

BDS 통계: 수문자료에의 응용

BDS Statistic: Applications to Hydrologic Data

김 형 수* / 강 두 선** / 김 종 우*** / 김 중 훈****

Kim, Hung Soo / Kang, Doo Sun / Geem, Zong Woo / Kim, Joong Hoon

Abstract

In this study, various time series are analyzed to check nonlinearities of the data. The nonlinearity of a system can be investigated by testing the randomness of the time series data. To test the randomness, four nonparametric test statistics and a new test statistic, called the BDS statistic are used and the results are compared. The Brock, Dechert, and Scheinkman (BDS) statistic is originated from the statistical properties of the correlation integral which is used for searching for chaos and has been shown very effective in distinguishing nonlinear structures in dynamic systems from random structures. As a result of application to linear and nonlinear models which are well known, the BDS statistic is found to be more effective than nonparametric test statistics in identifying nonlinear structure in the time series. Hydrologic time series data are fitted to ARMA type models and the statistics are applied to the residuals. The results show that the BDS statistic can distinguish chaotic nonlinearity from randomness and that the BDS statistic can also be used for verifying the validity of the fitted model.

keywords: BDS statistic, correlation integral, chaos, nonlinearity

요 지

본 연구에서는 다양한 특성을 가지는 시계열 자료들을 분석하여 자료의 비선형성 여부를 판단하였다. 시계열 자료의 무작위성을 분석하면 시스템의 비선형 구조를 알아낼 수 있다. 무작위성을 조사하는 통계기법으로는 전통적인 비모수 통계기법과 새로운 통계기법인 BDS 통계를 사용하였으며, 그들의 해석결과를 비교하였다. BDS 통계는 카오스 분석을 위해 이용되는 상관적분의 통계학적 특성을 바탕으로 한 검정방법으로서 무작위성과 비선형 동역학 시스템을 구분하는데 탁월한 능력이 있는 것으로 알려져 왔다. 이미 자료의 특성이 알려진 선형, 비선형 시스템에 BDS 통계를 적용한 결과, 비모수 통계기법에 비해 더욱 정확한 해석결과를 나타내었다. 실제 수문 시계열 자료를 이용하여 선형 추계학적 모형인 ARMA 형태의 모형을 구축한 후, 이 모형으로부터 계산된 잔차를 BDS 통계를 사용하여 분석하였다. 분석결과, BDS 통계는 시계열자료의 무작위성과 카오스 시스템의 비선형성을 판단하여 줄 뿐만아니라, 추계학적 모형의 잔차 분석을 통한 모형의 적합성 판단에도 유용한 방법임을 알 수 있었다.

핵심용어 : BDS 통계, 상관적분, 카오스, 비선형성

* 전문대학교 건설공학부 전임강사

Full-Time Lecturer, Div. of Construction Engrg., Sunmoon University, Chonan, Chungnam 330-150, Korea

** 고려대학교 토목환경공학과 석사과정

Graduate Student, Dept. of Civil and Envir. Engrg., Korea Univ., Seoul 136-701, Korea

*** 고려대학교 토목환경공학과 박사수료

Doctoral Candidate, Dept. of Civil and Envir. Engrg., Korea Univ., Seoul 136-701, Korea

**** 고려대학교 토목환경공학과 부교수

Associate Prof., Dept. of Civil and Envir. Engrg., Korea Univ., Seoul 136-701, Korea

1. 서 론

과거로부터 사용되어온 전통적인 통계기법들은 비선형적 구조를 가지는 시계열 자료를 해석함에 있어 그들의 비선형성을 고려하지 못하고, 대부분 선형적인 시계열로 해석하는 오류를 범하였다. 또한 그러한 비선형 자료를 표현하거나 해석하는 모형들도 대부분 선형적인 추계학적 모형에 국한되어 있었다. 그러나, 비선형 자료들을 선형적인 기법으로 해석하여 모형화하거나 예측한다는 것은 적절하지 못한 방법일 것이다. 선형 추계학적 모형은 Box와 Jenkins(1976) 이후로 여러분야에서 시계열 자료의 분석에 이용되어 왔으며, 수문 시계열 자료의 해석에 있어서도 선형 추계학적 모형인 자동회귀/이동평균(autoregressive/moving average, ARMA) 형태의 모형이 널리 사용되어져 왔다(Salas 등, 1980). 그러나, 자연현상은 통상 비선형적인 구조를 갖는 것으로 알려져 있다. 수문현상에 있어서 이러한 비선형성을 표현하기 위한 최근의 연구동향은 카오스이론을 사용한 비선형 확정론적 모형과 비선형 추계학적 모형이 수문학자들과 통계학자들에 의해 연구되어지고 있다(Rodriguez-Iturbe 등, 1989; Wilcox 등, 1991; Sangoyomi 등, 1996; Jeong 와 Rao, 1996; Porporato 와 Ridolfi, 1997).

