

불규칙한 관측주기를 갖는 지하수자료를 이용한 지하수위 변동의 시계열 분석

이명재 · 이강근*

주식회사 지오그린21 기술연구소 (yimj2@snu.ac.kr)

*서울대학교 지구환경과학부 수리지구환경연구실 (kklee@snu.ac.kr)

요 약 문

장기간 관측된 지하수위 자료를 시계열분석 중의 하나인 전이함수 모형(Transfer Function - Noise model)을 이용하여 분석하였다. 일반적으로 전이함수 모형은 입력 변수와 출력변수와의 관계가 선형적일 때 적용이 가능하며, 자료가 시간에 대해 연속적으로 존재해야 하는 제한이 있다. 강수량과 지하수위의 변동은 비선형적인 관계를 가지고 있어 이러한 전이함수 모형을 직접 적용하는데는 어려움이 있다. 이러한 비선형성의 정도를 감소시키기 위해 물리모형(HYDRUS)을 이용하여 침투량을 계산하고 이를 입력변수로 사용하여 전이함수 모형을 적용하였다. 침투량을 입력변수로 모형을 추정하였을 때, 강수량을 직접 입력자료로 사용했을 경우보다 ME(mean error), RMSE(root-mean-square error), MAE(mean absolute error)에서 상대적으로 작은 값을 보여주고 있다. TFN 모형의 모수를 추정하기 위해서 Kalman 필터 알고리즘과 최우추정법(Maximum Likelihood Estimation)을 이용하였다. Kalman 필터 알고리즘을 이용하여 불규칙한 관측주기를 갖는 시계열이나 결측값이 있는 시계열에 대해서도 전이함수 모형을 구하였으며, 이를 통해 결측값에 대한 추정이 가능하였다.

Key words : 시계열 분석, 전이함수 모형(TFN 모형), 비선형, 강수량, 침투량, 모수 추정, 물리모형, Kalman 필터

I. 도입

장기간 관측된 지하수위, 함양량과 같은 수리지질학에서 얻는 자료를 분석하기 위해 시계열분석 기법이 사용되어 왔다. 지하수위의 변동은 여러 가지 요인에 의해서 영향을 받는다. 이러한 영향요인을 모두 물리모형으로 계산하기 어렵기 때문에 선형 모형인 ARMA 모형이나 전이함수 모형(TFN 모형)과 같은 시계열 모형을 많이 사용

하였다 [Salas 등, 1982; Adamowski 등, 1983; Furbish, 1991; Knotters 등, 1999].

그러나 강수량과 지하수위와 같이 비선형적인 관계를 갖는 시계열자료의 경우는 일반적인 선형모형을 적용하는 경우 적절한 모형을 추정하기가 어렵다. Knotters 등 [1997]은 SWATRE와 같은 물리모형으로 지하수위를 예측하고 실제값과의 차이를 ARMA 모형으로 추정하는 방법을 사용하여 개선된 결과를 얻었으나, 지하수위를 물리모형으로 직접 예측하기 위해서 물리상수를 알고 있어야하고 광역적인 지하수의 흐름을 추정해야 하는 등 여전히 불확실성을 가지고 있는 단점을 가지고 있다.

이 논문의 목적은 침투량을 입력자료로 사용하여 강수량과 지하수위와의 TFN 모형을 더욱 개선시키고, 아울러 불규칙적인 관측주기나 결측값을 가진 시계열에 대한 모형을 Kalman 필터를 이용하여 추정하는 것이다.

II. 전이함수 모형과 Kalman 필터

분석하려는 시계열에 밀접한 관계가 있어 이의 해석에 도움이 되는 다른 시계열변수를 입력변수로 사용하는 방법으로 대표적인 모형이 전이함수 모형이다. 전이함수는 선형결합형태의 충격가중함수를 유한개의 모수를 가진 분수형태로 나타내어 모수의 숫자를 줄이고, 여기에 발생하는 오차가 백색잡음이 아닐 경우 이를 다시 ARMA 모형으로 식별하는 단계를 거친다(식(1)).

$$Y_t = \frac{w_0 + w_1B + \dots + w_sB^s}{1 + \delta_1B + \dots + \delta_rB^r} X_{t-b} + \frac{1 + \theta_1B + \dots + \theta_qB^q}{1 + \phi_1B + \dots + \phi_pB^p} a_t \quad (1)$$

여기서 X_t , Y_t , a_t 는 각각 시간 t 에서의 입력값(침투량), 출력값(지하수위), 백색잡음(white noise)을 나타내며, b 는 반응지체 시간을 나타낸다. w_s , δ_r , θ_q , ϕ_p 는 모형에서 사용되는 모수이고, B 는 후진연산자로 $X_tB^d = X_{t-d}$ 를 나타낸다.

