〈論 文〉

다변량 모형에 의한 하천유량의 모의 발생
A Multivariate Model Development For Stream Flow Generation

鄭 相 萬
Jeong, Sang Man*

Abstract □ Various modeling approaches to study a long term behavior of streamflow or groundwater storage have been conducted. In this study, a Multivariate AR (1) Model has been applied to generate monthly flows of the one key station which has historical flows using monthly flows of the three subordinate stations. The Model performance was examined using statistical comparisons between the historical and generated monthly series such as mean, variance, skewness. Also, the correlation coefficients (lag-zero, and lag-one) between the two monthly flows were compared. The results showed that the modeled generated flows were statistically similar to the historical flows.
* Senior Researcher, Water Resources Divison, Korea Institute of Construction Technology, Socho-gu, Umyon-dong 142, Seoul 137-140, Korea

요 지 : 단일지점(Single Site)에 대한 하천유량의 추계학적인 모의발생을 위해서는 간단한 모델중의 하나로 Univariate AR(1) 모델이 흔히 쓰여왔다. 그러나 다지점(Multi Sites)에 대한 하천유량에 관한 추계학적인 모의발생은 지점간 서로의 연관성 때문에 단일지점을 위한 모의발생처럼 쉽게 해결되지 않았다. 본 연구에서는 미국 아이다호주의 Camas Creek 유역에 대하여 하나의 키이지점(Key Station)과 주변에 세개의 종속지점(Subordinate Station)을 설정하고 다변량 AR(1) 모델을 적용하여 모의발생된 월유량과 실측치를 통계적으로 비교, 분석하였다.

모의발생된 월유량과 실측치를 평균, 분산, 왜곡도, 상관관계 등에 의해 비교, 분석한 결과 모의발생된 월유량과 실측치는 통계적으로 서로 유사성을 보였다.

1. Introduction

Since the early 1960's, extensive research efforts have been concentrated on developing methods to analyze the stochastic characteristics of hydrologic series and to devise generating schemes for univariate hydrologic series. Periodic multivariate models are necessary to generate multiple periodic series at several sites. For instance, a water resource system of several reservoirs may require the generation of monthly streamflows at several sites in order to simulate the operation of such a system. If monthly streamflow data are available at the site of interest, a periodic multivariate model can be used to model and generate monthly streamflows at those sites (1).

The primary objective of this study was to develop a Multivariate monthly flow model for Camas Creek, Idaho, U.S.A, based on stream flows in neighboring basins, and to test whether such a model is sensitive enough to detect, at some statistical level of sighificance.

2. Multivariate Models and Model Selection

The multivariate modeling of surface streamflow, using subordinate station data to predict or extend the streamflow record at key station, is well documented and has been extensively applied. In 1964, Fiering (2) proposed multivariate analysis for the simultaneous generation of synthetic flow sequences at a key station X and a subordinate station Y to preserve the lag-one serial and lag-zero cross correlation coefficient of the

* Senior Researcher, Water Resources Divison, Korea Institute of Construction Technology, Socho-gu, Umyon-dong 142, Seoul 137-140, Korea

two stations. However, subsequent examination of this model by others (3,4) showed that these correlation coefficients were not preserved unless some constraint conditions were met. Lawrance (4) presented a modification of the Fiering model that preserves the lag—one serial and lag—zero cross correlation coefficients with much less stringent constraints. In 1973, Yevjevich (5) presented an improved multisite model which can reduce time dependence, skew coefficient, and variance of the model residuals, while preserving all of the individual station statistics.

The general model structure of and AR(1) Model with monthly parameters is given by:

$$Z_{\nu,\tau} = A_{1,\tau} Z_{\nu,\tau-1} + B_{\tau}\,\varepsilon_{\nu,\tau} \quad \cdots\cdots\cdots (1)$$

where $A_{1,\tau}$ is an n x n monthly coefficient matrix associated with the time interval τ, $Z_{\nu,\tau-1}$ is an n x 1 vector of standardized monthly values, B_{τ} is an n x n monthly matrix for time interval τ, and $\varepsilon_{\nu,\tau}$ is an n x 1 vector of time independent, normally distributed random variables. The sequences, $Z_{\nu,\tau}$ and $\varepsilon_{\nu,\tau}$ have expected values equal to zero, and variances equal to one.

