

하천유량의 모의발생을 위한 추계학적 모형의 적용에 관한 연구

A Study on the Stochastic Modeling for Stream Flow Generation

이 주 현*

Lee, Joo Heon

Abstract

The purpose of the synthetic generation of monthly river flows based on the short term observed data by means of stochastic models is to provide abundant input data to the water resources systems of which the system performance and operation policy are to be determined beforehand. In this study, a multivariate autoregressive model has been applied to generate monthly flows of the multi sites considering the correlations between each site. The model performance was examined using statistical comparisons between the historical and generated monthly series such as mean, variance, skewness and correlation coefficients.

The results of this study showed that the modeled generated flows were statistically similar to the historical flows.

keywords : stochastic modeling, multivariate modeling, AR model, monthly flows

요 지

실측자료가 충분하지 못한 단기간의 유출량 자료로부터 추계학적 모형에 의해 장기간의 자료를 모의발생시키는 목적은 수공구조물의 설계에 필요한 설계홍수량의 산정 및 수자원 시스템의 운영조작 방침을 결정하기 위한 풍부한 입력자료를 제공하는데 있다. 특히 본 연구에서는 단일지점이 아닌 다지점에 대한 지점간 서로의 연관성을 고려한 하천유량의 추계학적인 모의 발생기법인 다변량 자기회귀 모형을 적용하고자 한다. 따라서 본 연구에서는 낙동강유역의 2개 지점에 대하여 다변량 모형을 적용하여 모의 발생된 월유량과 실측치를 통계적으로 비교, 분석하였다.

모의발생된 월유량과 실측치를 평균, 분산, 왜곡도, 상관관계 등에 의해 비교, 분석한 결과 모의발생된 월유량과 실측치는 통계적으로 매우 유사하게 나타났다.

핵심용어 : 추계학적 모의모형, 다변량 모형, 자기회귀 모형, 월유량

* 정희원, 중부대학교 건설공학부 조교수

1. 서 론

최근 국내에서 발생하는 수해의 형태는 우리나라가 대륙과 해양성 기후의 교차점인 다우지역에 위치하고 있다는 기상학적 취약성외에도 도시화와 산업화에 따른 개발사업의 증가로 인하여 하천의 홍수량을 증가함으로써 그 원인이 되는 경우가 많이 발생하고 있으며 엘리뇨와 라니냐 등 이상기후현상과 국지성 집중호우로 인한 피해도 점차 증가하고 있는 추세이다. 특히 최근에 발생하는 수해가 점차적으로 대규모화되고 있는 실정에서 합리적인 수공구조물의 설계를 통한 적극적인 수해예방만이 그 어느 때보다도 절실히 요구되고 있는 실정이다.

그러나 수공구조물의 설계 및 운용에 사용되는 과거 수문자료는 설계수문량의 재현기간에 턱없이 부족한 기록연환을 갖고 있는 관계로 인하여 결국 과거 수문 관측자료로부터 통계학적 특성치를 분석하여 이 수문시계열에 적합한 모의기법을 선정된 후에 선정된 모의기법을 이용하여 장기간의 수문자료를 모의발생하는 추계학적 방법을 이용하고 있다. 이와 같은 추계학적 모의발생기법은 과거 30여년간 많은 수문학자에 의해 연구가 진행되어 왔으며 초기에는 일변량(univariate), 정상(stationary) 시계열 모형으로서 단일지점에 대하여 자료의 주기성(periodicity)을 고려하지 않은 일정한 평균과 표준편차 및 자기상관함수(auto-correlation)만을 함수로 하는 모형의 개발 및 적용이 주를 이루었으나 갈수록 복잡해지는 수자원계획 및 수문특성으로 인하여 다변량(multivariate) 모형의 필요성이 증대되었다.

다변량모형은 단일지점이 아닌 2개 이상의 서로 인접해 있는 지점을 대상으로 하는 모형으로서 자기상관성분뿐만 아니라 서로 인접해 있는 지점간의 상호상관성분(cross correlation)을 함수로 하고 있다. 또한 다변량 모형은 이용하고자 하는 수문시계열이 단기간의 자료일 경우 일변량모형을 이용할 경우 야기되는 표본화(sampling)과정에서의 오차를 많이 줄일 수 있어 보다 정도 높은 결과를 유도할 수 있다고 알려져 있다.