추계학적 모형을 통해 계산된 잔차(residual)를 해석하는 방법은 그 모형의 적합성을 판단하는데 중요한 과정이라고 할 수 있을 것이다. 만일 계산된 잔차가 일정한 경향을 포함하고 있다면, 이러한 사실은 적용된 모형이 시계열 자료를 표현하기는 적합하지 않다는 결론이다. 반대로 적용된 모형이 시계열 자료의 특성을 정확히 표현하고 있다면 모형을 통해 구해진 잔차는 패턴을 갖지 않는 무작위한 분포를 따라야 할 것이다. 자료의 무작위성을 판단하기 위한 연구는 긴 역사를 통해 이루어져 왔으며 많은 통계적 기법들이 소개되어 왔다. 그러나, 지금까지 전통적으로 사용되어온 통계적인 방법들로는 이러한 자료의 비선형성을 정확히 구분하여 주지는 못하였다(Brock 등, 1991). 본 연구에서는 전통적으로 사용되어온 네가지 비모수 통계기법들과 새로운 통계기법인 BDS(Brock-Dechert-Scheinkman) 통계를 사용하여 추계학적 모형으로부터 계산되어진 잔차를 분석하여 그 적합성 여부를 판단하고자 한다.

BDS 통계는 Grassberger와 Procaccia(1983)에 의한 상관적분(correlation integral)의 통계학적 특성을 기초로 하여 개발되었으며(Brock 등, 1991), 카오스

와 비선형 추계학적 모형 해석에 사용되어지고 있다(Hsieh, 1989; Lai, 1996). 즉, BDS 통계는 무작위성을 따르는 자료와 카오스 혹은 비선형 추계학적 구조를 구분하는데 탁월한 능력을 가지고 있을 뿐만 아니라 적용된 모형으로부터 계산된 잔차를 해석함으로써 적용한 모형의 적합성을 판단하는데도 유용하게 사용되어지고 있다.

제2절에서는 전통적인 비모수 통계기법과 BDS 통계를 소개하고 정규분포의 무작위 성분과 카오스 반복 사상인 Logistic map에 적용하여 BDS 통계의 적용 가능성을 검증하였다. 제3절에서는 추계학적 모형에 적용하여 BDS 통계와 전통적인 통계기법의 결과를 서로 비교하였으며, 제4절에서는 실제 수문 시계열 자료에 적용하여 자료의 특성을 해석하였다.

2. 비모수 통계 기법과 BDS 통계

2.1 비모수 통계 기법

자료의 독립성 또는 무작위성을 판단하기 위한 여러 가지 통계 기법들이 있으나 본 연구에서는 수문 분야에 널리 이용되고 있는 4가지의 비모수 통계 기법들을 선택하였다. 이들 통계 기법들의 자세한 설명은 Salas 등(1993)에서 찾아볼 수 있으며 여기에서는 간단히 그 기법들을 소개한다.

1) Anderson's Correlogram Test

Anderson은 자료의 개수가 무한한 경우에 자기상관계수(r_k)는 평균이 0이고, 분산이 $1/n$ 인 정규분포를 따른다고 하였다. 따라서 자기상관계수가 아래의 구간에 포함될 경우에 자료는 상관성이 없는 무작위분포를 따른다고 하였다.

$$\left[\frac{-u_{1-\alpha/2}}{\sqrt{N}}, \frac{u_{1-\alpha/2}}{\sqrt{N}} \right] \quad (1)$$

Yevjevich(1972)는 위의 식 (1) 대신에 다음과 같은 분포구간을 제시하였으며, 마찬가지로 구간내에 자기상관계수가 분포할 경우에 자료는 상관성이 없다고 하였다.

$$\left[\frac{-1 - u_{1-\alpha/2}\sqrt{N-k-1}}{\sqrt{N-k}}, \frac{-1 + u_{1-\alpha/2}\sqrt{N-k-1}}{\sqrt{N-k}} \right] \quad (2)$$

2) Run Test

Run Test는 '0'과 '1'의 연속된 자료로써 정의된다.

관측치로부터 0과 1의 연속은 다음과 같은 식으로 정의할 수 있다.

$$\begin{cases} u_i = 1 & \text{for } x_i > \bar{x} \\ u_i = 0 & \text{for } x_i < \bar{x} \end{cases}$$

여기서, \bar{x} 는 자료의 평균을 의미한다. m 을 0의 연속구간의 개수, n 을 1의 연속구간의 개수로 정의하고 $R = m + n$ 이라고 하면, 다음과 같은 표준정규분포를 구할 수 있다.

$$R_c = \frac{R - 1 - 2mn/(m+n)}{\sqrt{(2mn(2mn - (m+n)))/((m+n)^2(m+n-1))}} \quad (3)$$

위에서 계산한 R_c 가 표준정규분포의 유의수준내에 포함될 경우, 시계열자료는 무작위분포를 따른다.