전이함수 모형을 구하기 위해서 먼저 어떤 형태의 모형을 적합시킬 것인가를 결정하고 해당 형태에 필요한 모수를 추정한다. 마지막으로 여러 가지 통계량을 이용하여 구해진 모형이 적절한지를 시험하는 단계를 거친다. 여러 가지 후보 모형 중에서 가장 적절한 모형을 선택하는 방법으로 AIC와 SBC와 같은 통계기준값을 이용하였다.

Kalman 필터를 이용하여 불규칙한 관측주기를 갖는 시계열에 대해서도 TFN 모형의 모수를 추정할 수 있으며, 또 입력자료의 관측주기와 같은 출력자료를 얻어낼 수 있다(Harvey, 1989; Box 등, 1994; Bierkens 등, 1999). 관측값과 예측값과의 차이인 예측오차, ν_t 와 그 분산 $\sigma_{\nu_t}^2$ 는 Kalman 필터의 반복 계산을 통해 구할 수 있으며, 이때의 우도함수(likelihood function)는 식 (2)와 같이 나타낼 수 있다.

$$\log L = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^T \log(\sigma_{v,t}^2) - \frac{1}{2} \sum_{i=1}^T \left(\frac{v_{i,T}^2}{\sigma_{v,t}^2} \right) \quad \text{관측값의 수} \quad (2)$$

우선 TFN 모형에 사용되는 모수를 임의로 선택하여 Kalman 필터를 돌린 후, 우도함수 값을 구한다. 우도함수를 최대화시키는 모수를 찾을 수 있는 알고리즘을 이용하여 적절한 모수를 구할 수 있다. 여기에서는 최적화 알고리즘으로 Downhill Simplex method를 이용하였는데, 이 방법은 최적화시킬 함수의 도함수를 필요로 하지 않는 장점이 있다.

III. Case Study

약 10년간(1988-1997)의 장기 지하수위 시계열과 같은 기간의 강수량을 이용하여 전이함수 모형을 적용시켰다. 1988년 1월 1일부터 1991년 8월 31일(구간 1)까지는 매일 지하수위가 관측되었고 이후는 일주일에 두 번씩 불규칙하게 측정되었다. 먼저 비선형성의 정도를 감소시키기 위해 HYDRUS를 이용하여 침투량을 계산하였으며 이를 전이함수 모형의 입력자료로 사용하였다 (그림 1).

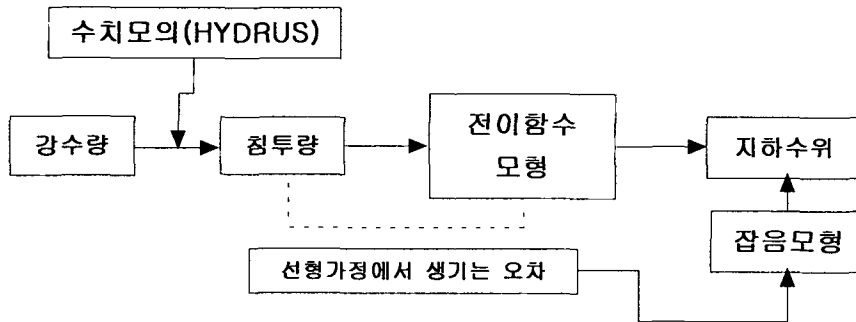


그림 1. 침투량을 입력자료로 사용한 전이함수 모형의 개념도

관측시작일에서 2000일 동안을 보정기간(calibration period)으로 두고 모형을 추정하였고, 이후 1000일 동안을 유효성평가 기간(validation period)으로 설정하여 추정된 모형의 ME, MRSE, MAE 등 통계량을 구하였다. 또 Kalman 필터가 결측값들을 추정할 수 있는지를 알아보기 위해서 <구간 1>에서 임의로 선택한 400개의 관측값을 이용하여 모형을 추정하고, 추정된 모형으로 사이의 결측값들을 예측하여 모형의 예측값을 관측값과 비교하였다.

$$y_t = 7.567615 + \frac{0.02134118 + 0.03133794B}{1 - 0.95719487B} x_t + \frac{1 - 0.09653339B}{1 - 0.94671990B} a_t \quad (3)$$

보정기간에서 TFN 모형이 식(3)과 같이 추정되었다. 이 TFN 모형을 이용하여

validation 기간에서의 예측값과 관측값을 <그림 2>에 나타내었다. 보정구간에서 구한 표준오차는 0.0967 m로 강수량을 입력자료로 사용했을 때의 0.2153 m보다 현저하게 감소했음을 알 수 있었다. 또 validation 기간에서의 ME는 -0.0175 m, RMSE는 0.1351 m, MAE는 0.0915 m로 계산되었다. 같은 기간에서 강수량을 입력자료로 사용하지 않았을 때는 ME = -0.0343 m, RMSE = 0.1492 m, MAE = 0.1052 m로 나타나 침투량을 사용했을 때 좀더 개선된 결과를 보여주고 있다.

