

# Wavelet transform을 이용한 물 수요량의 특성분석 및 예측 모형의 개발

조용준\*, 문영일\*\*, ○김종문\*\*\*

## 서론

時系列자료의 예측모형은 多重回歸分析系列과 時系列分析系列로 분류할 수 있다. 다중회귀분석모형의 경우 시계열자료에 영향을 미치는 인자를 산정하여 향후시계열자료를 추정하게 되는 데 급수량의 경우 인구수, 기온, 운도, 일조시간과 일사량에 선형적으로 영향을 받는 것으로 보고되어 있다 [현인환 1995]. 그러나 이런 水文因子들간의 독립성에 의문이 들고 시계열자료의 예측을 위해서는 수문인자의 예측이 선행되어야하며 또한 선형회귀분석모형의 경우 단기예측에는 상당한 오차가 수반된다. 일반적인 시계열자료에는 장기간에 걸친 變化推移를 설명하는 trend, solar 혹은 lunar component와 같은 주기적 성분 그리고 相互 獨立적인 無作位 變動性分등이 선형적으로 혼재하며, 시계열분석모형은 세가지 성분중 trend와 주기적성분을 決定論的인 인자로 취급하여 시계열자료에서 遊離하면 나머지 자료는 定常時系列을 구성한다는 가정에서 출발한다 [Kottegoda 1980]. 장주기변화거동과 주기성분은 least square계열의 회귀분석과 spectral analysis를 통해 類推하게 되는 데 회귀분석과정에서 도입되는 인위적인 filtering으로 인해 자료가 왜곡될 수있다. 무작위변동성분의 예측모형으로는 定常時系列 [stationary time series] 에 활용되는 ARMA모형, 장기간에 걸친 변화거동이 존재하는 非定常時系列 [nonsationary time series] 에 활용되는 ARIMA모형이 있으나 時系列에 내재한 非定常性을 제거하는데 활용되는 差分化과정 혹은 類似差分化과정에서 시계열자료에 왜곡을 야기할 수 있어 주의가 요구된다. 또한 spectral analysis를 위해 수행되는 Fourier transform의 경우 기저함수로 sinusoidal 함수를 사용함으로써 단지 주파수영역에서의 정보만을 제공할 뿐 주파수영역으로 전환되는 과정에서 시간영역에서의 정보는 유실되는 단점이 있다. 따라서 국지적인 기후 혹은 주거형태의 변화등으로 야기된 장주기의 변화거동이 존재하는 수문자료에는 직접적으로 활용할 수 없다. 최근 이러한 단점을 보완하고자 Fourier transform系로는 windowed Fourier method [Krauss 1994], Burt-Adelson의 pyramidal algorithm, quadrature mirror filter, sub-band coding 등을 활용한 wavelet transform등이 제안되었다 [Ruskai 1992, Walter 1994]. wavelet transform 해석방법은 무한한 support를 갖는 기존의 sinusoidal 함수대신 유한한 support를 갖고 상호 orthogonal한 wavelet 函數를 기저함수로 채택함으로써 주파수영역에서의 정보뿐만아니라 장기간에 걸친 변화추이와 같은 실시간영역에서의 거동도 유리해낼 수 있다. 본 연구에서는 서울소재 정수장중 가장 규모가 큰 구의정수장을 대상으로 정수장시간별생산량으로 時系列을 구성하여 시계열자료의 분해능력이 뛰어난 wavelet 해석기법을 사용하여 시간별생산량자료의 장기간에 걸친 변화거동을 규명하고 기존의 추계학적 모형에서 결정론적 인자로 취급해온 일별, 요일별, 계절별변화등과 같은 solar와 lunar component의 變化推移를 實時間領域에서 분석하여 이를 바탕으로 보다 정확한 예측모형을 개발하고자 한다.