3) Spearman's Rank Correlation Coefficient Test

Spearman의 순위상관계수 R 은 다음과 같은 식으로 나타낼 수 있다.

$$R = 1 - \frac{6 \sum_{i=1}^N (i - u_i)^2}{N(N^2 - 1)} \quad (4)$$

위의 식으로부터 계산된 R 은 시계열 자료가 무작위할 경우 표준정규분포를 따르고, $1 - R^2$ 은 자유도가 2인 χ^2 -분포를 따른다. 따라서 student's t -분포는 다음과 같다.

$$T_c = \frac{R\sqrt{N-2}}{\sqrt{1-R^2}} \quad (5)$$

위의 식에서 계산된 T_c 값이 유의수준내에 존재할 경우 시계열 자료는 무작위한 분포를 따른다.

4) Turning Point Test

시계열 자료에서 첨점(peak)과 골점(trough)은 다음과 같이 정의할 수 있다.

$$\begin{cases} x_{i-1} < x_i > x_{i+1} & \text{for a peak} \\ x_{i-1} > x_i < x_{i+1} & \text{for a trough} \end{cases}$$

여기서, 첨점과 골점의 총 갯수를 M 이라 할 때, 만일

시계열자료가 무작위한 분포일 경우 M 값은 정규분포를 따른다.

$$U_c = \frac{M - 2(N-2)/3}{\sqrt{(16N-29)/90}} \quad (6)$$

위의 식에서 계산된 U_c 값이 유의수준내에 분포할 경우, 시계열 자료는 무작위한 분포를 갖는다.

2.2 BDS 통계

BDS 통계는 시계열 자료가 무작위한 분포를 따른다는 가설을 검증하기 위해 상관적분(Grassberger와 Procaccia, 1983)을 토대로 개발된 방법이다. 특히, 무작위한 시계열자료와 카오스계 혹은 비선형 추계학적 계를 구분하는데 매우 유용한 통계 기법이다. BDS 통계를 설명하기 위한 상관적분은 다음과 같은 식으로 나타낼 수 있다(Brock 등, 1991, 1996).

$$C(m, N, r) =$$

$$\frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(r - \|\vec{x}_i - \vec{x}_j\|), \quad r > 0 \quad (7)$$

$$\begin{aligned} \text{여기서,} \quad \theta(a) &= 0, & \text{if } a < 0 \\ \theta(a) &= 1, & \text{if } a \geq 0 \end{aligned}$$

N 은 자료의 수, $M = N(m-1)$ 은 m 차원(embedding dimension) 공간상에서 상태벡터점들의 수를 의미하며, r 값은 상태벡터점들의 수를 결정하기 위한 반경(radius)을 나타낸다.

시계열 자료가 무작위한 분포를 따른다는 가설하에 $m > 1$ 인 경우의 BDS 통계는 다음과 같은 식으로 나타낼 수 있다.

$$BDS(m, M, r) = \frac{\sqrt{M}}{\sigma} [C(m, M, r) - C^m(1, M, r)] \quad (8)$$

여기서,

$$\begin{aligned} \sigma^2(m, M, r) &= \\ &= 4m(m-1)C^{2(m-1)}(K-C^2) + K^m - C^{2m} \\ &+ 2 \sum_{i=1}^{m-1} [C^{2i}(K^{m-i} - C^{2(m-i)}) - mC^{2(m-i)}(K-C^2)] \end{aligned} \quad (9)$$

$$K(m, M, r) = \frac{6}{M(M-1)(M-2)} \sum_{1 < (i, j) \leq M} [\theta(r - \|\vec{x}_i - \vec{x}_j\|) \theta(r - \|\vec{x}_j - \vec{x}_k\|)] \quad (10)$$

BDS 통계를 통해 비선형 확정론적 시스템과 비선형 추계학적 시스템을 구분할 수는 없으나, 무작위한 시계열 자료와 카오스 시스템 혹은 비선형 추계학적 시스템을 구분하기 위해서는 매우 유용한 통계 기법이다.

BDS 통계를 이용하기전에 m 과 r 값의 선택이 중요한데, Brock 등(1991)이 여러 가지 실험을 통하여 분석한 결과, m 의 경우에는 $2 \leq m \leq 5$ 의 값과, r 의 경우에는 $0.5s \leq r \leq 2.0s$ 값이 가장 적합하였으며, 본 연구에서는 $2 \leq m \leq 5$ 의 정수값과 $r = 1.0s$ 를 선택하였다. 여기서 s 는 자료의 표준편차이다. 본 연구에서 r 값으로 $1.0s$ 를 사용한 이유는 BDS 분포의 평균에 해당하는 지점의 r 값이 $1.0s$ 이며 따라서, 이 값을 대표값으로 사용하였다.