This model has the property to preserve the historical means, standard deviations, the lag—zero and the lag—one cross correlation and the lag—one autocorrelation. A recent study (6) has shown that the Yevjevich model was clearly superior to the other earlier models in properly modeling the flows on several Idaho streams. For this reason, the multisite AR(1) model was applied to this study.

For the record period of 1925 to the cut—off date of 1952, the multivariate AR(1) model can be developed to relate the Camas Creek monthly flows to the flows on the other three streams. When the monthly flows at stations of Big Wood Slough, Big Lost River, Goose Creek, and Camas Creek were used, the above four sites numbered as 1,2,3, and 4, respectively.

Using the representation of equation 1, the generating AR(1) model for one key station and three subordinate stations can be written as:

$$\begin{aligned}
Z_{\nu,\tau}^{(1)} &= a^{11}Z_{\nu,\tau-1}^{(1)} + a^{12}Z_{\nu,\tau-1}^{(2)} \\
&\quad + a^{13}Z_{\nu,\tau-1}^{(3)} + a^{14}Z_{\nu,\tau-1}^{(4)} + b^{11}\varepsilon_{\nu,\tau}^{(1)} \\
&\quad + b^{12}\varepsilon_{\nu,\tau}^{(2)} + b^{13}\varepsilon_{\nu,\tau}^{(3)} + b^{14}\varepsilon_{\nu,\tau}^{(4)} \\
Z_{\nu,\tau}^{(2)} &= a^{21}Z_{\nu,\tau-1}^{(1)} + a^{22}Z_{\nu,\tau-1}^{(2)} \\
&\quad + a^{23}Z_{\nu,\tau-1}^{(3)} + a^{24}Z_{\nu,\tau-1}^{(4)} + b^{21}\varepsilon_{\nu,\tau}^{(1)} \\
&\quad + b^{22}\varepsilon_{\nu,\tau}^{(2)} + b^{23}\varepsilon_{\nu,\tau}^{(3)} + b^{24}\varepsilon_{\nu,\tau}^{(4)} \\
Z_{\nu,\tau}^{(3)} &= a^{31}Z_{\nu,\tau-1}^{(1)} + a^{32}Z_{\nu,\tau-1}^{(2)} \\
&\quad + a^{33}Z_{\nu,\tau-1}^{(3)} + a^{34}Z_{\nu,\tau-1}^{(4)} + b^{31}\varepsilon_{\nu,\tau}^{(1)} \\
&\quad + b^{32}\varepsilon_{\nu,\tau}^{(2)} + b^{33}\varepsilon_{\nu,\tau}^{(3)} + b^{34}\varepsilon_{\nu,\tau}^{(4)} \\
Z_{\nu,\tau}^{(4)} &= a^{41}Z_{\nu,\tau-1}^{(1)} + a^{42}Z_{\nu,\tau-1}^{(2)} \\
&\quad + a^{43}Z_{\nu,\tau-1}^{(2)} + a^{44}Z_{\nu,\tau-1}^{(4)} + b^{41}\varepsilon_{\nu,\tau}^{(1)} \\
&\quad + b^{42}\varepsilon_{\nu,\tau}^{(2)} + b^{43}\varepsilon_{\nu,\tau}^{(3)} + b^{44}\varepsilon_{\nu,\tau}^{(4)}
\end{aligned} \quad \cdots\cdots\cdots (2)$$

where:

$Z_{\nu,\tau}^{(i)}$ = standardized monthly flows for station (i = 1,2,3,4)

a's and b's = the elements of the estimated matrix parameters A1 and B

$\varepsilon_{\nu,\tau}^{(i)}$ = independent and identically distributed normal (0,1) deviate for i = 1, 2, 3, 4

ν = 1, 2, ..., N (N is the number of years of records)

τ = 1, 2, ..., W (W is the number of time periods within the year)

Because it is assumed that is a lower triangular matrix, equation 2 can be changed as follows to generate monthly flows of the one key station (Camas Creek), using monthly flows of the three subordinate stations(Big Wood Slough, Big Lost River, Goose Creek).

