard normal variables, Figure 2 shows the plot of Var(D_n) against n for n up to 250. The straight line fit shows to be a good approximation.

In the case of the first-order Markov model with $\rho = 0.2, 0.4, 0.6$ and 0.8, the same procedure used for the independent standard normal variable is followed to estimate the values of the mean and variance of deficit. Figure 3 shows the linear regression parameters of Eq. (9) for various values of ρ used for the estimation of expected deficit. The plot is also particularly useful for finding the values of A and B for other values of ρ not used in the simulation.

Experimental results also show a good straight line relationship between $E(D_n)$ and $E(R_n)$ for the first-order Markov model. For instance, Figure 4 shows the linear relationship between $E(D_n)$ and $E(R_n)$ for $\rho = 0.20$.

Figure 5 through 8 show the comparison of the expected deficit determined by data generation and by the approximate expression (Eq. 9) for various values of ρ and sample size n. The plots indicate that for $\rho \leqslant 0.60$, Eq. (9) is a good approximation although for practical purposes it can be used for any value of ρ.

As regards the variance of deficit, experimental data obtained by simulation for the variance of deficit plotted against n indicate that a straight line fit is adequate in cases of $n \geqslant 5$. Therefore, the variance of deficit is approximated as a linear function of n as in Eq. (10) where the linear regression coefficients A and B are functions of ρ. Figures 9 through 12 show the plots of Var(D_n) against n up to 250 for various values of ρ. It is evident that the straight line fit is a good approximation.

As far as the ARMA(1,1) model is concerned, O'Connell (1971) pointed out that part of the parameter space of the ARMA(1,1) model that is of particular interest in synthetic hydrology is when $\varphi > \theta > 0$. Thus, three combinations of the parameter set within the parameter space of the ARMA(1,1) model are considered. The plots of $E(D_n)$ obtained from simulated samples and estimated by Eq. (9), and Var (D_n) against n as estimated by Eq. (10) are shown in Figures 13 through 17. It is shown that Eq. (9) is reasonable approximation to $E(D_n)$ of the ARMA (1,1) model. Furthermore, the plots also show that the linear function approximation of the Var(D_n) as function of n is also quite good.

5. *Summary and Conclusions*

Range analysis and deficit analysis are two analytical approaches in determinig the storage capacity of a reservoir. The range is used for full regulation of the river discharges while the deficit is used in the case of partial regulation.

Exact expressions for the mean and the variance of the range and the deficit of partial sums of stationary random variables are not yet available for moderate sample sizes. However, approximate expression for $E(R_n)$ of linearly dependent normal variables has been proposed elsewhere. Inasmuch as the range and the deficit both relate to the same stochastic property of the cumulative sums, this paper aims to determine a simple relation between the derived variables. The $E(D_n)$ and Var(D_n) are estimated as simple linear function fo $E(R_n)$ and n, respectively. Results indicate this approach is a reasonable approximation.

It has been found elsewhere (Gomide, 1975) that the asymptotic distribution of deficit, when corrected for the mean and the variance of the deficit, provides a good approximation of the exact distribution of the deficit of stationary series. Thus, once the mean and the variance of deficit have been estimated, the distribution of deficit for given sample size can be approximated by correcting the standardized asymptotic density for the mean and variance of deficit. Thus, any design criteria other than the expected value can then be used in the design of storage capacity considering the associated risk and uncertainty.

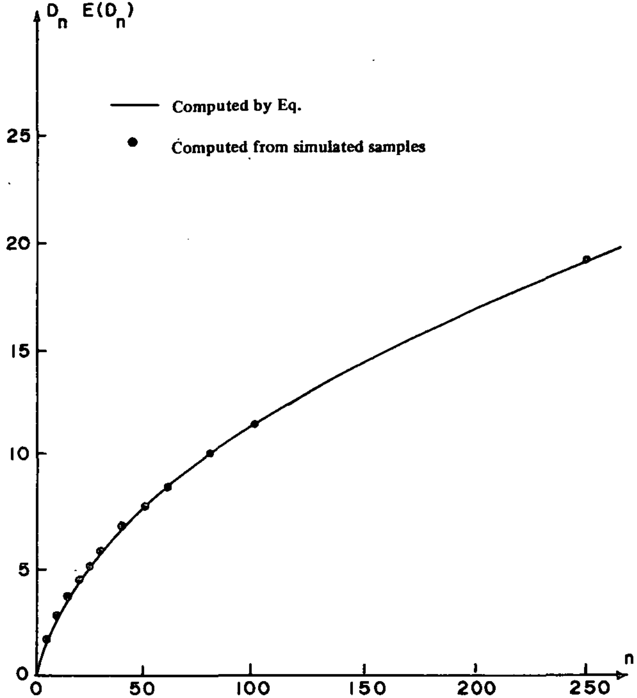

Figure 1. Comparison of the expected deficit obtained from simulated samples and computed by Eq. 9 for independent standard normal net inputs.

