

# 허스트 지수 산정 방법에 대한 고찰

## On the Estimation Techniques of Hurst Exponent

김 병 식\* / 김 형 수\*\* / 서 병 하\*\*\*

Kim, Byung Sik / Kim, Hung Soo / Seoh, Byung Ha

### Abstract

There are many different techniques for the estimation of the Hurst exponent. However, the techniques can produce different characteristics for the persistence of a time series each other. This study uses several techniques such as adjusted range, rescaled range(RR) analysis, modified rescaled range(MRR) analysis, 1/f power spectral density analysis, Maximum Likelihood Estimation(MLE), detrended fluctuations analysis(DFA), and aggregated variance time(AVT)method for the Hurst exponent estimation. The generated time series from chaos and stochastic systems are analyzed for the comparative study of the techniques. Then this study discusses the advantages and disadvantages of the techniques and also the limitations of them.

**Keywords** : Hurst exponent, long & short term memory, RR, DFA, AVT

### 요 지

허스트 지수를 산정하기 위하여 기존에 여러 방법론들이 제안되어 왔다. 그러나, 이들 방법론들은 시계열들의 지속성에 대하여 각기 다른 특성들을 보이고 있음을 기존의 연구에서 알 수 있다. 따라서 본 연구에서는 수문학에서 주로 이용하고 있는 보정용량, 조정용량, 수정조정용량 방법 이외에 생리학 분야와 전자 분야 등에서 이용되고 있는 1/f 파워 스펙트럼 밀도 분석, DFA, AVT 방법, 최우도법 등을 이용하여 허스트 지수를 산정하여 보았다. 즉, 단기간과 장기간 기억을 가진 카오스와 추계학적 시계열들에 대하여 각각의 방법들을 적용하여 비교 분석하고자 하였으며, 각 방법론들에 대한 장점 및 단점 그리고 한계에 대하여 논의하였다.

**핵심용어** : 허스트 지수, 장·단기 기억, 조정용량, 수정조정용량, DFA, AVT

### 1. 서 론

대부분의 관측 수문 시계열 자료는 평균값을 기준으로 상하에 연속적으로 존재하는 경우가 많다. 이러한 성향을 수문 시계열에서는 장기간의 지속성(long term persistence)이라고 부른다. 그러나, 시계열 집합의 크기가 지정되지 않는다면 시계열에서의 지속성의

의미는 명확하지 않으며 또한, 지속성의 의미가 없다고 할 수 있다(Kashyap 등, 1976). Hurst(1951)는 경험적 연구에 의해 어떤 시계열 평균과 시계열 값간의 누가최대편차와 누가최소편차간의 범위 또는 용량  $R(N)$ 은  $(N/2)^\beta$ 에 비례한다고 결론지었으며, 여기서,  $N$ 은 표본의 크기,  $\beta$ 는 지수이고 이 지수의 평균은 0.729이고 표준편차는 0.092라는 값을 얻었다. 허스트 현상

\* 한국건설기술연구원 수자원연구부 선임연구원  
Senior Researcher, Water Resources Research Dept., KICT, Goyang-Si, 2311, Korea (E-mail: hydrokbs@kict.re.kr)  
\*\* 인하대학교 토목공학과 교수, 공학박사  
Prof., Dept. of Civil Engrg., Inha Univ., Incheon, 402-751, Korea (E-mail: sookim@inha.ac.kr)  
\*\*\* 인하대학교 토목공학과 교수, 공학박사  
Prof., Dept. of Civil Engrg., Inha Univ., Incheon, 402-751, Korea

을 설명하기 위하여 Mandelbrot 등(1969)은 Joseph호과를 예로 들었고, 표본의 표준편차,  $\sigma(n)$ 로 용량(range)을 나누어 이를 조정용량(Rescaled Range)이라고 하였으며,  $n$ 이 무한히 증가할 경우  $R(n) / \sigma(n)$ 이  $n^H$ 과 같아짐을 증명하였다. 여기서,  $H$ 를 허스트 지수라고 하였으며, 허스트 지수를 추정하기 위해 대수치를 취하여 도시하였다. 조정용량  $\log(R(n))$ 과  $\log(n/2)$ 를 도시하여 그려진 직선의 기울기로부터 허스트 지수가 추정된다. 정규 독립 시계열에서는  $h$ 는 0.5에 근접하며(Mandelbrot 등, 1969) 통계학적으로 상관관계가 존재하는 시계열의 경우에는 허스트 지수가 0.5 이상( $h > 0.5$ )이 된다고 하였다. 최근 들어 수문 시계열자료에서 장기간 지속성에 관한 연구(Hipel 등, 1994)가 계속 진행되어 왔다. 그러나, 지금까지 수문분야에서 허스트 지수를 추정하기 위해 가장 보편적으로 이용되고 있는 조절용량(adjusted range) 방법은 시계열자료를 장기간 기억과 단기간 기억의 특성으로 구분하는 데는 문제가 있음이 Lo(1991)에 의해 제시되었으며 Rao와 Bhattacharya (1999; 2001)는 수문시계열 자료를 분석해 본 결과 조절용량 방법은 단기간 기억을 지닌 수문자료의 특성을 추출하기에는 적절하지 못하므로 Lo(1991)에 의해 제시된 수정조정용량(Modified Rescaled Range) 방법을 이용하여 수문자료의 장기간 기억과 단기간 기억의 특성을 추출하고자 하였다.

생리학 분야와 전자 컴퓨터 분야 그리고 물리학분야에서는 최근 들어 허스트 지수를 뇌파, 심전도, 전자파의 신호 등의 지속성을 판단하는 지표로 사용할 뿐 아니라 카오스의 특성과 자기 유사성(self-similarity)을 지니고 있는지를 판단하는 지표로도 이용하고 있다(Herman, 1999; Pallikari, 1999; Kantelhardt, 2001). 그러므로, 허스트 지수의 추정 방법은 더욱 중요하다고 할 수 있다.

허스트 지수를 산정하기 위해 생리학(Physiology), 금융(Finance), 물리학(Physics), 전기(Electricity) 분야에서 사용되고 있는 허스트 지수의 추정 방법들을 살펴보면 1/f Power Spectral Density(Bigger 등, 1996), Aggregated Variance Time Method(Taqqu와 Teverovsky, 1995; Zhao, 1998; Zafer와 Sirin, 1999), Detrended Fluctuations Analysis (Ausloos, 2000; Kantelhardt, 2001), Maximum Likelihood Estimation (Kendziorzski, 1999)등이 있다. 본 논문에서는 이 방법들에 대하여 살펴보고 카오스 및 추계학적 시스템의 시계열 자료에 적용하여 허스트 지수를 추정한 후 그 결과를 비교 및 분석하여 각 방법들의 장단점과 한계성을 파악하고자 하였다.

## 2. 허스트 지수 산정을 위한 대상 자료

본 논문에서는 허스트 지수를 일부 시스템에 대하여 미리 설정함으로써 단기간 및 장기간 기억의 특성을 갖도록 하였다. 즉, 허스트 지수를 이미 알고 있거나 설정한 시스템의 자료들과 그렇지 않은 시스템의 자료들에 대해 여러 기법들을 적용, 허스트 지수를 산정하여 기법들을 비교하고자 한다. 식(1)에서 (7)까지 각 시스템들을 나타내었고 이들 시스템들로부터 각 1,000개의 카오스 및 추계학적 시계열 자료들을 모의 발생하였으며, 그림 1은 이들 시계열자료를 보이고 있다. 식(1)은 정규난수를 의미하고, 식(2), (3), (4)는 선형추계학적 시스템(Salas 등, 1995), 식(5)는 준주기(quasi-periodic) 시스템, 식(6)은 카오스 시스템, 그리고 식(7)은 비선형 추계학적 시스템의 한 예(Tong, 1990)를 나타낸다.

