# AN ANALYSIS OF A UNIVARIATE MULTI-LAG AUTO-REGRESSIVE STOCHASTIC STREAMFLOW MODEL

by

German T. Velasquez*

## ABSTRACT

An analysis of a single-site auto-regressive multiple correlation is presented for the purpose of justifying a streamflow generator model. The multi-lag analysis as presented by Fiering and Jackson (1971) is studied through its application to two streamflow records and through the use of different correlation coefficients. The simple hydrologic model by Fiering (1967) is also studied to determine its ability to mimic the actual flow instabilities. Various model parameters are tested and their effect on the general characteristics of the resulting generator are determined.

This analysis yielded the following conclusions: The Fiering model for estimate correlation between flows does not mimic actual correlograms in the general sense. Computer modeling presents instabilities in generation especially for AR models with flows lagged greater than seven. Care must be taken in the choice of a random number generator so that the proper statistics will be preserved. Data generation should only be made after a careful analysis of the stochastic part of the model because short term generations are greatly affected by this part of the model.

## INTRODUCTION

It is usual for planning engineers to require a historical study of the flows of a prospective project domain. Historical data, however, are oftentimes too short and inadequate to produce the necessary statistical and hydrological parameters required for a confident decision. It is necessary therefore to produce replicate synthetic data records that are statistically equivalent to the historical sequences. Of the most typical modeling methods used in the generation of synthetic flow series, the basic auto-regressive model is a common initial tool in an analysis. Although more complex models are now basically used for data generation, an analysis of the basic structure of the *AR* model would be beneficial for the planning engineer especially when model modifications are necessary to fit a specific project.

* Graduate student of Hydraulic Engineering at Nagoya University, Nagoya, Japan

This study is based on the initial efforts of Thomas in 1965 regarding rainfall and was later modified by Fiering in 1967 as a model capable of mimicking the historical streamflow correlation coefficients. Although typical models are already studied during those times, it was not until Matalas (1967) worked out the basic transformation equations for non-normal distributions that developments took to a start. In 1971 Fiering and Jackson produced a general *AR* model monograph that compiled their recommendations for typical modeling procedures. Box and Jenkins, however expanded these for a general synthetic hydrology theory wherein basic models were later known after them. Current research are centered on long memory models that are capable of reproducing the critical returning conditions such as the Hurst effect. General application programs are currently available to suit various purposes of modeling.

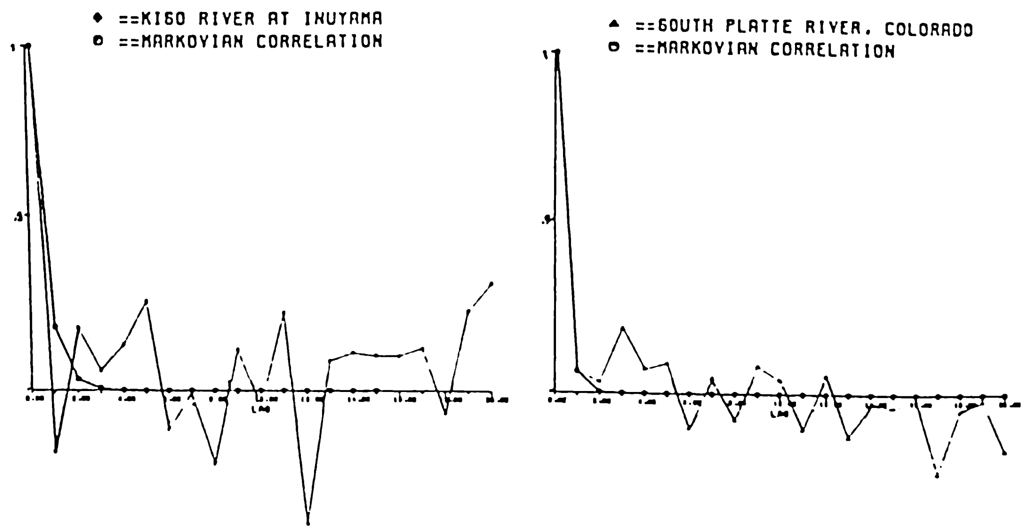



Figure 1: Multiple correlograms of the historical data at Kiso River at Inuyama (left) and data at South Platte River, Colorado (right)

## MODEL CONCEPT

Contrary to the Markovian assumption of exponential decay between the correlation coefficients and lags, actual plots of relations between the two display wide fluctuations in magnitude and abrupt changes in sign (Figure 1). It is quite logical therefore to include this property through a multi-lag analysis of a stochastic streamflow model to enabie the resulting synthetic data to assume a more realistic stance. The question however of including this property in a stochastic model brings about the suggestion of Fiering to utilize the generalized Thomas (1965) model instead of using the historical coefficients. The implications of this is studied to determine the effectivity of the Thomas model in generating a more realistic generated data set. For the determination of the number of lags to include in the $AR(i)$ streamflow-model[1], the step outlined by Feiring and Jackson is also analyzed through generation of synthetic data from $AR(1)$ to $AR(10)$ streamflow models.

---

[1] Auto-Regressive model that includes flows up to lag i

The general model used is an elaboration of the Markovian lag-one model. Changes arise primarily to accommodate run-off conditions in which the groundwater aquifer stores water from season to season and then contributes a fraction of it each season for part of the total run-off. The behavior of the groundwater storage is represented by a multi-lag model; a model with a long memory. It is assumed that more than one past flow matters, thus the following deterministic part of the model is obtained:

$$d_i = \beta_1 q_{i-1} + \beta_2 q_{i-2} + \cdots + \beta_m q_{i-m}, \tag{1}$$

with $m > 1$. The justification of the above model comes from the fact that it is capable of reproducing important statistical characteristics of the historical flow patterns, specifically the jagged correlograms of the historical data.

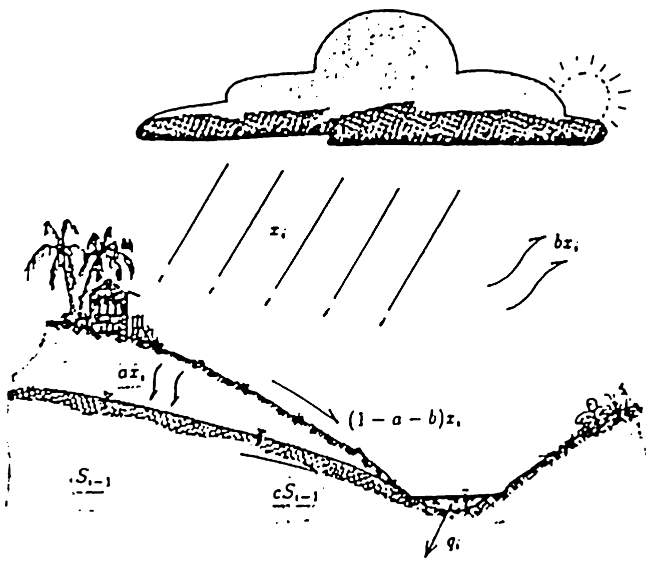



Figure 2: Hydrologic model by Fiering(1967)

For the number of lags to be included in the model, a reasonable procedure would be the one presented by Fiering (1967)[2]. Starting with a deterministic linear regression model.

$$\hat{q}_i = \beta_0 + \beta_1 q_{i-1}, \tag{2}$$

estimates of $\beta_0$ and $\beta_1$ are found using linear regression of the historical data. The goodness of fit of our model is checked through the calculation of the coefficient of determination between

[2] later extended by Fiering and Jackson 1971)

the actual values, $q_i$ , and the computed values, $\hat{q}_i$ , using the linear model. The same procedure is followed for lags greater than one.