2.3 백색잡음과 Logistic map

2.3.1 백색잡음

통계 기법들을 비교하기 위하여 우선 평균이 0인 정규분포를 따르는 백색잡음을 이용하였으며 잡음 자료의 발생은 다음식과 같다.

$$x_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (11)$$

총 1000개의 잡음 자료를 발생시켜 각 통계 기법들을 이용하여 검정을 실시하였으며, 95 % 유의수준으로 가설의 적정성 여부를 판단하였다.

표 1에서 보는 바와같이 모든 통계결과들이 유의수준내에 분포하며, 예상했던 대로 무작위한 특성이 모든 통계기법들을 통해서 증명되었다.

2.3.2 Logistic map

Logistic map은 다음과 같은 식으로 표현할 수 있으며, 카오스 시계열을 발생시키기 위해 조정계수 $r = 4$ 를 사용하여 1000개의 자료를 발생시켰다.

$$x_{t+1} = r x_t (1 - x_t) \quad (12)$$

표 1. 백색잡음 자료에 대한 검정결과

통계 기법	통계치	95% C.I.
Run	0.182	[-1.96, 1.96]
Anderson's	0.025	[-0.063, 0.061]
Spearman's	0.256	[-1.96, 1.96]
Turning Point	0.250	[-1.96, 1.96]
BDS(2)	-0.679	[-1.96, 1.96]
BDS(3)	-0.374	[-1.96, 1.96]
BDS(4)	-0.722	[-1.96, 1.96]
BDS(5)	-1.339	[-1.96, 1.96]

표 2. Logistic map에 대한 검정결과

통계 기법	통계치	95% C.I.
Run	1.521	[-1.96, 1.96]
Anderson's	-0.032	[-0.063, 0.061]
Spearman's	1.156	[-1.96, 1.96]
Turning Point	1.856	[-1.96, 1.96]
BDS(2)	122.31	[-1.96, 1.96]
BDS(3)	128.24	[-1.96, 1.96]
BDS(4)	129.45	[-1.96, 1.96]
BDS(5)	133.35	[-1.96, 1.96]

발생시킨 1000개의 자료를 통해 검정을 실시한 결과는 아래의 표와 같다. 잘 알려져 있는 것처럼 Logistic map으로부터 발생된 자료는 비선형이며 확정론적 시스템이다. 그러나 표 2의 결과에서 보면 4가지의 비모수 통계 기법들은 모두 유의수준내에 통계치들이 존재하며, 이는 Logistic map이 무작위한 특성을 지니고 있다는 결과를 나타내는 것으로 사실에 위배되는 결과이다. 반면, BDS 통계결과는 모두 유의수준을 벗어나고 있으며, 자료가 비선형적인 특성을 갖고 있음을 잘 나타내고 있다.

3. 추계학적 모형

3.1 자동회귀 모형

자동회귀모형(autoregressive model, AR(p))중 다음식과 같은 1차 자동회귀 모형(AR(1))을 이용하여 선형 추계학적 자료를 발생시켰다.

$$x_t = \phi_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (13)$$

여기서, ϕ_1 은 1차 자기상관계수로써 $\phi_1 = 0.7$ 를 사

표 3. AR(1)모형의 잔차에 대한 검정결과

통계기법	통계치	95% C.I.
Run	0.2124	[-1.96, 1.96]
Anderson's	6.648 × 10E-3	[-0.063, 0.061]
Spearman's	0.7819	[-1.96, 1.96]
Turning Point	0.6255	[-1.96, 1.96]
BDS(2)	1.0551	[-1.96, 1.96]
BDS(3)	0.9723	[-1.96, 1.96]
BDS(4)	-0.8348	[-1.96, 1.96]
BDS(5)	-1.0146	[-1.96, 1.96]

용하였으며, ϵ_t 는 평균이 0인 정규분포를 따르는 무작위 자료이다. 여러가지 통계기법들을 이용하여 식 (13)으로부터 계산된 잔차를 분석하였다.

표 3은 4가지 비모수 통계기법과 BDS 통계기법에 의해 계산된 결과이다. 모두 유의수준내에 분포하여 잔차의 무작위성을 잘 나타내주고 있으며, 이러한 사실은 적용한 AR(1)모형이 적절한 모형을 의미한다.

3.2 임계 자동회귀 모형

임계 자동회귀모형(threshold autoregressive model, TAR(K; p))중 TAR(2; 1)는 다음과 같은 식으로 나타낼 수 있다.

$$\begin{aligned}
 x_t &= -0.5 x_{t-1} + \epsilon_t & \text{if } x_{t-1} < 1 \\
 x_t &= 0.4 x_{t-1} + \epsilon_t & \text{if } x_{t-1} \geq 1
 \end{aligned}
 \tag{14}$$

여기서, ϵ_t 는 평균이 0인 정규분포를 따르는 무작위 자료이다. 이렇게 발생된 시계열자료를 여러 가지 통계기법을 사용하여 검정을 실시하였으며, 그 결과는 표 4와 같다.