① Gaussian white noise ;  $H \sim 0.5$

$$x_t = \epsilon_t \sim N(0,1) \quad (1)$$

② Fractional Gaussian noise with long range correlaton(FGN) ;  $H \sim 0.8$

$$x_t = B_H(i+1) - B_H(i), i \geq 1 \quad (2)$$

여기서,  $B_H$  ; Fractional Brownian motion

③ Autoregressive model(AR(1)) with  $\rho_1 = 0.7$

$$x_t = \rho_1 x_{t-1} + \epsilon_t \sim N(0,1) \quad (3)$$

여기서,  $\rho_1$  ; 1차 자기상관 계수,  $\epsilon_t$  ; 정규분포난수

④ Fractionally Differenced ARMA (FARIMA);  $H \sim 0.8$

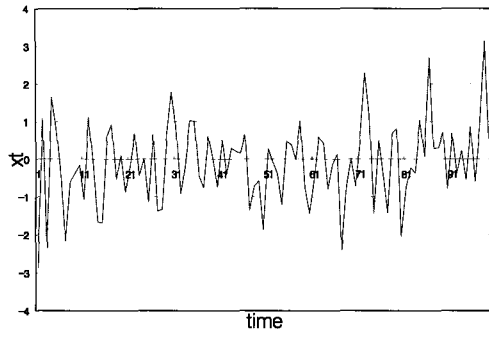
$$\Phi(B)(1-B)^d x_t = \Theta(B)\epsilon_t \sim N(0,1) \quad (4)$$

여기서,  $\Phi(B)$  ; AR 모형 변수

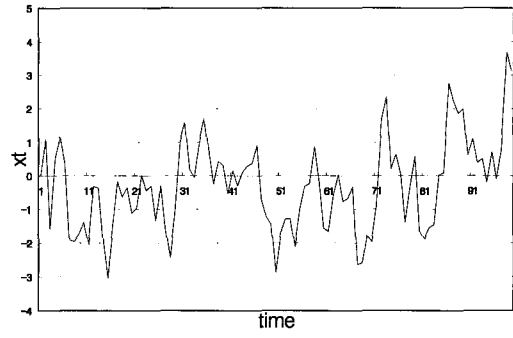
$\Theta(B)$  ; MA 모형 변수

$B$  ; back shift operator

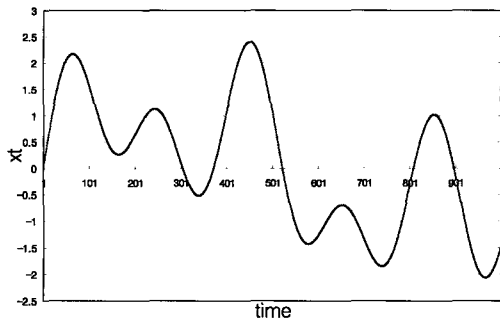
$d$  ;  $H - 1/2$ ,  $\epsilon_t$  ; 정규분포난수



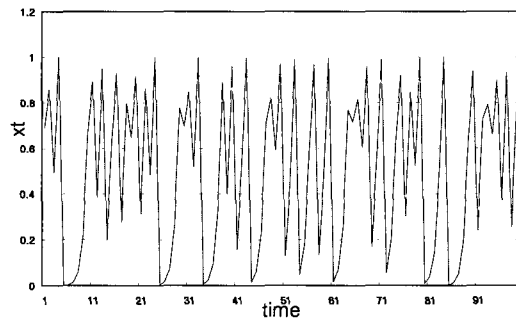
(a) white noise



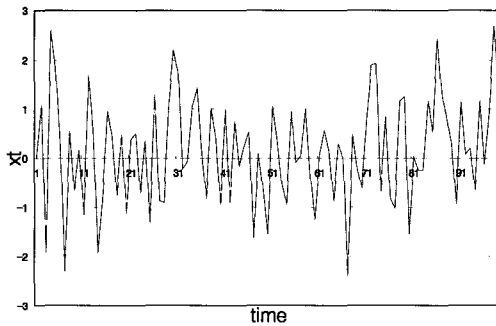
(b) AR(1)



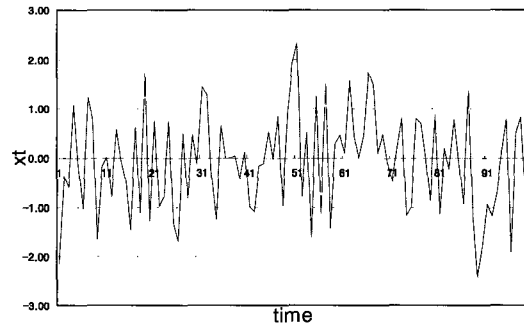
(c) Three torus



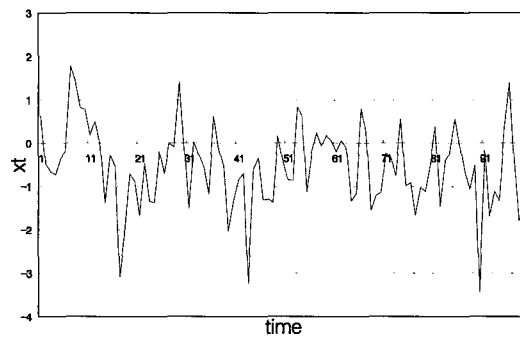
(d) Logistic map



(e) TAR(2,1)



(f) FARIMA



(g) FGN

그림 1. 카오스 및 추계학적 시계열

⑤ Three torus(준주기 시스템)

$$x_t = \sin\left[\frac{3t}{500}\right] + \sin\left[\frac{3\sqrt{2}t}{250}\right], \quad t \geq 1 \quad (5)$$

$$+ \sin\left[\frac{9\sqrt{3}t}{500}\right]$$

⑥ Logistic map (카오스 시스템)

$$x_{t+1} = 4x_t(1-x_t) \quad (6)$$

⑦ Threshold autoregressive model(TAR(2,1))

$$x_t = -0.5x_{t-1} + \epsilon_t \quad \text{if } x_{t-1} < 1 \quad (7)$$

$$x_t = 0.4x_{t-1} + \epsilon_t \quad \text{if } x_{t-1} \geq 1$$

여기서,  $\epsilon_t$ ; 정규분포난수

### 3. 허스트 지수 산정 방법

허스트 지수 산정을 위하여 여러 가지 방법들이 존재하는 데, 본 연구는 이들 방법들을 이용하여 앞절에서 언급한 자료들에 대한 허스트 지수를 산정하여 상호 비교함으로써 각 방법들의 적정성 여부를 논의하고자 한다. 본 연구에서 이용한 각각의 허스트 지수의 추정 방법들에 대해 간단히 살펴보면 다음과 같다.

#### 3.1 보정용량 방법

Hurst(1951)에 의해 처음 제안된 방법으로 나무의 나이테, 강우량, 유량, 기압 등 많은 자연시계열을 분석한 결과 다음과 같은 식을 알아내었다.

$$R = \left[ \max \sum_{t=1}^N (X_t - \bar{X}_n) - \min \sum_{t=1}^N (X_t - \bar{X}_n) \right] \quad (8)$$

$$R/\sigma = kN^\beta \rightarrow R/\sigma = \left(\frac{N}{2}\right)^H \quad (9)$$

$$H = \frac{\text{Log}(R_N)}{\text{Log}\left(\frac{N}{2}\right)} \quad (10)$$

여기서,  $X_t$ ; 시계열 자료  $t = 1, 2, 3, \dots, N$

$N$ ; 전체 자료의 크기

$\sigma$ ; 표준편차,  $H$ ; 허스트 지수

#### 3.2 조정용량 방법

Mandelbrot(1969)에 의해 처음 제안된 방법으로 기존의 보정용량 방법을 수정한 것이다.

$$\hat{Q}(n) = \frac{1}{\sigma(n)} \left[ \max \sum_{t=1}^k (X_t - \bar{X}_n) - \min \sum_{t=1}^k (X_t - \bar{X}_n) \right] \quad (11)$$

$$\sigma(n) = \sqrt{\frac{\sum_{t=1}^n (X_t - \bar{X}_n)}{n-1}} \quad (12)$$

여기서,  $X_t$ ; 시계열 자료  $t = 1, 2, 3, \dots, N$

$N$ ; 전체 자료의 크기

$n$ ; 표본의 크기 ( $1 \leq n \leq N$ )

$\sigma(n)$ ; 표준편차

Mandelbrot(1969)는 40년에서 2000년 범위의 800개 연 시계열자료를 이용하여 각각의  $n$ 에 대한  $\hat{Q}(n)$ 를 계산하여 다음과 같은 관계가 성립함을 알아내었다.