Because changes in $R^2$ reflect changes in the explanatory power of the model, Fiering and Jackson (1971) suggested to stop including lags when $R^2$ values reaches a relatively flat plateau. When the coefficient's value exceeds unity, then we should also stop including lags because flows cannot be correlated by more than a value of one; also, because of the immense computational load involved, Fiering suggested an upper limit of 8 to 10 previous flows, reasoning that it is dangerous to include lags that approach the length of the historical record because estimates of the correlation coefficients becomes highly unstable and imprecise.

Since historical flows may not produce realistic estimate multiple correlation coefficients, Fiering (1967) suggested using the generalized Thomas model to mimic actual correlograms. Starting with a simple basin model in figure 2; let the precipitation record be denoted by a set of annual values($x_i$ ), $i = 1,2,\ldots, n$ . Assume that $ax$ is the amount of rain that percolates through the soil to the groundwater during any year $i$ and that $bx$ is the amount of rain lost directly to the atmosphere by evaporation and transpiration. Assume further that during any year $i$, the amount of groundwater that leaves the aquifer storage and drains into the stream is $cS_{i-1}$ where $S_{i-1}$ is the amount of groundwater held in storage at the start of the year $i$. It is assumed that $x_i$ derive from a normal distribution and that their pattern of time dependence is characterized by a Markovian process in which the lag-j serial correlation coefficient is written $\pi_j = \pi^j$ where $\pi$ is the lag-one serial correlation coefficient. Values of a, b, and c are known (or estimated) and the values of the mean and the variance of the historical flows are given.

Fiering showed that for serially uncorrelated rainfall, the coefficient of correlation between lagged flows may be expressed as,

$$r(q_i, q_{i+k}) = \frac{\sigma_x^2}{\sigma_q^2}\left[\pi^k(1-a-b)^2 + (1-c)^{k-1}(1-a-b)ac\sum_{l=0}^{k-1}\left(\frac{\pi}{1-c}\right)^l \right.$$
$$+ac(1-c)^k(1-a-b)\left(\frac{\pi}{[1-\pi(1-c)]}\right)$$
$$+a^2c^2(1-c)^{k-2}\sum_{l=0}^{k-1}(1-c)^{-2l}\left[\frac{1}{1-\pi(1-c)} - \sum_{m=0}^{l}\pi^m(1-c)^m\right]$$
$$+ac(1-a-b)(1-c)^{-(k+1)}\left[\frac{1}{1-\pi(1-c)} - \sum_{m=0}^{k}(\pi(1-c))^m\right]$$
$$\left. +c^2(1-c)^k\frac{a^2}{c(2-c)}\left[\frac{1+\pi(1-c)}{1-\pi(1-c)}\right]\right], \qquad (3)$$

In the stochastic part of the model, Markovian analysis requires the use of the lag-one product moment correlation coefficient for the standard error of estimate. This can be obtained through regression of the historical data. Fiering and Jackson however, suggests using the derived coefficient of determination for this estimate. For the purpose of analysis, this study uses the Markovian approach applied to higher lags for the unexplained variance.

To produce the required white noise for the stochastic part of the model, the application of the Central limit theorem utilizing uniformly distributed rectangular random numbers is advocated. Starting with a multiplicative linear congruential formula,

$$(R.N.)_{i+1} = \text{Fractional part of } K \times (R.N.)_i, \tag{4}$$

rectangular deviates from zero to one are generated. Initial estimates of the constant K and the initial seed (R.N.) are made to start the generation. The result of the above generation can now be applied using a simple formula based on the statistics of the Central limit theorem,

$$v_i = \frac{\sigma_v}{(n/12)^{0.5}} \sum_{j=1}^{n} (R.N.)_j + \left[\mu_v - \frac{n\sigma_v}{2(n/12)^{0.5}}\right] \tag{5}$$

where $\sigma_v$ and $\mu_v$ are the required standard deviation and mean of the generated normal deviates.

## MODEL PARAMETERS

### Case Study Data Sets

Two cases are studied for the purpose of determining the effects of record length to the model generating accuracy. It is also intended to produce results which are not biased nor are location dependent thus the variety of the sites chosen.

The first data set is from the main artery of the Kiso sansen[3], which is Kiso river itself at Inuyama (see figure 3). This is 32 years long, from 1956 to 1987, and is expressed in cubic meters per second. The catchment area above the Inuyama station is about 5,275 square kilometers. The average annual precipitation in the basin is from 2000mm to 3400mm, which is rather large[4]

---

[3] The Kiso river, the Nagara river and the Ibi river run separately now but in the old days these three rivers flowed turbulently through the Nohbi Plain as one river and the river course changed everytime there was a flood; so the people there think of these three rivers as one river, and called them Kiso Sansen which means the three rivers of Kiso.

[4] in Japan the average precipitation is 1800mm

and the flood discharge is also large compared with the catchment area. Of the three main rivers comprising Kiso sansen, Kiso river is the largest, and it flows down concentrating the water from mountains of 3000 meter class, such as Mount Ontake in Nagano Prefecture. There are diversions above the Inuyama station but mainly for power development and generation. Table 1(a) gives a maximum discharge for the period shown of 397.92 cubic meters per second in 1976 and a minimum discharge of 175.66 cubic meters per second in 1984. The mean of the flows for the record is 300.7 cubic meters per second and the standard deviation is 57.666 cubic meters per second. The skewness is only -0.0986, a very low value, and the lag-one serial correlation is a high negative coefficient -0.179 which means that there is a negative influence in the water carry-over from year to year.

The second data set is from the South Platte river near Kersey in Colorado (see figure 4). The drainage area at this point is approximately 60,000 square kilometers. The record shown is from 1906 through 1974 for which a complete and homogeneous record is available. Table 1(b), expressed in cubic meters per second as the mean annual discharge, shows a maximum value during the period of 62.5 cubic meters per second in 1973 and a minimum value of 6.2 cubic meters per second in 1955. There are diversions above the stations for irrigations of about