본 연구에서는 단일지점이 아닌 유역의 전반적인 수문특성 또는 서로 다른 지점간의 상호상관성을 고려한 수문특성을 파악하기 위하여 월별

주기성을 고려한 1차 자기회귀(autoregressive) 2지점 모형(bivariate model)을 낙동강 유역의 왜관과 진동 2지점에 대하여 적용한 후 100년간의 월유량을 모의발생하여 모의발생된 자료의 통계학적 특성치를 과거실측자료의 통계학적 특성치와 비교함으로써 다변량 모형을 이용한 수문자료의 확충 가능성을 검토하고자 한다.

2. 다변량모형의 구조

본 연구에서 하천유량의 모의발생에 이용하는 다변량 모형은 많은 수문학자에 의해 연구되어져 왔으며 Fiering(1964)은 주지점(key station)과 종속지점(subordinate station)간의 상호상관계수를 함수로 하여 하천유량의 모의발생을 시도하는 2지점(two station) 모형을 개발하여 제시하였다. 한편 Kahn(1974)과 Lawrance(1976) 등은 Fiering의 2지점 모형을 실제의 유역에 적용하여 보다 개선된 모형을 제시하기도 하였으며 Yevjevich(1975)는 각각의 단일지점의 통계학적 특성치를 보전하면서도 2개 이상의 다지점(multi site)에 적용이 가능한 다지점 모형을 개발하여 제시하였다.

하천유량의 모의발생에 적용하게 될 일반적인 다변량 1차 자기회귀 모형을 월별 주기성을 고려하여 n 개의 지점에 적용하면 다음과 같은 일반적인 구조를 갖게 된다.

$$Z_{v,t} = A_{1,t} \cdot Z_{v,t-1} + \epsilon_{v,t} \quad (1)$$

여기서, $Z_{v,t}$ 는 n 개 지점의 표준화(standardized), 정규화(normalized) 되어있는 월유량값의 ($n \times 1$) 행렬이며, $\epsilon_{v,t}$ 는 계열간에 있어서 독립적이고 공간에 대해서는 종속적인 잔차항의 ($n \times 1$) 행렬로서 다음 식 (2)에 의하여 산정되어 진다.

$$\epsilon_{v,t} = B_t \cdot \zeta_{v,t} \quad (2)$$

여기서, $\zeta_{v,t}$ 는 시간과 공간에 대해 독립적인 무작위 변수이며 일반적으로 원시계열의 확률분포형과 동일한 분포형을 갖는 난수를 적용한다. 한편 식(1)과 식(2)에서의 $A_{1,t}$ 와 B_t 는 n 개 지

점의 상호상관성을 나타내는 매개변수의 (n×n) 행렬이다.

따라서 식(1)과 식(2)를 1차 자기회귀 모형인 AR(1) 모형에 대입하여 2개의 지점에 적용하면 다음과 같은 구조를 갖게 된다.

$$Z_{v,t}^{(1)} = a^{11} Z_{v,t-1}^{(1)} + a^{12} Z_{v,t-1}^{(2)} + b^{11} \zeta_{v,t}^{(1)} + b^{12} \zeta_{v,t}^{(2)} \quad (3)$$

$$Z_{v,t}^{(2)} = a^{21} Z_{v,t-1}^{(1)} + a^{22} Z_{v,t-1}^{(2)} + b^{21} \zeta_{v,t}^{(1)} + b^{22} \zeta_{v,t}^{(2)} \quad (4)$$

여기서, $Z_{v,t}^{(i)}$ 는 i지점에 대한 표준화되어있는 월유량 자료를 나타내며 a^{km} 과 b^{km} 은 다변량 모형의 매개변수인 $A_{1,t}$ 와 B_t 행렬의 요소(elements)를 나타내고 있다.