## Wavelet Transform 과 단주기성분의 통계특성

時系列분석에 사용되는 wavelet함수는 時系列자료의 장기간에 걸친 變化推移가 n차다항식으로 模擬되는 경우 n차이상의 moment가 零인 wavelet 함수가 선정되어야 자료의 인위적인 歪曲을 방지

\* 서울시립대 토목공학과 조교수 \*\* 서울시립대 토목공학과 전임강사 \*\*\* 서울시 상수도 사업본부

할 수 있다 [Krauss 1994]. 구의정수지 生産量時系列 자료를 육안으로 판별시 장기간에 걸친 변화가동이 비교적 고차의 다항식으로 模擬되는 것으로 판단되어 1st, 2nd, 3rd와 4th moment가 零이고  $2^6-1$ 의 support를 갖는 Coiflets5 함수를 wavelet 해석의 基底函數로 선정하였으며 Coiflets5 함수는 Fig. 1에 도시하였다.

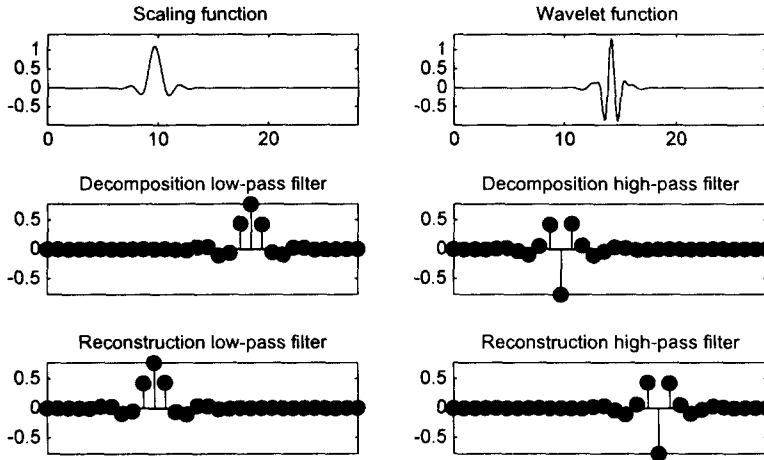


Fig. 1 Coiflets 5 Wavelet Function

Continuous wavelet transform과 Fourier transform에서의 주기에 해당되는 scale를 2의 지수승  $[2^j, j=1,2,\dots,12]$ , 즉, dyadic scale로 변화시키면서 수행한 discrete wavelet transform의 결과는 각 Fig. 2와 Fig. 3에 圖示하였다. wavelet transform은 다음과 같이 정의되며

$$c = \int_{-\infty}^{\infty} z(t)\psi dt \quad (1)$$

여기서  $\psi$ 와  $z(t)$ 는 wavelet 함수, 시계열 자료를 나타내며 wavelet coefficient  $c$ 는 wavelet 함수가 有限한 support를 갖기 때문에 scale과 시간  $t$ 의 함수가 된다. low band와 high band filtering을 연속해서 실시한 결과  $2^{12}$  scale에서 장기간에 걸친 변화추이를 원시계열자료에서 분리해 낼 수 있었으며 Fig. 3에 도시된 대로 1996년에 구의정수장에서 공급된 수량은 hyperbolic tangent함수의 형태로 전기간에 걸쳐 꾸준한 증가세를 보였다. 이는 같은 기간에 상계동지역의 입주가 시작된 사실과 밀접한 관계가 있을 것으로 추측되나 소득수준의 향상으로 인한 수요량 증가일 가능성도 배제할 수 없어 장기에측도형의 개발을 위해서는 이에 대한 보다 정밀한 원인분석과 증가량 起因因子에 대한 향후예측도 병행되어야 할 것으로 판단된다. 또한 節氣 혹은 景氣週期和 밀접한 관련이 있을 것으로 생각되는 추가수요가 6月末과 12月末을 정점으로 발생하며 이 추가수요량은 하절기의 경우 1700 ton/hr, 12月末의 경우 500 ton/hr 정도인 것으로 관측되었다. 또한  $2^j [j=3,4,5,6,7,8]$  scale에서 관측된 주기성분을 전시계열구간에서 발생한 波峰數로 분석한 결과 주기가 각각, 12, 23.98, 46, 75.18, 149.25와 303 hr인 것으로 분석되었다. 본시계열자료에서  $2^{12}$  scale에서 관측된 장기간에걸친 변화추이를 분리한 후 일반적인 수문자료에 내재하는 solar component와 lunar component와 같은 주기성분을 wavelet tranform과의 비교를 위해 Hanning window와 256 point에서의 Fast Fourier Transform algorithm을 이용해 spectral analysis를 수행하였으며 그 결과는 Nyquist