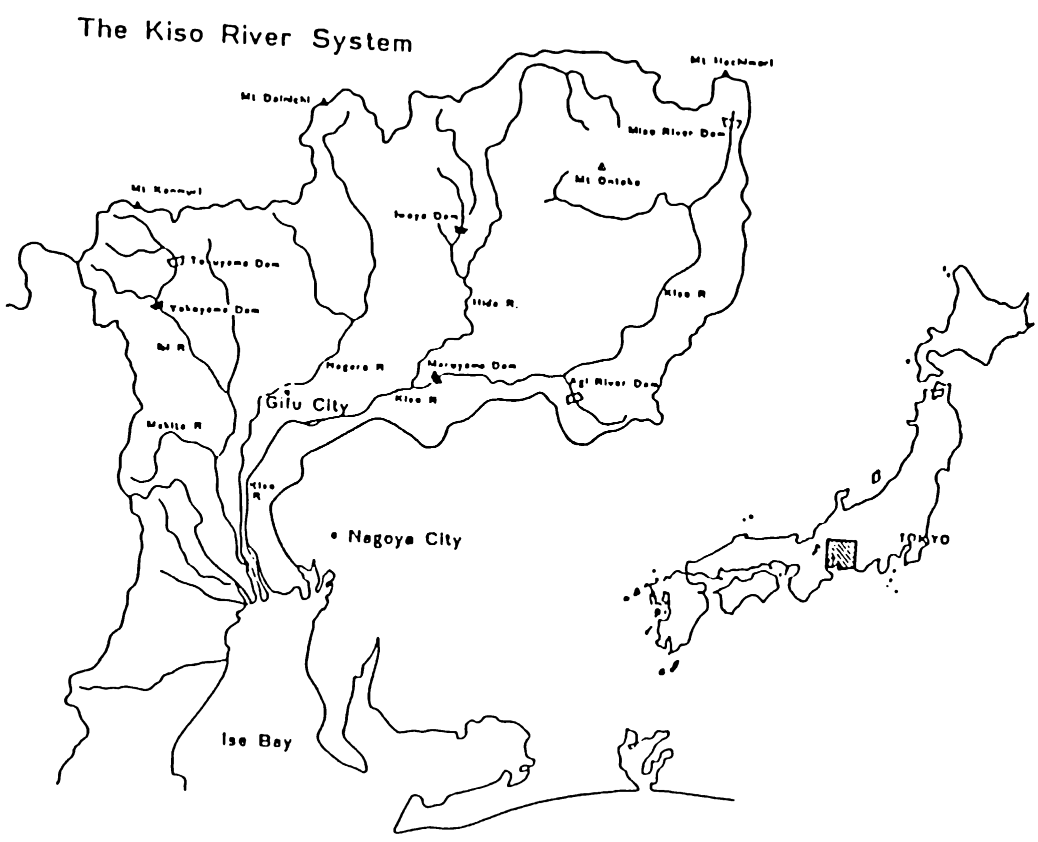



**Figure 3 : Kiso river catchment area above Inuyama station**

MEAN ANNUAL FLOWS IN CMS

| | | | | |
|---|---|---|---|---|
| 1956-1959 | 361.040 | 347.730 | 332.210 | 337.680 |
| 1960-1963 | 294.000 | 396.000 | 240.000 | 304.200 |
| 1964-1967 | 284.800 | 300.100 | 282.500 | 258.680 |
| 1968-1971 | 238.210 | 312.670 | 256.470 | 323.070 |
| 1972-1975 | 357.240 | 220.580 | 311.910 | 322.750 |
| 1976-1979 | 397.920 | 230.110 | 246.210 | 292.870 |
| 1980-1983 | 357.700 | 312.640 | 261.130 | 366.340 |
| 1984-1987 | 175.660 | 370.430 | 266.130 | 211.540 |

MEAN ANNUAL FLOWS IN CMS

| | | | | |
|---|---|---|---|---|
| 1906-1909 | 14.200 | 31.700 | 10.600 | 34.800 |
| 1910-1913 | 16.700 | 7.100 | 20.700 | 16.600 |
| 1914-1917 | 61.600 | 32.300 | 15.900 | 39.900 |
| 1918-1921 | 21.800 | 15.600 | 20.700 | 47.300 |
| 1922-1925 | 11.500 | 30.300 | 52.600 | 9.600 |
| 1926-1929 | 32.500 | 15.200 | 22.800 | 15.200 |
| 1930-1933 | 16.200 | 11.500 | 8.400 | 15.300 |
| 1934-1937 | 8.900 | 9.300 | 11.200 | 11.600 |
| 1938-1941 | 21.300 | 18.300 | 7.100 | 12.900 |
| 1942-1945 | 56.900 | 18.500 | 17.400 | 14.500 |
| 1946-1949 | 11.200 | 34.800 | 26.100 | 33.900 |
| 1950-1953 | 9.900 | 13.800 | 21.800 | 10.300 |
| 1954-1957 | 7.100 | 6.200 | 7.300 | 33.100 |
| 1958-1961 | 36.100 | 15.700 | 19.100 | 30.000 |
| 1962-1965 | 32.900 | 11.500 | 9.600 | 31.800 |
| 1966-1969 | 13.700 | 17.900 | 12.700 | 35.000 |
| 1970-1973 | 48.100 | 36.900 | 18.100 | 62.500 |
| 1974 | 26.300 | | | |

**Table 1**

**Mean annual flows, Kiso river at Inuyama (left) and South Platte river, Colorado (right)**

South Platte river basin

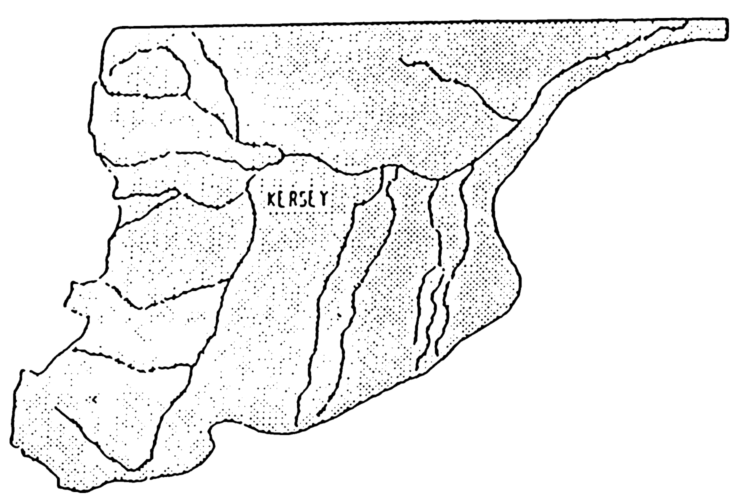



**Figure 4: South Platte river catchment area near Kersey, Colorado**

360,000 hectares and by transmountain and transbasin diversions, storage reservoirs, power development, groundwater withdrawals, and return flows from irrigated areas. The mean flow for the period is 22.2 cubic meters per second, and the standard deviation is 13.7 cubic meters per second.

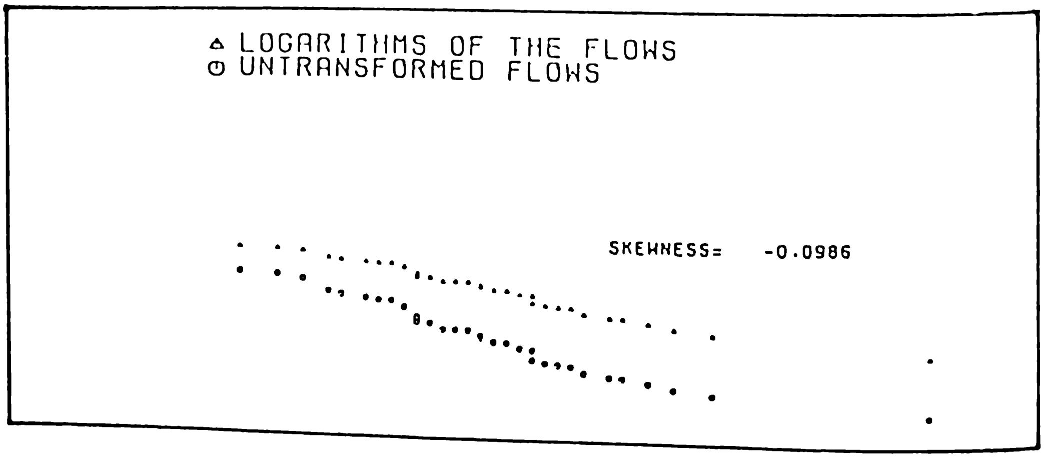



**Figure 5: Cumulative probability plot for the untransformed and log-transformed data set from Kiso river**

The skewness coefficient is 1.19 a high positive value. The lag-one serial correlation coefficient is very low, only 0.06. This means that there is relatively small water carry-over in the river basin from year to year.

