

## TFNM, ANN, ANFIS를 이용한 국가지하수관측망 지하수위 변동 예측 비교 연구

윤필선<sup>1</sup> · 윤희성<sup>1\*</sup> · 김용철<sup>1</sup> · 김규범<sup>2</sup>

<sup>1</sup>한국지질자원연구원 지구환경연구본부

<sup>2</sup>K-water연구원

## A Comparative Study on Forecasting Groundwater Level Fluctuations of National Groundwater Monitoring Networks using TFNM, ANN, and ANFIS

Pilsun Yoon<sup>1</sup> · Heesung Yoon<sup>1\*</sup> · Yongcheol Kim<sup>1</sup> · Gyoo-Bum Kim<sup>2</sup>

<sup>1</sup>Korea Institute of Geoscience and Mineral Resources

<sup>2</sup>Geowater+ Research Center, K-water Institute

### ABSTRACT

It is important to predict the groundwater level fluctuation for effective management of groundwater monitoring system and groundwater resources. In the present study, three different time series models for the prediction of groundwater level in response to rainfall were built, those are transfer function noise model (TFNM), artificial neural network (ANN), and adaptive neuro fuzzy interference system (ANFIS). The models were applied to time series data of Boen, Cheolsan, and Hongcheon stations in National Groundwater Monitoring Network. The result shows that the model performance of ANN and ANFIS was higher than that of TFNM for the present case study. As lead time increased, prediction accuracy decreased with underestimation of peak values. The performance of the three models at Boen station was worst especially for TFNM, where the correlation between rainfall and groundwater data was lowest and the groundwater extraction is expected on account of agricultural activities. The sensitivity analysis for the input structure showed that ANFIS was most sensitive to input data combinations. It is expected that the time series model approach and results of the present study are meaningful and useful for the effective management of monitoring stations and groundwater resources.

**Key words :** Groundwater level, Transfer function noise model, Artificial neural network, Adaptive neuro fuzzy interference system

### 1. 서 론

최근 가뭄에 따른 지표수 물 공급의 부족 및 지역 불균형으로 인한 피해 사례가 증가하면서 안정적인 수자원 공급의 필요성이 대두되고 있다. 지하수는 일반적으로 급격한 기상 변화에 의한 영향을 비교적 느리고 적게 받기 때문에 이러한 문제에 대한 중요한 해결 방안이 될 수 있다. 또한 지표수에 비해 오염에 대한 취약성이 적어 물 안보 측면에서도 주요한 역할을 할 수 있는 국가 자원이다. 이러한 지하수자원의 분포 및 변화를 파악하기 위해

다양한 목적의 관측망이 설치, 운영되고 있다.

지하수 관측망 및 지하수자원을 효율적으로 관리, 운영하기 위해서는 지하수위 변화를 정확히 예측하는 것이 필요하다. 지하수위 변화 예측 방법으로 크게 물리 모델링 기법과 시계열 모델링 기법을 들 수 있다. 물리 모델링 기법에서는 물리적인 개념을 바탕으로 수학적 지배방정식을 세우고 대상 지역에 대한 해석해 또는 수치해를 구하게 된다(Rai and Singh, 1995; Knotters and Bierkens, 2000). 물리 모델링 기법은 대상 지역 전체에 대한 지하수위의 장기적인 시공간적인 분포의 변화를 예측할 수 있

\*Corresponding author : hyoon@kigam.re.kr

Received : 2014. 5. 2 Reviewed : 2014. 5. 22 Accepted : 2014. 5. 22

Discussion until : 2014. 8. 31

는 반면 정확한 예측을 위해서는 매질의 물성값과 관련된 방대한 입력 자료가 요구된다는 특징이 있다. 시계열 모델링 기법은 대상 지역의 지하수위 시계열 자료와 지하수위에 영향을 주는 입력 변수의 시계열 자료 간의 반응함수를 구성하고 이를 이용하여 예측을 수행한다. 시계열 모델링 기법은 입·출력 자료에 대한 주기적인 관측 자료가 필요하다는 제한점이 있지만 그 외의 자료가 필요하지 않고 특히 단기 예측에 대해 일반적으로 예측 능력이 뛰어나다. 또한 입·출력 자료 취득되는 지점에 대한 예측이기 때문에 공간적 대표성에 대한 유효 범위 및 장기 예측과 관련된 고찰이 필요한 방법이다. 그러나 최근 관측망의 규모가 지속적으로 확대되면서 시계열 모델링 기법의 활용성이 커지고 있다.