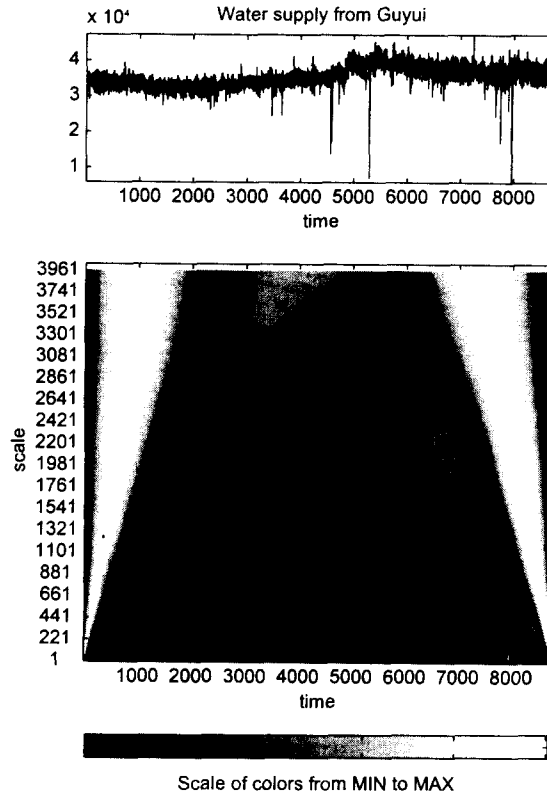


Fig. 2 Continuous Wavelet Transform

frequency까지 Fig. 4 (a) 에 도시하였다. Fig. 4 (a)에서 관찰되는 바와 같이 본시계열자료에는 주파수가 0.005, 0.0133, 0.03, 0.0417와 0.0833 hr<sup>-1</sup> 인 주기성분이 내재하는 것으로 판단되며 이를 주기로 환산하면 8.33 day, 3.13 day, 33.33 hr, 23.98 hr 와 12 hr가 된다. 진폭은 예상되로 주기가 23.98 hr인 성분이 가장 크고 그외 주기성분의 진폭은 비슷한 것으로 나타났으며 개개파의 진폭과 위상은 다음과 같으며

$$P_{8.33 \text{ day}} = 5.6229 \cdot 10^3 e^{-2.0966i}$$

$$P_{3.13 \text{ day}} = 7.8290 \cdot 10^3 e^{0.1769i}$$

$$P_{33.33 \text{ hr}} = 7.7890 \cdot 10^3 e^{1.4152i}$$

$$P_{23.98 \text{ hr}} = 1.5758 \cdot 10^4 e^{-2.6341i}$$

$$P_{12 \text{ hr}} = 6.0089 \cdot 10^3 e^{1.5574i}$$

여기서  $i$ 는 imaginary unit 을 나타낸다.  $2^j$  [ $j=1,2,\dots,12$ ] scale에서 관측된 단주기성분 [이하 detail]의 확률밀도함수 [이하 PDF]와 누가 확률밀도함수는 Fig. 5에 도시하였다. 일반적인 추계학적 모형에서는 주기적성분을 결정론적 인자로 취급하나 wavelet transform을 이용한 해석한 결과 spectral analysis를 통해서도 주기적으로 보이는 성분도 實時間領域에서 관찰할 경우 sinusoidal wave의 형

