EVAPORATION DATA STOCHASTIC GENERATION FOR KING FAHAD DAM LAKE IN BISHAH, SAUDI ARABIA

Abdulmohsen A. Al-Shaikh
Associate Professor, Civil Engineering Department, College of Engineering, King Saud University

Abstract: Generation of evaporation data generally assists in planning, operation, and management of reservoirs and other water works. Annual and monthly evaporation series were generated for King Fahad Dam Lake in Bishah, Saudi Arabia. Data was gathered for period of 22 years. Tests of homogeneity and normality were conducted and results showed that data was homogeneous and normally distributed. For generating annual series, an Autoregressive first order model AR(1) was used and for monthly evaporation series method of fragments was used. Fifty replicates for annual series, and fifty replicates for each month series, each with 22 values length, were generated. Performance of the models was evaluated by comparing the statistical parameters of the generated series with those of the historical data. Annual and monthly models were found to be satisfactory in preserving the statistical parameters of the historical series. About 89% of the tested values of the considered parameters were within the assigned confidence limits

Key Words: Evaporation, stochastic generation, time series

1. INTRODUCTION

In hydrologic practices, it is often necessary to estimate evaporation rates that will occur from reservoirs for operational and management purposes. The available evaporation data are often quite short, therefore there is a need for stochastic generation of such data for better understanding of possible future values of evaporation.

Time series analysis has been used in stochastic generation of hydrologic data sets for different sites. For examples, Nicks and Harp (1980) used Thomas and Fiering models to generate temperature and solar radiation data in Oklahoma, U.S.A. They tested these models against historical records and found them adequate for generating simulated data for periods of the same record length. Srikanthan and McMahon (1983) applied several stochastic models to simulate annual, monthly and daily evaporation data for Australia. For annual and monthly synthesis, the evaporation values were generated independently and dependently of rainfall. They found AR(1) satisfactory for annual evaporation and method of fragments satisfactory for monthly evaporation with an adequate length of historical data. Tsakiris (1986) used Thomas-Fiering model, Two-Tier model, and method of fragments to generate monthly potential evapotranspiration data in Greece. He found

Two-Tier model suitable for stochastic generation of monthly potential evapotranspiration values when modeled independently of rainfall.

Lag-one Markov model was used by Al-Solai (1991) to generate annual rainfall sequences and method of fragments was used to generate monthly rainfalls from generated annual rainfall sequences for eight selected rainfall stations in Saudi Arabia. After that, hecompared the statistical parameters of generated sequences with those of historical data and found both methods were simple and very efficient in terms of computer time, and satisfactory in preserving the statistical parameters of the historic data. Koutsoyiannis (1992) used a nonlinear disaggregation method with a reduced parameter set. The general idea for his study was the conversion of a sequential stochastic simulation model, such as a seasonal AR(1), into a disaggregate model. Two different configurations were used, lower-level AR(1) and higher-level AR(1). Both configurations were found to perform well with regard to the correlationof consecutive lower-level variables each located in consecutive higher-level time steps for ifferent hydrologic approaches. Al-Eid (1993)investigated stochastic generations of evaporation sequences by using time series. He generated annual and monthly evaporation sequencesfor ten locations in Saudi Arabia. He used first order Markov model for generating annual evaporation and three auto-regressive models (Fragments, Thomas-Fiering monthly model, and the two-tier model) for generating monthlyevaporation. After that, he compared the statistical parameters of the generated sequences with those of the historical data. Annual model was found to be satisfactory in preserving the statistical parameters of the historical series. Among the three monthly models used the Thomas-Fiering model gave the best presentation of gave the best presentation of the statistical parameters.