표 4. TAR(2;1)모형에 대한 검정결과

통계기법	통계치	95% C.I.
Run	2.7914	[-1.96, 1.96]
Anderson's	-0.0497	[-0.063, 0.061]
Spearman's	1.0507	[-1.96, 1.96]
Turning Point	1.5512	[-1.96, 1.96]
BDS(2)	5.0933	[-1.96, 1.96]
BDS(3)	5.1914	[-1.96, 1.96]
BDS(4)	4.7732	[-1.96, 1.96]
BDS(5)	4.3381	[-1.96, 1.96]

표 5. AR(2)모형의 잔차에 대한 검정결과

통계기법	통계치	95% C.I.
Run	1.5219	[-1.96, 1.96]
Anderson's	-4.157 × 10E-5	[-0.063, 0.061]
Spearman's	0.1668	[-1.96, 1.96]
Turning Point	0.3253	[-1.96, 1.96]
BDS(2)	3.5989	[-1.96, 1.96]
BDS(3)	3.1047	[-1.96, 1.96]
BDS(4)	2.3838	[-1.96, 1.96]
BDS(5)	2.0050	[-1.96, 1.96]

위의 표에서 보는 바와 같이 Run test를 제외한 세 개의 비모수 통계결과는 자료가 선형분포라는 결과를 보인 반면, Run test를 포함하여 BDS 검정결과는 모두 자료의 비선형성을 정확히 구분하고 있다.

다음으로, 식 (14)를 통해 발생된 시계열 자료에 선형 AR(2)모형을 fitting하여 잔차를 계산하였으며, 계산된 잔차를 검정하였다.

표 5에서 보는 바와 같이 잔차를 검정한 결과 네 개의 비모수 통계치는 유의수준내에 분포하며 이는 잔차의 무작위성을 나타내는 것으로 AR(2)모형이 적합한 모형을 의미한다. 하지만, BDS 통계결과는 모두 유의수준을 벗어나고 있다. 실제로 식 (14)에 의해 발생된 자료는 비선형 자료가므로, 선형 AR모형으로 표현하기에는 부적절하다고 할 수 있다. 이러한 사실로 미루어보아 모형의 잔차 해석결과에서 BDS 통계가 더욱 정확한 해석능력을 보임을 알 수 있다.

4. 수문자료의 적용

4.1 대청댐 일유입량 자료

대청댐의 일 유입량자료를 사용하여 실제 수문자료의 무작위성 여부를 검토하였다. 1990년부터 1995년까지의 총 6개년 자료를 사용하였으며, 시계열자료의 분포는 그림 1과 같다. 원자료에 로그를 취하고 평균을 제거하여 정상시계열로 변환한 후, ARMA모형에 fitting 하였다. 적절한 차수를 결정하기 위하여 AIC(Akaike Information Criterion)와 BIC(Bayesian Information Criterion)를 계산하였으며, 두 값이 가장 작은 값(AIC=12342.99, BIC=12376.03)을 보이는 AR(26)모형을 선택하였다. AR(26)모형은 다음과 같은 식으로 표현할 수 있다.

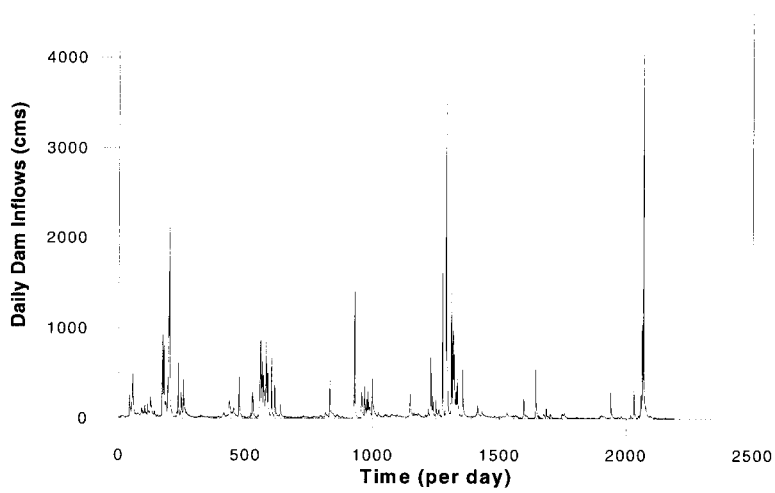


그림 1. 대청댐 일유입량 자료 (1990~1995)