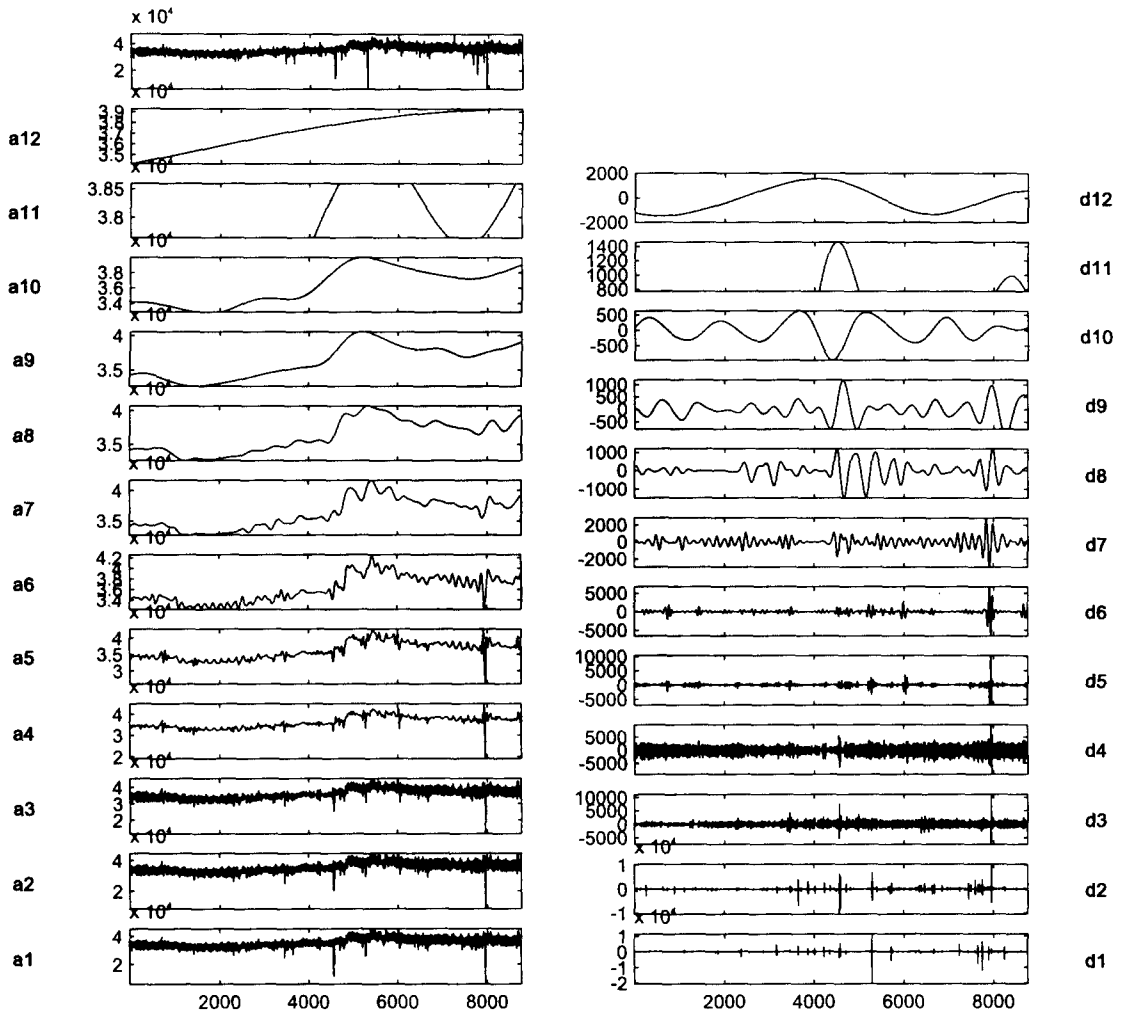


Fig. 3 Discrete Wavelet Transform and its Approximation and Detail Components

태로 반복되는 것은 아니라는 사실이 입증되어 개선이 요구된다. 본 연구에서 해석한 결과  $2^j$  [ $j=1,2,\dots,12$ ] scale 에서의 단주기성분들이 Gaussian PDF를 따른다는 사실에 착안하면 향후 보다 실제에 근접한 예측모형개발도 가능하리라 판단된다; 진폭 A는 Raleigh PDF, 위상  $\epsilon$ 은 uniform PDF를 따르는 무작위변량으로 취급할 경우 Gaussian PDF를 따르는 無作為變量을 얻을 수 있다. 즉, spectral analysis를 통해 주기적으로 보이는 성분  $z$ 를 sinusoidal wave로 모형화하되 진폭과 위상을 각각 Raleigh PDF, 위상은 uniform PDF를 따르는 無作為變量으로 취급하면  $z$ 와  $z$ 의 Hilbert transform  $\hat{z}$ 는 다음과 같고