Bender, M. (1994) used Seasonal Auto-Regressive Integrated Moving-Average modeling (SARIMA) for long-range stream flow forecasting at Manitoba Hydro, Canada, and compared it with statistical time series tools. SARIMA models were found to be more flexible for natural inflow with low upstream storage capacity and high variability. He found that deseasonalized auto-regressive (AR) and moving-average (MA) combination models may be better suited to natural inflow systems that have a large storage capacity, lower variability, and greater response lags to precipitation events. Elsner et al. (1999) examined the annual record of hurricane activity in the North Atlantic basin for the period (1986-1996) from the perspective of time series analysis. They used singular spectrum analysis combined with the maximum entropy method on the time series of annual hurricane occurrences over the entire basin to extract the dominant modes of oscillation.

This paper is aimed toward generating annual and monthly evaporation series for King Fahad Dam Lake in Bishah, Saudi Arabia. Auto-regressive first order model, AR(1), is used to generate annual evaporation series while method of fragments is used to generate monthly evaporation series. Those two methods are used in particular because of their popularity in dealing with cases similar to the case considered here. Validity of these two methods will be tested for the site studied in this paper.

2. DESCRIPTION OF METHODS, SITE AND DATA USED

The formulation of AR(1) for annual evaporation results comprising deterministic and random parts can be written as (Gupta, 1989):

$$E_i = \overline{E} + r_i\left(E_{i-1} - \overline{E}\right) + t_i\, S\sqrt{1 - r_i^2} \qquad (1)$$

where

E_i = evaporation at ith time
$\overline{E}$ = mean of recorded evaporation.
r_1 = lag 1 series or autocorrelation coefficient
S = standard deviation of recorded evaporation
t_i = random variate from an appropriate distribution with a mean of zero and variance of unity.
i = ith position in series from 1 to N years.

Disaggregation models have recently become a major technique for modeling hydrologic time series. Disaggregation models is a process by which divide known higher-level values (e.g. annual) into lower-level ones (e.g. monthly). Thus, they have the advantage of providing the ability to transform a time series form a higher time scale to a lower one. Disaggregation modeling has two advantages. First, it is a technique which often allows for a reduction in the number of parameters with little or no corresponding loss of desirable properties in the generated data. Second, disaggregating allows for increased flexibility in the method used for generation (Salas et al., 1980). A simpler disaggregation scheme is called method of fragments.

According to the method of fragments, the observed monthly values are standardized year by year by dividing monthly evaporation values in a year by the corresponding annual evaporation value, so that the sum of standardized monthly values in any year equals unity. By doing so, one will have N set of fragments of correlated monthly evaporation from N years of record. The sets of fragments are arranged in the same order as they occurred historical, and numbered from 1 to N. The fragments are selected at random by the number computed from INTEGER (N*U+1), where N is a uniformly distributed random number between 0 and 1 the monthly evaporation values are obtained by multiplying the generated annual evaporation values by the fragment chosen (Srikanthan and McMahon, 1983). Therefore

$$\text{Fragment}_{(i,j)} = \frac{E_{i,j(\text{historical monthly})}}{E_{i(\text{historical annual})}}$$

M = INTEGER (N * U+1)

$$E_{i,j(\text{gen. monthly})} = \text{Fragment}_{M,j} * E_{i(\text{gen. annual})} \qquad (2)$$

where i = 1,2,3,,,.n years, j = 1,2,3,...,12 months

AR(1) and method of fragments are used to generate evaporation data which conserve the main statistical characteristics of the historical data. This is verified through comparing values of mean, standard deviation, and correlation coefficient of generated evaporation data with those of historical data. Methods of verification are given below.

For the mean, Standard error of the mean is given as (Mood et al., 1974) $S_x = S/\sqrt{n}$, where S is the sample standard deviation and n is the number of observation. The quantity ($\overline{E}$ - μ) / S_x has a t-distribution with n-1 degrees of freedom, then

$$t = (\bar{E} - \mu) / S_X$$

And for 95% confidence limit

$$E - t_{(0.975,n-1)} S_x < \mu < E + t_{(0.975,n-1)} S_x \quad (3)$$

The value on the left side of the inequality yields the lower limit, and on the right side yields the upper limit for the mean. From the results of Eq. 3 above we are 95% confident that the limits contain the true population mean.