**Fitting Estimate Correlograms**

Not knowing the actual values of the basin parameters for both data sets, we venture to determine the capability of the model by Fiering in mimicking the historical multiple-correlation coefficients as we increase lags. Since the main objective of this study is to extend the Markovian assumption for models which can reproduce jagged correlograms, we wish to generate in this, estimate coefficients of the same characteristics. Looking back at figure 1, it is intended to produce a plot that at least gives the same fluctuations for each data set.

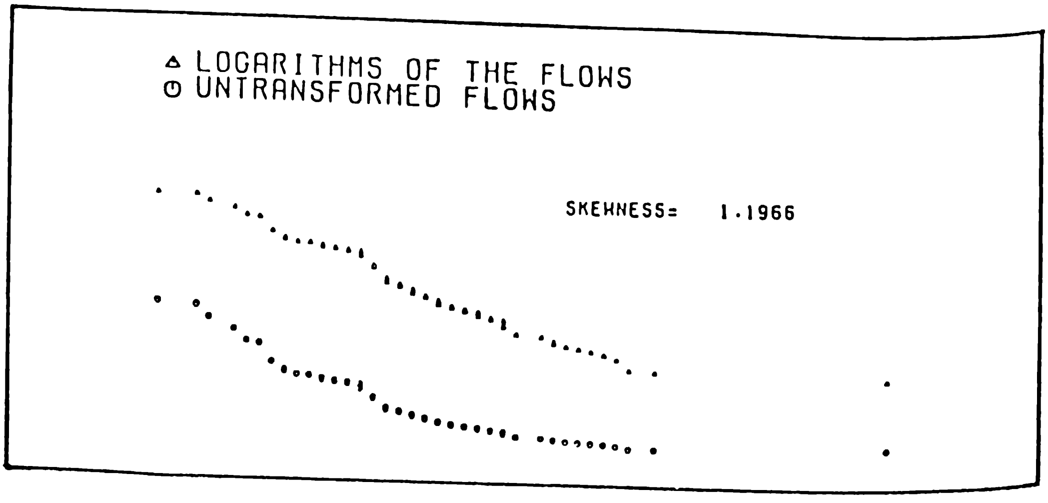



**Figure 6: Cumulative probability plot for the untransformed and log-transformed data set from South Platte river**

Applying equation 3 to generate plots of correlations against lags, figure 7 is arrived at. This figure shows the limitations of the estimate-correlation coefficient generating model by Fiering. Not only that its jaggedness is limited to fluctuations about the datum of zero, but its requirement of a large negative serial correlation coefficient for the precipitation seems unreasonable. Nevertheless, fitting this estimate to the historical correlogram for the data set from Kiso river, the last plot of figure 7 is arrived at. This produces estimate correlations of -0.176, 0.104, and -0.297 for the first three lags. This can be compared to the historical values of -0.179, 0.178, and 0.056 also for the first three lags.

The values of the estimated basin parameters are given as $a = 0.1, b = 0.1, c = 0.2, \pi = -0.5$. Although highly unlikely to be the actual values of these variables, this fit is used in the analysis of this model. Also because of the difficulty of modeling a separate plot that will fit the correlogram for the data from South Platte river, the estimate plot for Kiso river is also used for the data set from South Platte river. This assumption is a precondition by which became one of the basis for the unlikeliness of the generality of the model presented by Fiering.

## GENERATED DATA ANALYSIS

### Cummulated Assumptions of Model Parameters

Initial assumptions are made based on the previous analysis of the historical data sets. Starting with the type of general population of the historical flows, on the basis of the Chi-squared test and the Kolmogorov-Smirnov one-sample test, the data set from Kiso river and South Platte are both assumed to be derived from a normally distributed population. For the number of lags to be included in the generator model, we follow the results as found by the lag-inclusion analysis and generate synthetic flows using an $AR(4)$ model for Kiso river and an $AR(6)$ model for South Platte river. With regards to the estimate multiple correlation coefficients, we model using the estimate fit for Kiso river for both the data set regardless of the unnaturality of the estimate values of the basin parameters. On the case of the standard error of estimate, we apply the Markovian assumption to higher lags and use the lagged product-moment correlation coefficient in the unexplained variance of the model. For the random number generator, the constant is assumed to have a value of 997 and the initial seed is taken to be 0.5284163. On the strength of two out of three tests made for the determination of the existence of trend, we neglect its initial effects and try to determine its overall effect on the generator model. Since this study is not aimed at generating actual synthetic streamflow sequences for the use in planning, but rather it is just aimed to analyze the multiple-lag model and the effects of variations of various parameters to the general credibility of the resulting model, the preceding assumptions are deemed acceptable and workable.