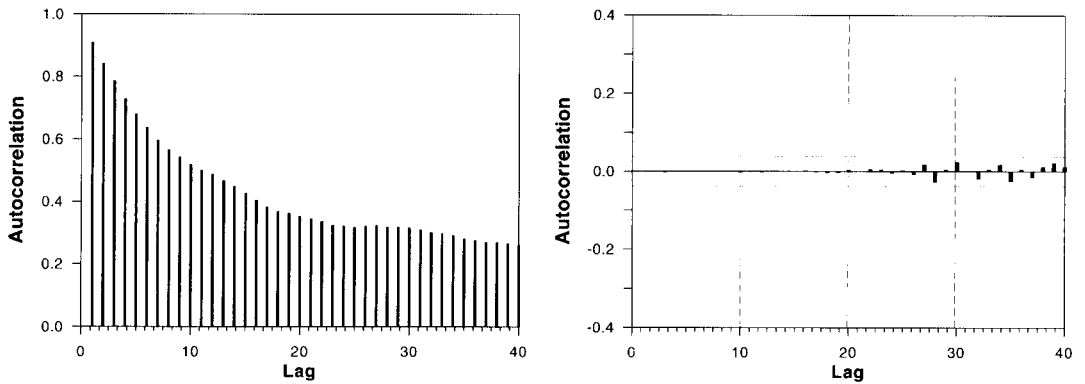


그림 2. (a) 표준화된 일유입량 자료의 자기상관함수, (b) AR(26)모형의 잔차에 대한 자기상관함수

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) \cdots + \phi_{25} x(t-25) + \phi_{26} x(t-26) + \varepsilon(t), \varepsilon(t) \sim N(0, \sigma^2) \quad (15)$$

그림 2는 정상시계열로 변환한 자료의 자기상관함수를 계산한 결과와 AR모형에 fitting하여 구해진 잔차의 자기상관함수를 계산한 결과이다. 그림 2(b)에 나타난 잔차의 자기상관도에서는 상관성이 거의 없는 무작위한 분포를 보이며, 이는 AR(26)모형이 적합한 모형임을 의미한다. 이러한 결과는 표 6의 Spearman test를 제외한 세가지 비모수 통계결과에서도 같은 결과를

표 6. 대청댐 일유입량 자료의 잔차에 대한 검정결과

통계기법	통계치	95% C.I.
Run	0.2287	[-1.96, 1.96]
Anderson's	-5.612×10E-4	[-0.0423, 0.0414]
Spearman's	2.1840	[-1.96, 1.96]
Turning Point	0.3211	[-1.96, 1.96]
BDS(2)	14.550	[-1.96, 1.96]
BDS(3)	16.520	[-1.96, 1.96]
BDS(4)	17.930	[-1.96, 1.96]
BDS(5)	19.130	[-1.96, 1.96]

나타내었다. 하지만, BDS 통계결과는 모두 유의수준을 벗어나고 있으며, 이는 잔차의 무작위성에 위배되는 결과이다. 따라서, AR(26)모형은 결국 대청댐의 일유입량 자료의 해석에 부적합함을 의미하며 이러한 결과는 다른 차수의 AR모형의 적용에서도 같게 나타났다. 즉, 선형 추계학적 모형인 AR모형은 대청댐 일 유입량 자료의 해석에 적절하지 못한 모형임을 의미하며, 조금 더 복잡한 비선형 추계학적 모형이나 혹은 카오스 모형을 적용하는 것이 바람직할 것으로 판단된다.

4.2 Great Salt Lake 용적자료

Great Salt Lake(GSL)는 평균수심이 3-5 m 정도로 매우 얇고, 대단히 넓은 표면적(6400 km²)을 가진 세계에서 네 번째로 큰 호수이다. 1847년부터 1992년까지 15일 단위의 용적자료(volume data)중 앞선 2300개의 자료를 사용하였다 (Sangoyomi 등, 1996). 시계열자료의 분포는 그림 3과 같다. 그림에서 보는 바와같이 건기와 우기가 수년에 걸쳐서 지속되는 현상을 보이고 있음을 알 수 있다.

그림 3의 원자료에 로그를 취하고 평균을 제거하여 정상시계열로 변환한 후, 선형 시계열 모형인 ARMA 모형에 fitting 하였다. 앞서 4.1절에서 언급한 바와같이 AIC와 BIC를 계산하여 가장 작은 값 (AIC=-13518.58, BIC=-13506.22)을 보이는 AR(25)모형을 선택하였으며, 다음과 같은 식으로 표현할 수 있다.

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) \cdots \phi_{24} x(t-24) + \phi_{25} x(t-25) + \varepsilon(t), \quad \varepsilon(t) \sim N(0, \sigma^2) \quad (16)$$

그림 4는 정상시계열로 변환한 자료의 자기상관함수를 계산한 결과와 AR모형에 fitting하여 구해진 잔차의 자기상관함수를 계산한 결과이다. 그림 4(a)로부터 시계열간에 상당히 긴 지체시간까지 큰 상관성을 가짐을 알 수 있다. 그림 4(b)에 나타난 자기상관도에서는 AR모형으로부터 계산된 잔차가 상관성이 없는 무작위한 분포를 가지며, 이는 AR(25)모형이 적합한 모형임을 의미한다.