$$\begin{aligned}
\varepsilon_{\nu,\tau}^{(1)} &= 1/b^{11}\,(Z_{\nu,\tau}^{(1)} - a^{11}Z_{\nu,\tau}^{(1)} \\
&\quad - a^{12}Z_{\nu,\tau}^{(2)} - a^{13}Z_{\nu,\tau-1}^{(3)} - a^{14}Z_{\nu,\tau-1}^{(4)})
\end{aligned}$$

$$\varepsilon_{\nu,\tau}^{(2)} = 1/b^{22}(Z_{\nu,\tau}^{(2)} - a^{21}Z_{\nu,\tau-1}^{(1)} - a^{22}Z_{\nu,\tau-1}^{(2)} - a^{23}Z_{\nu,\tau-1}^{(3)} - a^{24}Z_{\nu,\tau-1}^{(4)} - b^{21}\varepsilon_{\nu,\tau}^{(1)})$$

$$\varepsilon_{\nu,\tau}^{(3)} = 1/b^{33}(Z_{\nu,\tau}^{(3)} - a^{31}Z_{\nu,\tau-1}^{(1)} - a^{32}Z_{\nu,\tau-1}^{(2)} - a^{33}Z_{\nu,\tau-1}^{(3)} - a^{34}Z_{\nu,\tau-1}^{(4)}) - b^{31}\varepsilon_{\nu,\tau}^{(1)}) - b^{31}\varepsilon_{\nu,\tau}^{(1)} - b^{32}\varepsilon_{\nu,\tau}^{(2)}$$

$$\varepsilon_{\nu,\tau}^{(4)} = \text{random number (normal deviate)}$$

$$Z_{\nu,\tau}^{(4)} = a^{41}Z_{\nu,\tau-1}^{(1)} + a^{42}Z_{\nu,\tau-1}^{(2)} \cdots\cdots (3)$$

$$+ a^{43}Z_{\nu,\tau-1}^{(3)} + a^{44}Z_{\nu,\tau-1}^{(4)} + b^{41}\varepsilon_{\nu,\tau}^{(1)} + b^{42}\varepsilon_{\nu,\tau}^{(2)} - b^{43}\varepsilon_{\nu,\tau}^{(3)} + b^{44}\varepsilon_{\nu,\tau}^{(4)} \quad (4)$$

To generate standardized monthly values of the key station ($Z_{\nu,\tau}^{(4)}$) using the standardized monthly values of the three subordinate stations, the following two steps are necessary: the first step requires the calculation of four standardized normal random numbers, $\varepsilon_{\nu,\tau}^{(1)}$, $\varepsilon_{\nu,\tau}^{(2)}$, $\varepsilon_{\nu,\tau}^{(3)}$ and $\varepsilon_{\nu,\tau}^{(4)}$ using equation 3; and the second step is to generate a $Z^{\tau(4)}$ value, using the above calculated standardized normal random numbers and equation.

3. Preliminary Analysis

The main purposes of this analysis were to check the normality of the original monthly time series, and to make appropriate transformations to normal, if necessary. The normality of the raw monthly series and lognormal transformed monthly series for the four stations was checked by determining if the coefficient of skewness of each series did not differ from zero. Based on each normalized series (four stations) from the previous step, the monthly mean, monthly standard deviation, and the monthly correlation coefficients (lag-zero, and log-one) for each series were computed.

4. Estimation of Parameters

The standardized variate $Z_{\nu,\tau}^{(i)}$ is:

$$Z_{\nu,\tau}^{(i)} = \frac{X_{\nu,\tau}^{(i)} - \overline{X}_{\nu,\tau}^{(i)}}{S_{\nu,\tau}^{(i)}} \quad i = 1, 2, 3, 4 \cdots\cdots (5)$$

where $X_{\nu,\tau}^{(i)}$ is the monthly stream flows, and $\overline{X}_{\nu,\tau}^{(i)}$ and $S_{\nu,\tau}^{(i)}$ are the estimated values of the mean and standard deviation of the series.