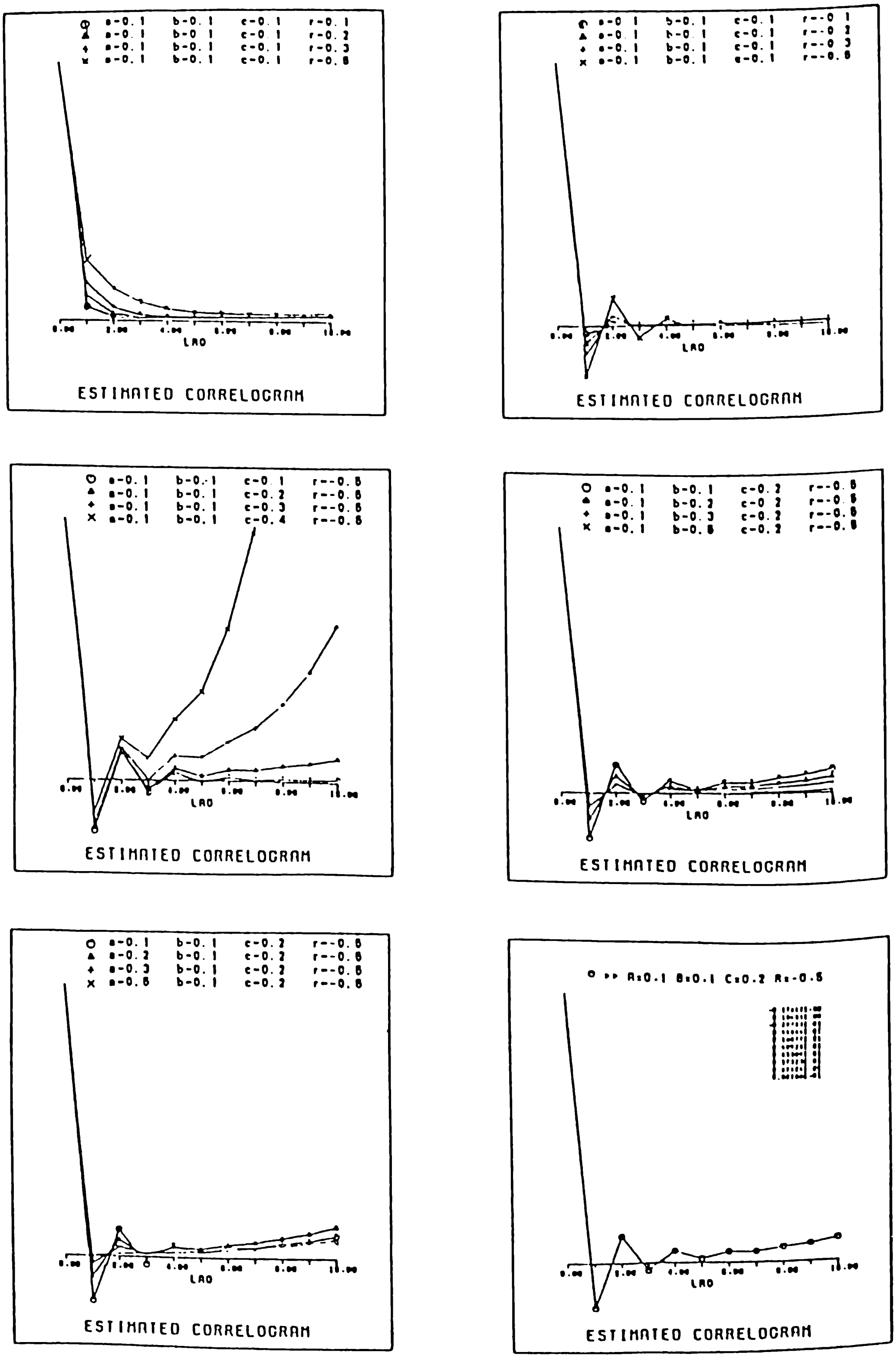



Figure 7: Estimate correlograms generated by the Fiering model for the Kiso river data set fit

## Product Moment Analysis

Before higher moment analysis on the generated data, the first two moment must first be checked. The first two moments, namely, the mean and the variance (or in this study the standard deviation), are the statistics that we are intending to keep in our generated data; the third moment's ratio, the skewness coefficient is only kept if there is a general willingness to use the Gamma distribution.

To determine the effectivity of the choice as forced by the lag-inclusion analysis and the effect of the estimate correlation coefficient, 200 years of synthetic data from typical models of $AR(1)$ to $AR(10)$ are generated using two different sets of correlation coefficients. From these 10 two hundred year data sets, the first 100 years of data are disregarded to eliminate starting condition bias[5]. The remaining one hundred years of data are then used to compare the mean and standard deviation with that of the historical data. The resulting plots can be seen in figures 8 and 9.

Analysis of these plots suggests that there is not much of a difference as far the mean and the standard deviation are concerned concerning the number of lags to be included in the model except for some instability probably due to computer inherent errors for lags bordering 8 to 10. Although these are not the only aspects desirable in a general model, an empirical choice, especially when materially constrained, is not unredeemable in a typical initial analysis. With regards to the estimate correlation coefficients, the stability is greatly dependent on the best few fits of the correlograms and becomes highly unstable when the coefficients depart from the historical values. This points to the fact that these estimate correlation coefficients are greatly dependent on the goodness-of-fit of the correlograms. This brings us back to convenience of use and efficiency of the historical correlation coefficients.

## Unexplained Variance Comparison

Because of the assumption of the application of the Markovian approach applied to higher lags instead of the recommendation of Fiering and Jackson of using the computed correlation coefficients used in the lag inclusion analysis, plots of generated data sets similar to that of figures 8 and 9 are drawn but this time using the correlation coefficients suggested by Fiering and Jackson. This is shown on figure 10. These plots use the historical correlation coefficient for briefness of the comparison. It will be noticed that the statistics do not vary greatly for both models of different assumptions. This justifies the use of the historical coefficients especially in computer encoding because of its convenience.

## Lag Stability and Trend

Checking the plots of figures 8 and 9, this gives the impression that there is an increasing trend in effect because of the computed mean of the generated flows persistently lie above the

---

[5] Fiering and Jackson (1971) suggested eliminating from 50 to 100 years of the first generated synthetic data.

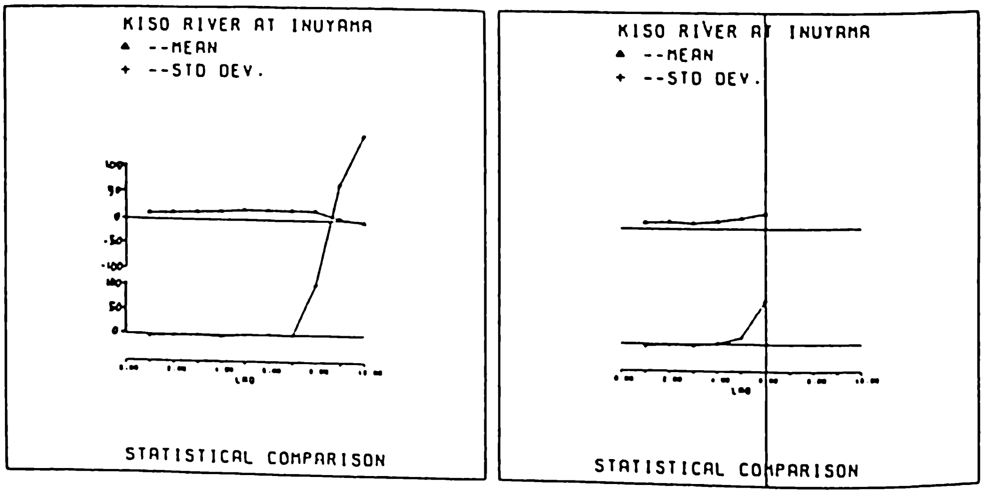



Figure 8: Statistical comparison of the mean and standard deviation of the historical and generated data for the Kiso river data set for an $AR(i, i = 1, \ldots, 10)$ model whose correlation coefficients are historical (left), and estimated (right).

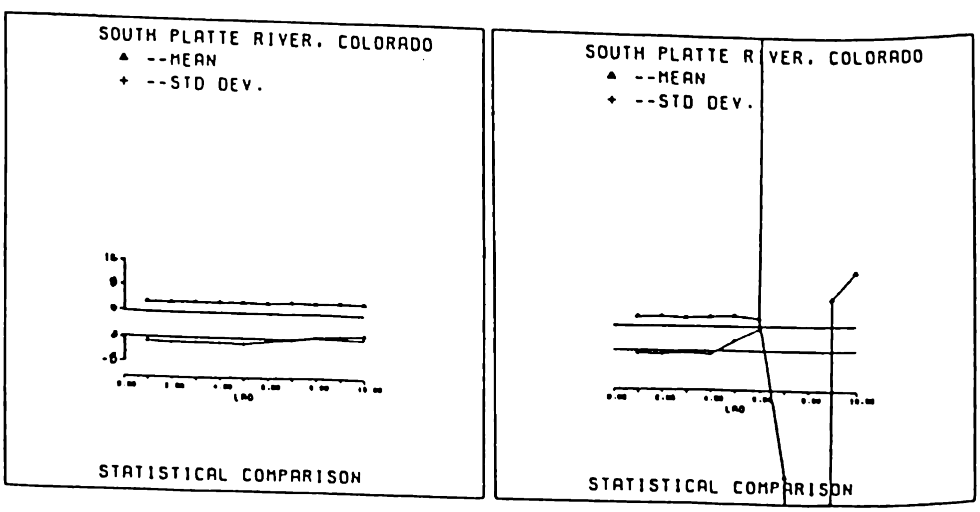



Figure 9: Statistical comparison of the mean and standard deviation of the historical and generated data for the South Platte river data set for an $AR(i, i = 1, \ldots, 10)$ model whose correlation coefficients are historical (left), and estimated (right).

historical mean; this is also reflected in figure 16 thus eliminating the possibility of having the unexplained variance as the source of such. For the standard deviation part, the computed values from the synthetic flows fluctuate gently about the historical value so there seems to be no problem in this regard.