3. 다변량모형의 적용

3.1 기본통계량 산정 및 예비해석

다변량모형의 적용에 앞서서 원자료에 대한 예비해석이 필요하며 평균 및 표준편차와 같은 기본적인 통계량의 산정과 함께 원 수문시계열의 정규성(normality)을 검정하게 되고 또한 필요에 따라서 적절한 정규화(normalization) 방법에 의해 정규분포를 갖는 자료로 변환하게 된다. 한편

본 연구에서 사용하게 될 왜관과 진동 지점의 월유출량 자료(건설부, 한국수문조사년보)는 1964년부터 1983년까지의 자료로서 기본통계량의 산정 결과는 표 1과 같으며 원자료에 대하여 확률지도 시법(plotting position technique)에 의한 적합도검정(goodness of fit test)에서 원자료가 대수정규분포에 가장 근접한 것으로 분석되어 하천유량의 정규화에 일반적으로 이용되는 2변수 대수변환(two parameter logarithmic transform)을 이용하였다.

3.2 매개변수의 산정

다변량 1차 자기회귀 모형에서 추정해야 할 매개변수는 식 (1)과 식 (2)의 $A_{1,t}$ 행렬과 B_t 행렬이 있으며 매개변수의 추정에 앞서서 원자료는 다음 식 (5)에 의해서 표준화되어진다.

$$Z_{v,t}^{(i)} = \frac{X_{v,t}^{(i)} - \bar{X}_{v,t}^{(i)}}{S_{v,t}^{(i)}} \quad (i=1,2,\dots,n) \quad (5)$$

여기서, $X_{v,t}^{(i)}$ 는 월유량 자료를 나타내며 $X_{v,t}^{(i)}$ 와 $S_{v,t}^{(i)}$ 는 월유량자료의 평균과 표준편차를 나타낸다. 한편 다변량 모형의 매개변수인 $A_{1,t}$ 행렬과 B_t 행렬은 서로 다른 지점간의 상호상관계수(cross correlation)를 요소(elements)로 하는 상관 행렬(correlation matrix)에 의해 구할 수 있으며 다음 식 (6) 및 식(7)에 의하여 산정할 수 있다.

표 1. 월유량의 기본통계량 산정결과

월	평균(cms)		표준편차(cms)		상호상관계수	
	왜관	진동	왜관	진동	lag 0	lag 1
1	39.7	66.7	38.2	25.3	-0.016	-0.012
2	49.5	102.2	37.6	74.7	0.640	0.221
3	75.4	179.3	88.2	224.3	0.799	0.560
4	114.2	286.6	92.7	212.1	0.922	0.257
5	106.1	248.7	96.8	193.7	0.885	0.809
6	92.2	241.5	88.6	245.1	0.818	0.071
7	445.7	1092.3	335.3	779.8	0.920	0.295
8	323.6	877.5	205.6	511.3	0.873	-0.114
9	254.3	61.5	204.3	566.8	0.910	-0.123
10	70.6	195.7	36.4	88.9	0.859	0.577
11	50.7	139.8	27.2	95.6	0.863	0.543
12	40.3	94.1	16.4	36.9	0.657	0.514

$$A_{1,t} = M_{1,t} \cdot M_{0,t-1}^{-1} \quad (6)$$

$$B_t \cdot B_t^T = M_{0,t} - M_{1,t} \cdot M_{0,t-1}^{-1} \cdot M_{1,t}^T = D_t \quad (7)$$

여기서, $M_{0,t}$ 행렬과 $M_{1,t}$ 행렬은 표준화 되어 있는 $Z_{v,t}$ 계열의 상관행렬이고 이를 n개의 지점에 적용한 행렬형태로 나타내면 식 (8)과 같다.

$$M_{k,t} = \begin{bmatrix} \gamma_{k,t}^{11} & \gamma_{k,t}^{12} & \dots & \gamma_{k,t}^{1n} \\ \gamma_{k,t}^{21} & \gamma_{k,t}^{22} & \dots & \gamma_{k,t}^{2n} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{k,t}^{n1} & \gamma_{k,t}^{n2} & \dots & \gamma_{k,t}^{nn} \end{bmatrix} \quad (8)$$

여기서, k는 지체량(lag), t는 주기, n을 지점수를 각각 나타내고 있다. 한편 B_t 행렬은 $A_{1,t}$ 행렬과 같이 일정한 정해를 갖고 있지 않기 때문에 Young (1968)은 B_t 행렬을 하위삼각행렬(lower triangular matrix)로 가정하므로써 B_t 행렬의 0이 아닌 대각 요소(diagonal elements)를 구하는 방법을 제시하였으며 다음과 같은 행렬형태로 나타낼 수 있다(Salas et al, 1980).