$$\frac{\hat{Q}(n)}{\sigma(n)} = an^H \quad (13)$$

여기서,  $H$ ; 허스트 지수

$n$ ; 표본의 크기( $1 \leq n \leq N$ )

$a$ ; 상수

위 식을 대수변환을 이용하면 식(14)와 같은 회귀식으로 표현될 수 있으며 이를 통해 허스트 지수를 추정할 수 있다.

$$\text{Log}(R(n)/\sigma(n)) = \text{Log}(a) + H(\text{Log}(n)) \quad (14)$$

#### 3.3 수정조정용량 방법

Lo(1991)는 기존의 조정용량에서 회귀관계식을 이용하여 허스트 지수를 추정할 때 문제점이 있음을 제시한 바 있으며, 그 이유로써 회귀계수(regression coefficients)가 자기상관성에 의해 왜곡된다는 점이다. 이로 인해 기존의 조정용량에 의한 허스트 지수의 추정은 장기간 기억을 지닌 시계열 자료의 경우에는 적합하지만 단기간 기억을 지닌 시계열자료에는 적합하지 않다고

하였다. 이러한 문제점을 해결하고자 Lo(1991)는 수정 조정용량 방법을 제시하였으며, Rao와 Bhattacharya (1999; 2001)는 이 방법을 관측 수문시계열 자료에 적용하였다. 이 방법과 조정용량 방법의 차이는 조정용량 식에서 분모항의 표준편차가 자기상관계수를 고려할 수 있도록 수정되었다는 점이다. 수정조정용량을 수식으로 써 표현하면 다음과 같다.

$$\hat{Q}(n, q) = \frac{1}{\sigma_n(q)} [\max_{t=1}^k (X_t - \bar{X}_n) - \min_{t=1}^k (X_t - \bar{X}_n)] \quad (15)$$

$$\hat{\sigma}_n(q) = \hat{\sigma}_x^2(n) + 2 \sum_{j=1}^q \omega_j(q) \hat{\gamma}_j \quad (16)$$

$$\omega_j(q) = 1 - \frac{j}{q+1} \quad q < n, \quad j = 1, 2, \dots, n \quad (17)$$

여기서,  $\omega_j(q)$  ; 가중자기공분산 함수

$q$  ; 가중 자기공분산함수(weighted auto-covariance function)의 절단 지체시간

$\hat{\gamma}_j$  ; 자기공분산 추정치

수정 조정용량 방법을 실제로 적용할 경우 가장 주의해야 할 점은 가중자기공분산함수의 절단지체시간,  $q$  를 결정하는 것이다. 여기에서  $q$ 의 의미는 시계열에서 초기의 지속성이 전체 시계열에 미치는 범위(windows)를 의미한다고 할 수 있다. 그러므로, 만일  $q$ 가 너무 작게 설정된다면 단기간 기억의 영향을 고려하지 못하게 되며  $q$ 가 너무 클 경우에는 장기간 기억의 영향을 무시하게 된다(Teverovsky 등, 1998). 수정조정용량 방법을 적용하기 위해서는 최적 절단 지체시간,  $q_{opt}$ 의 추정이 선행되어야 한다. 이를 위해 Phillis(1987)는  $q = N^{0.25}$ 를 제시하였으며 Teverovsky 등(1998)은 식 (18)을 제시하였다. 그러나, Rao와 Bhattacharya(2001)은  $q = N/10$ 와  $q = N/5$ 라고 가정한 후 수문자료에 수정조정용량 방법을 적용하였다. 본 연구에는 식 (18)을 이용하여  $q_{opt}$ 을 결정하고자 하였다.

$$q_{opt} = \left(\frac{3N}{2}\right)^{1/3} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3} \quad (18)$$

여기서,  $N$  ; 전체 자료의 크기

$\hat{\rho}$  ; 1차 자기상관계수 (first-order autocorrelation)

### 3.4 1/f 파워스펙트럼 밀도분석

1/f 파워스펙트럼밀도(Power spectral density, PSD) 방법은 주로 생리학 분야에서 심장박동, 뇌파의 자기유사성 분석을 목적으로 허스트 지수를 추정하기 위해 이용되는 방법이다. 1/f 파워스펙트럼밀도 방법에서 경향성은 주기,  $f$ 에 대해 역비례하며  $1/f^\alpha$ 를 따른다. 1/f의 경향성은  $\alpha$ 를 기울기로 가지는 직선으로 표현되며  $-\alpha$ 의 값은 저주파폭(low frequency band)의 회귀분석을 통해 쉽게 계산할 수 있다. 기울기,  $\alpha$ 와 허스트 지수를 식(20)과 (21)으로 나타낼 수 있다 (Bigger 등, 1996).

$$h(f) = \sigma^2 (2\pi)^{-1} \sum_{k=-\infty}^{+\infty} \rho(k) e^{ikf} \quad (19)$$

여기서,  $h(f)$  ; 1/f 파워스펙트럼밀도

$\rho(k)$  ; 자기상관계수

$k$  ; 지체시간,  $i = 1, 2, \dots$

$$\alpha > 1 \quad \longrightarrow \quad H = \frac{\alpha - 1}{2} \quad (20)$$

$$\alpha < 1 \quad \longrightarrow \quad H = \frac{\alpha + 1}{2} \quad (21)$$

### 3.5 Detrended Fluctuations Analysis 방법

Detrended Fluctuations Analysis(DFA) 방법은 규모 (scaling) 지수로써 시계열의 자기상관성을 나타내는 방법으로써, 금융 분야나 생리학 분야에서 널리 사용되고 있다. 이 방법에 대하여 간단히 요약하자면, 먼저 분석하고자하는 시계열 자료를 부분합산을 한 후, 합산된 자료를 동일한 크기로 평균을 한 후 최소자승 직선 (least squares line)에 일치시킨다. 여기서,  $\text{Log}(F(n))$ 와  $\text{Log}(n)$ 으로 나타나며 이때의 기울기를 규모 지수,  $d$ 라 부르며  $d$ 는 1/f 파워스펙트럼밀도의 기울기  $\alpha$ 와 식 (22)의 관계가 있다. DFA 방법을 수식으로 표현하면 식(23)과(24)와 같다.

$$\alpha = 2d - 1 \quad (22)$$

$$y(k) = \sum_{i=1}^k [X(i) - X_{mean}] \quad (23)$$

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^n [y(k) - y_n(k)]^2} \quad (24)$$

### 3.6 Aggregated Variance Time 방법

Aggregated Variance Time(AVT) 방법은 먼저, 원 자료  $X_i, i=1,2,\dots,N$  를 식 (27)과 (28)를 이용하여  $N/m$  블록으로 표본자료의 크기를 변경하면서 평균과 분산(variance)을 산정한 후  $\text{Log}(\text{Var}X^{(m)})$ 와  $\text{Log}(m)$  대하여 도시했을 때의 기울기,  $\beta$ 를 추정한다. 여기서  $N$ 은 전체 자료의 크기이고,  $m$ 은 표본자료의 크기를 의미한다. 이때 기울기,  $\beta$ 와 허스트 지수,  $H$ 는 식 (27)의 관계가 있다.