Checking the possibility of a long term increasing trend in effect for both data sets, we plot figure 11 which uses only the first 100 years of synthetic data generated to compute the statistics. Inspection of these plots shows that indeed in the first 100 years of data, the mean

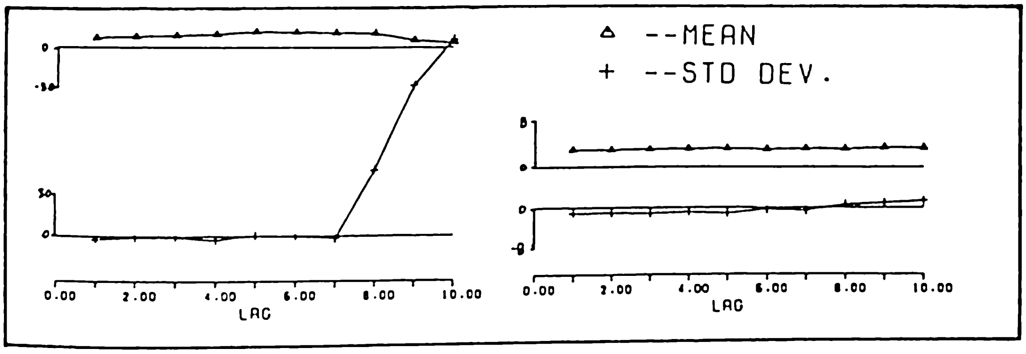



Figure 10: Comparison of statistics for models using coefficients from the lag inclusion analysis for $R^2$ for the data set from Kiso river (left), and the data set from South Platte river (right). Multiple correlation coefficients are computed from the historical data.

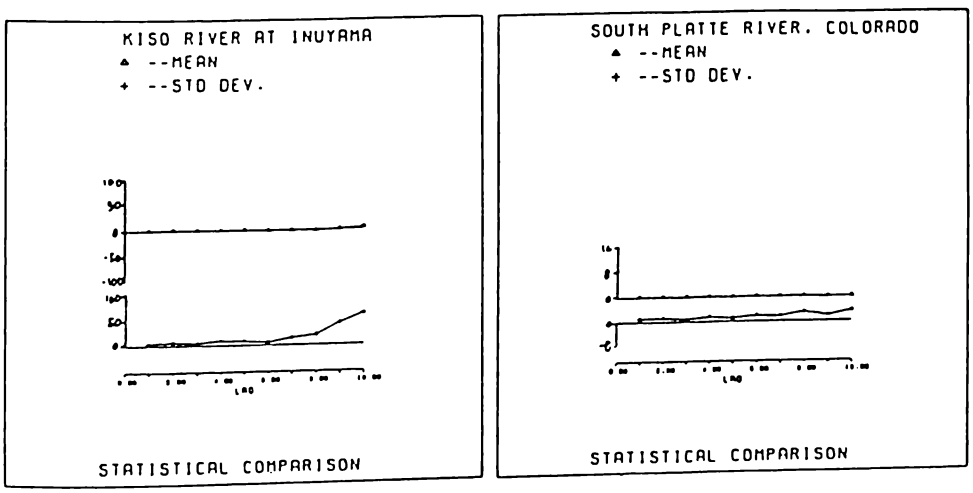



Figure 11: Statistical comparisons for the first 100 years of generated data using models utilizing historical correlation coefficients for both the serial coefficients and the unexplained variance. Data set on the left is from Kiso river, to the right is from South Platte river.

fluctuates evenly about the historical mean. This clearly shows that the model generates data sets which preserve generally the standard deviation, but gently increases the mean in proportion to the generated length of time.

This suggest increasing trends for both data sets; but this is contrary to the statistics found in previous tests made specifically for this purpose, especially in the case of Kiso river where it was found to exhibit a decreasing trend, if there is one affecting the model, instead of an increasing one. With regards to the possibility of a skewed population fit instead of the assumed normal distribution for the data sets involved, it also violates the statistical tests made again especially for Kiso river which to add, has also a very low skewness coefficient. On the other hand, this possibility should not be completely disregarded for the case of South Platte river

| SERIAL COEFFICIENTS | |
|---|---|
| BETA 1 = | -0.0506 |
| BETA 2 = | 0.0718 |
| BETA 3 = | 0.0391 |
| BETA 4 = | 0.0555 |

| SERIAL COEFFICIENTS | |
|---|---|
| BETA 1 = | -0.0492 |
| BETA 2 = | 0.1547 |
| BETA 3 = | 0.0826 |
| BETA 4 = | 0.0011 |
| BETA 5 = | -0.1808 |
| BETA 6 = | -0.0183 |

Table 2: Serial coefficients of the generating models for Kiso river (left) data set, and South Platte river (right) data set

because not only does it yield a high positive skewness, it also fails on the probability fitting test made previously.

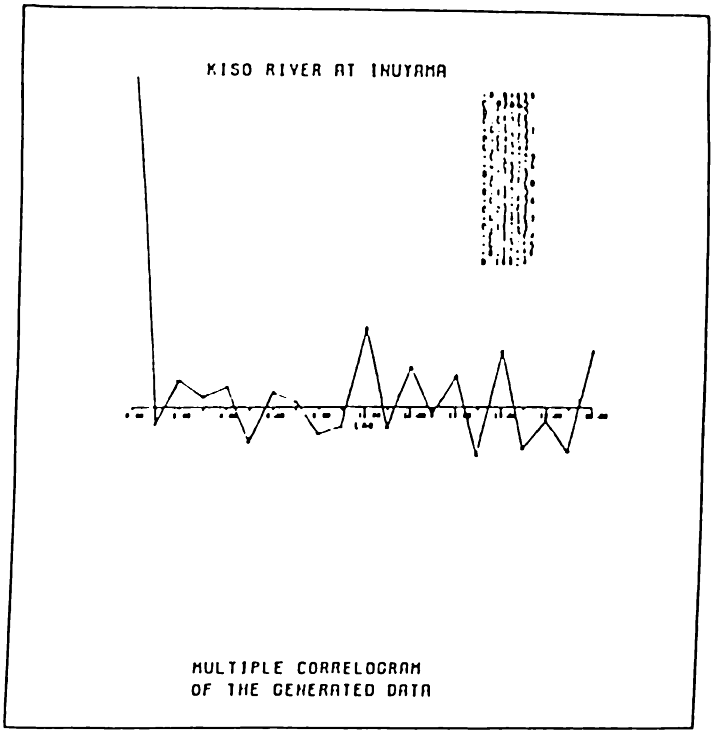
### Correlogram of the Generated Data

Following the initial assumptions for the general model, we generate 200 years of synthetic data using the historical correlation coefficients for the serial coefficients and the unexplained variance. With these 200 years of data, the multiple correlation coefficients are computed and these are plotted against lags to produce the desired correlograms. These can now be compared to the historical correlograms to see if the model produces data that exhibit the same characteristics as the historical flows. These plots can be seen in figure 12. A quick check shows that the lag-one serial correlation coefficients of the generated data are very low, which means that the generated data are almost independent of each other or there is not much persistence between flows from year to year. Further check of this conclusion are the values of the serial coefficients of the data which can be seen on the listings of table 2.

