

## SSA를 이용한 일 단위 물수요량 단기 예측에 관한 연구

### A Study of Short Term Forecasting of Daily Water Demand Using SSA

권현한<sup>1</sup> · 문영일<sup>2,\*</sup>

Kwon, Hyun-Han<sup>1</sup> · Moon, Young-Il<sup>2,\*</sup>

1 Columbia University in the City of N.Y., Dept. of Earth & Environmental Engineering, Postdoctoral Research Associate

2 서울시립대학교 토목공학과 부교수

(2004년 9월 13일 논문접수; 2004년 12월 1일 최종수정논문 채택)

#### Abstract

The trends and seasonalities of most time series have a large variability. The result of the Singular Spectrum Analysis(SSA) processing is a decomposition of the time series into several components, which can often be identified as trends, seasonalities and other oscillatory series, or noise components. Generally, forecasting by the SSA method should be applied to time series governed (may be approximately) by linear recurrent formulae(LRF). This study examined forecasting ability of SSA-LRF model. These methods are applied to daily water demand data. These models indicate that most cases have good ability of forecasting to some extent by considering statistical and visual assessment, in particular forecasting validity shows good results during 15 days.

**Key words:** Water demand forecasting, singular spectrum analysis, linear recurrent formulae

**주제어:** 물수요량 예측, Singular spectrum analysis, linear recurrent formulae

#### 1. 서론

경제규모의 확대, 도시로의 인구집중 등으로 물 수요량이 증가되고 장기적으로는 절대량 부족과 수자원의 합리적 이용, 상수도 시설의 계획적 관리가 되지 않아 문제로 지적되고 있는 현실이다.

상수도는 현대 생활의 기반시설로 인간의 생존뿐만 아니라 도시의 경제 및 사회활동을 위한 가장 기

본적인 시설이다. 그러나 현재까지의 상수도는 보다 더 많은 사람들의 풍요로운 생활을 유지하기 위해 상수도 보급률 향상과 증가되는 수요량에 대한 기본적인 대처 등이 주된 목적이 되어 양적인 정비로 일관되어진 것이 사실이다. 그러나 현재 우리나라의 상수도 보급률은 거의 100%에 육박하고 있으며 최근에는 양적인 개선 이외에 배수시스템의 질적 향상을 위한 관심이 증가하고 있다. 이러한 관점에서 원활한 물공급을 위한 기초자료로써 물수요량의 시간적, 공간적

\*Corresponding author Tel: +82-2-2210-2764, E-mail: ymoon@uos.ac.kr (Moon, Y.I.)

변동특성을 해석하는 것은 상수도 계획과 설계에 있어서 매우 중요한 요소 중 하나이다.

공학자들이 관심을 가지는 통계학적인 추정 문제는 주로 공간적인 내삽, 시공간적 경향의 평가, 여러 매개변수들 사이에서 함수적 상관관계, 그리고 유사한 특성을 가진 추계학적 시계열자료에 대한 예측 등이 될 수 있다. 물 사용량은 대표적인 시계열 자료로서 각각의 사용자가 요구하는 양의 총합으로 정의할 수 있으며 물수요량의 변동은 사용자의 일상적인 생활습관과 온도, 습도 등의 외부요인에 영향을 받는다.

물수요량 및 수자원과 관련해서 시계열을 모의하거나 예측하기 위한 방법으로 많은 추계학적 모형들이 개발되고 있으며 다수의 연구자들에 의해서 모형의 보완과 검토가 이루어지고 있다. 그러나 불규칙한 특성을 갖는 시계열에 대해서 통계학적인 기법을 통해서 정확하게 실제현상을 예측하는데 무리가 따르며 현재까지도 매우 어려운 문제로 인식되고 있다. 이러한 시계열을 모의하기 위해 어떤 사상의 값과 이전시점 사상의 값 사이에 상관성이 존재하는 사실로 기술할 수 있는 자기회귀과정(autoregressive process) 또는 Markov 과정(Markov process)이 대표적인 방법으로 주로 사용되고 있다.

최근에 시계열분석에서 주된 관심사는 비선형의 특성을 갖는 시계열의 보다 합리적인 방법을 통한 예측이라 할 수 있다. 유량이나 물수요량 등의 시계열 분석에서 단기 예측은 다양한 방법들에 의해서 수행되고 있으며 수자원계획 및 조정을 위해서 주로 이용되고 있다. 최근에 Abarbanel 등(1996), Kember 등(1993), Smith(1991), Yakowitz와 Karlsson(1987)은 비매개변수적 회귀모형을 이용한 시계열의 예측기법을 제안하였다. 문영일(1997)은 시계열자료의 비선형 상관관계를 검토하기 위해서 Mutual Information(MI)을 이용하여 예측하였다. 박무종과 윤용남(1989)은 Multiplicative ARIMA 모형을 사용하여 주기성과 경향성을 가지는 월유량계열을 예측하였으며 안상진과 이재경(2000)은 ARIMA 모형과 상태공간모형을 이론적으로 하여 계절별 월 유출량을 모의 하였다. 문영일(2000)은 지역가중다항식을 이용하여 시계열 예측을 위한 모형을 구성하였으며 김형수 등(2002)은 퍼지이론을 이용하여 강우량을 예측하였다.

상하수도 분야에서 물수요량의 변동을 예측하기 위한 많은 노력이 이루어져 왔으며 최재호 등(1999)은 도시지역의 재개발에 따른 장기 물수요량 예측 모형을 개발하였으며 김신걸 등(1999)은 지역별로 시간 변동에 따른 물수요량을 시간단위로 예측하였다. 주충남 등(1999a)은 배수관리를 위하여 인공신경망을 이용한 단기 물수요 예측모형을 개발하였고, 주충남 등(1999b)은 물 수요예측 모형을 이용한 배수불력의 최적화를 시도하였으며 최태용 등(1999)은 회귀분석 모형을 이용하여 서울시에 물수요량을 추정하였다. 최근에 유명진 등(2004)은 선형 및 비선형 모형을 이용하여 시간별 물수요량을 예측하였다.

물수요량은 시간, 일, 월, 년 단위로 다른 특성을 가지고 있으며 시간단위의 물수요량은 2중 첨두를 갖는 일정한 경향성을 가지고 있으며 ARMA, ARIMA, Fourier 급수 등과 같은 매개변수적 모형을 통해 비교적 용이하게 주성분을 추출할 수 있으며 예측이 가능하다. 그러나 일단위 수요량은 일정한 경향성을 추출하기가 곤란하며 따라서, 대부분 일단위 분석은 온도 등의 외부인자와 결합하여 시계열을 분석하고 있다. 이러한 문제점과 함께 기존의 시계열 모형은 대부분 매개변수적 모형으로서 매개변수 추정 및 모형 선택 등의 많은 어려움과 특히 시계열 예측 능력에 있어서 많은 문제점을 내재하고 있다.