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i \quad (25)$$

$$k = 1, 2, 3, \dots, [N/m]$$

$$\text{Var}X^{(m)} = \frac{1}{N/m} (X^{(m)}(k) - \bar{X})^2 \quad (26)$$

$$H = 1 - \beta/2 \quad (27)$$

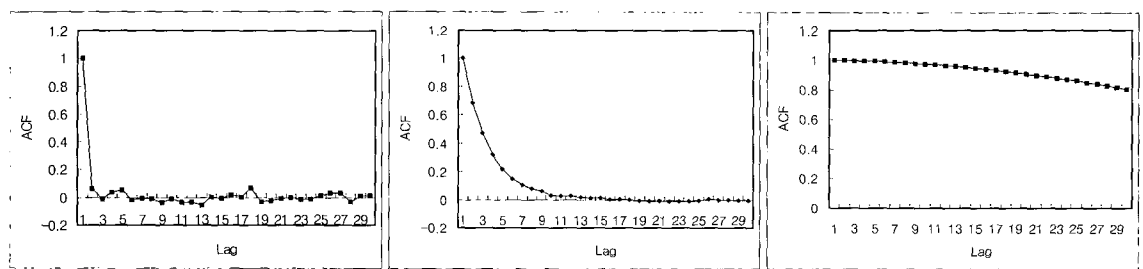
### 3.7 최우도법

최우도법(Maximum Likelihood Estimation, MLE)에 의해 허스트 지수를 추정하기 위하여 S-Plus의 S-MLE 함수(MathSoft, 2000)를 이용하였다. 차분계수,  $d$ 를 S-MLE 함수에 의해 추정하였으며 차분계수,  $d$ 는 허스트 지수와 다음의 관계가 있다.

$$d = H - 0.5 \quad (28)$$

### 4. 적용 및 비교

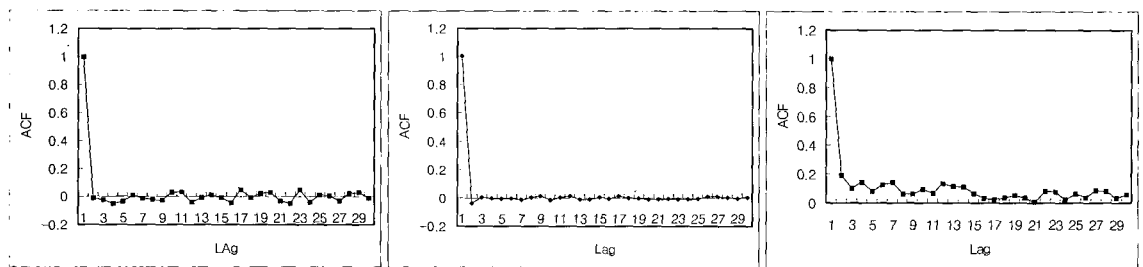
그림 2는 본 논문에서 적용한 카오스 및 추계학적 시계열 자료들의 자기상관도(autocorrelogram, ACF)를 나타낸 것으로써 자기상관도 통해 도시적으로 시계열의 개략적 장단기 기억 특성을 파악할 수 있다. 즉, 어떤 시계열 자료가 자기 상관도의 값이 0을 중심으로 진동을 하면서 수렴을 할 경우에는 단기간의 기억의 특성을 지닌 시계열이라고 볼 수 있으며 반대로 시간이 흐름에 상관없이 진동하지 않을 경우는 장기간의 기억을 가진다고 볼 수 있다. 그림 2를 보면 white noise와 logistic map의 시계열은 단기간 기억의 특성을 지니고 있으며 그 이외의 5가지 시계열 자료는 장기간의 기억 특성을 보이고 있다. 본 절에서는 앞에서 설명한 각각의 방법에 의하여 허스트 지수를 추정하였으며 각각의 결과를 그림과 표로써 나타내었다. 각 방법에 대한 분석과 고찰은 비교 부분에서 다루고자 한다.



(a) White noise

(b) AR(1)

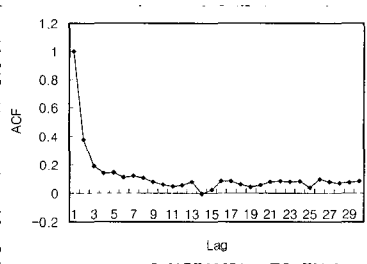
(c) Three torus



(d) Logistic map

(e) TAR(2,1)

(f) FARIMA



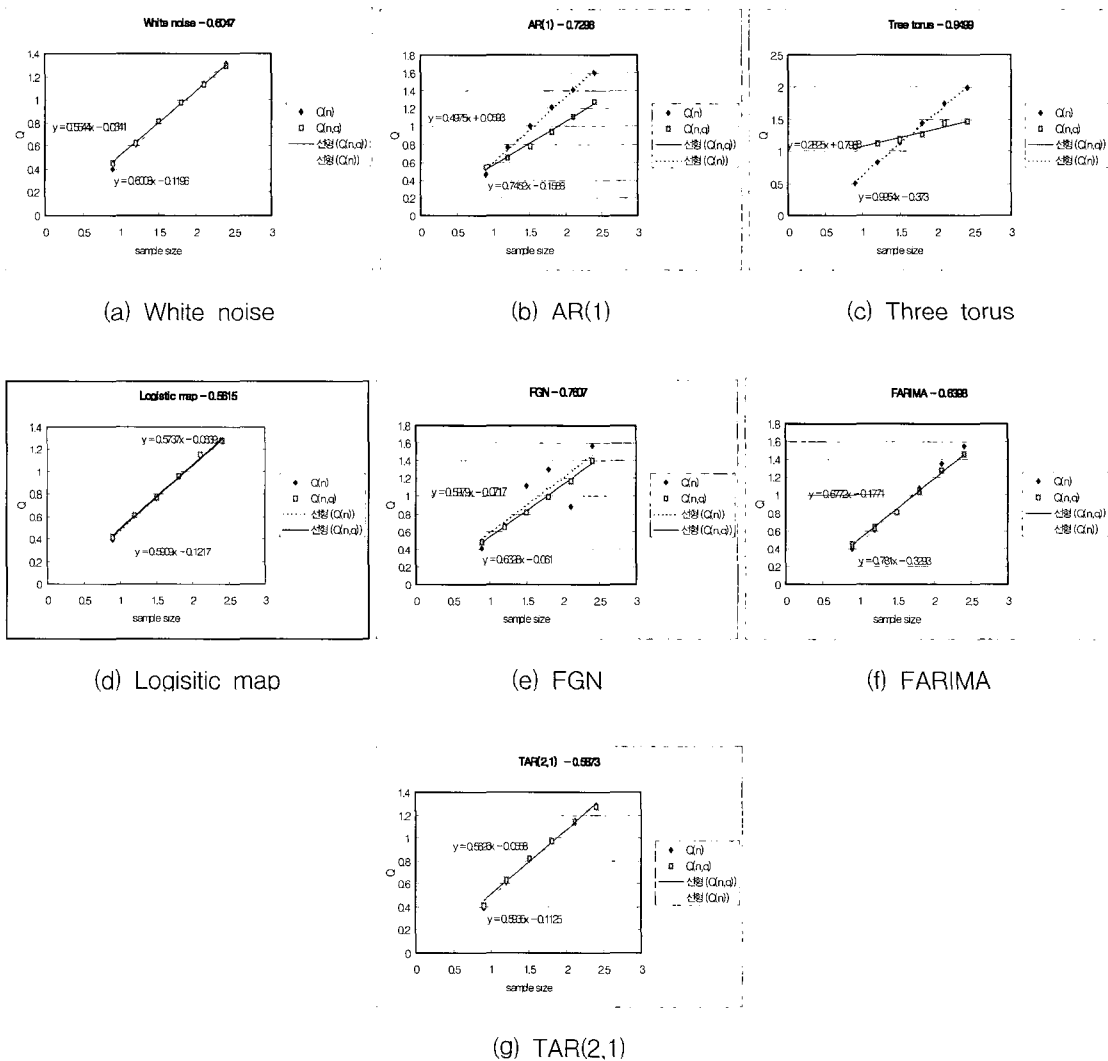
(g) FGn

그림 2. 각 시계열 자료에 대한 자기상관도

#### 4.1 조정용량과 수정조정용량 방법에 의한 허스트 지수 산정

그림 3은 앞 절에서 설명한 보정용량, 조정용량 그리고 수정조정용량에 의해 산정된 허스트 지수 값을 비교

한 것으로 그림을 통해 알 수 있듯이 시계열이 단기간의 기억특성을 지니고 있는 경우에는 조정용량과 수정 조정용량방법에 의해 추정된 기울기가 비슷하지만, 반면에 시계열이 장기간의 기억 특성을 지닌 경우에는 회식의 기울기가 차이가 두드러짐을 확인 할 수 있다.