The quantity (n – 1) S^2 / σ^2 has a Chi-square distribution with n – 1 degrees of freedom (Abraham and Ledolter, 1983). For the 95% confidence limits (Mood et al. 1974).

$$\frac{(n-1)S^2}{X^2_{(0.975,n-1)}} < \sigma^2 < \frac{(n-1)S^2}{X^2_{(0.025,n-1)}} \quad (4)$$

Where $X^2_{(0.975,\, n-1)}$ and $X^2_{(0.025,\, n-1)}$ are the 0.0975 and 0.025 quantile points, respectively, of the Chi-square distribution. The value on the left side of the inequality should yield the lower limit, and the right side yields the upper limit for the variance. The square root of these limits should yield the confidence limits for the standard deviation.

For correlation coefficient, the following approximate relation is used to get the 95% confidence limit for the lag-one correlation coefficient (Abraham and Ledolter, 1983)

$$r_1 - Z_{\alpha/2}\ 1/\sqrt{n} < \rho < r_1 + Z_{\alpha/2}\ 1/\sqrt{n} \quad (5)$$

Where n is the number of observations, and $1/\sqrt{n}$ is the standard error of the lag-one sample correlation coefficient r_1, $Z_{\alpha/2}$ is the normal deviate with significance level α.

Selected site is King Fahad Dam Lake located

Table 1. Historical Data Description

Month	Evaporation (mm/day)
January	6.00
February	7.13
March	8.95
April	10.00
May	11.21
June	13.57
July	14.31
August	13.52
September	12.86
October	10.44
November	7.67
December	5.89
Mean	10.10
Standard Deviation	3.10
Coefficient of Variation	0.31

in Bishah, Aseer region, Southwest of Saudi Arabia. The dam has 103m length from its foundation and about 70m from valley level with 507m width and has maximum water surface area about $18km^2$. Its capacity is 325 million m^3. In view of this huge quantity of water, it becomes clear how important is studying of evaporation at the dam to know the quantity of water will evaporate as step in management of this quantity.

Evaporation data used in this study were obtained from Ministry of Agriculture and Water, Saudi Arabia (Al-Qahtany, 1998). Records started from 1975 to 1996 in (mm/day). This period of record is a short record to depend on for future hydrologic planning and management. Table 1 shows the average values of monthly evaporation data of 22 years of records and some basic statistics. Large values of evaporation occur in summer (June, July and August) with maximum value equals 14.31 mm/day in July. On the other hand, low values occur in winter with minimum value equals 5.89 mm/day. These figures show reasonable values but analysis will be conducted to check reliability of data.

2.1 Homogeneity Test

The nature of homogeneity in a hydrologic process is examined statistically with respect to space and time. Space homogeneity means that the occurrence of particular hydrologic events at all places within a statistically homogeneous area are equally likely. Time homogeneity, which concerns this study, is satisfied when the identical events under consideration in the series are equally to occur at all time. Hydrologic characteristics are generally subject to changes due to non-homogeneity. Non-homogeneity in a data is common in hydrologic time series; it is induced by humans or produced by significant natural disruptive factors .

To check the homogeneity the historical data are divided into two or more sub-series and some statistical characteristics are tested for each sub-series (Salas et al. 1980). For this study the 22 values are divided into two sub-series each of length equals 11 values of evaporation. Test of homogeneity depends on comparing z calculated with z tabulated at some confidence level (95% is assumed). z calculated is given as

$$z_{cal} = \frac{\overline{E}_1 - \overline{E}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

Where:

$\overline{E}_1, \overline{E}_2$ = mean of sub-series 1 and 2, respectively.

S_1, S_2 = standard deviation of sub-series 1 and 2, respectively.

N_1, N_2 = sample size of sub-series 1 and 2, respectively.

z tabulated is ranged between ± 1.96. If z_{cal} is within the range of z_{tab}, data is homogeneous, otherwise it is not.