본 논문에서는 일단위 수요량과 같은 불규칙한 성

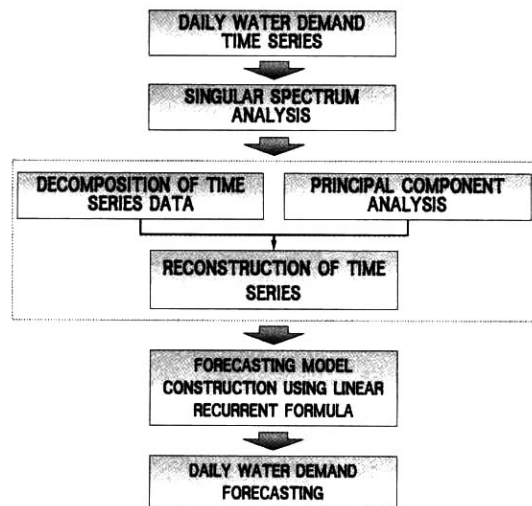


Fig. 1. Forecasting of water demand using SSA.

분으로 인해 해석상의 어려움이 따르는 문제점을 개선하기 위해서 시계열에 내재해 있는 구성성분을 추출할 수 있고 비매개변수적 방법인 Singular Spectrum Analysis(SSA)기법을 이용하여 일단위 수요량을 예측하고자 한다. 따라서 본 연구에서는 주성분분석과 상관성 분석에 주로 적용되고 있는 SSA와 시계열 예측을 위해서 Linear Recurrence Formula(LRF)를 결합한 예측 모형을 일단위의 물수요량 시계열 자료를 대상으로 적용성 및 타당성을 검토해 보았다.

## 2. 본 론

### 2.1. Singular Spectrum Analysis

SSA는 시계열상에서 주기적 변동과 관련된 지연 상관관계를 고려하여 상관 행렬을 만들고 이를 고유값 문제로 그 해를 구하여, 원 시계열에 내재되어 있는 서로 독립적인 모드의 주기적 변동 및 변화 경향들을 지배적인 모드 순으로 추출하여 내는 분석 방법이다. 독립적인 변동 특성을 지배적인 모드 순으로 추출하여 내는 고유 분석의 하나라는 점에서 경험적 직교함수(Empirical Orthogonal Functions: EOF)와 유사하나, EOF가 공간내 각 지점과의 동시상관에 분석 기능이 집중되어 있는 반면에 SSA는 어느 한 지점에서 지연 상관까지 고려한, 시간에 따른 변화 경향 및 주기적 변동에 분석기능이 집중되어 있다는 점에서 차이가 있다.

SSA는 Principal Component Analysis(PCA)에서 파생된 기법으로, SSA기법의 기본적인 배경은 원자료의 잡음들로부터 비조화성분을 제거함으로써 중요한 성분들(principal component: PC)만 분리하여 원자료에서 얻지 못하는 주기성과 경향성(trend)을 쉽게 파악할 수 있는 방법이다(Moon과 Lall, 1996). 즉, SSA는 주어진 임의의 시계열자료를 여러 개의 조화성분(대략 자료 개수의  $1/5 \sim 1/3$ )으로 나누어 자료가 갖고 있는 주기성과 경향성을 파악할 수 있다. 여기서 파악된 자료의 경향성을 통하여 앞으로 발생할 사상에 대해 예측하는 능력 또한 다른 기법들에 비해 우수하다.

Karhunen-Loeve(K-L) Expansion이라고도 알려진 SSA기법은 PCA기법에서 단변량을 적용시켜 사용하는데 최근의 Vaurard와 Ghil(1989), Rasmusson 등

(1990), Vautard 등(1992)에 의해 특히 기상학 분야의 연구에 많이 사용되는 기법이다. Colebrook(1989)에 의해 해양생물학에 처음으로 사용되었으며, 80년대의 Broomhead와 King(1986), Fraedrich(1986)을 거쳐 90년대의 Rasmusson(1990), Vautard와 Ghil(1989), Moon과 Lall(1996), 문영일 등(1999)이 엘니뇨현상을 연구하기 위해 사용하였다.

이 기법은 EOF에 기초를 두고 이를 단변량의 시계열에 적용시킨 것이며, 시간적 분포가 길고 표본율(sampling rate)이 높을 때 특히 사용이 용이하다(Mo와 Ghil(1992), Vautard 등(1992)). 원자료에 대해 필터링하여 잡음을 제거하는 연구는 Vautard 등(1992)에 의해 정리된 바 있다.

SSA는 다른 신호(signal)에 비하여 특수한 빈도의 구간(band)의 신호를 선택한다거나 어떤 지배적인 변동성을 파악할 때 이용된다. Vautard 등(1992)의 연구에 의하면 재구성성분은 주요성분의 최적의 선형조합임을 알 수 있다. 또한 SSA를 사용하여 비조화성분을 제거함으로써 원자료에서 얻지 못하는 진동의 모드와 경향성을 쉽게 파악할 수 있다(Moon과 Lall, 1996).

시간영역에서의 EOF를 알아보기 전에 먼저 공간적 EOF에 대해 살펴보면, 이는 지수  $i$ 와  $j$ 가  $-1 \leq i \leq M$ 이고,  $1 \leq j \leq N$ 이며, 각각 공간과 시간에 대한 방향을 나타낼 때, History of discrete field  $x_{i,j}$ 를 그것의 EOF와 PC의 집합들에 대입하게 된다. 이 때 SSA는 공간적 방향인  $i$ 가 시간지체로 대체, 즉,  $x_{i,j} = x_{i+j}$ 로 대체되고,  $M$ 이 지체의 수가 된다. 그러므로 수식상으로는 똑같은 형식을 취하게 되지만 시간적 주요성분(Temporal Principal Component: T-PC)은  $M-1$ 만큼 원시계열보다 적게 된다(Keppen 등, 1992). 이 때  $M$ 을 선택하는 것은 얻고자 하는 정보의 양과 필요로 하는 통계학적 신뢰도의 정도에 따라 좌우된다.

SSA를 통하여 EOF와 PC들은 자료에 대한 분산을 나타내게 되는데, 이러한 각각의 EOF와 PC들은 재구성되어  $N$  길이의 시계열을 가지게 된다. 그러므로 재구성된 성분을 전부 합하게 되면 원 시계열의 값을 얻는 것이다. 이렇게 SSA를 거쳐 나타난 EOF에서의 고유치는 이에 상응하는 PC들의 분산값을 나타내며, 이 고유치의 제곱근 값은 Singular Value(SV)를 나타

낸다.