표 7과 표 8은 각각 정상시계열로 변환한 자료와 변환된 자료를 AR(25)모형에 fitting하여 계산된 잔차를 앞서 언급했던 여러 가지 통계기법들을 통해 검정한 결과이다. 표 7의 결과를 통해서 볼 때, 원자료계열은 무작위한 분포가 아닌 상관성을 갖는 비선형적인 구조를 가지고 있음을 짐작할 수 있다. 이는 그림 4-(a)의 자기상관도와 같은 결과이다.

표 8의 잔차를 분석한 결과에서도 마찬가지로 Anderson's test를 제외한 모든 통계결과가 유의수준을 벗어나고 있다. 이는 잔차의 무작위성에 위배되는 결과로써 원자료계열을 모형한 AR(25)모형이 원자료가 가지고 있는 비선형적인 특성을 충분히 반영하지 못하였으며, 이는 잔차의 자기상관도에서는 알 수 없었던 결과이다. 표 8의 결과로부터 선형 AR모형은 GSL

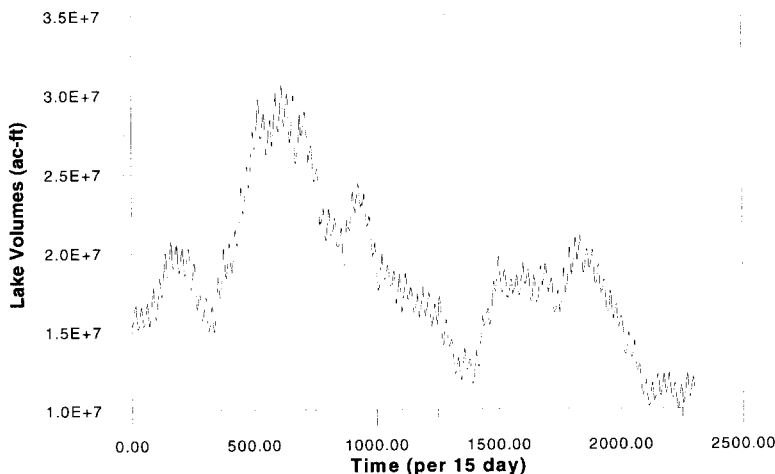


그림 3. Great Salt Lake 용적자료

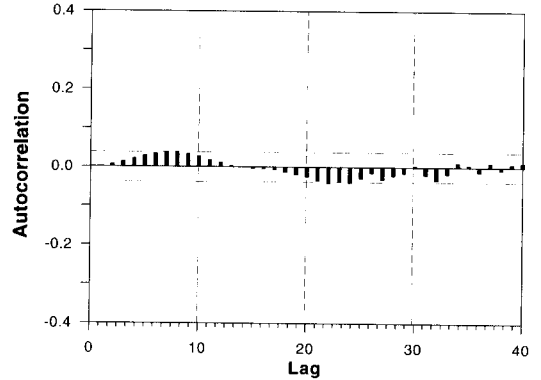
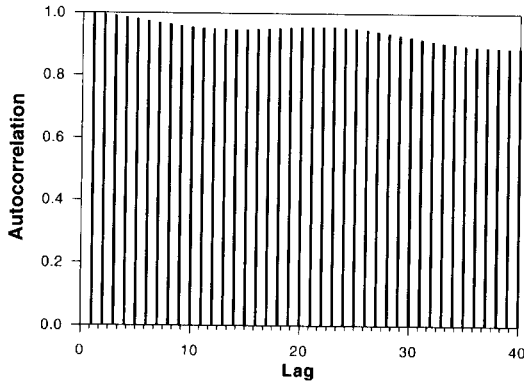


그림 4. (a) 표준화된 GSL 용적자료의 자기상관함수, (b) AR(25)모형의 잔차에 대한 자기상관함수

표 7. 표준화된 Great Salt Lake 용적자료에 대한 검정결과

통계기법	통계치	95% C.I.
Run	46.05	[-1.96, 1.96]
Anderson's	0.998	[-0.0413, 0.0404]
Spearman's	28.84	[-1.96, 1.96]
Turning Point	63.08	[-1.96, 1.96]
BDS(2)	140.78	[-1.96, 1.96]
BDS(3)	175.81	[-1.96, 1.96]
BDS(4)	225.93	[-1.96, 1.96]
BDS(5)	302.59	[-1.96, 1.96]

표 8. Great Salt Lake 용적자료의 잔차에 대한 검정결과

통계기법	통계치	95% C.I.
Run	7.247	[-1.96, 1.96]
Anderson's	$2.643 \times 10E-3$	[-0.0413, 0.0404]
Spearman's	4.264	[-1.96, 1.96]
Turning Point	14.20	[-1.96, 1.96]
BDS(2)	15.14	[-1.96, 1.96]
BDS(3)	19.45	[-1.96, 1.96]
BDS(4)	20.87	[-1.96, 1.96]
BDS(5)	22.42	[-1.96, 1.96]

용적자료에 적합하지 않은 모형임을 알 수 있다. 실제로, Sangoyomi 등(1996)은 GSL 용적자료가 비선형 확정론적 특성을 보인다는 사실을 카오스 분석으로부터 증명한 바있다.