With the AR(1) model for four stations, the monthly matrix parameter estimates for the model of equation 1.1 can be obtained from (7):

$$A_{1,\tau} = M_{1,\tau}\, M^{-1}_{0,\tau-1} \cdots\cdots (6)$$

and

$$B_\tau B_\tau^T = M_{0,\tau} - M_{1,\tau}\, M^{-1}_{0,\tau-1}\, M^T_{1,\tau} \cdots\cdots (7)$$

where $M_{0,\tau}$, $M_{0,\tau-1}$, and $M_{1,\tau}$, are the monthly matrix correlations given by:

$$M_{\kappa,\tau} = \begin{matrix} r^{11}_{\kappa,\tau}, r^{12}_{\kappa,\tau}, r^{13}_{\kappa,\tau}, r^{14}_{\kappa,\tau} \\ r^{21}_{\kappa,\tau}, r^{22}_{\kappa,\tau}, r^{23}_{\kappa,\tau}, r^{24}_{\kappa,\tau} \\ r^{31}_{\kappa,\tau}, r^{32}_{\kappa,\tau}, r^{33}_{\kappa,\tau}, r^{34}_{\kappa,\tau} \\ r^{41}_{\kappa,\tau}, r^{42}_{\kappa,\tau}, r^{43}_{\kappa,\tau}, r^{44}_{\kappa,\tau} \end{matrix} \cdots\cdots (8)$$

for k = 0,1 and = 1, ..., 12. The monthly correlation coefficients $r_{\kappa,\tau}^{ij}$ are obtained by correlating $Z_{\nu,\tau}^{(i)}$ with $Z_{\nu,\tau-\kappa}^{(i)}$. The elements of the matrix B_τ can be obtained by use of lower triangular matrix.

5. Data Generation

The multivariate AR(1) model for the synthetic generation of the monthly series for the key station (Camas Creek) can be obtained by substituting the estimated parameters into their corresponding model equations. Thus, the generation of the monthly series for the key station $X_{\nu,\tau}^{(4)}$ can be made by:

$$X^{(4)}_{\nu,\tau} = \overline{X}_{\nu,\tau} + S^{(4)}_{x,\tau} Z^{(4)}_{x,\tau} \cdots\cdots (9)$$

for raw monthly series and

$$X^{(4)}_{\nu,\tau} = \exp\left[\overline{Y}^{(4)}_{\nu,\tau} + S^{(4)}_{y,\tau} Z^{(4)}_{\nu,\tau}\right] \cdots\cdots (10)$$

for log transformed monthly series

where $\overline{X}_{\nu,\tau}^{(4)}$ and $S_{x,\tau}$ are the estimates of the monthly mean and monthly standard deviation for the raw monthly series, and $\overline{Y}_{\nu,\tau}$ and $S_{y,\tau}$ are the estimate of the monthly mean and monthly standard deviation for the log-transformed monthly series. $Z_{\nu,\tau}^{(4)}$ is the standardized, normalized monthly flow generated by equation 1.4. The following relationships (8) were used to relate the characteristics of the untransformed variate $\overline{X}_{\nu,\tau}$ and the transformed variate $\overline{Y}_{\nu,\tau}$

$$\overline{Y} = 1/2\log\left[\frac{\overline{X}^2_{\nu,\tau}}{[C^2_{\nu,\tau}+1]}\right] \cdots\cdots\cdots\cdots (11)$$

$$S_{Y,\tau}{}^2 = \log(C^2_{\nu,\tau}+1) \cdots\cdots\cdots\cdots (12)$$

where $C_{\nu,\tau}$ is the coefficient of the variation of the raw data ($C_{\nu,\tau} = S_{X,\tau} / \overline{X}_{\nu,\tau}$).

For the period of natural flow (1925–1952) on the four streams, the historical data from the subordinate stations were used to generate monthly flows for Camas Creek. The historical monthly flows and the generated monthly flows for Camas Creek for the period of natural flow (1925–1952) were plotted. In addition, the historical annual flows and the generated annual flows (monthly sums) at Camas Creek were poltted. Although the model appeared to simulate overall historical time series reasonably well, its performance varied by month and season. The poorest agreement between modeled and observed flows occurs in the peak spring runoff months and is probably attributable to variations in snowpack among the four basins used in the model development. As typical results of monthly flows and the results of annual flows included here as Fig.1 and Fig.2.