만일 주어진 시계열의 값이  $x_i (1 \leq i \leq N)$ 라고 하면 아래와 같이 표현할 수 있다.

$$x_{i+(j-1)\tau} = \sum_{k=1}^M a_i^k E_j^k, 1 \leq j \leq M, 1 \leq i \leq N-M+1 \quad (1)$$

여기서  $a_i^k$ 는 주요성분이라고 불리는 투영계수이다.  $E^k$ 는 EOF ( $1 \leq k \leq M$ )이며,  $M$ 은  $M$ 차의 공간 (smoothing window의 역할)을 의미하고,  $\tau$ 는 추출비 (sampling rate)이다.  $k$ 번째 주요성분은  $k$ 번째 경험적 직교함수 (EOF)에 대하여 시계열의 정투영 (orthogonal projection)으로 정의된다.

$$a_i^k = \sum_{k=1}^M x_{i+(j-1)\tau+l} E_j^k, 0 \leq i \leq N-M \quad (2)$$

따라서 고유치(eigenvalue)의 크기에 따라 자료의 비조화성분을 제거하여 필요한 조화성분을 분리할 수 있고 아래의 식과 같이 원 자료를 재구성할 수 있으며 이를 재구성성분(Reconstructed Component, RC)이라 한다. 따라서  $M$ 개의 재구성성분들을 전부 합하면 원 시계열을 나타낸다.

$$(R_{,x})_i = \frac{1}{i} \sum_{j=1}^i \sum_{k \in A} a_{i-j}^k E_j^k, 1 \leq i \leq M-1 \quad (3)$$

$$(R_{,x})_i = \frac{1}{M} \sum_{j=1}^M \sum_{k \in A} a_{i-j}^k E_j^k, M \leq i \leq N-M+1 \quad (4)$$

$$(R_{,x})_i = \frac{1}{N-i+1} \sum_{j=1-N+M}^M \sum_{k \in A} a_{i-j}^k E_j^k, N-M+2 \leq i \leq N \quad (5)$$

**Fig. 2**는 SSA를 이용한 시계열의 주성분 분석을 이용한 시계열의 분해 과정을 주기성이 강한 자료에 대해서 예를 들어 설명하였다. 첫 번째 그림은 시계열의 원자료를 나타낸다. 그림에서 보듯이 직감적으로 일정하게 증가하는 성분과 주기적으로 변동하는 성분을 확인할 수 있다. 이를 SSA를 이용해서 경향성분, 계절성분 등으로 분해할 수 있으며 가운데 그림이 증

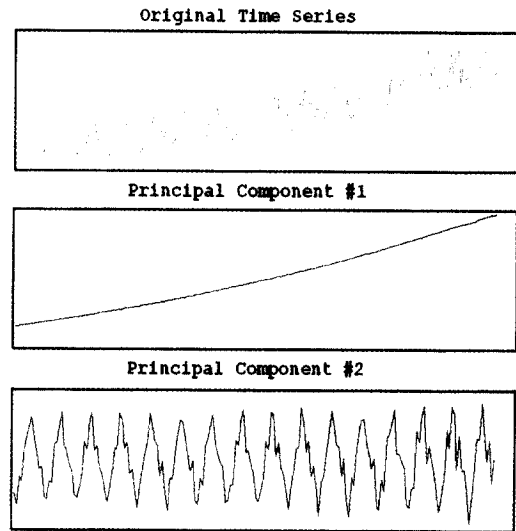


Fig. 2. Principal component analysis using SSA.

### Original Time Series & Reconstruct Time Series

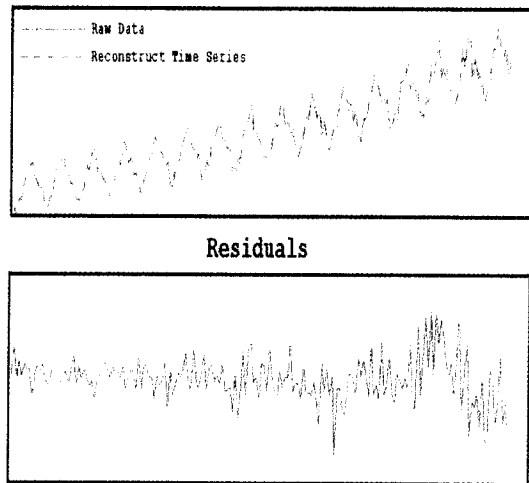


Fig. 3. Reconstructed time series using SSA.

가 경향을 나타내는 가장 큰 주성분을 나타내며 (98.939%) 마지막 그림은 주기적으로 진동하는 2번째 주성분의 합(0.99%)를 나타낸다. **Fig. 3**은 주성분 분석을 통해서 추출된 2가지 주성분을 가지고 재구성한 시계열과 실측자료와의 잔차를 나타내며 실측자료와 거의 일치하는 결과를 얻을 수 있으며 SSA 기법의 주성분 분석의 타당성을 나타낸다 할 수 있다.

## 2.2. Linear Recurrent Formula를 이용한 시계열 예측

Linear Recurrent Formula와 특성 다항식(characteristic polynomial)의 이론은 Gelfond(1967), Buchstaber(1994), Gantmacher(1998)에 의해서 제안되고 수정되었다.

SSA를 이용한 시계열 분해의 중요한 성질은 원자료  $x_t$ 가 차수(order)  $d$ 와 계수  $a_1, \dots, a_d$ 를 가지는 Linear Recurrent Formula(LRF)의 형태를 갖는 다음식 (6)을 만족한다면

$$x_t = a_1 x_{t-1} + \dots + a_d x_{t-d} \quad (6)$$

어떤  $N$ 과  $M$ 에 대해서 Trajectory 행렬(행렬의 대각선을 따라  $i + j$ 의 값이 일정한 행렬)  $X$ 의 Singular Value Decomposition(SVD)에서 차수  $d$ 에 해당하는 영(zero)이 아닌 SV가 존재하게 된다는 것이다. 그러므로  $M > d$ 가 만족한다면 단지 차원  $d$ 를 갖는  $X_d$  행렬만으로 시계열을 재구성할 수 있다. 시계열  $x_t$ 가 LRF를 만족한다는 것은 지수함수(exponential), 다항식(polynomial), 조화성분(harmonic) 등의 곱으로서 시계열의 대표성을 표현할 수 있다는 것을 의미한다. 즉

$$x_t = \sum_{k=1}^q \alpha_k(t) e^{\mu_k t} \sin(2\pi \omega_k t + \phi_k) \quad (7)$$

여기서,  $\alpha_k(t)$ 는 시간  $t$ 에서의 다항식을 의미하며  $\mu_k, \omega_k$ 와  $\phi_k$ 는 매개변수를 나타낸다. 선형적인 독립적인 항  $q$ 는 차수  $d$ 와 같거나 작아야 한다.  $N$ 의 길이를 갖는 시계열에 비해 상대적으로 작은 차수  $d$ 의 LRF에 의해서 시계열이 표현될 수 있다면 시계열  $x_1, \dots, x_N$  ( $N \leq \infty$ )은 동질성을 나타낸다고 할 수 있다.