$$B_t \cdot B_t^T = \begin{bmatrix} b_t^{11} & 0 & 0 & \dots & 0 \\ b_t^{21} & b_t^{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & 0 \\ b_t^{n1} & b_t^{n2} & \dots & b_t^{nn} \end{bmatrix} \quad (9)$$

$$D_t = \begin{bmatrix} d_t^{11} & d_t^{12} & \dots & d_t^{1n} \\ d_t^{21} & d_t^{22} & \dots & d_t^{2n} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ d_t^{n1} & d_t^{n2} & \dots & d_t^{nn} \end{bmatrix} \quad (10)$$

이상과 같은 방법에 의해서 월별로 산정된 다변량 AR(1) 모형의 매개변수 추정결과를 표 2와 같다.

표 2. 다변량 AR(1) 모형의 매개변수 추정결과

월	$A_{1,t}$		B_t	
1	0.7018	-0.4731	0.8486	0.0000
	0.0804	0.2063	0.0530	0.9625
2	0.7949	0.2332	0.5655	0.0000
	0.2202	0.7377	0.5353	0.3549
3	0.4477	-0.1714	0.7535	0.0000
	0.0774	0.2368	0.5793	0.5368
4	0.5358	-0.1714	0.9112	0.0000
	0.0543	0.2368	0.4307	0.3451
5	0.6680	0.1929	0.5281	0.0000
	-0.1513	0.9714	0.4307	0.3451
6	1.2380	-1.0245	0.8145	0.0000
	0.8766	-0.6632	0.7066	0.5673
7	-0.5819	0.7709	0.8951	0.0000
	-0.5545	0.6943	0.8294	0.3907
8	-0.1462	0.0200	0.9918	0.0000
	-0.8195	0.7751	0.8644	0.3862
9	0.3280	-0.4099	0.9794	0.0000
	-0.0715	0.1358	0.9434	0.3208
10	0.6484	-0.0131	0.7713	0.0000
	0.1071	0.5675	0.6011	0.4411
11	0.8272	-0.1674	0.7250	0.0000
	0.3366	0.2302	0.7003	0.4584
12	1.5116	-0.7904	0.3899	0.0000
	0.3375	0.5450	0.2496	0.4577

3.3 난수의 발생

난수(random number)란 자연계에서 일어나는 일련의 수문현상에 대한 무작위성(randomness)을 대변해 주는 추계학적 요소로서 컴퓨터에 의해서 발생된 평균이 0.5이고, 분산이 0.083인 등분포 난수(uniform random number)를 중심극한정리(central limit theorem)에 의해 표준정규난수로 변환하였으며 중심극한정리의 기본식은 식 (11)과 같다

$$R = \frac{(\sum_{i=1}^n r_i) - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} \quad (11)$$

여기서 r_i 는 평균이 μ 이고 분산이 σ^2 인 무작위 난수를 의미하며 무작위난수의 개수 n이 무한대에 가까워지면 식 (11)에 의해 산정된 무작위 변수 R 값들은 점차 평균이 0이고 분산이 1인 정규분포를 갖게 된다.

4. 월유량자료의 모의발생 및 결과의 분석

4.1 월유량자료의 모의발생

낙동강유역의 왜관과 진동 2개 지점에 대한 모의발생을 위하여 구축된 다변량 AR(1) 모형은 표준화된 자료의 모의발생을 하게 되며 최종적인 월유량자료의 생성은 다음 식 (12)에 의해서 가능하다.

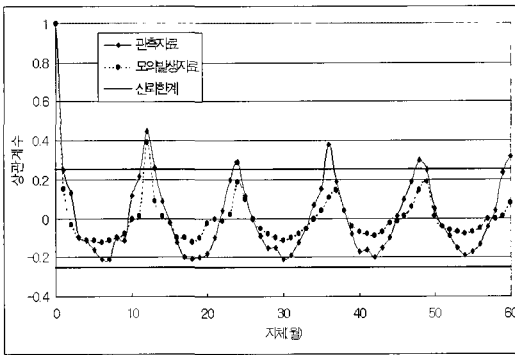


그림 1. 월유량자료의 상관도(왜관)

4.2 상관도 분석

월유량 자료에 대한 지속성과 정상성의 존재여부를 판단하기 위하여 상관도 분석을 실시하였으며 왜관과 진동의 실측자료와 모의발생된 월유량 자료에 대하여 자기상관계수 (autocorrelation coefficient)를 산정하여 각각 5년간의 값을 지체량에 대하여 도시한 것이 그림 1과 그림 2이다.