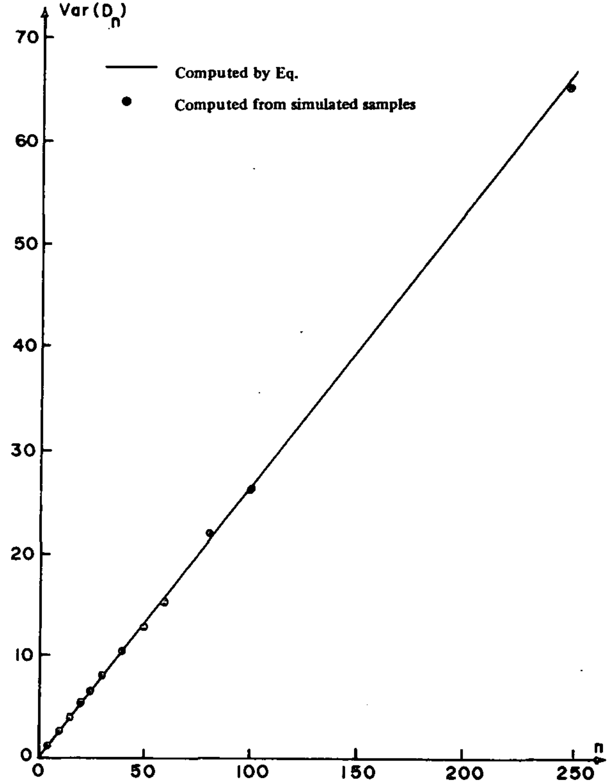

Figure 2. Variance of deficit computed from simulated samples and the fitted linear function for independent standard normal net inputs.

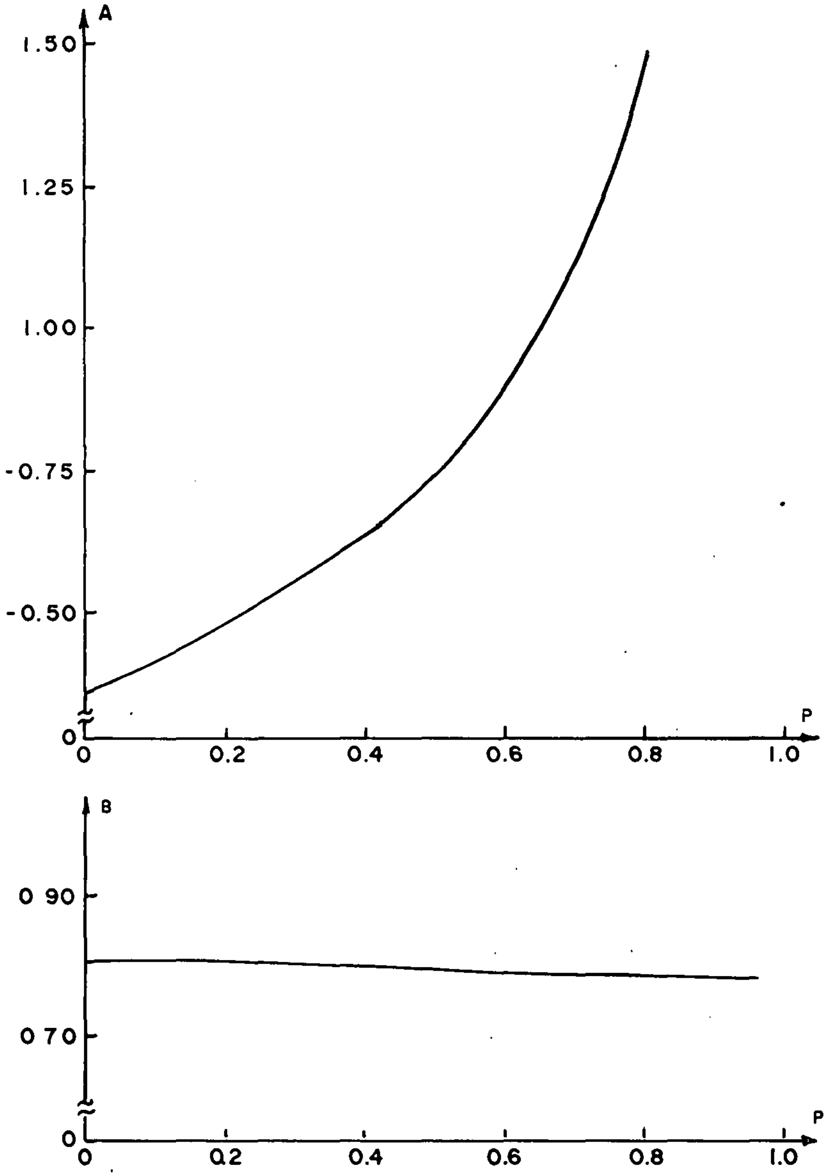

Figure 3. Linear regression parameters for the expected deficit as function of expected range for the first-order linear Markov model.