물론, 실제 시계열 자료 중에서 LRF에 의해서 표현될 수 있는 동질성을 갖는 순수시계열은 극히 드물다. 그러나, LRF와 작은 차수를 갖는 형태의 확정론적 시계열들에 의해서 추정될 수 있는 시계열들의 종류는 의외로 많다고 할 수 있다. 즉 원시계열자료로부터 주기성과 경향성을 갖는 요소들을 중심으로 시계열 자료를 검토해보면 이러한 주성분을 제외한 잡음(noise)과 진동성분은 시계열 분석에서 제거될 수 있으며 따라서 주성분만을 가지고 LRF를 통해서 시계열을 예측할 수 있다.

한 쌍의 켈레복소수 근은 지수(exponent)에 의해서 조정되는 사인함수의 형태를 나타낸다. 이 요소의 주기는  $2\pi/\phi$ 와 같으며, 여기서  $\phi$ 는 극좌표에서의 양의 각을 나타낸다. 지수의 감쇠율(rate)은  $LN(R)$ 로 나타낼 수 있으며 여기서  $R$ 은 근의 절대값을 나타낸다. 그러므로  $R$ 이 1과 같다면 이 요소는 순수한 사인함수적인 거동을 나타내며  $R$ 이 1보다 작은 경우는 감소하는 경향을 이끄는 반대로  $R$ 이 1보다 큰 경우에는 지수 함수적으로 증가하는 사인함수의 거동을 나타내게 된다. 양의 실근은 경향성을 나타낸다. 여기서 차수 1의 근  $R$ 을 갖는다면 감쇠율  $LN(R)$ 을 가지는 지수적인(exponential)의 요소를 발생시키며 1과 같은 이중근은 선형적인 요소를 나타내는 항으로 작용한다. 작은 절대값을 가지는 근은 단기적인 예측에 있어서 중요하지 않다.

## 2.3. SSA를 이용한 일단위 물 수요량 예측

본 논문에서 제시한 방법론을 요약하면 원시계열 자료를 SSA 기법을 이용하여 자료를 가장 잘 표현해주는 몇 개의 주성분으로 시계열을 분리하여 재구성하게 되며 재구성된 시계열자료를 LRF를 통하여 예측모형을 구성하게 된다. 일단위 물수요량을 예측하기 위한 대상지역으로 구의정수장내 아차산, 월계,

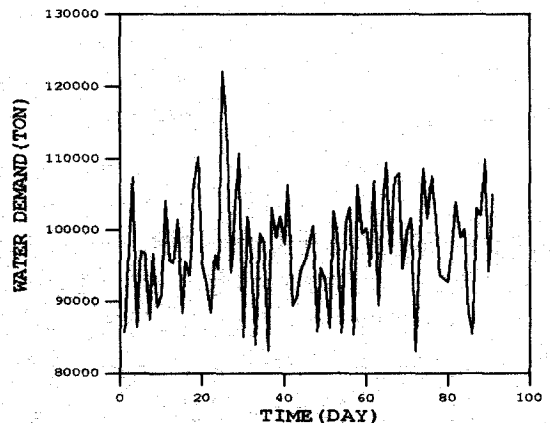


Fig. 4. Water demand data of Achasan distribution sect(2004, January-May).

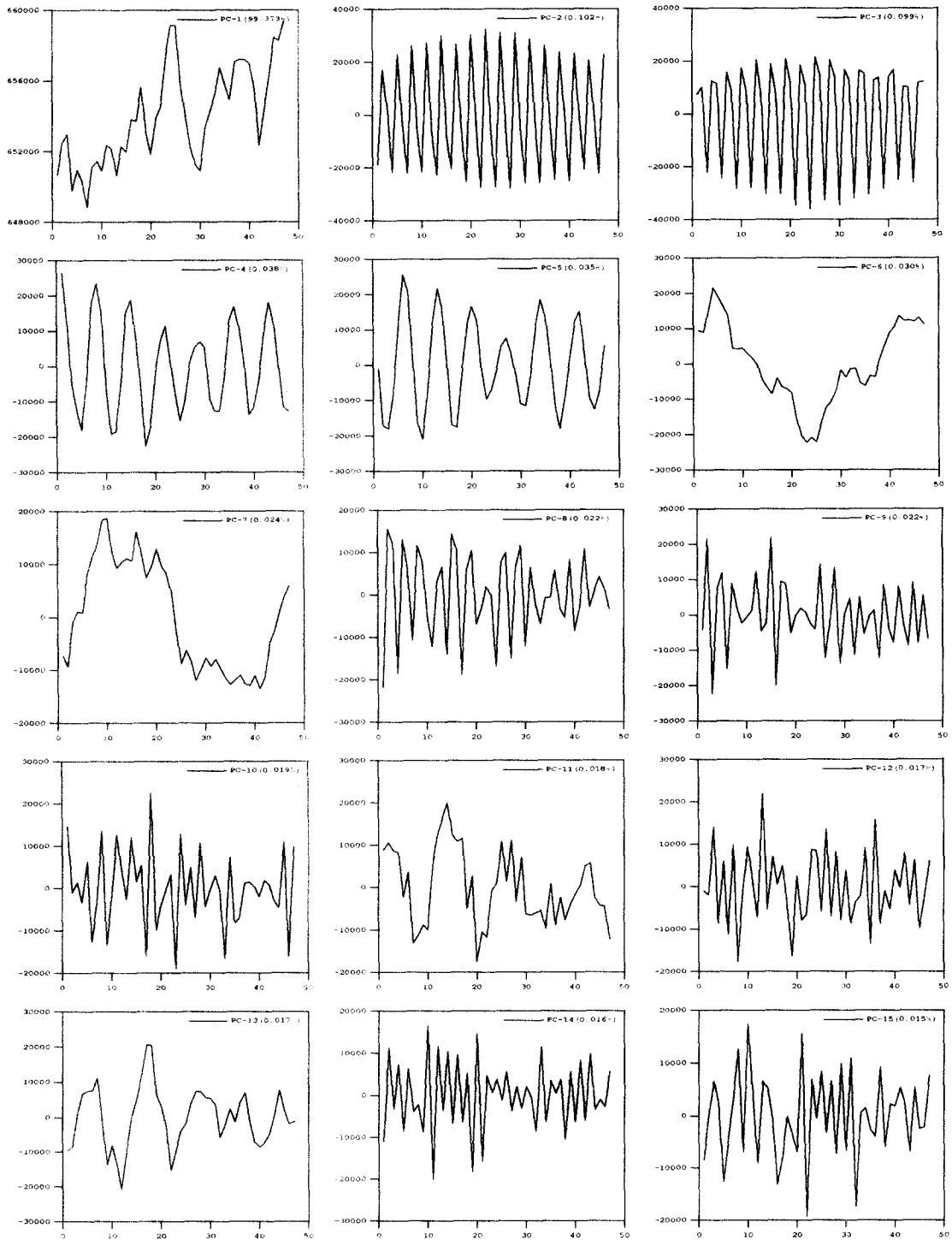


Fig. 5. Principal component extraction using SSA.