$$z = A \cos(\omega t + \epsilon)$$

$$\hat{z} = A \cdot \sin(\omega t + \varepsilon)$$

일반적인 無作為變量的 變換기법을 이용하면  $J$  차위변량  $z$  와  $\hat{z}$  의 結合確率密度函數  $f_{z\hat{z}}(z, \hat{z})$  는 다음과 정의된다.

$$f_{z\hat{z}}(z, \hat{z}) = 1/2\pi \cdot f_A(A) \left| J\left(\frac{A, \varepsilon}{z, \hat{z}}\right) \right| \quad (2)$$

$$= 1/2\pi \cdot \exp[-1/2(z^2 + \hat{z}^2)]$$

여기서  $J$ 는 無作為變량의 變換에 따른 Jacobian matrix를 의미하며, 식 (2)에서  $z$  는 Gaussian PDF 를 따르는 것을 확인할 수 있다.

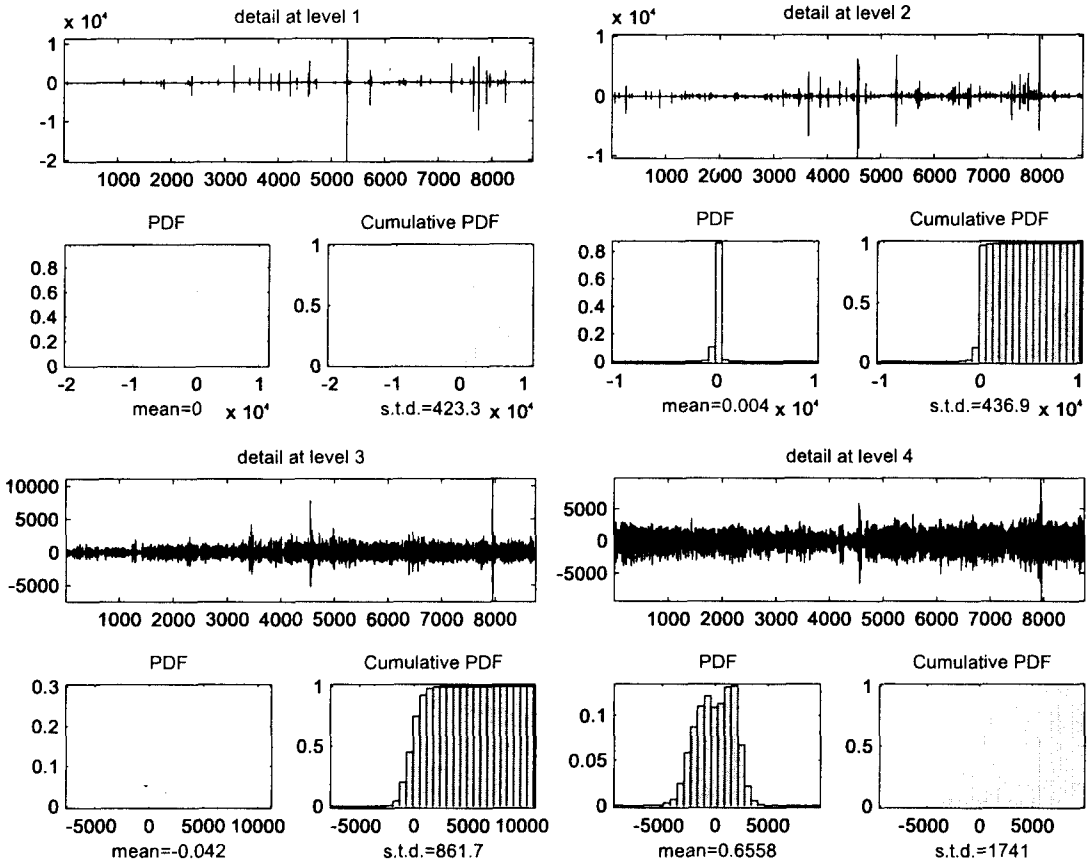


Fig. 4 PDF and Cumulative PDF of Detail Components at  $2^j$  [ $j=1,2,3,4$ ] Scale