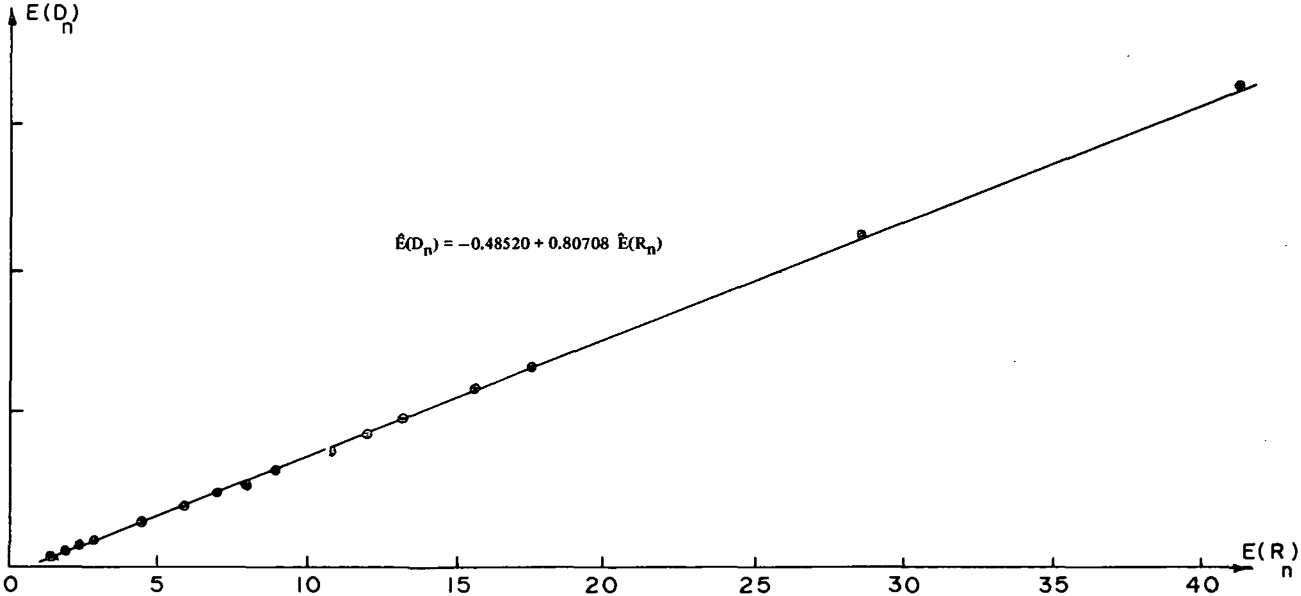

Figure 4. **Relation between mean deficits and mean ranges obtained from simulated samples for the first-order autoregressive (AR) model with $\phi = 0.20$.**

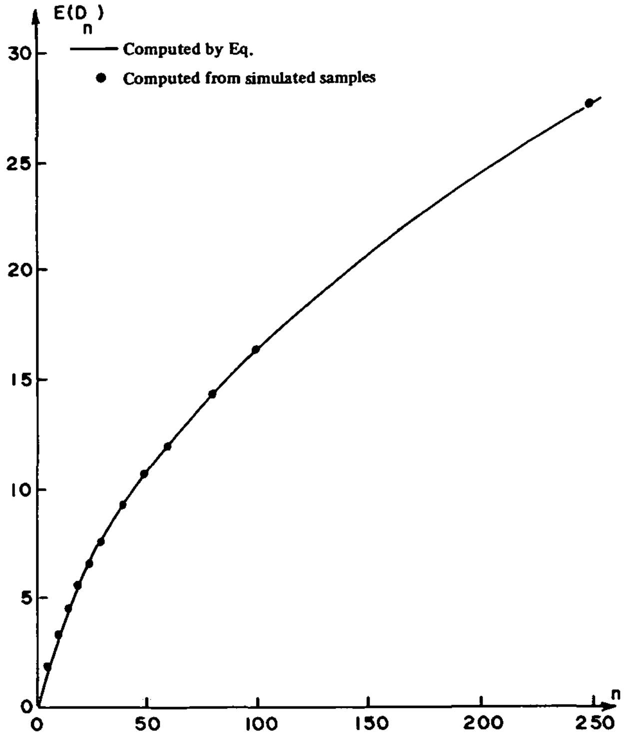

Figure 5. Comparison of the expected deficit obtained from simulated samples and computed by Eq. for the AR (1) model with $\phi = 0.20$.