용마 배수구역의 2004년 1월부터 7월 수요량 자료를 이용하여 모의를 실시하였다. 이 중 1월부터 3월까지 자료를 이용하여 4월의 15일을 예측하였으며 4월에서 6월까지의 자료를 통하여 7월의 15일을 예측하였다. 3개의 배수구역 중에서 아차산 배수구역을 중심으로 1~3월의 모의과정을 서술하였다. Fig. 4는 아차산 배수구역의 1~3월까지의 물수요량 시계열 자료를 나타낸다.

Fig. 5에서 나타냈듯이 Fig. 4의 시계열을 총 15개의 주성분으로 분리가 가능하며 총 99.847%의 성분을 표현해 주고 있다. 15개의 주성분과 나머지 백색잡음(white noise)과의 상관성을 검토한 결과 상관계수가 0.01의 아주 작은 값으로 추정되었으며 따라서, SSA를 이용한 주성분의 분석과 시계열의 재구성 과

정이 합리적인 것으로 판단되었다. 분리된 주성분을 대상으로 시계열을 구성할 수 있으며 Fig. 6에서 보듯이 원자료와 거의 일치하는 결과를 나타내고 있다. Fig. 7은 주성분 1~15를 이용하여 재구성된 자료의 잔차(residual)를 나타낸다.

SSA에 의해서 재구성된 자료와 LRF에 의해서 추정된 시계열 자료의 적합성을 평가하기 위해서 잔차분석(residual analysis)을 실시하였으며 자기상관함수(autocorrelation function)와 정규성 검토를 실시하였다. Fig. 8-9는 SSA를 통해 재구성된 시계열의 자기상관함수와 정규성 검토 결과를 나타내며 Fig. 8에서 보듯이 상관성이 존재하지 않으며 잔차의 정규성을 Fig. 9로부터 확인할 수 있다. Fig. 10-11은 LRF를 통해 추정된 시계열의 자기상관함수와 정규성 검토

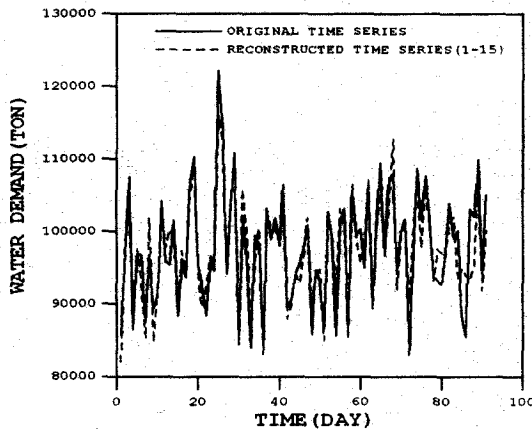


Fig. 6. Reconstructed times series(PC 1-15).

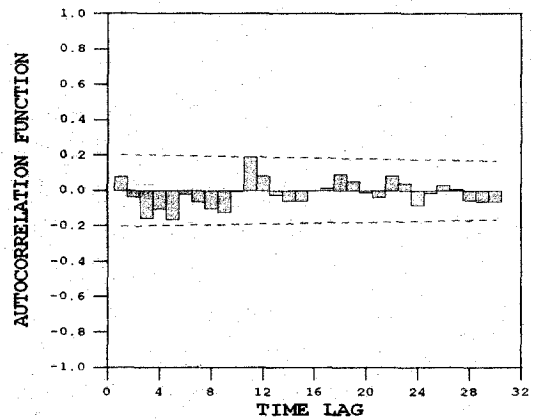


Fig. 8. Autocorrelation function of residuals of reconstructed time series.

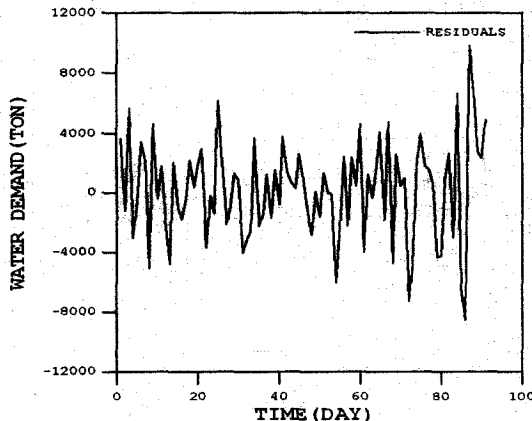


Fig. 7. Residuals of reconstructed time series.

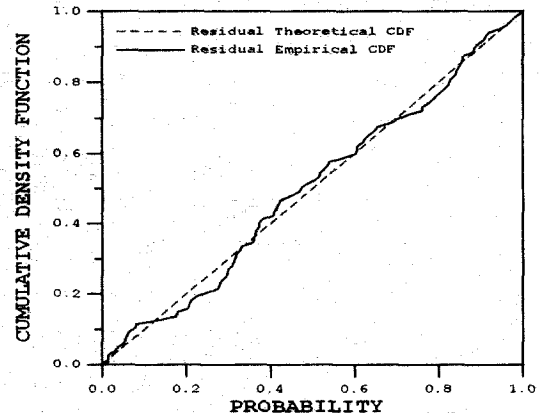


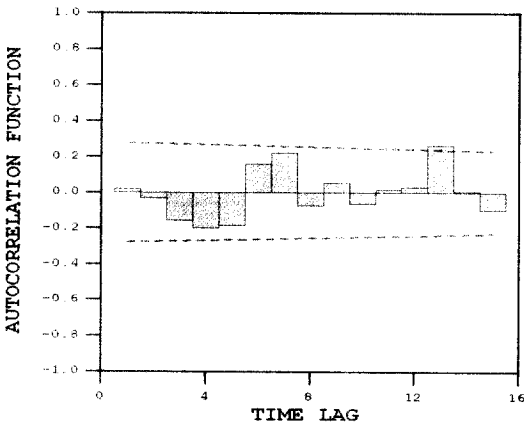
Fig. 9. Normality test of residuals of reconstructed time series.

**Table 1.** Comparison of statistic characteristics between observed data and simulated data in Achasan distribution sect during January-May

Statistic Characteristics	Observed data	Forecasting data
Average	97636.3	97410.2
Standard deviation	7800.28	6821.27
Correlation coefficient		0.904
Mean absolute error(%)		2.752
Skewness	0.221	0.063

**Table 2.** Comparison of statistic characteristics between observed data and forecasting data in Achasan distribution sect using January-May data

Statistic Characteristics	Observed data	Forecasting data
Average	99853.3	99481.5
Standard deviation	4395.59	4247.16
Correlation coefficient		0.588
Mean absolute error(%)		2.988
Skewness	-0.3209	-0.466



**Fig. 10.** Autocorrelation function of residuals of approximated time series using LRF.

결과를 나타내며 Fig. 10에서 보듯이 상관성이 존재하지 않으며 잔차의 정규성을 Fig. 11로부터 확인할 수 있다.