(a) White noise

(b) AR(1)

(c) Three torus

(d) Logistic map

(e) FGn

(f) FARIMA

(g) TAR(2,1)

그림 3. 조정용량과 수정 조정용량에 의한 허스트 지수 산정

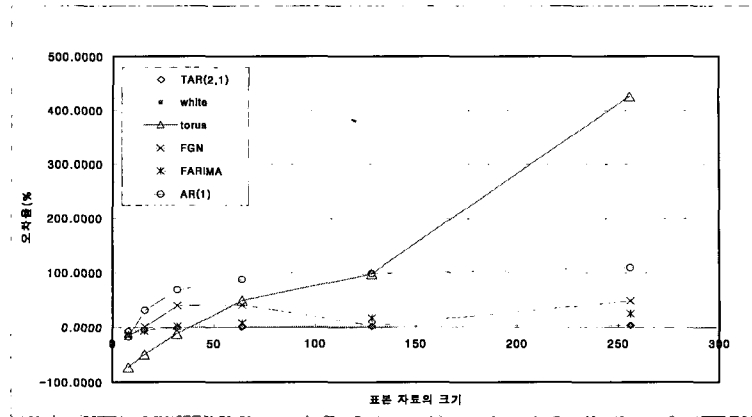
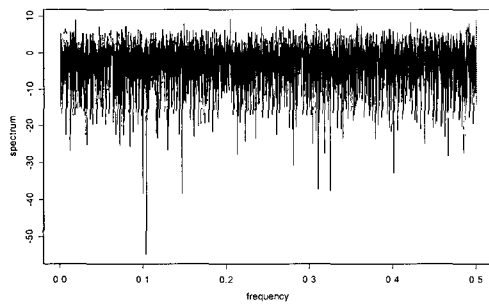


그림 4. 조정용량과 수정 조정용량방법의 오차 비교

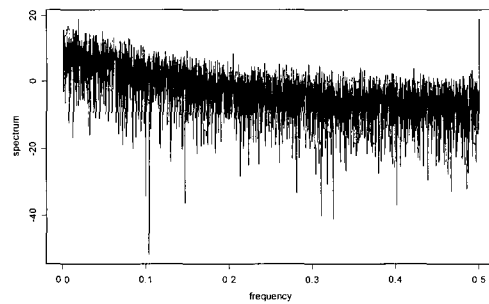
그림 4는 조정용량과 수정 조정용량 방법의 오차를 비교한 것으로서 표본자료의 크기가 작을 때에는 두 방법의 오차가 적었으나 표본자료 크기가 커질수록 오차가 증가함을 알 수 있다. 특히 자기상관성이 강한 시계열 자료일수록 오차가 크게 나타남을 확인할 수 있었다. 조정용량 방법의 경우 white noise가 완전한 무작위적 자료임에도 불구하고 허스트 지수가 0.6047로 추정되었으며 이 허스트 지수를 기준으로 삼는다면 장기간 기억의 특성을 지닌 시계열 자료로 분류될 수 있다. 그러나, 수정조정용량 방법을 적용하여 허스트 지수를 산정하면 0.55로 추정되며 이는 white noise가 단기간 기억 시계열 자료임을 나타내는 것이라고 할 수 있다.

#### 4.2 $1/f$ 파워스펙트럼 밀도 분석에 의한 허스트 지수 산정

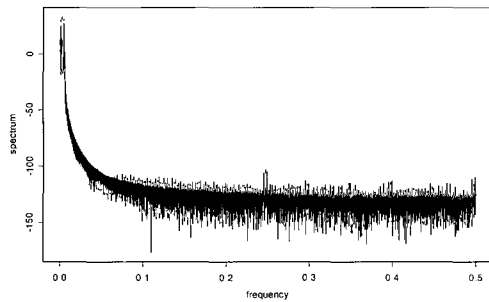
본 절에서는 앞에서 기술한  $1/f$  파워스펙트럼밀도분석을 실시하였으며 이를 통해 먼저 기울기  $\alpha$ 를 산정한 후 식 (20)과 (21)을 이용하여 허스트 지수를 산정하였다. 그림 5는 각 시계열자료의 허스트 분석 결과이며 표 1은  $1/f$  스펙트럼 기울기와 허스트 지수를 나타낸 것이다. 표 1을 통해 알 수 있듯이 이 방법은 비교적 양호한 허스트 지수 산정 결과를 보이고 있으나 Tree torus 자료와 같이 강한 지속성을 지닌 카오스 시스템의 시계열 자료에 적용하는데 문제가 있음을 확인 할 수 있다.



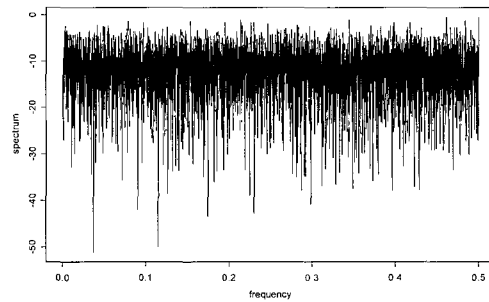
(a) White noise



(b) AR(1)

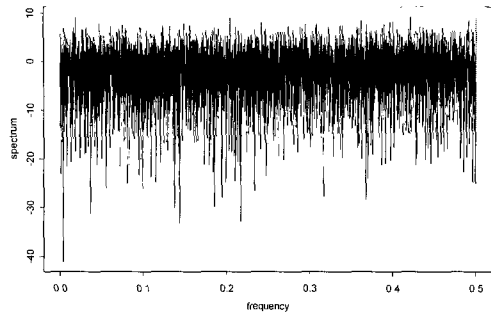


(c) Three torus

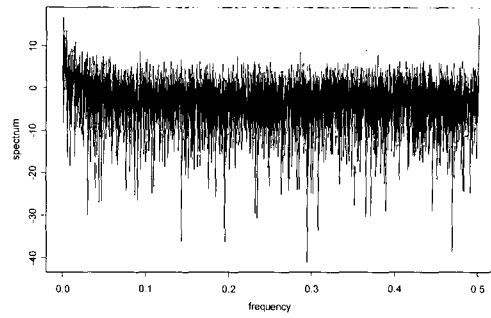


(d) Logistic map

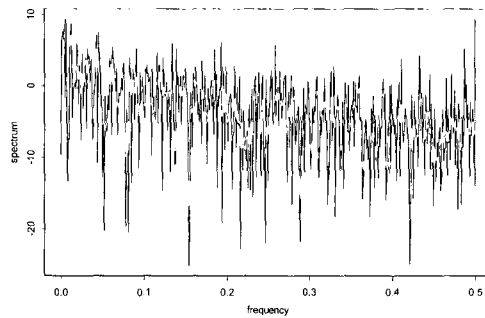




(e) TAR(2,1)



(f) FARIMA



(g) FGN

그림 5.  $1/f$  파워스펙트럼 분석 결과

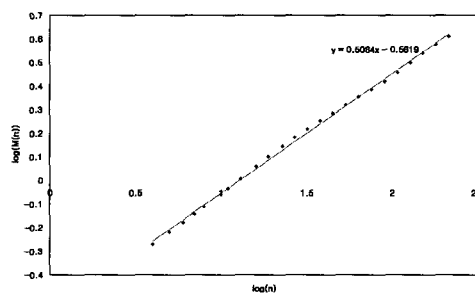
표 1. 기울기( $\alpha$ ) 와 허스트 지수

	White noise	AR(1)	Three torus	Logistic map	TAR(2,1)	FARIMA	FGN
$\alpha$	0.180	0.716	5.287	0.746	0.323	0.519	0.746
허스트 지수	0.590	0.858	$\infty$	0.873	0.661	0.760	0.873

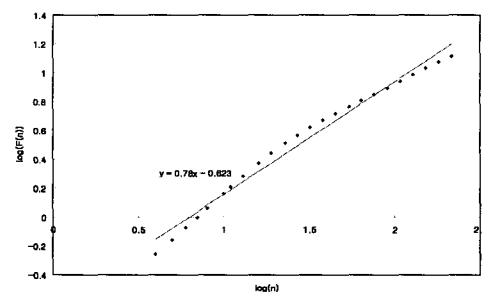
### 4.3 Detrended Fluctuations Analysis에 의한 허스트 지수 산정

본 절에서는 앞에서 설명한 DFA를 이용하여 허스트 지수를 산정하였다. 그림 6은 DFA를 이용하여 허스트

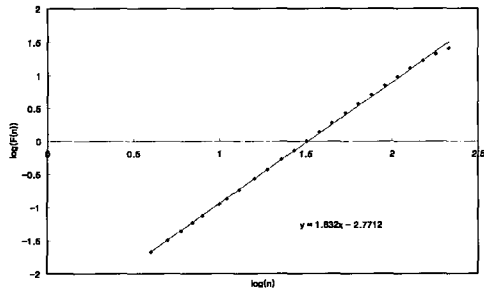
지수를 산정한 결과로서 여기서, 기울기는 규모지수,  $d$ 를 의미하며 식 (22)를 이용하여  $1/f$  파워 스펙트럼 기울기,  $\alpha$ 를 추정된 후 식 (20)과 (21)의 관계를 통하여 허스트 지수를 산정하였다. 표 2는 그 결과를 수록한 것이다.