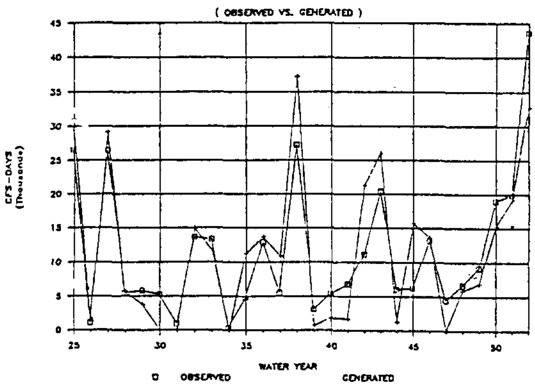

Fig. 1 Camas Creek May Streamflows

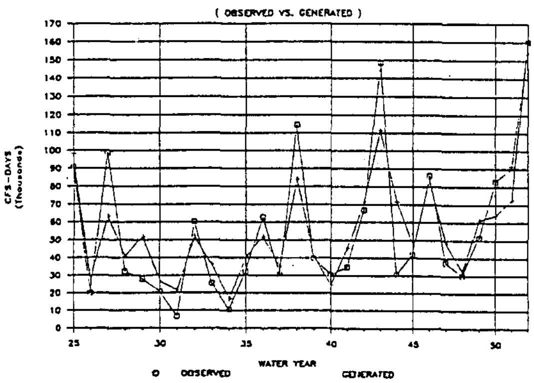

Fig. 2 Camas Creek Annual Streamflows

6. Analyses and Conclusion

Statistical comparisons (mean, variance, skew) were performed between the historical and generated monthly series. Hypothesis tests were used to compare the statistics of the two series. Hypothesis tests were used to compare the statistics of the two series (historical vs. generated) for each month and annual flows (9,10). A paired mean difference t–statistic (Table 1) was considered in oredr to compare the statistics of the two series. The variability of the two series were considered to understand the results by the paired mean difference testing. Therefore, equality of the variances between two series was first tested by using an F–statistic (Table 2). Then the means of the two series were tested using the paired difference t–statistic. The results of these tests are presented in Table 3

From Table 2, it was seen that at the 5% level of significance, the null hypotheses of equality of variance between the two monthly series were not rejected, except for September flows. Unequality of variance for the september flows is caused by an extremely large value was generated by the model. Likewise,from Table 3 no significant differences between the means for the two monthly series for all months were found. Coefficients of skew for the historical monthly streamflows and the generated monthly flows were compared on a monthly basis. The comparison showed that the skew coefficients of all but two months (February, September)were similar to each other. It was concluded that the differences of the skew coefficients of the two months were caused by the relaxed normality assumption for those months of Goose Creek station. Also, the correlation coefficients between his-

Table 1 Hypothesis Tests using F-and Paired Difference t-statistics

F-statistic

Null Hypothesis : $\sigma h^2 = \sigma g^2$

Alternative Hypothesis : $\sigma h^2 \neq \sigma g^2$

Test Statistics : F = (larger of s_h^2 and s_g^2) / (smaller of s_h^2, s_g^2)

Degrees of Freedom : n-1

(n = sample size of s_h^2 and s_g^2)

t-statistics

Null Hypothesis : $\mu_h - \mu_g = \mu_d = D_0 = 0$

Alternative Hypothesis : $D_0 \neq 0$

Test Statistic :

$$t = \frac{d - d_0}{s_d \quad n}$$

Degrees of Freedom : n-1

where the subscripts h and g represent the two series of each length, n. The Greek letters represent the population statistics while the lowercase letters represent the sample estmates of these statistics. d and s_d are sample mean and standard deviation of the n differences.