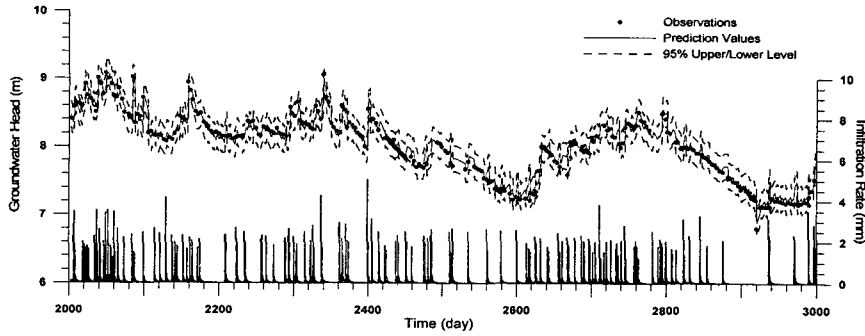


그림 2. Validation 기간에서의 예측 및 예측값의 95% 신뢰구간

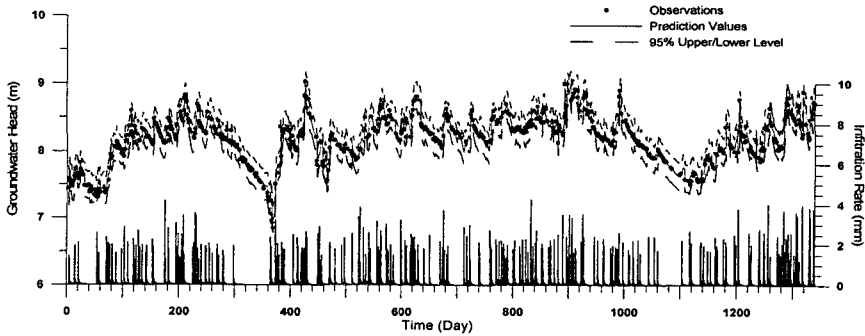


그림 3. 임의로 선택된 400개의 관측값을 이용한 TFN 모형의 예측 및 신뢰구간

한편, 구간 1의 1339개의 관측값 중에서 임의로 선택된 400개의 관측값을 이용한 모형은 식 (4)로 구해졌으며, 구간 1에서의 예측 및 신뢰구간의 시계열도를 <그림 3>에 나타내었다. 모형에서 추정된 표준오차값이 0.1025 m 였으며, ME는 -0.0013 m, RMSE는 0.1024 m, MAE는 0.0661 m로 얻어졌다.

$$y_t = 7.461832 + \frac{0.02296922 + 0.03256981B}{1 - 0.96048692B} + \frac{1 - 0.19351755B}{1 - 0.92939336B} a_t \quad (4)$$

IV. 요약

1. 불규칙적인 자료간격을 가진 지하수위자료에 대한 TFN 모형의 모수를 추정하기 위하여 Kalman 필터와 최우추정법을 사용하였다.
2. 또한 강수량과 지하수위간의 비선형성을 감소시키기 위해 입력자료로 사용된 침투량과의 TFN 모형이 강수량을 사용하였을 때보다 개선된 결과를 나타내었다.
3. Kalman 필터를 이용한 TFN 모형으로 결측값에 대한 예측이 가능하였다.

참 고 문 헌

- Adamowski, M. and Hamory, T. 1983, A stochastic systems model of groundwater level fluctuations, *Journal of Hydrology*, vol 62. 129~141.
- Bierkens, M.F.P., Knotters, M., and Van Geer, F.C. 1999. Calibration of transfer function-noise models to sparsely or irregularly observed time series, *Water Resources Research*. vol 35. 1741~1750.
- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. 1994. *Time series analysis: Forecasting and control*. 3rd ed. New Jersey: Prentice Hall.
- Chang, I., Tiao, G.C., and Chen, C. 1988. Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*. vol 30. 193~204.
- Furbish, D.J. 1991, The Response of Water Level in a Well to a Time Series of Atmospheric Loading Under Confined Conditions, *Water Resource Research*. vol 27. 557~568.
- Harvey, A.C. 1989. *Forecasting, structural time series models and the Kalman filter*. Cambridge Univ. Press.
- Knotters, M., and van Walsum, P.E.V. 1997. Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component, *Journal of Hydrology*, vol 197. 25~46.
- Salas, J.D., and Obeysekera, J.T.B. 1982. ARMA model identification of hydrological time series. *Water Resource Research*. vol 18. 1011~1021.