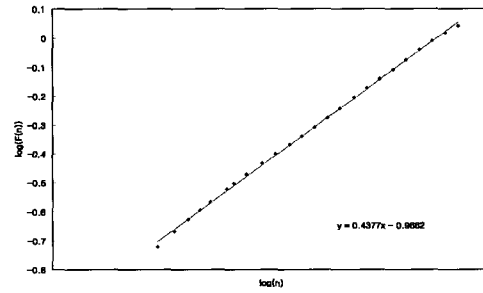
(a) White noise



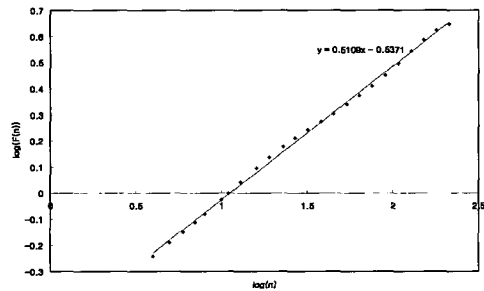
(b) AR(1)



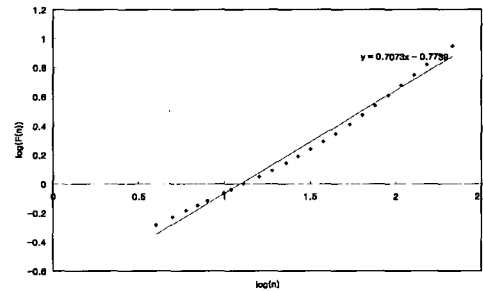
(c) Three torus



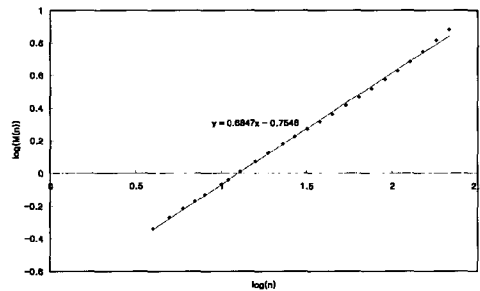
(d) Logistic map



(e) TAR(2,1)



(f) FARIMA



(g) FGN

그림 6. Detrended Fluctuations Analysis 분석 결과

표 2. DFA scaling 지수( $d$ )와 Hurst 지수

	White noise	AR(1)	Three torus	Logistic map	TAR(2,1)	FARIMA	FGN
$d$	0.506	0.78	1.832	0.438	0.510	0.703	0.684
허스트 지수	0.506	0.78	0.832	0.562	0.510	0.703	0.684

주)  $a = 2d - 1$

$a > 1$  일 때  $H = (a - 1)/2$ ,  $a < 1$  일 때  $H = (a + 1)/2$

#### 4.4 Aggregated Variance Time 방법에 의한 허스트 지수 산정

본 절에서는 Aggregated Variance Time(AVT) 방법을 이용하여 각 시계열의 허스트 지수를 산정하였다. 그림 7과 8은 그 분석결과를 나타낸 것으로 그림에서 볼 수 있듯이 자료의 산점도가 표본자료의 크기가 어느

정도 이상이 되면 산개되는 특성을 나타내었으며, 이때 이 산점도를 전체 사용하여 직선 회귀식의 기울기를 구한 후 식 (29)를 이용하여 허스트 지수를 추정하면 잘못된 허스트 지수 값을 추정함을 확인하였다. 그러나, 임의로 산개되는 곳의 자료를 제외한 후 산점도의 직선 기울기를 구하면 기지의 알고 있는 허스트 지수의 값을 추정할 수 있었다.

(1) 산개되는 곳의 자료를 포함할 경우

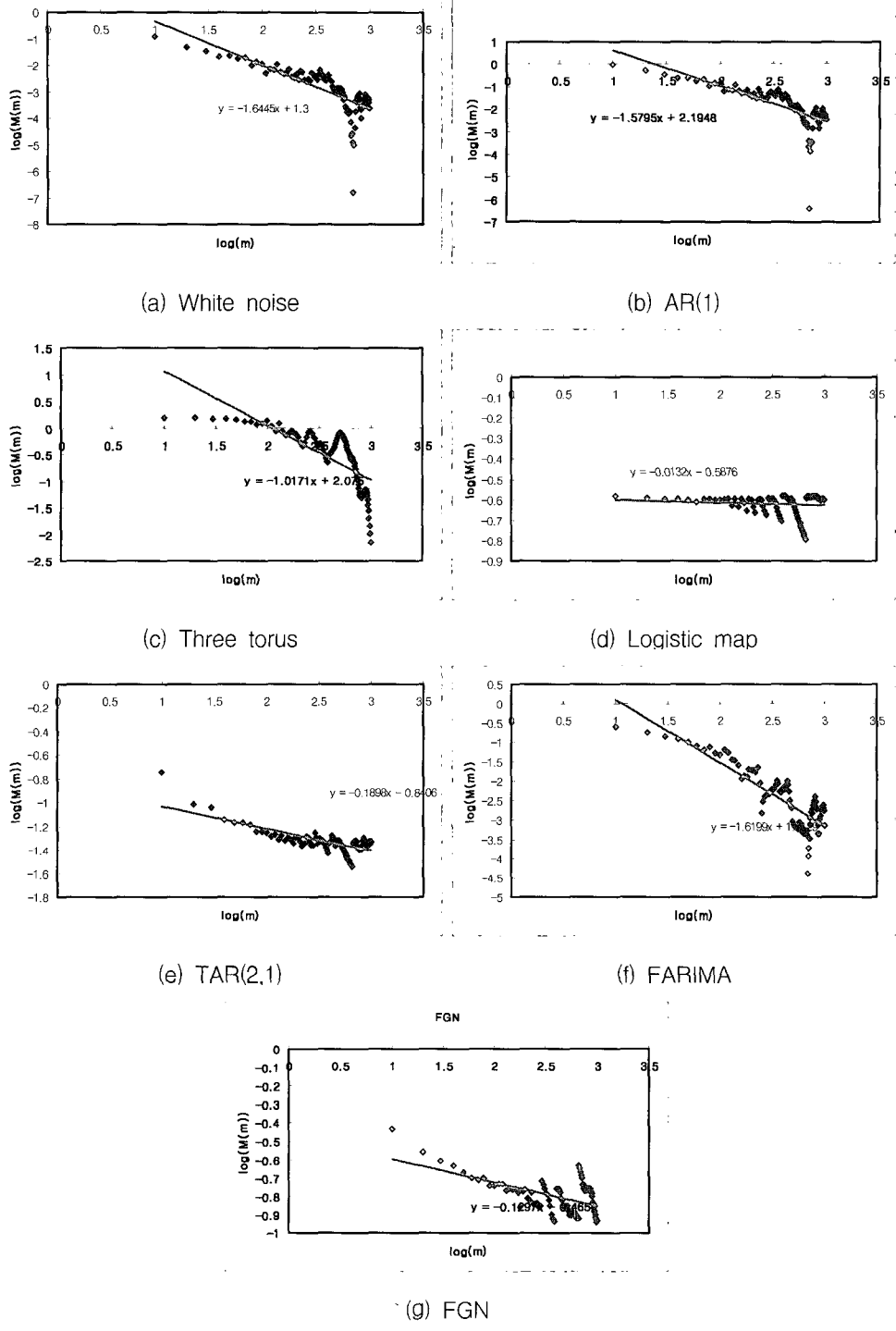


그림 7. Aggregated Variance Time 방법의 분석 결과

표 3. 기울기 ( $\beta$ )와 허스트 지수

	White noise	AR(1)	Three torus	Logistic map	TAR(2,1)	FARIMA	FGN
$\beta$	1.645	1.580	1.017	0.013	0.189	1.619	0.129
허스트 지수	0.170	0.210	0.491	0.934	0.905	0.190	0.930