5. 결 론

BDS 통계를 여러 가지 시계열 자료에 적용하여 자료의 무작위성 여부를 해석하였다. 적용결과, BDS 통계는 자료의 비선형성 해석에 매우 정확한 결과를 나타내었으며, 특히 Logistic map이나 TAR모형에 의해 발생된 비선형 시계열의 해석에 있어서 기존의 비모수 통계기법은 자료의 비선형성을 구분하지 못한 반면, BDS 통계는 정확히 구분하였다. 실제 수문 시계열 자료의 적용에서는 대청댐의 일유입량 자료와 미국 GSL의 격주 용적자료를 사용하였다. 자료를 선형 AR모형에 fitting하여 계산된 잔차를 해석하여 모형의 적합성 여부를 판단하였다. 모형에 fitting한 결과 대청댐 일유입량 자료는 AR(26)모형, 그리고 GSL 격주 용적자료의 경우에는 AR(25)모형을 선택하였다. 잔차의 해석결과 대청댐의 일유입량 자료에서는 기존 비모수 통계기법으로는 모형이 적합하다는 결과가 나왔으나 BDS 통계결과 AR모형은 적절하지 않은 것으로 나타났다. GSL 용적자료의 잔차해석 결과는 비모수 통계결과와 BDS 통계결과가 모두 유의수준을 벗어나 역시 선형 AR모형의 적용은 적절치 못함을 판단할 수 있었다. 따라서, 두 경우 모두 비선형 추계학적 모형이나 카오스 모형과 같은 비선형 모형의 선택이 이루어져야 할 수 있었다. BDS 통계를 사용한 시계열 자료의 해석을 통해 자료의 비선형성을 판단할 수 있음은 물론, 적용모형의 적합성 여부를 판단하여 가장 적합한 모형의 선택이 가능하리라 생각된다.

참 고 문 헌

- Box, G.E.P., and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. Revised edition, Holden-Day, San Francisco.
- Brock, W.A., Heish, D.A. and LeBaron, B. (1991). *Nonlinear dynamics, chaos, and instability: Statistical theory and economic evidence*. The MIT Press.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A. and LeBaron, B. (1996). "A test for independence based on the correlation dimension." *Econ. Rev.*, Vol. 15, No. 3, pp. 197-235.
- Grassberger, P. and Procaccia, I. (1983). "Measuring the strangeness of strange attractors." *Physica 9D*, pp. 189-208.
- Hsieh, D.A. (1989). "Testing for nonlinear dependence in daily foreign exchange rates." *J. Business*, Vol. 62, No. 3, pp. 339-368.
- Jeong, G. D. and Rao, A. R. (1996). "Chaos characteristics of tree ring series." *J. Hydrology*, Vol. 182, pp. 239-257.
- Lai, D. (1996). "Comparison study of AR models on the Canadian lynx data: A close look at the BDS statistic." *Computational Statistics & Data Analysis*, 22, pp. 409-423.
- Porporato, A. and Ridolfi, L. (1997). "Nonlinear analysis of river flow time sequences." *Water Resour. Res.*, Vol. 33, No. 6, pp. 234-247.
- Rodriguez Iturbe I., Febres de Power, B., Sharifi, M. B. and Georgakakos K. P. (1989). "Chaos in rainfall." *Water Resour. Res.*, Vol. 25, No. 7, pp. 1667-1675.
- Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L. (1980). *Applied modeling of hydrologic time series*. Water Resources Publications, Littleton, Colorado.
- Salas, J.D., Smith, R.A., Tabios, G.Q. III and Heo, J.H. (1993). *Statistical Computer techniques in hydrology and water resources*. Unpublished text book in CE 622, Colorado State University.
- Sangoyomi, T.B., Lall, U. and Abarbanel, H.D. I. (1996). "Nonlinear dynamics of the Great Salt Lake: Dimension estimation." *Water Resour. Res.*, Vol. 32, No. 1, pp. 149-159.
- Wilcox, B.P., Seyfried, M.S. and Matison, T.H. (1991). "Searching for chaotic dynamics in snowmelt runoff." *Water Resour. Res.*, Vol. 27, No. 6, pp. 1005-1010.
- Yevjevich, V.M. (1972). *Probability and statistics in hydrology*. Water Resources Publications, Fort Collins, Colorado.

(논문번호:98-058/접수:1998.09.09/심사완료:1998.11.16)