Table 2 Hypothesis Test for Equality of Variances between Two Monthly Series (for 1925-1952 water Years)

Month	Years	Std(h) (Cfs-days)	Std(g) (Cfs-days)	Sample F	5% F	Unequal / Equal
Oct.	28	119	97	1.48	1.96	equal
Nov.	28	267	224	1.42	1.96	equal
Dec.	28	438	375	1.36	1.96	equal
Jan.	28	218	253	1.34	1.96	equal
Feb.	28	1048	809	1.68	1.96	equal
Mar.	28	3950	3485	1.28	1.96	equal
Apr.	28	28690	22098	1.69	1.96	equal
May.	28	10069	11235	1.24	1.96	equal
Jun.	28	2891	3564	1.52	1.96	equal
Jul.	28	696	779	1.25	1.96	equal
Aug.	28	113	97	1.36	1.96	equal
Sep.	28	69	202	8.57	1.96	unequal
Ann.	28	39693	31124	1.63	1.96	equal

* "Std(h)" = Standard Deviation of the Historic Monthly Streamflows
* "Std(g)" = Standard Deviation of the Generated Monthly Streamflows

torical monthly flows and generated monthly flows were compared. Generally, the correlation coefficient between historical monthly flows and generated monthly flows for each month preserved the correlation coefficient between the key station and the three other subordinate stations. From these results, it was concluded that the modeled generated flows were statistically similar to the historical flows.

Table 3 Hypothesis Test of Paired Mean Difference between Two Monthly Series (for 1925-1952) Water Years, Streamflow in cfs-days)

Month	Years	Mean (Cfs-days)	Std Error (Cfs-days)	Sample t	5% t	Unequal / Equal
Oct.	28	-2	13	-0.19	-2.05	equal
Nov.	28	10	31	0.33	2.05	equal
Dec.	28	11	83	-0.14	-2.05	equal
Jan.	28	15	30	0.53	2.05	equal
Feb.	28	-197	211	-0.93	-2.05	equal
Mar.	28	-419	417	-1.00	-2.05	equal
Apr.	28	1070	3829	0.28	2.05	equal
May.	28	-285	952	-0.30	-2.05	equal
Jun.	28	-638	361	-1.77	-2.05	equal
Jul.	28	-19	09	-0.18	-2.05	equal
Aug.	28	2	16	0.16	2.05	equal
Sep.	28	-33	32	-1.02	-2.05	equal
Ann.	28	-508	3190	-0.16	-2.05	equal

REFERENCES

1. Salas, J.D., Delleur, J. W.(1980), Yevjevich, V., and Lane, W.L., Applied Modeling of Hydrologic Time Series, Water Resources Publications, pp. 142.
2. Fiering, M.B.(1964), Multivariate Technique for Synthetic Hydrology, J. Hydraulic Div. ASCE, HY5: 43−49.
3. Kahn, J.P.(Dec. 1974), A Method for Maintaining Cross and Serial Correlations and the Coefficient of Skewness under Generation in a Linear Bivariate Regression Model, Water Resources Research, Vol.10,no.6,pp. 1245−1248.
4. Lawrance, A.J.(March 1976), A Reconsideration of the Fiering Two−station Model, Journal of Hydrology, Vol.29,no.1/2,pp.77−85.
5. Yevjevich, V.(May 1975), Generation of Hydrologic Samples, Case Study of the Great Lakes, Hydrology Papers, no.72, Colorado State University, Fort Collins, Col.
6. Horn, D.(Sept. 1985), Analysis and Generation of Low−Flow Sequences in Idaho Streams using Disaggreation Modeling, IWRRI Final Report.
7. Salas, J.D., and Pegram, G.G.S.(1978), A Seasonal Multivariate Multilag Autoregressive Model in Hydrology, In Modeling Hydrologic Processes, Edited by H. Morel−Seytoux, J.D.Salas,T.G. Sanders and R.E.Smith, Water Resources Publications, Fort Collins, Colorado.
8. Chow, V.T.(1954), The Log−probability Law and Its Engineering Applications. Preceedings American Society of Civil Engineers 80:536−1 to 536−25.
9. Ott, L.(1988), An Introduction to Statistical Methods and Data Analysis, 3rd Edition, PWS−KENT.
10. Dowdy, S.M.(1983), Statistics for Research, John Wiley & Sons Inc.

−1991. 10. 1 접수−