지하수위 시계열 예측 모델은 크게 선형 모델과 비선형 모델로 나눌 수 있다. 전통적인 선형 모델로 Box and Jenkins(1976)에 의해 제안된 자기회귀누적이동평균(autoregressive integrated moving average: ARIMA) 모형과 전이함수 잡음 모형(transfer function noise model: TFNM)이 이용되어 왔다(Tankersley et al., 1993; van Geer and Zuur, 1997; Yi et al., 2004; Yi and Lee, 2004). 비선형 시계열 모델은 일반적으로 비선형 학습 알고리즘을 이용하여 구성되고 비교적 최근 개발 및 적용 연구가 진행되고 있다. 가장 대표적인 모델은 인간의 뇌 구조를 모사하여 고안된 인공신경망(artificial neural network, ANN) 모델로 수자원 변수와 관련하여 초기에 주로 지표수를 대상으로 연구가 진행되었고(French et al., 1992; Karunanithi et al., 1994; Zealand et al., 1999; Hu et al., 2005) 지하수 분야에 대한 적용 연구로 확대되고 있다(Coulibaly et al., 2001; Mohanty et al., 2010). 최근 인공신경망과 퍼지 로직을 결합하여 입력 설정에 있어 설명적 요소를 고려할 수 있게 하는 적응형 뉴로 퍼지 추론 시스템(adaptive neuro fuzzy interference system)이 고안되었고(Jang, 1993) 수자원 문제에 대한 적용성 평가 연구가 진행되고 있다(Hong and White, 2009; Kisi and Shiri, 2012).

지하수위 시계열 예측 모델 개발 및 국내 적용 연구를 살펴보면 Yi et al.(2004)은 ARIMA 모형을 이용하여 국가 지하수관측망의 지하수위 자료 분석 연구를 수행하였고 Yi and Lee(2004)는 국내 가스 저장소 내 지하수위 변화를 예측하는데 TFNM을 활용하였다. Yoon et al.(2011)은 비선형 시계열 모델 및 직접 예측 방법을 이용하여 동해 연구지역의 해안 대수층 지하수위 변화 예측 연구를 수행하였고 Yoon et al.(2013)은 비선형 시계열

모델 및 반복 예측 방법의 국가지하수 관측망 적용 연구를 진행하였다. 현재까지 연구에서는 선형 모델과 비선형 모델의 비교 연구 및 직접 예측 방법을 이용한 국가지하수 관측망 자료에의 적용 연구는 이루어지지 않아 이에 대한 고찰이 필요하다고 판단된다. 이에 본 연구의 목적은 강우에 대한 지하수위 변화 예측에 대해 선형 모델인 TFNM과 비선형 모델인 ANN 및 ANFIS 모델을 구성하고 이를 강우-지하수위 관계가 비교적 뚜렷한 국가지하수 관측망 자료에 적용하여 모델들의 예측 결과를 비교 평가하는 것이다. 또한 입력 구조에 따른 모델의 민감도를 분석하고 ANFIS의 국내 적용성을 평가하고자 하였다.

## 2. 연구방법 및 자료

### 2.1. 시계열 모델의 예측 방법

시계열 모델을 이용한 예측 방법은 크게 직접 예측(direct prediction)과 반복 예측(recursive prediction)의 두 가지로 나눌 수 있다(Ji et al., 2005; Herrera et al., 2007). 과거 특정 시점부터 현재까지의 자료를 이용하여 시계열 모델을 구성하였을 때 얼마만큼의 시점 이후부터 미래를 예측할 것인지를 나타내는 용어는 일반적으로 Lead time이라 정의된다. 직접 예측 방법은 각 Lead time에 해당하는 모델들을 만들고 이를 이용해 예측하는 방법이다. 항상 실제 관측값을 입력으로 사용하기 때문에 예측력이 뛰어난 특징이 있지만 다수의 모델을 만들어야 한다는 제한점을 가지고 있다. 반복 예측 방법은 Lead time이 1인 모델을 만들고 이를 이용하여 예측된 값을 반복적으로 입력 성분으로 이용하는 방법이다. 반복 예측 방법은 하나의 모델만을 반복적으로 이용할 수 있지만 예측의 각 시간 단계마다 관측값이 아닌 예측값을 입력으로 사용하면서 오차가 누적되어 장기 예측 시 오류를 범할 가능성이 있다.

### 2.2. 전이함수 잡음 모델

시계열 자료의 분석 및 예측에 대한 전통적인 방법으로 단변량 시계열 모형인 ARIMA(AutoRegressive Integrated Moving Average, Box and Jenkins, 1976) 모형이 이용되어 왔다. 일반적으로 지하수위는 강우에 의해 변동하게 되므로 지하수위 변화를 예측하는데 있어 강우의 영향을 고려한 다변량 시계열 모형을 고려하는 것이 보다 효과적일 수 있다.