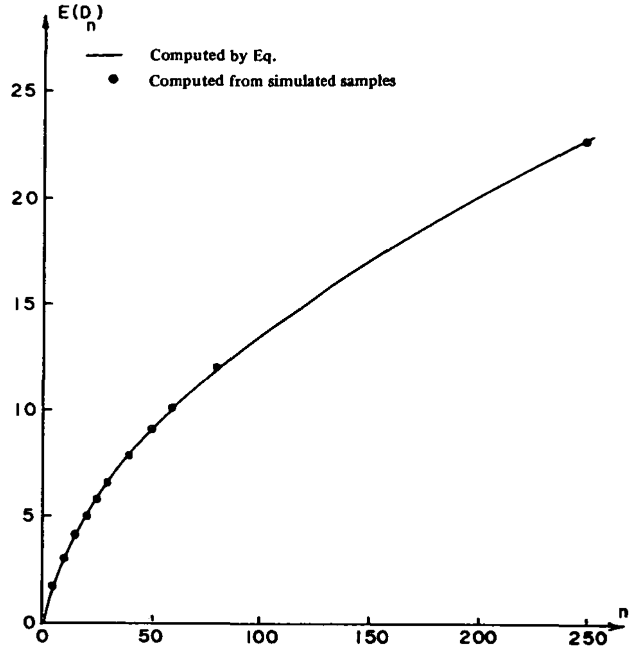

Figure 6. Comparison of the expected value of deficit obtained from simulated samples and computed by Eq. for AR (1) model with $\phi = 0.40$.

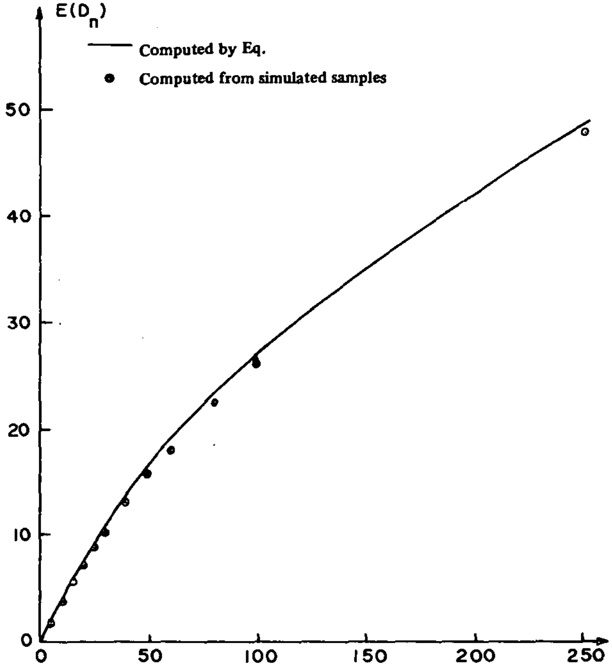

Figure 7. Comparison of the expected value of deficit obtained from simulated samples and computed by Eq̂. for the AR (1) model with $\phi = 0.60$.

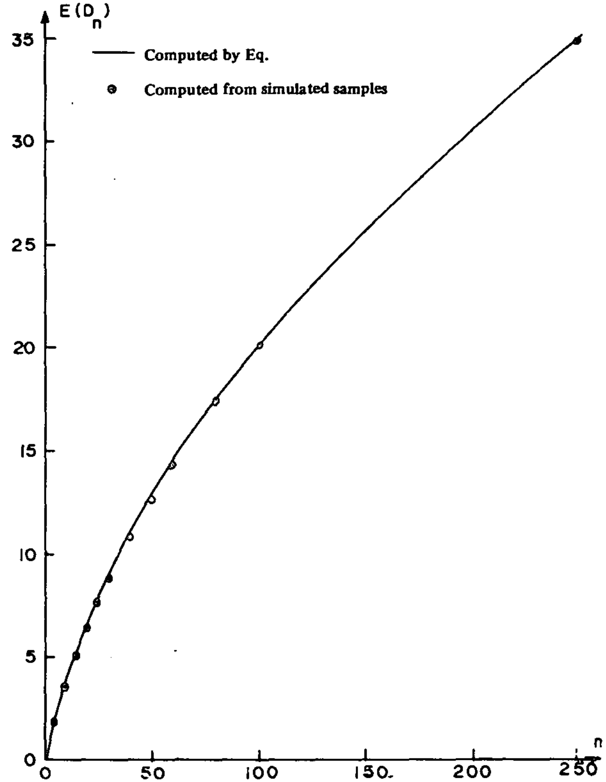

Figure 8. Comparison of the expected deficit obtained from simulated samples and computed by Eq. for the AR (1) model with $\phi = 0.80$.