그림에서 알 수 있듯이 실측자료의 두 개 지점 모

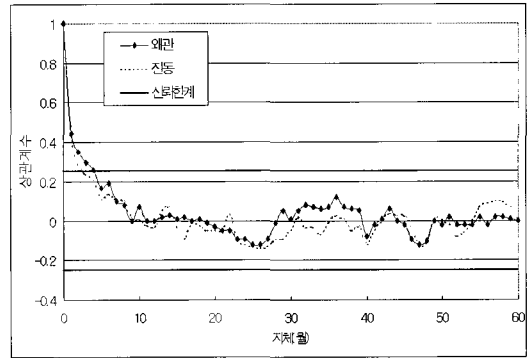


그림 2. 월유량자료의 상관도(진동)

$$X_{v,t} = \exp[\bar{Y}_{v,t} + S_{Y,t} \cdot Z_{v,t}] \quad (12)$$

여기서, $\bar{Y}_{v,t}$ 와 $S_{Y,t}$ 는 대수변환에 의해 정규화된 월유량자료의 평균과 표준편차를 나타내며 $Z_{v,t}$ 는 식 (1)에 의해서 모의발생된 표준화, 정규화 되어있는 월유량자료를 나타낸다.

두 12개월의 뚜렷한 주기성을 나타내고 있으며 동시에 신뢰한계를 벗어나고 있는 비정상시계열로 나타나고 있다. 따라서 모의발생에 앞서서 원자료의 정상화 작업이 요구되었으며 식 (5)에 의해 표준화된 자료로부터 변량간의 자기상관계수를 구하여 지체량에 대해 도시한 것이 그림 3이다. 그림에서 알 수 있듯이

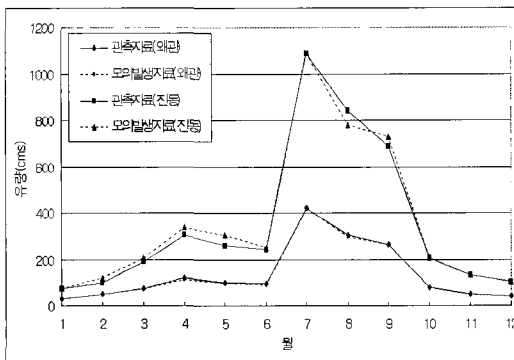


그림 4. 월유량자료의 평균값 비교

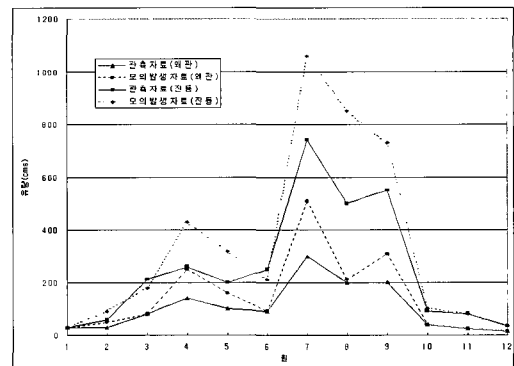


그림 5. 월 유량자료의 표준편차값 비교

표준화 작업을 통하여 원자료의 주기성이 완전히 제거되었으며 이와 같은 과정을 통하여 다변량 AR(1) 모형에 의하여 모의발생된 자료는 그림 1 및 그림 2에서와 같이 실측된 월유량 자료의 월별 주기성향을 매우 잘 재현하고 있음을 알 수 있다.

4.3 기본통계량의 비교

왜관과 진동 두지점에 대하여 다변량 AR(1) 모형을 적용하여 모의발생된 월유량의 평균과 표준편차를 각각 산정하였으며 이러한 통계학적 특성치들이 어느만큼 과거의 수문현상을 잘 재현하고 있는가를 알아보기 위하여 실측자료의 기본통계량과 비교하여 그림 4와 그림 5에 나타내었다.