For annual values, z_{cal} = 1.19. Since 1.19 is within ± 1.96, annual data is homogeneous. For monthly values, same procedures are followed. Each sub-series contains 22 x 12/2 = 132 values of evaporation. Results showed z_{cal} = 0.99 so, it can be concluded that monthly data are also homogeneous.

2.2 Normality Test

Prior to using time series models, historical data should be tested for normality. If not normal, data should be transformed using different techniques available in the literature. To facilitate this test, MINITAB computer program was used. MINITAB program uses, among other methods, Anderson and Darling method. The procedures for performing the Anderson and Darling method are given below as described by D'Agostino and Stephens (1986). This test defines a statistical quantity A*. So, in case studied here according to the test, reject the null hypothesis of normality if A* exceeds 0.752 at level of significance = 0.05, on the other hand, accept the null hypothesis of normality if A* less than 0.752.

For annual, A* = 0.678 which is less than 0.752. For monthly values A* values of each month were estimated from Jan. till Dec. as: 0.433, 0.125, 0.155, 0.227, 0.176, 0.500, 0.225, 0.568, 0.564, 0.386, 0.179, and 0.728. All values are less than 0.752. So, data are normally distributed in both annual and monthly basis.

3. APPLICATUON

3.1 Generation of Annual Evaporation Data

Annual evaporation data were generated by using AR(1) model. AR(1) model depends on three parameters: mean, standard deviation, and correlation coefficient for historical data. Values of these three parameters as calculated from historical data are: 3696.94, 278.46 and 0.14 respectively. So recalling Eq. (1)

$$E_i = E + r_1\ (E_{i-1} - \bar{E}) + t_i\ S(1 - r_1^2)^{1/2}$$

or

$$E_i = 3696.94 + 0.14\ (E_{i-1} - 3696.94) + t_i\ 278.46\ (1\text{-}0.14)^{1/2}$$

Fifty replicates, each of length equal to the historical data length (that is 22) were generated at various randum values of t_i. Various parameters were estimated for each replicates and the average values were computed. Overall mean, standard deviation and correlation coefficient were 3691.68, 269.08 and –0.03 respectively.

Maximum generated evaporation value was equal 4375 mm/year. When comparing this value with maximum historical evaporation value, ratio of 4375/4276 = 1.023 is obtained. Ratio of 2539/3148 = 0.81 is obtained for case of minimum values. These two ratios are of certain importance for future planning and operation of the selected site.

3.2 Check of Model Validity

AR(1) model described and used before in generating evaporation data needs a validity check. Validity check is required to insure that generated data preserve the same statistical features as those of historical data.

For comparison of mean, using Eq. (3):

$$S_x = 278.46\ \sqrt{22} = 59.73$$

$$t_{(0.975,21)} = 2.056$$

$$\bar{E} = 3696.94$$

$$\mu = 3691.68$$

Substituting in Eq. 3, we have

$$\bar{E} - 2.056\ S_x < \mu < \bar{E} + 2.056\ S_x$$

$$3574.8 < \mu < 3818.98$$

We are 95% confident that the interval 3574.8 to 3818.98 contains the true population mean (average of means of generated data).

For comparison of standard deviation, using Eq. (4)

$$X^2_{(0.975,\ 21)} = 35.479$$

$X^2_{(0.025, 21)} = 10.283$

$S = 278.46$

$\sigma = 269.08$

$(22\text{-}1) * (278.46)^2 / 35.479 < \sigma^2 < (22\text{-}1) * (278.46)^2 / 10.283$

$214.26 < \sigma < 398$

We are 95% confident that the interval 214.26 to 398 contains the true population standard deviation (average of std. dev.'s of generated data).