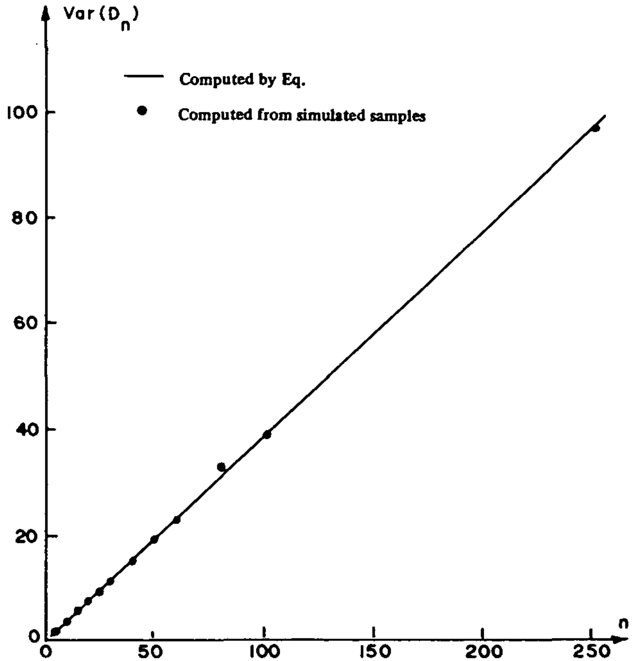

Figure 9. Variance of the deficit obtained by simulation and the fitted linear function for the first-order AR model with $\phi = 0.20$

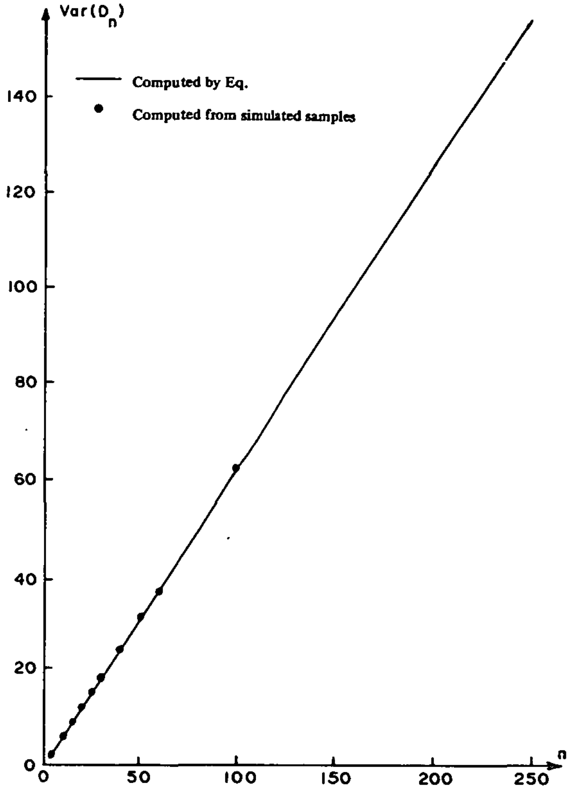

Figure 10. Variance of the deficit obtained from the simulated samples and the fitted linear function for the AR (1) model with $\phi = 0.40$.