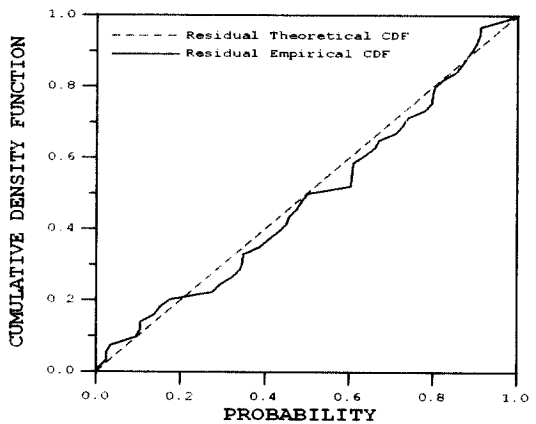
Table 1은 시계열예측에 앞서 모형의 추정과정을 나타내며 통계적인 특성을 중심으로 나타내었다. 모든 통계치가 실측치가 거의 같은 값을 나타내고 있으며 시계열의 특성을 효과적으로 반영한 결과로 판단할 수 있다. Table 2는 15일간의 모의된 자료와 실

**Table 3.** Comparison of statistic characteristics between observed data and simulated data in Achasan distribution sect during April-June

Statistic Characteristics	Observed data	Forecasting data
Average	96529.7	96157.7
Standard deviation	13178.8	11221.5
Correlation coefficient		0.824
Mean absolute error(%)		6.497
Skewness	-1.611	-1.204

**Table 4.** Comparison of statistic characteristics between observed data and forecasting data in Achasan distribution sect using April-June data

Statistic Characteristics	Observed data	Forecasting data
Standard deviation	13178.8	11221.5
Average	94813.3	84342.4
Standard deviation	10116.53	11181.41
Correlation coefficient		0.607
Mean absolute error(%)		1.820
Skewness	-1.14638	0.045728



**Fig. 11.** Normality test of residuals of approximated time series using LRF.

측자료와의 통계치를 나타내며 평균 및 표준편차 등의 일반적인 통계치는 실측치와 거의 같은 값을 나타내고 있으나 상관계수가 0.588로 정도가 약간 떨어지는 결과를 나타내었다. 그러나 평균절대오차는 2.958%로서 비교적 정확한 값을 나타내고 있다. 이러한 결과는 15일간의 전체적인 예측오차는 크지 않으나 예측능력에 있어 실측치에 거동을 정확하게 반



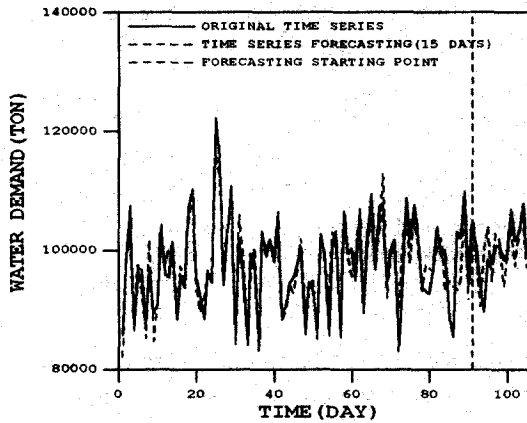


Fig. 12. 15 days forecasting of Achasan distribution sect using January-May data.

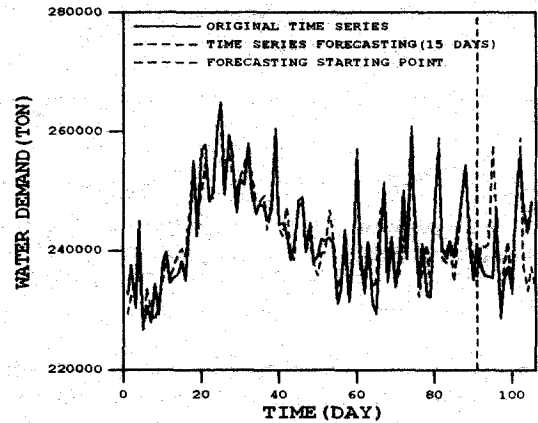


Fig. 14. 15 days forecasting of Yongma distribution sect using January-May data

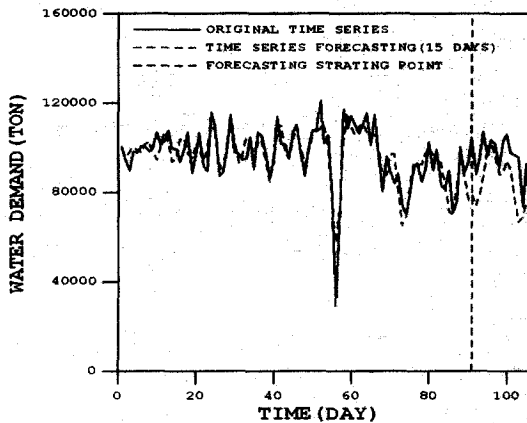


Fig. 13. 15 days forecasting of Achasan distribution sect using April-June data

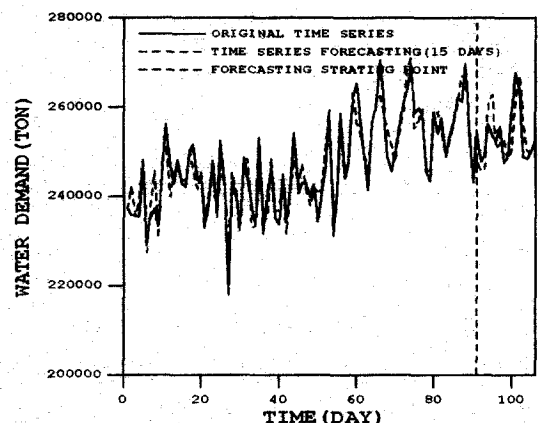


Fig. 15. 15 days forecasting of Youngma distribution sect using April-June data

영하지 못하고 시간에 따라 반응이 약간 상이하게(지체) 발생하는 원인으로 사료된다.

Fig. 12는 실측치와 예측치를 비교한 그림으로서 예측시점인 4월 이후에서 보듯이 실측치와 거의 유사한 거동을 보이고 있다. 따라서 통계적 특성 및 시각적인 결과를 바탕으로 추정해보면 분산이 매우 큰 일수요량의 예측 모형으로서 타당성이 입증될 수 있다고 사료된다.

Table 3은 시계열예측에 앞서 4~6월의 자료를 바탕으로 모형의 추정과정을 나타내며 통계적인 특성치를 중심으로 나타내었다. 모든 통계치가 실측치가 거의 같은 값을 나타내고 있으며 시계열 예측을 위한 모의과정으로서 타당한 결과를 제시해주고 있다.