For comparison of correlation coefficient, using Eq. (5)

$Sn = 1/\sqrt{22} = 0.213, \quad Z_{0.05/2} = 1.96,$

$r_1 = 0.14, \quad \rho = -0.03$

$r_1 - Z_{a/2}/\sqrt{n} < \rho < r_1 + Z_{a/2}/\sqrt{n}$

$0.14 - 1.96/\sqrt{22} < \rho < 0.14 + 1.96/\sqrt{22}$

$0.27 < \rho < 0.56$

We are 95% confident that the interval –0.27 to 0.56 contains the true population correlation coefficient (average of corr. coef.'s of generated data). So, it can be seen concluded that the suggested model, that is AR(1) model, performs satisfactory.

3.3 Generation of Monthly Evaporation Data

Monthly evaporation data were generated by using method of fragments as was presented. Fifty replicates, each of length equal to the historical data record, were generated foreach month. Table 2 shows the averages of parameters for generated series.

Maximum as well as minimum generated evaporation for each month were compared to those of historical of the same month. Results are presented in Table 2. October has the greatest value of the ratio max. gen. / max. his., while June has the lowest value of the ratio min. gen. / min. his. Month of April has a ratio of minimum > 1, which means that no value of the whole generated values exceeded the minimum historical value in April. This may be due to the presence of some low historical values of evaporation in this month.

Table 2. Average Statistics for Monthly Generated Series

Month	Mean	Std. Dev.	Corr. Coeff.	Max.	Min.
January	185.5	27.9	0.02	1.15	0.86
February	200.0	29.1	-0.35*	1.07	0.80
March	281.2	39.4	-0.08	1.10	0.85
April	315.1	35.7*	-0.13*	1.07	1.04
May	352.1	41.3	-0.12	1.04	0.79
June	402.5	40.0	0.03	1.15	0.76
July	442.2	49.7	-0.11	1.18	0.81
August	421.2	49.6	-0.10	1.14	0.89
September	379.4	34.6	0.03	1.13	0.76
October	308.2	32.8	-0.08	1.22	0.80
November	226.3	28.6	-0.04	1.13	0.90
December	186.2	30.7	-0.23*	1.20	0.85

3.4 Check of Model Validity

Method of fragments should conserve the main statistical characteristics of the historical data. For the purpose of comparison of means, the average of means of generated data for all months were listed in Table 2. 95% confidence intervals of means were calculated based on Eq. (3). From Table 2 it can be seen that all means of generated data were inside the 95% confidence interval.

Confidence interval of standard deviation were calculated based on Eq. (4). Table 2 also shows the average of standard deviations of monthly generated data. According to thisTable, standard deviations of monthly generated data are within the confidence interval, except that of April.

Table 8 presents the average correlation coefficients of all months. Confidence intervals were calculated based on Eq. (5). The correlation coefficient of February, April and December were out of the 95% confidence interval.

4. SUMMARY AND CONCLUSIONS

In this paper, evaporation data time series were generated for King Fahad Dam in Bishah, Saudi Arabia. Historical data were collected from Ministry of Agriculture and Water for the period of 22 years (1975 – 1996). Homogeneity of data was tested. Result showed that data are homogeneous for both annually and monthly basis. Prior to time series generation, normality of annual and monthly data was also examined using Anderson and Darling method. Data were found to be normally distributed.

An autoregressive first order model AR(1) was used to generate annual evaporation data. Method of fragments was used to generate monthly evaporation data by disaggregating generated annual evaporation data into monthly values. In each model, fifty replicates, each with twenty two evaporation events were generated.

For evaluating, historical and annual generated data were compared where confidence intervals for statistical parameters (mean, standard deviation, and correlation coefficient) were calculated at 95% significant level. These statistical parameters were all within the range of confidence intervals. So, it can be concluded that the parameters of historical data are well preserved and annual model AR(1) works satisfactorily with involved minimum set of parameters and simplicity of calculation. Maximum and minimum values of generated data were compared to these of historical. Maximum generated value can be 1.023 as to that of historical, while minimum can be 0.81.