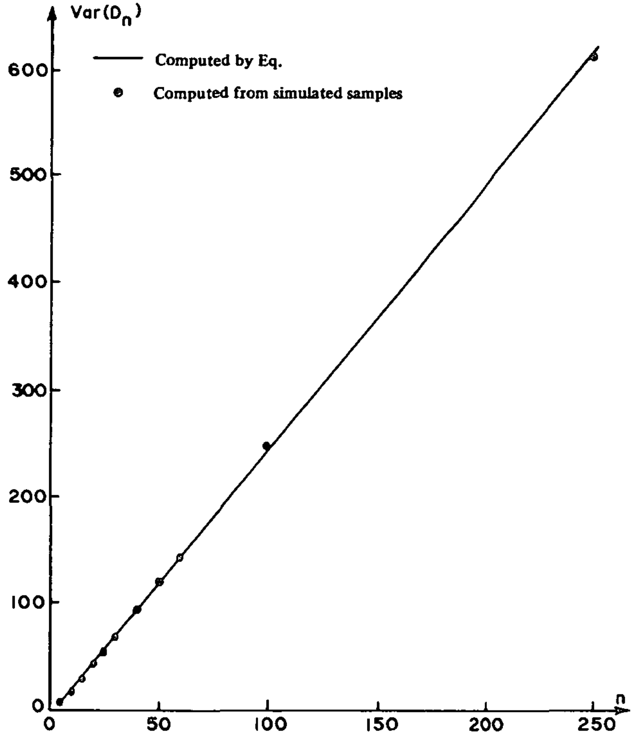

Figure 11 Variance of the deficit obtained from simulated samples and the fitted linear function for the AR (1) model with ϕ = 0.06

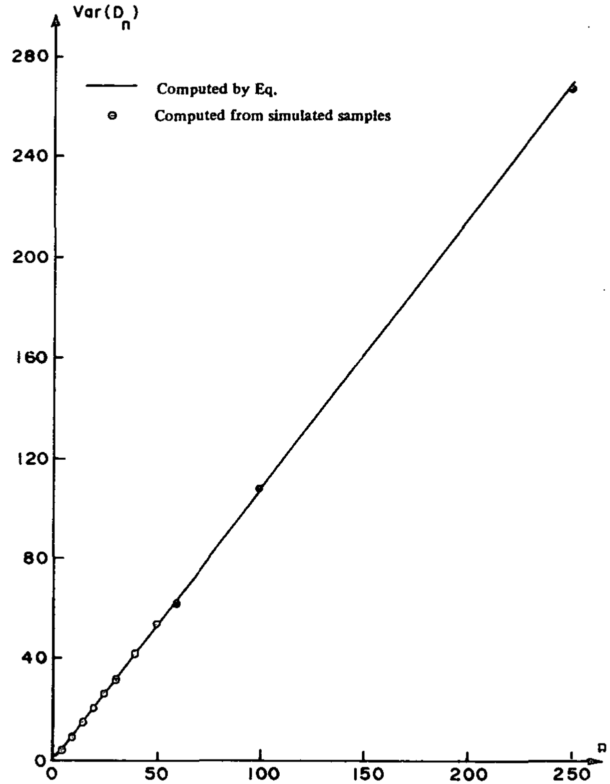

Figure 12 Variance of the deficit obtained from simulated samples and the fitted linear function for the AR (1) model with ϕ = 0.80

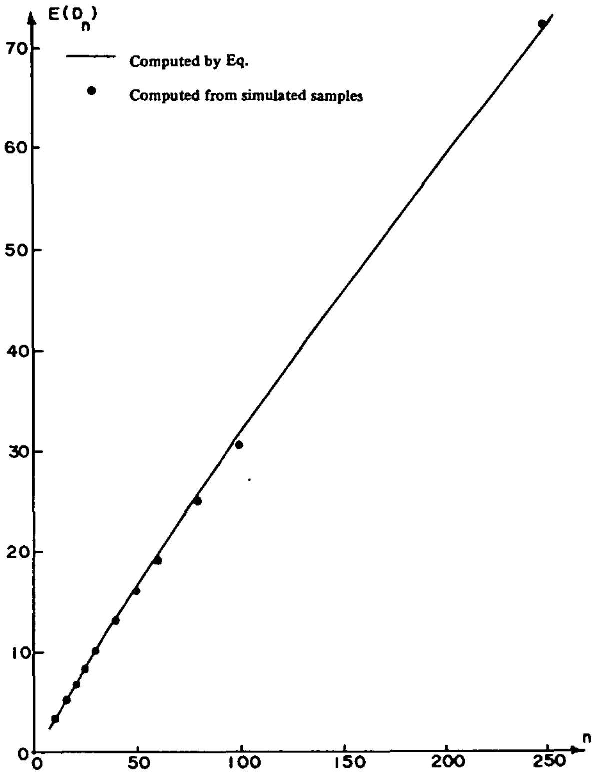

Figure 13. Comparison of the expected deficit obtained from simulated samples and computed by Eq. for the ARMA (1.1) model with parameters $\phi = 0.990$, $\theta = 0.368$.