주)  $H = 1 - \beta/2$

(2) 산개되는 곳의 자료를 제외 할 경우

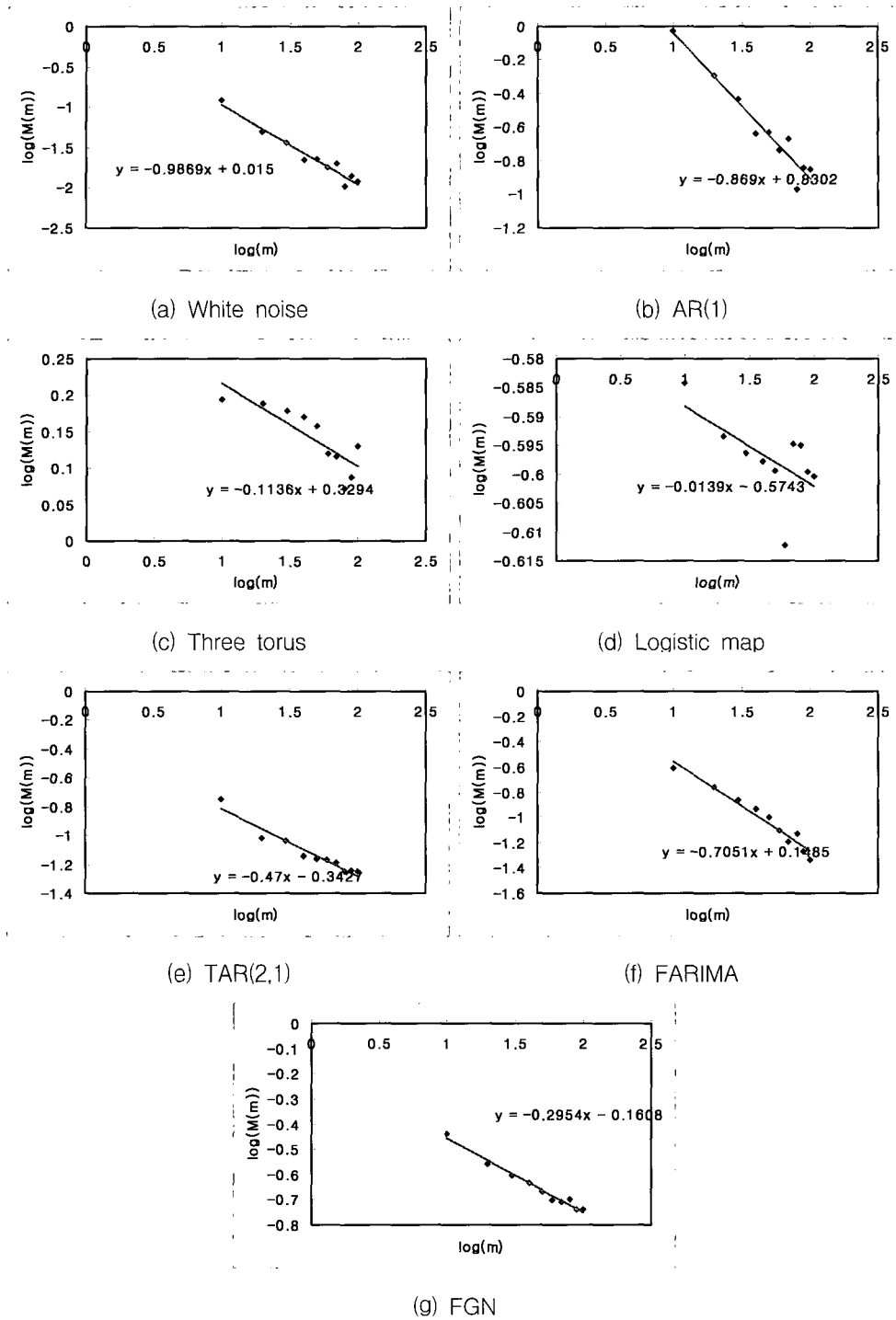


그림 8. Aggregated Variance Time 방법의 분석 결과

표 4. 기울기 ( $\beta$ )와 허스트 지수

	White noise	AR(1)	Three torus	Logistic map	TAR(2,1)	FARIMA	FGN
$\beta$	0.987	0.869	0.114	0.014	0.470	0.705	0.295
허스트 지수	0.506	0.566	0.943	0.993	0.765	0.648	0.852

주)  $H = 1 - \beta/2$

#### 4.5 최우도 추정법에 의한 허스트 지수 산정

본 절에서는 통계 패키지인 S-PLUS 내장되어 있는 S-MLE 함수를 이용하여 허스트 지수를 산정하였으며 그 결과를 표 5에 수록하였다.

#### 4.6 비교 및 고찰

앞에서 설명한 7가지의 카오스 및 추계학적 시계열 자료에 대한 각각의 허스트 지수 추정 방법들의 결과를 비교해 보면 보정용량과 조정용량 방법은 white noise 와 같은 단기간 기억 시계열자료의 허스트 지수가 0.6 이상으로 과대하게 추정되었으며 반면에 FGN, FARIMA, three torus 등과 같이 장기간 기억 시계열 자료는 비교적 정확한 허스트 지수 값을 산정함을 확인 할 수 있었다. 또한, 수정조정용량 방법은 단기간 기억 시계열자료의 허스트 지수는 적절하게 추정하였지만, 장기간 기억 시계열 자료에는 적합하지 않음을 확인하였다. AVT 방

법은 그림 7과 8에서 볼 수 있듯이 자료의 산점도가 표본자료의 크기가 어느 정도 이상이 되면 산개되는 특성을 나타내었으며, 이때 이 산점도를 전체 사용하여 직선 회귀식의 기울기를 구한 후 식 (29)를 이용하여 허스트 지수를 추정하면 왜곡된 허스트 지수 값을 추정함을 확인하였다. 그러나, 임의로 산개되는 곳의 자료를 제외한 후 산점도의 직선 기울기를 구하면 기지의 살고 있는 허스트 지수의 값을 추정할 수 있었다.

1/f 파워스펙트럼밀도 방법은 6가지의 자료에 대하여 적절한 허스트 지수를 추정하였지만 three torus의 경우 기울기의 산정이 곤란함을 보였다.

DFA 방법과 최우도법은 단기간 기억 시계열 자료와 장기간 기억 시계열 자료 자료에 모두 적합한 허스트 지수를 산정하였으며 특히 DFA 방법의 경우 최우도법에 비해 그 식이 간단하며 계산이 쉽다는 장점이 있다. 그림 9는 각각의 허스트 지수 추정방법에 의한 결과를 비교하여 나타낸 것이다.

표 5. 기울기 (d)와 허스트 지수

	White noise	AR(1)	Three torus	Logistic map	TAR(2,1)	FARIMA	FGN
d	0.0005	0.0004	0.364	0.023	0.199	0.326	0.193
허스트 지수	0.5005	0.5004	0.864	0.502	0.699	0.826	0.693