Monthly historical and generated data were also compared to evaluate results of fragments method. That was achieved through calculating confidence intervals for statistical parameters for all the 12 months of the year. For mean, all months were within the confidence interval, while there was one month, April, out of interval for standard deviation. Correlation coefficient of February, April, and December were found to be out of the range. Generally speaking, it can be concluded that method of fragments gave accepted results and preserved the statistical features for (12 + 11 + 9) / 36 = 88.9% of the values of parameters tested for all the months. Maximum values generated compared to that of historical was found to be 1.22 in October, while that value of minimum was 0.76 in June. Main recommendations are:

1. Time series generation for evaporation data is highly required in order to get better future operation and management for dams and other hydrologic purposes.

2. Data gathered should be first tested for homogeneity and normality using any suitablestatistical technique.
3. AR(1) model was found to perform satisfactorily in generating annual series. But othermodels, such as AR(2) could be also used and evaluated.
4. Methods of fragments was generally suitable for generating monthly time series. Othermethods, such as Two-tire model, and Thomas-Fiering model, may give better results.
5. Seasonality was ignored in this study. Future study may consider seasonality and generate time series accordingly to get better results.
6. Forecasting of future evaporation values canbe conducted for future study. The last tworecommendations are to be considered in anext paper by the author.

REFERENCES

① Abraham, B. and Ledolter, J. (1983) *Statistical Method for Forecasting*, New York, John Wiley & Sons, Inc

② Al-Eid, A. (1993). *Stochastic Generation of Evaporation Sequences*, M.S. Theses, King Saud University, Riyadh.

③ Al-Qahtany, S. (1998). *Evaporation Analysis As Affected by Climatic Factors in Bishah, Saudi Arabia*, Senior Project in Civil Engineering Department in King SaudUniversity, Riyadh, Saudi Arabia.

④ Al-Solai, A. (1993). *Stochastic Generation of Annual and Monthly Rainfall Sequences*, M.S. Theses, King Saud University, Riyadh, June.

⑤ Bender, M. and Simonovic, S. (1994). *Time Series Modeling for Long Range Stream-Flow Forecasting*, Journal of Water Resources Planning and Management, Vol. 120, No. 6, Nov-Dec. p. 857-870.

⑥ D'Agostino, R.B. and Stephens, M.A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, Inc., New York.

⑦ Elsner, J., Kara, A.., and Owens, M. (1999). *Fluctuations in North Atlantic Hurricane Frequency*, Journal of Climate, Vol. 12,No.2, Feb. p. 427-437.

⑧ Gupta, R. (1989). *Hydrology and Hydraulic Systems*, Prentice Hall, Inc., Englewood, New Jersey.

⑨ Kontsoyiannis, D. (1992). *A Non-Linear Dis- Aggregation Method with a Reduced Parameter Set for Simulation of Hydrologic Series*, Journal of Water Resource Research, Vol. 28, No. 12, Dec. p. 3175-3191.

⑩ Mood, A., Graybill, F. and Boes, D., (1974). *Introduction to the Theory of Statistics*, McGraw-Hill International Editions, p. 564, .

⑪ Nicks, A., and Harp, J., (1980). *Stochastic Generation of Temperature and Solar Radiation Data*, Journal of Hydrology, Vol. 48, Jan. p. 1-17.

⑫ Salas, J., Delleur, J., Yevjevich, V. and Lane, W., (1980). *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Colorado, USA.

⑬ Srikaanthan, R. and McMahon, T. (1983). *Stochastic Simulation of Evaporation Data for Australia*, Nordic Hydrology, p. 207-228.

⑭ Tsakiris, G. (1986). *Stochastic Generation of Monthly Potential Evaporation in Arid and Semi-Arid Regions*, Arabian Journal for Science and Engineering, Vol. 11, No. 4, p. 371-380.

Abdulmohsen A. Al-Shaikh, Associate Professor, Civil Engineering Department, College of Engineering, King Saud University, Riyadh

11421, Saudi Arabia
(Received September 7. 2001; accepted September 17, 2001)