TFNM은 대표적인 다변량 시계열 모형으로 어떤 동적 시스템에 입력 시계열  $\{X_t\}$ 가 주어졌을 때, 출력 시계열

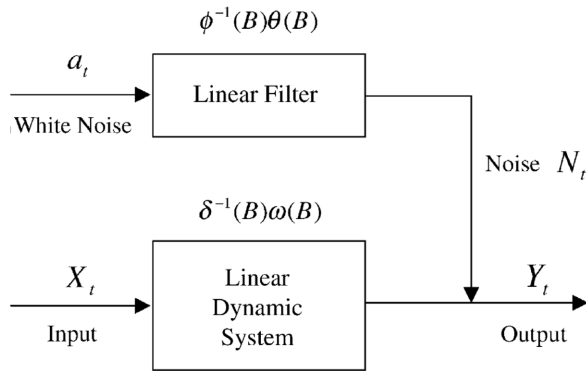


Fig. 1. A schematic diagram of TFNM (Box and Jenkins, 1976).

$Y_t$ 을 충격반응함수에 의한 필터 전이함수모형과 잡음모형의 결합으로 표현한다(Fig. 1). 지하수위를  $g$ , 강우를  $p$ 라 했을 때 본 연구에 적용된 TFNM은 다음과 같이 표현된다.

$$g_t = \frac{\omega(B)}{\delta(B)}p_{t-b} + \frac{\theta(B)}{\phi(B)}a_t \quad (1)$$

여기서,  $t$ 는 시간,  $b$ 는 지연 매개변수,  $a_t$ 는 백색잡음을 나타낸다.  $B$ 는 후진연산자로  $B^i = B_{i-}$ 로 정의된다. 전이함수 모형 및 잡음모형의 각 함수들은 다음과 같이 표현된다 (Box and Jenkins, 1976).

$$\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s \quad (2)$$

$$\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r \quad (3)$$

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (4)$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (5)$$

일반적으로 시계열 모형의 전개는 모형 식별, 추정, 진단의 세 가지 단계로 진행되며 시계열 모형의 구조는 모형 식별 단계에서 시계열 분석을 통해 정해진다(Box and Jenkins, 1976; Hipel et al, 1977). 그러나 본 연구의 주목적 중 하나는 시계열 모델들 간의 비교 평가이므로 동일한 구조의 모델을 임의로 설정하여 구성하였으며 이 후 입력 구조에 대한 민감도 분석을 수행하였다.

### 2.3. 인공신경망

일반적으로 ANN은 입력층, 은닉층, 출력층으로 구성되어 있고 각 층은 다수의 노드들로 이루어져 있으며 각 층간의 노드들은 일정한 연결강도로 연결되어 있다(Fig. 2). ANN 모델 구성 과정은 학습 단계라 하며 최종 목적은 주어진 입출력 자료에 가장 적합한 모델 파라미터들과 각 연결강도들을 결정하는 것이다. 본 연구에서는 가장 대

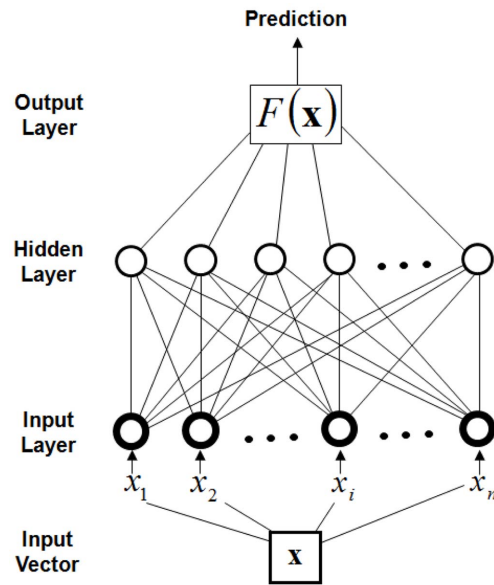


Fig. 2. Model structure of ANN (Yoon et al., 2011).

표적인 ANN의 구조 및 학습 방법을 선택하여 적용하였다. ANN 구조로는 최근까지 수자원 변수 예측 관련 연구에서 가장 많이 사용되어온 1개 은닉층 구조를 이용하였으며(Maier and Dandy, 2000), 모델 학습 방법으로 역전파 알고리즘 (Back Propagation Algorithm) (Rumelhart et al, 1986)을 적용하였다.

ANN 구조에서 입력층의 입력 노드에 자료가 주입되었을 때 출력값이 연산되는 과정은 다음과 같이 표현된다.

$$y_n = F_n \left( \sum_{m=1}^L w_{nm} x_m + b_n \right) \quad (6)$$

여기서,  $m$ ,  $n$ 은 각각 이전 층과 현재 층의 노드를 나타내고,  $x$ 와  $y$ 는 각각 이전 층과 현재 층의 노드 값을 나타낸다.  $w$ 는  $x$ 와  $y$  사이의 연결강도를,  $b$ 는 현재 층에서의 편차 값을 나타낸다.  $F$ 는 은닉층과 출력층의 활성화함수로 본 연구에서는 은닉층의 활성화함수로 로그시그모이드 함수를, 출력층의 활성화함수로 선형함수를 활용하였다.