## MODEL MODIFICATIONS

### Effect of the Initial Seed

Because of the persistent increasing mean of the generated data, an analysis of the effect of the initial seed on the generating accuracy of the model is made. Statistical comparisons between generated models of different initial seeds were made by the author and were found to be influencing the model's generating accuracy. To make the analysis clearer, a plot of the deviations of the historical mean and standard deviation from those of the generated values against initial seeds of zero to one were made and the results are seen on figure 13. This shows that the effect of the initial seed on the model to be a complex non-linear function, which should not be the case. Logic dictates that the choice of the initial seed should not be this much of an influence on the generator model, and at the worst should display a much smaller fluctuation about the datum. This conclusion dictates that another factor other than the value of the initial seed is influencing the model's accuracy.



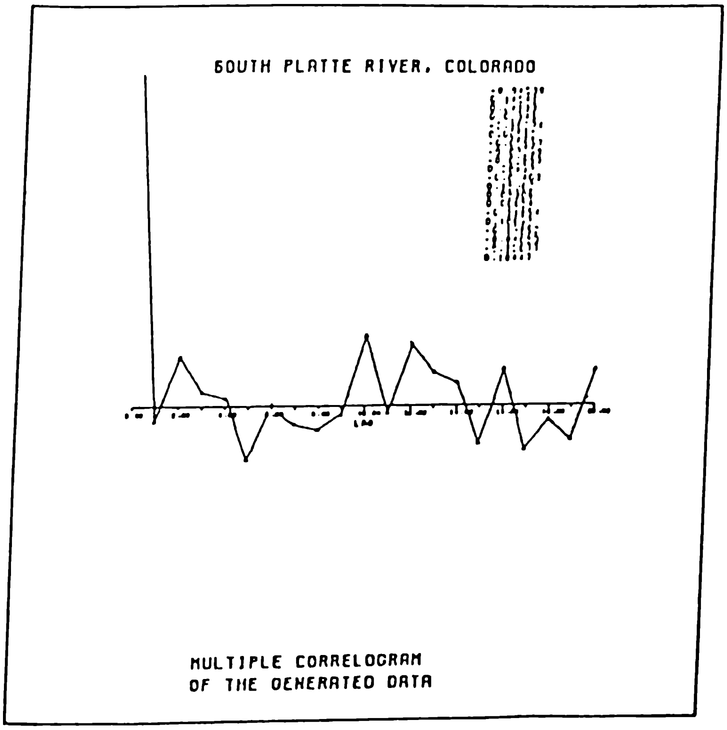




Figure 12: Correlograms of the generated data

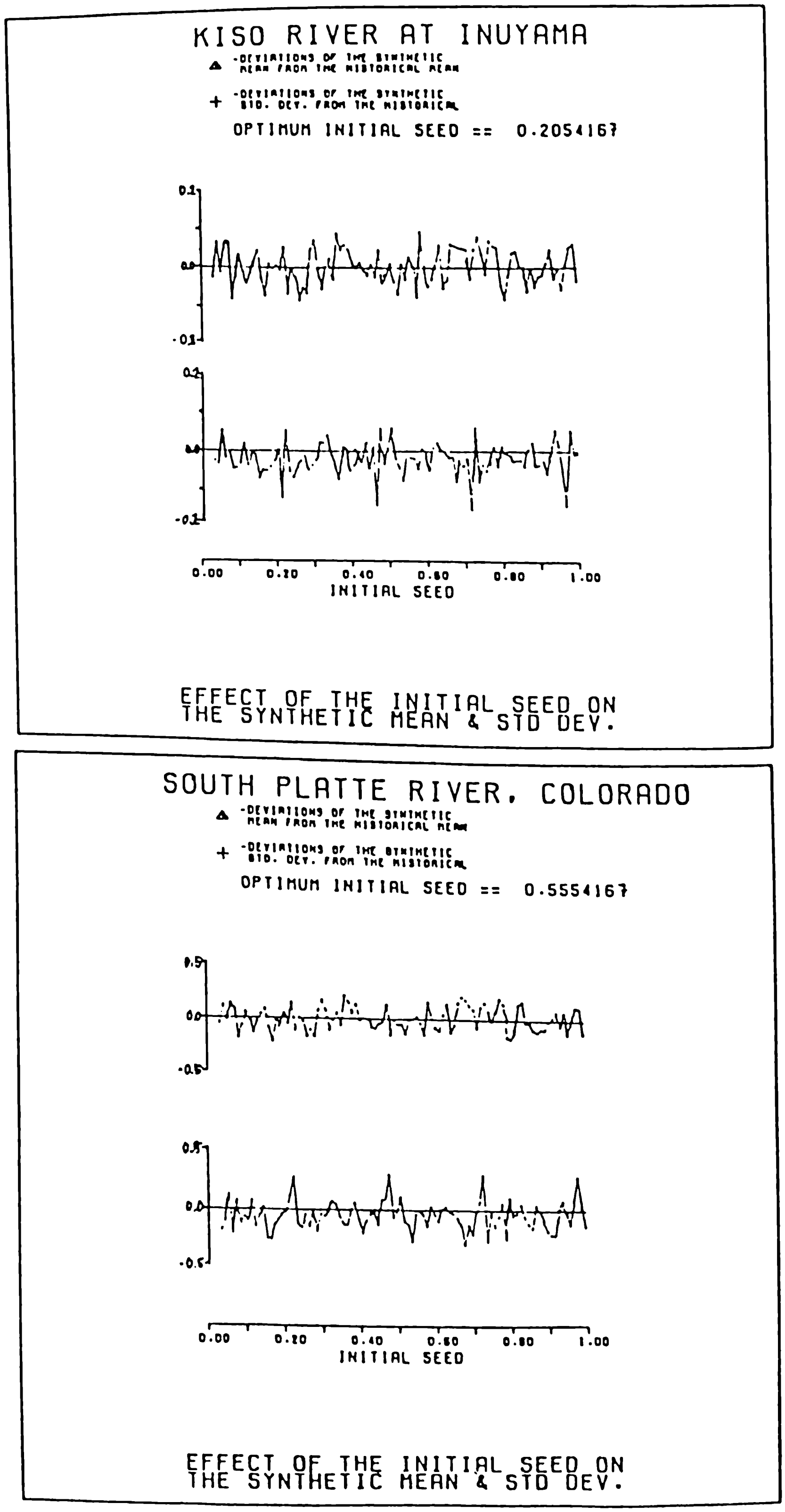



Figure 13: Plots of the deviations of the historical values from the generated mean and standard deviation from models using initial seeds of zero to one.