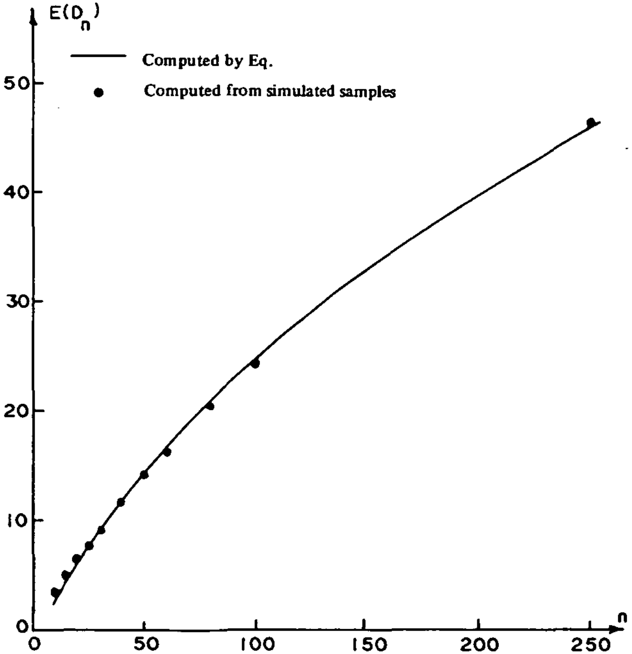

Figure 14. Comparison of the expected deficit obtained from simulated samples and computed by Eq. for the ARMA (1.1) model with parameters $\phi = 0.90$, $\theta = 0.627$.

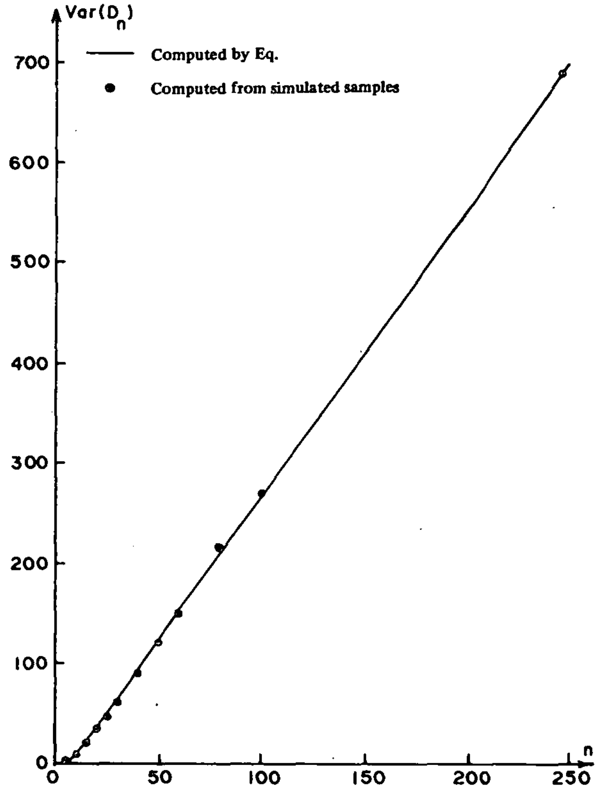

Figure 15. Variance of the deficit obtained from simulated samples and the fitted linear function for the ARMA (1.1) model with parameters $\phi = 0.900$, $\theta = 0.627$.

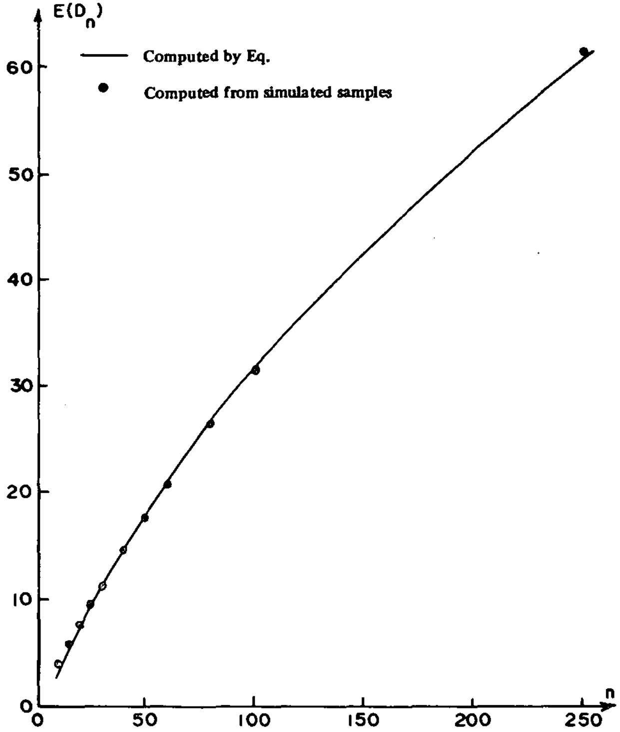

Figure 16. Comparison of the expected deficit obtained from simulated samples and computed by Eq. for the ARMA (1,1) model with parameters $\phi = 0.90$, $\theta = 0.10$.