그림에서 보면 왜관지점의 경우 월별 평균값은 1.99~23.66 %, 표준편차값은 4.84~46.74 %의 상대오차를 나타내고 있으며 진동지점의 경우 평균값은 2.01~23.10 %, 표준편차값은 8.12~39.91 %의 상대오차를 각각 보였다.

특히 다변량 AR(1)모형이란, 모형의 수학적인 구조상 일변량(univariate) 자기회귀 모형과 같이 지체량이 1일때의 계열상관(serial correlation) 성분을 갖고 있을 뿐만 아니라 서로 다른 계열간의 상호상관(cross correlation) 성분 또한 갖고 있다는 맥락에서 모의발생된 월유량자료로부터 지체량이 0일때와 지체량이 1일때의 두 지점에 대한 상호상관계수를 구하여 그림 6에 나타내었다.

상관도 분석시에도 알 수 있었듯이 다변량 모형은 실측 유량자료의 계열상관성분을 잘 재현하고 있었으며 그림 6의 상호상관성분 역시도 실측치와 거의 일치하는 것을 알 수 있다.

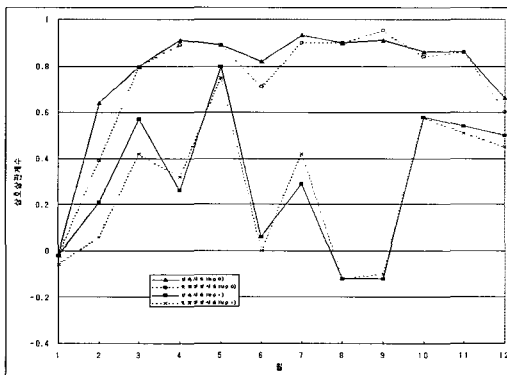


그림 6. 월 유량자료의 상호상관계수 비교

5. 결 론

본 연구에서는 다변량 1차 자기회귀 모형 (multivariate AR(1) model)을 월별 주기성을 고려하여 낙동강 유역의 왜관과 진동 두 지점에 적용하였으며 모의발생된 월유량 자료와 실측자료와의 통계학적 특성치를 비교, 분석한 결과 상관도 분석시 실측자료와 모의발생 자료 모두 강한 계절성을 나타냈으며, 실측자료를 정상시계열로 변환시키기 위해서는 1차 표준화작업이 필요한 것으로 나타났다. 특히 모의발생된 월유량의 통계학적 특성치는 실측자료의 특성치를 잘 재현하고 있었으며 지점간의 상호상관성분도 실측치의 특성과 거의 일치하고 있었다. 따라서 다변량 모형의 적용을 통한 장기간의 월유량자료의 확충이 가능할 것으로 판단되며 확충된 월유량자료의 신뢰성도 상당히 높을 것으로 기대된다.

한편 다변량 모형의 적용을 통한 보다 신뢰성 있는 결과의 유도를 위해서는 2개 지점을 대상으로 하는 것보다 더욱더 많은 지점을 대상으로 하는 다지점 모형의 적용과 함께 또한 장기간의 실측자료를 활용한 모델링 과정이 필요한 것으로 판단된다.

감사의 글

본 논문은 1999년도 중부대학교 학술연구개발비 지원에 의하여 이루어진 것임을 밝힙니다.

참 고 문 헌

- Fiering, M. B. (1964), Multivariate Technique for Synthetic Hydrology, Journal of Hydraulic Division, ASCE, Vol 90, HY5, pp 43-60
- Kahn, J. P. (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
- Kottogoda, N. T. (1980), Stochastic Water Resources Technology, The Macmillan Press.
- Lawrance, A. J. (1976), A Reconsideration of the

- Fiering Two-Station Model, Journal of Hydrology, Vol. 29, No. 1/2, pp 77-85
- Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane (1980), Applied Modeling of Hydrologic Time Series, Water Resources Publication.
- Yevjevich, V. (1975), Generation of Hydrologic Samples, Case Study of the Great Lakes, Hydrology Papers, No. 72, Colorado State University, Fort Collins.