예측모형

$2^{12}$  scale에서 遊離된 장기간에 걸친 變化推移는 外插하여, 주기가 3.13 day, 33.33 hr, 23.98 hr 와 12 hr인 주기성분에 대해서는 계속해서 반복되는 決定論的 인자로 취급하여 模擬하였다. Fig. 4 (b), (c), (d), (e)와 (f)에는 原時系列자료에서 장기간에 걸친 變化推移만을 분리한 자료의 autocorrelation function과  $2^j$  [ $j=1,2,3,4$ ] scale에서 유리된 단주기성분의 autocorrelation function을 도시하였으며 주기성분이 내재하고 있음을 보여준다. 최적 ARMA모형의 차수는 原時系列자료의 autocorrelation coefficient와 partial autocorrelation coefficient를 이용하여 원시계열자료가 AR[p] 모형

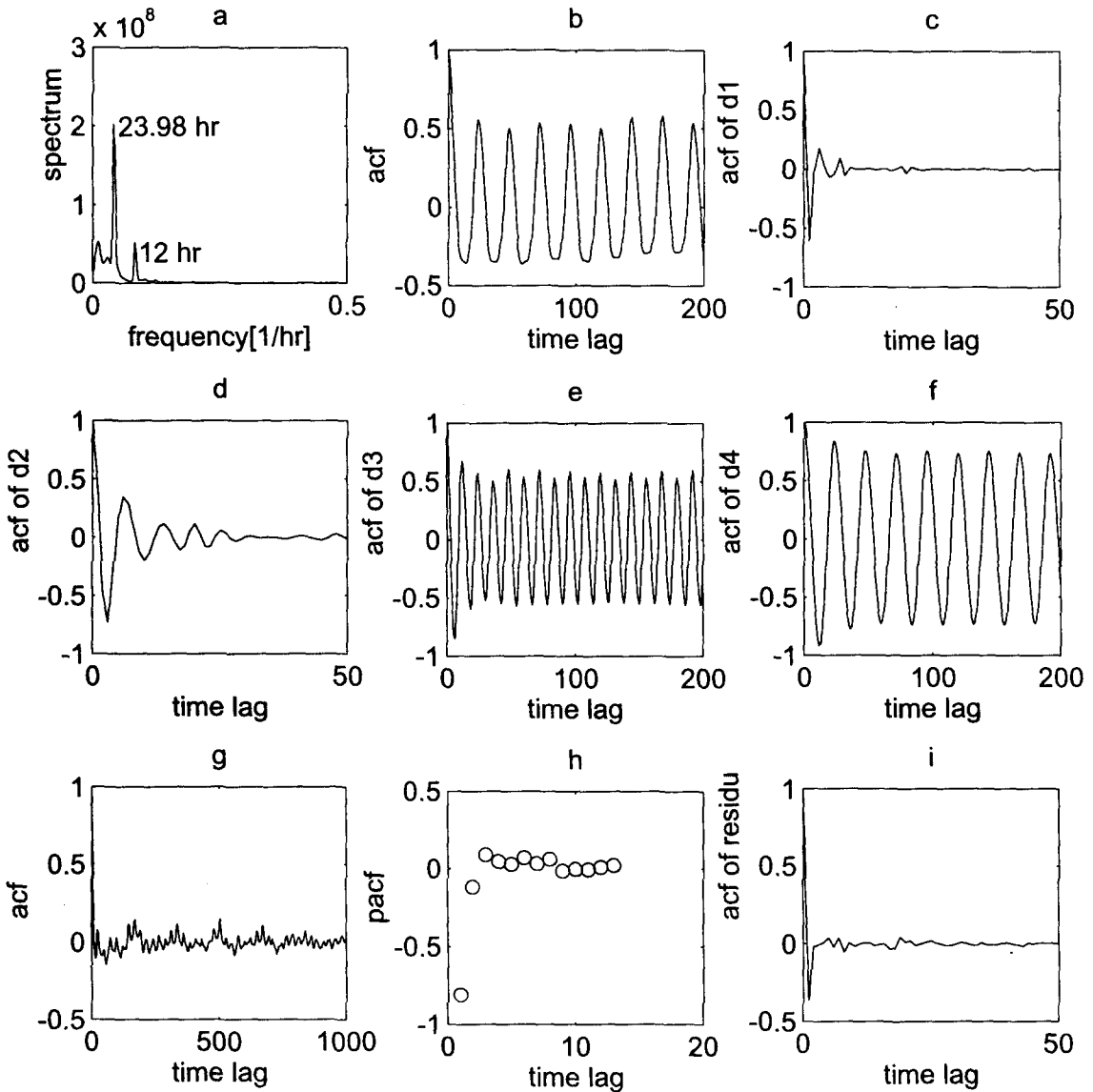


Fig. 5 Spectrum and ACF of Time Series after Removing Long Term Evolution, ACF of Detail and Time Series after Removing Long Term Evolution and Periodic Components, PACF and ACF of Residue Induced by ARMA[3, 2]

을 따른다면 지체시간이 p이상인 영역에서 partial autocorrelation coefficient는 零에 收斂한다는 점과 모형을 통한 예측치와 계측자료간의 차이로 정의되는 잔차성분  $[\epsilon]$ 의 상호독립성, 白色波여부와 FPE [Akaike의 final prediction error]의 최소화를 기준으로 선정하였다.  $\epsilon$ 은 다음과 같이 주어지며

$$\epsilon = z(t) - \overline{z(t)}$$

여기서  $z(t)$ 는 관측치,  $\overline{z(t)}$ 는 예측치를 나타낸다. 原時系列자료에서 장기간에 걸친 變化推移와 주기가 3.13 day, 33.33 hr, 23.98 hr 와 12 hr인 주기성분이 분리된 시계열자료의 autocorrelation function은 Fig. 4 (g)에 도시하였으며 지체시간이 증가함에 따라 autocorrelation function이 빠르게 零에 수렴하여 定常時系列이 구성되었음을 확인할 수 있다. 