Table 4에서는 아차산 배수구역의 4월~6월까지의 실측자료를 바탕으로 본 논문에서 제시한 방법론을 적용하여 15일에 대해서 Blind Forecasting을 실시한 결과와 실측자료의 통계치를 비교하였다. 실측자료와 예측자료에서 평균 등의 일반적인 통계치는 유사한 결과를 나타내고 있으며 상관계수는 0.607로서 앞서 제시한 예측과 마찬가지로 약간 정도가 떨어지는 값을 나타내고 있다.

Fig. 13에서 보듯이 그래프의 전체적인 경향성은 잘 표현해주고 있으나 실측치에 비해 약간 작은 값으로 모의가 이루어지고 있다. 일수요량은 자료의 특성상 매우 큰 변동폭을 나타내고 있으며 이러한 원인은

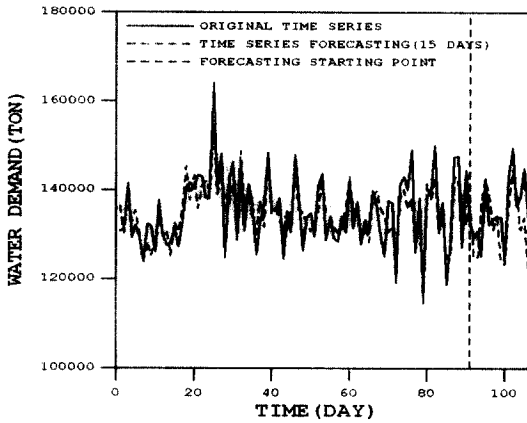


Fig. 16. 15 days forecasting of Wolgye distribution sect using January-May data

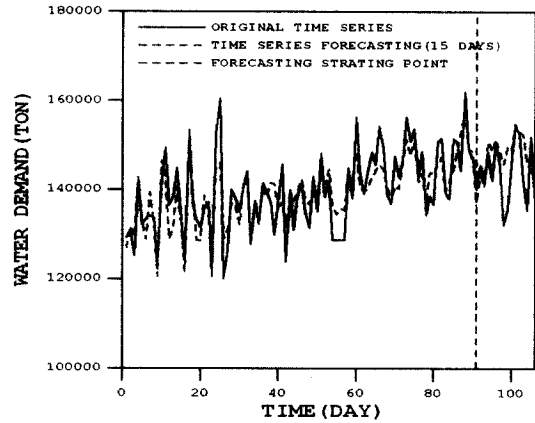


Fig. 17. 15 days forecasting of Wolgye distribution sect using April-June data

Table 5. Comparison of statistic characteristics between observed data and simulated data in Yongma and Wolgye distribution sect during January-May and April-June

Statistic Characteristics	Yongma 15 days forecasting using January-May data		Yongma 15 days forecasting using April-June data		Wolgye 15 days forecasting using January-May data		Wolgye 15 days forecasting using April-June data	
	Observed data	Forecasting data	Observed data	Forecasting data	Observed data	Forecasting data	Observed data	Forecasting data
Average	240685	240922	253446	253739	135348	132455	144080	148074
Average	242817	242629	247014	247067	135207	134998	139708	139199
Standard deviation	8654.28	8000.23	10444.5	9527.04	8064.81	5811.17	9034.22	6902.9
Correlation coefficient		0.933		0.939		0.817		0.854
Mean absolute error(%)		1.015		1.128		2.906		2.772
Skewness	0.452	0.305	0.113	0.260	0.388	0.328	0.113	-0.218

Table 6. Comparison of statistic characteristics between observed data and forecasting data in Yongma and Wolgye distribution sect using January-May and April-June

Statistic Characteristics	Yongma 15 days forecasting using January-May data		Yongma 15 days forecasting using April-June data		Wolgye 15 days forecasting using January-May data		Wolgye 15 days forecasting using April-June data	
	Observed data	Forecasting data	Observed data	Forecasting data	Observed data	Forecasting data	Observed data	Forecasting data
Average	240685	240922	253446	253739	135348	132455	144080	148074
Standard deviation	7304.5	8499.3	6094.6	6591.2	7245.1	7371.2	6838.6	4187.5
Correlation coefficient		0.447		0.681		0.594		0.507
Mean absolute error(%)		2.432		1.598		4.253		3.769
Skewness	0.4918	1.0238	1.2318	0.8052	0.303	0.0108	-0.185	-0.21

로 인해서 순환변동을 나타내는 시간수요량의 예측과 같이 정확하게 일치하는 결과를 나타내주지는 못하는 것으로 사료된다.

본 논문에서 제시한 방법론의 검증을 위해서 앞서 제시한 예와 더불어 용마, 월계 배수구역에 대해서 같은 방법으로 모의를 실시하였으며 자세한 모의 과정은 중복되는 관계로 제외하였다. 모형의 잔차에 대한 검정결과 4개의 예측 모두 시계열분석의 가정을 만족하였다. Fig. 14-17은 용마, 월계 배수구역의 예측결과를 실측자료와 함께 나타낸 그림이며 비교적 실측치에 근접한 거동을 나타내고 있으나 정확히 일치하지는 않고 있다. 이는 앞선 예측결과와 마찬가지로 일수요량 자체의 높은 분산과 변동성 때문으로 사료된다.

Table 5는 시계열예측에 앞서 1-3, 4~6월의 자료를 바탕으로 모형 추정과정에서의 통계적인 특성치를 나타내며 아차산 배수구역과 동일하게 모든 통계치가 실측치와 거의 같은 값을 나타내고 있다. 시각적인 검토와 더불어 4가지 예측결과에 통계분석 결과 Table 6과 같이 평균, 표준편차, 왜곡도 모두 실측치와 근접한 결과를 나타내고 있으며 상관계수는 0.45~0.68의 범위를 평균 절대 오차는 1.598%~3.769%의 범위를 갖는 값으로 모의가 이루어졌다.

### 3. 결론

최근 시계열을 모의하거나 예측하기 위한 방법으로 많은 추계학적 모형들이 개발되고 있으며 다수의 연구자들에 의해서 모형의 보완과 검토가 이루어지고 있다. 그러나 시계열에 대해서 통계학적인 기법을 통해서 정확하게 내재해 있는 통계적인 특성을 예측하는데 무리가 따르며 현재까지도 매우 어려운 문제로 인식되고 있다. 이러한 시계열을 모의하기 위해 어떤 사상의 값과 이전시점 사상의 값 사이에 상관성이 존재하는 사실로 기술할 수 있는 자기회귀과정 또는 Markov 과정이 대표적인 방법으로 주로 사용되고 있다. 그러나 기존의 시계열 모형은 대부분 매개변수적 모형으로서 매개변수 추정 및 모형 선택 등의 많은 어려움이 따르며 특히 시계열 예측 능력에 있어서는 문제점을 내재하고 있다.