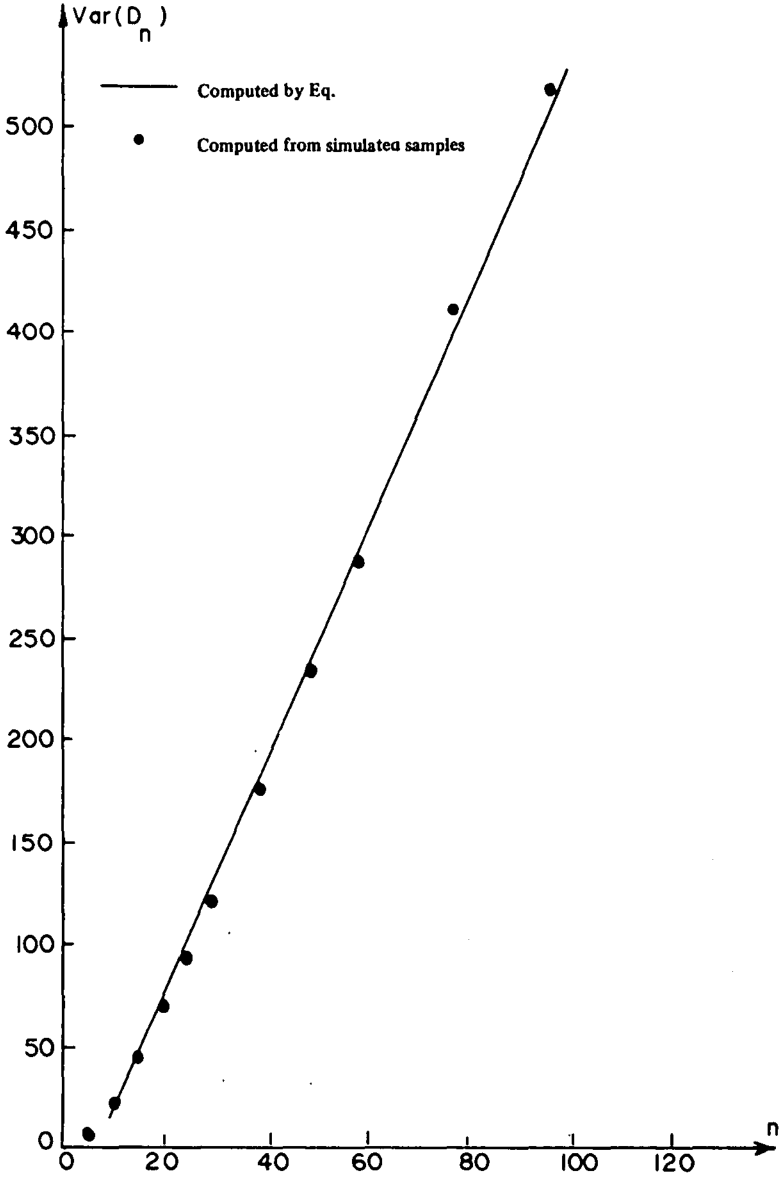

Figure 17. Variance of the deficit obtained from simulated samples and the fitted linear function for the ARMA (1,1) model with parameters $\phi = 0.900$, $\theta = 0.100$.

REFERENCES

Anis, A. A. and E. H. Lloyd, 1953. "On the range of partial sums of a finite number of independent normal variates", Biometrika, 40, pp. 35-42.

Gomide, F. L. S. 1975. "Range and deficit analysis using Markov chains", Hydrology Paper No. 79, Colorado State University, Fort Collins, Colorado.

O' Connel, P. E., 1971. "A simple stochastic modelling of Hurt's law", Int. Symp. on Mathematical Models in Hydrology, Warsaw, pp. 327-358.

Salas-La Cruz, J. D. 1972. "Range analysis for storage problems of periodic stochastic processes", Hydrology Paper 57, Colorado State University, Fort Collins, Colorado.

Troutman, B. M., 1976. "Limiting distributions in storage theory", Ph.D. Thesis, Colorado State University, Fort Collins, 162 pp.

Yevyevich, V. M., 1967. "Mean range of linearly dependent normal variables with applications to storage problems", Water Resources Res. Vol, 3, pp. 633-671.

ACKNOWLEDGMENTS

The author would like to thank the University of the Philippines at Los Baños – For Foundation Program on Environmental Science and Management (PESAM) for the fellowship grant and the U.S. National Science Foundation Project No. CEE-8110782 "Stochastic Modelling of Geophysical Time Series" at Colorado State University for the financial support.