주)  $H = d+0.5$

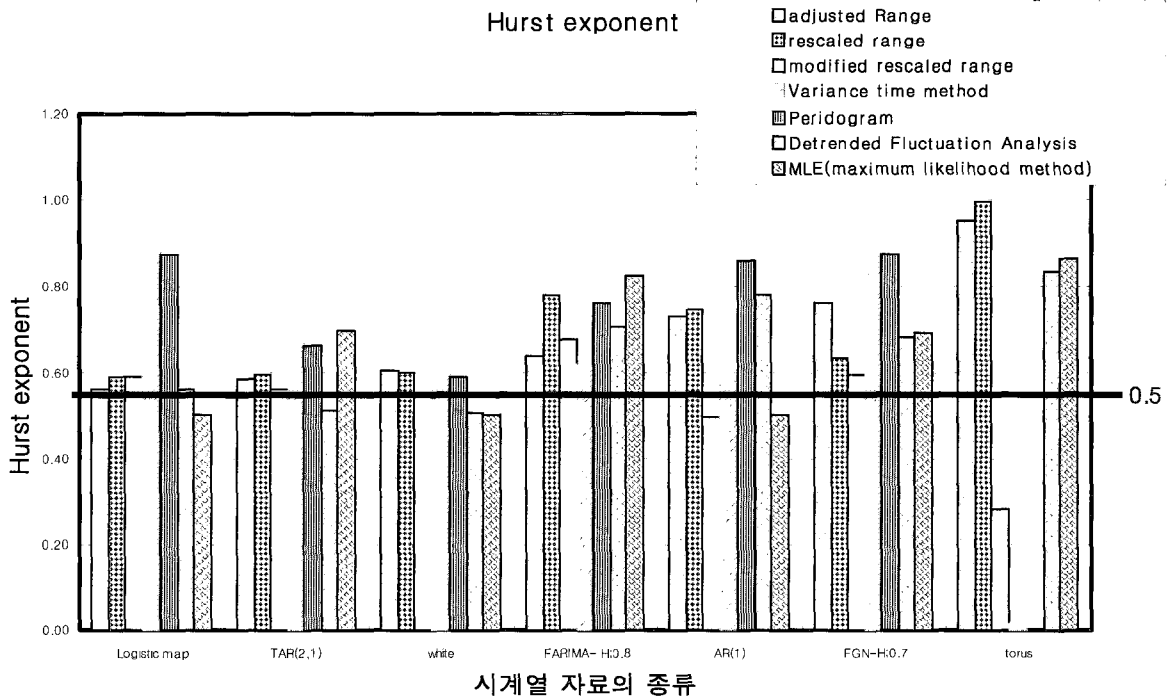


그림 9. 각 방법에 의해 추정된 허스트 지수의 비교

## 5. 결 론

본 연구에서는 수문학에서 주로 이용하고 있는 보정 용량, 조정용량, 수정조정용량 방법 이외에 생리학 분야와 전자 분야 등에서 이용되고 있는  $1/f$  파워 스펙트럼 밀도 분석, DFA 방법, AVT 방법, 최우도법 등을 이용하여 허스트 지수를 산정하여 보았다. 그 결과 각각의 방법들이 모두 장점들을 가지고 있기는 하지만 그 중에서 DFA와 최우도법은 단기간 기억 시계열자료와 장기간 기억 시계열자료, 두 가지 경우에 대하여 허스트 지수 추정에 가장 우수함을 확인 할 수 있었다. 그러므로 이 방법들을 수문자료에 적용한다면 기존에 수문학에서 사용하고 있는 방법보다는 장·단기간 기억 시계열자료에 대하여 더 정확한 허스트 지수를 추정할 수 있을 것이다. 본 논문에서의 결과를 요약하면 다음과 같다.

- (1) 보정용량과 조정용량 그리고 수정조정용량 방법은 각각 장기간 기억시계열 자료와 단기간 기억 시계열 자료의 허스트 지수 추정에만 적절함을 확인하였다. 또한, 조정용량 방법, 수정조정용량 방법의 오차를 비교해 보면 표본자료의 크기가 작을 때에는 두 방법의 오차가 적었으나 표본자료의 크기가 커질수록 오차가 증가함을 확인 하였으며 특히, 자기상관성이 강한 시계열 자료일수록 오차가 크게 나타남을 확인 하였다.
- (2) AVT 방법은 시계열 자료의 산점도가 표본자료의 크기가 어느 정도 이상이 되면 산개되는 특성을 나타냈으며 이로 인해 회귀직선의 기울기,  $\beta$  을 추정할 때 큰 오차가 발생할 가능성을 확인하였다.
- (3)  $1/f$  파워스펙트럼 밀도 방법은 거의 모든 시계열 자료에 대하여 적절한 허스트 지수를 추정할 수 있었지만 three torus의 경우처럼 자기상관성이 거의 1에 가까운 시계열 자료의 경우에는 적합하지 않음을 확인하였다.
- (4) DFA와 최우도법은 단기간 기억 시계열자료와 장기간 기억 시계열자료, 두 가지 경우에 대하여 허스트 지수 추정에 가장 우수함을 확인 할 수 있었다

## 감사의 글

본 연구는 건설교통부 건설핵심기술연구개발사업인 「도시홍수재해관리기술 연구단」의 연구비 지원에 의해 수행되었으며 이에 감사드립니다.

## 참 고 문 헌

- 김형수, 박진욱, 김중훈 (1998). "수문 시계열의 Hurst 현상.", *대한토목학회논문집*, 대한토목학회, 제 18 권, 제II-6호, pp. 571-582.
- Ausloos, M. A. (2000). "Statistical Physics in foreign exchange currency and stock markets.", *Physica A*, Vol. 285, pp. 48-65.
- Bigger, T., Richard, C., Steinman, A.B., Linda, M., Joseph, L., and Richard, J. (1996). "Power law behavior of RR-interval variability in healthy middle-aged persons, patients with recent acute myocardial infarction, and patients with heart transplants.", *Circulation* Vol. 93, pp. 2142-2151.
- Goldberger, A.L., Amaral, L.A.N., Hausdorff, J.M., Ivanov, P.C., Peng, C.K., and Stanley, H.E. (2002). "Fractal dynamics in physiology; Alterations with disease and aging.", *PANS*, Vol. 99., pp. 2466-2472.
- Herman, P. (1999). "Physiological time series; distinguishing fractal noises from motions.", *European Journal of Physiology*, Vol. 439, pp. 403-415.
- Hipel, K.W. and Mcleod, A.I. (1994). "Time series modelling of water resources and environmental system.", Elsevier Science, Amsterdam.
- Hurst, H. (1951). "Long-term storage capacity of reservoirs.", *Transcation of the American Society of Civil engineer*, Vol. 116, pp. 770-799.
- Kantelhardt, J.W., Koscielny, B.E., Rego, H.A., Havlin, S., and Bunde, A. (2001). "Detecting long range correlations with Detrended Fluctuation Analysis.", *Physica A* Vol. 295, pp. 441.
- Kashyap, R.L. and Rao, A.R. (1976). "Modeling river flows.", *Dynamic stochastic models from empirical data*, Academic, New York, pp. 238-281.
- Kendzioriski, C.M. (1999). "Evaluation maximum likelihood estimation methods to determine the Hurst coefficient", *Physica A* Vol. 273 pp. 439-451.
- Lo, A.W. (1991). "Long term memory in stock market prices.", *Econometrica*, Vol. 59, pp. 1279-1313.
- Mandelbrot, B. and Wallis, J. (1969). "Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence.", *Water Resour. Res.*, Vol. 5, pp. 967-988.

- Nakamura, Y., Yamamoto, Y., and Muraoka, I. (1993). "Autonomic control of heart rate during physical exercise and fractal dimension of heart rate variability." *J Appl Physiol*, Vol. 74, pp. 875-881.
- Pallikari, F. (1999). "Rescaled Range Analysis of Random Events.", *Journal of scientific Exploration*, Vol. 13, pp. 25-40.
- Philips, P. (1987). "Time series regression with a unit root." *Econometrica*, Vol. 55, pp. 703-705.
- Rao, A.R. and Bhattacharya, D. (1999). "Hypothesis testing for long term memory in hydrologic series.", *Journal of Hydrology*, Vol. 216, pp. 183-196.
- Rao, A.R. and Bhattacharya, D. (2001). "Effect of short-term memory on hurst phenomenon.", *Journal of Hydrologic Engineering*. March/April, pp. 125-131.
- MathSoft (2000). S-Plus2000 ; Guide to statistical and mathematical analysis. StatSci Division, Seattle, Washinton.
- Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1995). "Applied modelling of hydrologic time series" WRP.
- Taqqu, M.S., and Teverovsky, V. (1995). "Estimation for long range dependence; an empirical study.", *Fractals*, Vol. 3(4), pp. 785-798.
- Teverovsky, V., Taqqu, M.S., and Willinger, W. (1998). "Acritical look at Lo's modified R/S statistic.", *Journal of Statistical Planning and Inference*.
- Tong, H.(1990)."Non-linear Time Series; ADynamical System Approach." Clarendon press oxford.
- Zafer, S. and Sirin, T. (1999). "Traffic engineering for multimedia networks; data collection onthe internet, extensions to wireless.", NSF industry/ University co-operative research center for digital video & media.
- Zhao, Y. (1998). "Self-similarity in high performance network analysis.", University of missouri-Columbia.

(논문번호:04-93/접수:2004.10.04/심사완료:2004.11.03)