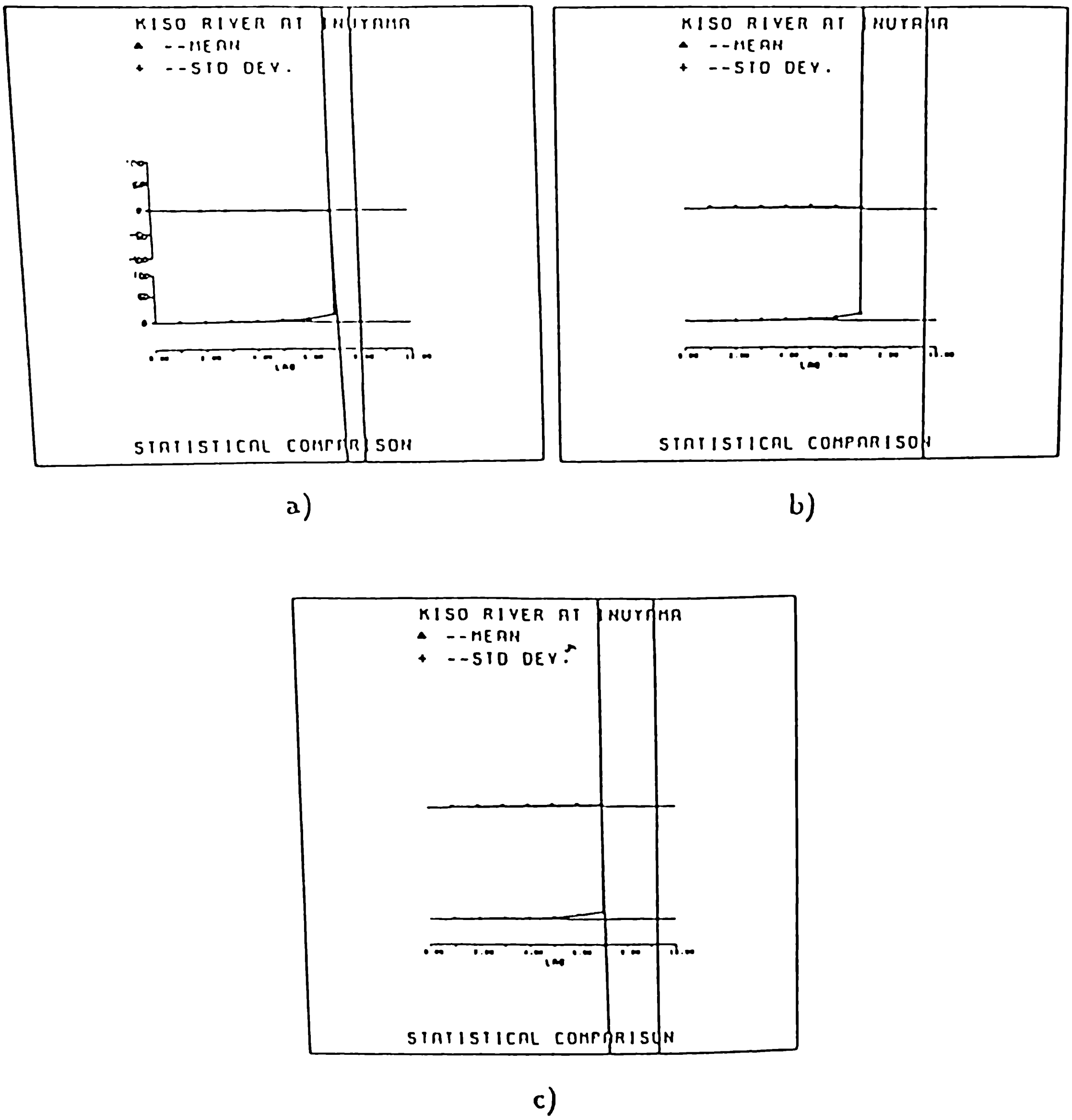



Figure 14: Statistical comparisons of the mean and standard deviations of the last 800 years of a 900-year generated data set using models utilizing initial seeds of a)0.2274167, b)0.5574167 and c)0.9974167

**Effect of the Length of the Generated Data**

The possibility of the length of the generated synthetic data affecting the model's generating accuracy is dependent on the distinct possibility that although normally distributed, the random numbers generated are not evenly scattered in its time scale. This would point out an increasing, or decreasing, generated data set if the first half of the modeled data are disregarded and not used. This can be checked by producing longer string of synthetic data and computing the statistics from such. For this study, a replicate plot of statistical comparisons are made using different values of initial seeds but this time utilizing the last 800 years of a 900-year series of generated data. The results can be seen in figure 14. This shows that for a considerably long synthetic data set, the effect of the initial seed can be considered negligible in the model's generating accuracy.

## CONCLUSIONS

After this brief analysis, the following conclusions are arrived at.

The estimate correlation coefficient generating model by Fiering is highly restricted by the data set being studied. Also, its fitting is limited to fluctuations about the zero value and fitting other than about zero is much likely improbable if not impossible. The assumptions made to make a respectable fitting are not realistic (i.e., assumptions of precipitation correlations) and are highly questionable. Finally, the reasons justifying the use of such assumptions are not good enough compared with the merits of using a simpler assumption (using historical correlation coefficients).

For this multiple lag analysis, an upper limit of an $AR(7)$ model seems to be more practical compared to Fiering and Jacksons $AR(8 \text{ to } 10)$ maximum. This can be attributed to correlation instabilities coupled with computer inherent errors in the computations for models using lags greater than 7.

The choice of an initial seed should be carefully made so as not to influence the model's generating accuracy. Although its choice should be impartial, the possibility of a shorter return period for a certain initial seed is not at all distinct and should be considered. This is especially so when using a random number generator which has an unequal scatter of its generated quantities.

The length of the generated synthetic sequence should be chosen in such a way that enough generated data can be disregarded to eliminate starting condition bias and still encompass possibilities of weakness in the stochastic part of the model. Although the final choice will be left to the planning engineer, an initially long estimate will be a good position in the analysis.

For very short generation of data, care must be taken in the amount of initial quantities eliminated because an unchecked stochastic part of the model can greatly influence the first few years of generated data.

Trends are negligible for short term data generation, but its effect cannot be disregarded for a longer analysis and generation. Statistical testing may be of help in detecting trends and cycles but the final choice is entirely upon the planning engineer.

The type of population should be chosen using regular statistical tests for the purpose, and for primary analysis, a simpler population estimate does not greatly affect the model's performance and can be taken as an initial assumption. However for the final data generation process, population types should be chosen depending on factors that are inherent on the project itself and should include decision and game theory in this regard. Although this is not taken in this study, its inception in the final analysis should be taken as a standard measure.

## REFERENCES


1. A. T. BARUCHA-REID. *Elements of the Theory of Markov Processes and Their Applications.* Mc-Graw Hill Book Company, New York, 1960

2. H. J. MOREL-SEYTOUX, et. al., editors. *Modeling Hydrologic Processes.* Water Resource Publications, Fort Collins, Colorado, 1979.

3. L. TAKACKS. *Stochastic Processes.* Methuen Company Limited, London, 1960.

4. L. M. POLINTAN. *Mathematical Synthesis of Monthly Streamflow Sequences.* Lecture in Advanced Hydrology, De La Salle University, Manila, 1988.

5. M. G. KENDALL, A. STUART. *The Advanced Theory of Statistics, Vol. 1* Charles Girffin and Company Limited, London, 1943.

6. MYRON B. FIERING, BARBARA B. JACKSON. *Synthetic Streamflows.* American Geophysical Union, Washington, D.C., 1971.

7. MYRON B. FIERING. *Streamflow Synthesis.* Harvard University Press, Cambridge, Massachusetts, 1967.

8. N. T. KOTTEGODA. *Stochastic Water Resources Technology.* The Macmillan Press Ltd., London, 1980.

9. RICHARD H. McCUEN, WILLARD M. SNYDER. *Hydrologic Modelling: Statistical Methods and Applications.* Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

10. T. W. ANDERSON. *The Statistical Analysis of Time Series.* John Wiley and Sons, New York, 1971.

11. V. VEVYEVICH, NEIL S. GRIGG. *General Aspects of Modelling and Time Series Analysis,* and *Generation of Hydrologic Samples.* Systems Analysis Lecture Series, Colorado State University, Fort Collins Colorado, 1981.