따라서 본 논문에서는 시계열 분석에 있어서 변동성 추출 및 매개변수의 선택에 문제점을 보완할 수 있는 비매개변수적 방법인 SSA 기법을 적용하였다. 물수요량과 같은 시계열은 변동성과 무작위성이 매우 큰 특성이 있다. 이러한 관점에서 본 연구에서는 기상현상과 수문사상의 상관성 분석에 주로 적용되고 있는 SSA와 LRF를 결합한 시계열 예측 모형을 일단위의 물수요량 시계열 자료를 대상으로 적용성 및 타당성을 검토해 보았다.

모형을 통해 일단위 수요량을 모의한 결과 전체적인 통계적인 특성 및 시각적인 검토에서 실측자료와 매우 유사한 모의가 가능하였으며 실측 자료를 바탕으로 Blind Forecasting을 실시한 결과 3개 배수구역에서 상관계수는 0.45~0.68의 범위를 평균절대오차는 1.598%~3.769%의 범위를 갖는 값으로 모의가 이루어졌다. 따라서 자료의 특성상 정확한 모의가 이루어지지 않았지만 변동성이 큰 일수요량에 있어서 15일 정도의 단기 예측모형으로서 타당성과 적용성이 있는 것으로 판단된다. 전체적으로 예측구간동안 실측치와 거의 같은 거동을 보이고 있으며 매개변수 추정 및 모형의 선택 등에 있어 많은 장점을 가지고 있다고 사료된다. 따라서 기존 모형과 비교하여 본 연구에서 제시한 방법론을 검토한다면 합리적인 상수도 계획이 가능하리라 사료되며 추후에 시간단위 및 월 단위 수요량 예측에도 이용이 가능하리라 판단된다.

### 참고문헌

- 김신걸, 주충남, 구자용(1999) 지역별 시간변동에 따른 물 수요량 예측, *대한환경공학회 추계학술연구발표회* 논문집, pp. 517-518.
- 김형수, 나창진, 김종훈, 강인주 (2002) 퍼지론에 의한 강수 예측 I, II, *한국수자원학회논문집*, 35, pp. 619-629, pp. 631-638.
- 문영일(1997) 시계열 수문자료의 비선형 상관관계, *한국수자원학회논문집*, 30, pp. 641-648.
- 문영일, 박대형, 황성환(1999) Singular Spectrum Analysis를 이용 필터링된 자료로부터의 엘니뇨/라니냐 빈도분석, *한국수자원학회 학술발표회논문집*, pp. 171-176.
- 문영일(2000) 지역가중다항식을 이용한 예측모형, *한국수자원학회논문집*, 33, pp. 31-38.
- 박무중, 윤용남(1989) Multiplicative ARIMA 모형에 의한 월량의 추계학적 모의 예측, *한국수자원학회논문집*.

- 22, pp. 331-339.
- 안상진, 이재경(2000) 추계학적 모의발생기법을 이용한 월 유출 예측. *한국수자원학회논문집*, **33**, pp. 159-167.
- 유명진, 구자용, 구윤희, 김신걸(2004) 선형 및 비선형 모델을 이용한 시간별 물수요량의 예측. *대한환경공학회 논문집*, **26**(3), pp. 277-283.
- 주충남, 구자용, 유명진(1999a) 인공신경망을 이용한 단기 물수요 예측. *대한환경공학회 추계학술연구발표회 논문집*, pp. 325-326.
- 주충남, 최재호, 구자용, 유명진(1999b) 물 수요예측 모델을 이용한 배수불력의 최적화. *대한상하수도학회 추계학술대회 논문집*, pp. 91-95.
- 최재호, 구자용, 유명진(1999) 도시의 재개발 계획에 따른 장기 물 수요 예측과 수리학적 검증에 관한 연구. *대한환경공학회 논문집*, **21**(9), pp. 1783-1796.
- 최태용, 구자용, 장석환, 전병호, 박상우(2001) 회귀분석에 의한 서울시 상수도 수요예측. *한국수자원학회 학술발표회 논문집*, pp. 327-334.
- Abarbanel, H.D.I., Lall, U., Moon, Young-Il, Mann, M., and Sangoyomi, T. (1996) Nonlinear dynamic of the Great Salt Lake: A predictable indicator of regional climate, *Energy*, **21**(7/8), pp. 655-665.
- Broomhead, D. S., and King G.P. (1996) Extracting Qualitative Dynamics from Experimental Data, *Phys. D Amsterdam*, **20**, pp. 217-236.
- Buchstaber, V.M. (1994) Time series analysis and grassmannians. In S Gindikin (Ed.), *Applies problems of Radon transform*, 162, pp. 1-17.
- Colebrook, J.M. (1989) Continuous Plankton Records: Zooplankton and Environment, Cline. Optimal kernel estimation of densities.
- Fraedrich, K. (1986) Estimating the Dimensions of Weather and Climate Attractors, *J. Atmos. Sci.*, **43**, pp. 419-432.
- Gantmacher, F.R. (1998) *The theory of Matrices*, Chelsea, Providence, RI.
- Gelfond, A.O. (1967) *Fimite Difference Calculus*, Moscow, Nauka.
- Kember, G., Flower, A.C., and Holubeshen, J. (1993) Forecasting river flow using nonlinear dynamics, *Stochob. Hydraul. Hydraul.*, **7**, pp. 205-212.
- Keppen C.L., and Ghil, M. (1992) Adaptive Filtering and Prediction of the Southern Oscillation Index, *Journal of Geophysical Research*, **97**(D18), pp. 449-454.
- Mo, K.C., and Ghil, M. (1992) Interannual Oscillation in the 700mb Geopotential Height Field over Northern Hemisphere, *J. Clim.*
- Moon, Young-Il and Lall, U. (1996) Atmospheric Flow Indices and Interannual Great Salt Lake Variability, *Journal of Hydrologic Engineering*, **1**, pp. 55-62.
- Rasmusson, E.M., Wang, X., and Ropelewski, C.F. (1990) The Biennial Component of ENSO Variability, *J. Mar. Sys.*, **1**, pp. 71-96.
- Smith, J.A. (1991) Long-range streamflow forecasting using nonparametric regression, *Water Resour. Bull.*, **27**(1), pp. 39-46.
- Vautard, R., and Ghil, M. (1989) Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time Series, *Phys. D, Amsterdam*, **35**, pp. 395-424.
- Vautard R., Yiou, S., and Ghil, M. (1992) Singular Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals, *Physica D*, **58**, pp. 95-126.
- Yakowitz, S., and Karlsson, M. (1987) Nearest neighbor methods with application to rainfall/runoff prediction, *Stochastic hydrology*, Edited by Macneil, J.B., and Humphries, G.J., D. Reidel, Hingham, MA, pp. 149-160.