장기간에 걸친 변화추이와 주기성분이 분리된 시계열자료를 사용하여 구한 partial autocorrelation function은 Fig. 4 (h)에 도시하였다. 3차 이상의 차수에서 partial autocorrelation function이 零에 收斂하여 ARMA모형중 auto regressive부분의 차수는 3차로 선정하였으며 moving average부분의 차수를 1차에서 3차까지 변화시켜가며 잔차분석을 수행한 결과 ARMA[3,2]모형에서 잔차의 분산이 최소가 되었으며 이때 잔차성분의 autocorrelation function은 Fig. 4 (i)에 도시하였다.

## 결론

서울소재 정수장중 가장규모가 큰 구의정수장을 대상으로 정수장시간별생산량으로 時系列을 구성하여 시계열자료의 분해능력이 뛰어난 wavelet transform을 사용하여 해석한 결과 장기간에 걸친 변화추이를 實時間領域에서 분리할 수 있었고 정수장생산량 時系列자료에 내재한 주기성분은 주기가 각각 3.13 day, 33.33 hr, 23.98 hr 와 12 hr인 것으로 判明되었으며 무작위변동성분의 최적예측모형으로는 ARMA[3,2]모형인 것으로 판단된다.

## 참고문헌

현인환, 목동우 [1995] 상기급수량의 단기예측 대한상하수도학회지, 제 9권, 제 1호

Kottegoda, N. T. [1980] Stochastic water resources technology, The Macmillan press LTD.

Krauss, T. P., Shure, L. and Little, J. N. [1994] Signal processing tool box, the math works Inc.

Ruskai, M. B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y. and Raphael, L. [1992] Wavelets and their applications, Jones and barlett publishers international.

Walter, G. G. 1994 Wavelets and other orthogonal systems with applications, CRC press Inc.