역전파 알고리즘을 이용한 ANN 학습과정, 즉 연결강도 보정 과정은 다음과 같이 표현된다(Rumelhart et al., 1986).

$$\Delta w^{i+1} = \beta(\Delta w^i) + (1-\beta)\gamma \left( -\frac{\partial E^i}{\partial w^i} \right) \quad (7)$$

여기서,  $\Delta w^{i+1}$ 은  $w^{i+1} - w^i$ 를 나타내고,  $E^i$ 는  $i$ 번째 반복 단계에서 관측값( $t_i$ )과 예측값( $y_i$ ) 사이의 오차로  $\sum_k (t_k^i - y_k^i)$ 로 정의된다.  $\gamma$ 는 학습속도 나타내고  $\beta$ 는 모멘텀으로

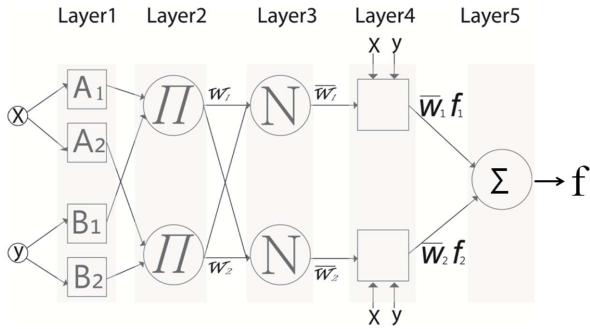


Fig. 3. Model structure of ANFIS (Jang, 1993).

ANN의 지역해 탐색 문제를 보완하기 위한 도구로 활용된다(Rumelhart et al., 1986).

2.4. 적응형 뉴로 퍼지 추론 시스템

ANFIS는 퍼지시스템의 장점인 추론 능력과 ANN의 장점인 학습 능력을 결합한 수학 구조로 Jang(1993)에 의해 제안되었다. ANFIS는 주어진 입·출력 자료를 이용하여 자료에 가장 적합한 룰과 소속 함수를 학습하게 된다. 학습이라는 측면에서는 인공신경망과 유사하지만, ANFIS는 if then 이라는 퍼지추론을 이용한다. 퍼지추론은 퍼지집합에 기초하는데, 퍼지집합은 명확히 구분할 수 없는 경계를 표현 할 수 있는 집합으로 원소는 소속의 부분적인 값을 정의하여 부정확한 사건들을 나타낼 수 있다. 이 소속 값을 정의 하는에는 소속함수(Membership function)가 쓰이며, 퍼지 집합은 0과 1사이의 소속함수 값을 이용하고, 소속함수는 다루고자 하는 문제의 형태에 따라서 다양하게 선택하고 정의할 수 있다.

ANFIS의 구조는 노드와 링크로 구성되어 있으며, 각 노드들의 내부에는 출력 값에 영향을 주는 파라미터들을 포함하고 있다. 이 네트워크들이 지속적으로 입·출력 자료의 관계를 학습하면서, 각 노드 내부의 파라미터들이 변하여 입·출력 자료를 설명하는 구조를 만들 수 있게 된다. 본 연구에서는 Jang(1993) 이 제안한 Sugeno 퍼지를 사용한 구조를 이용하였다(Fig. 3). Sugeno 퍼지를 이용한 ANFIS 구조는 다음과 같은 룰을 갖는다.

$$\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \quad \text{THEN } f_1 = p_1x + q_1y + r_1 \quad (8)$$

$$\text{If } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \quad \text{THEN } f_2 = p_2x + q_2y + r_2 \quad (9)$$

또한 ANFIS는 입력층 외에 5개의 Layer가 존재하고 각 Layer에서는 다음과 같은 연산이 이루어진다.

- Layer 1 : 주어진 변수들의 소속값을 계산한다.
- Layer 2 : 각 소속값의 곱을 통해 룰의 영향력을 계산한다.
- Layer 3 : 영향력의 강도를 정규화 한다.
- Layer 4 : 출력값 계산을 위한 선형 파라미터들을 계산한다.
- Layer 5 : 전체 출력값을 계산한다.

본 연구에서는 가장 단순한 삼각형 형태의 소속함수를 설정하고 모형을 구성하였다.

이상에서와 같이 TFN, ANN, ANFIS의 세 가지 모형을 이용하여 강우에 대한 지하수위 변화를 예측하는 시계열 모델을 작성하였다. 작성 도구로는 각각 SPSS, C언어, MATLAB을 활용하였다.

2.5. 연구자료

TFNM, ANN, ANFIS 시계열 모델의 비교평가를 위하여 국가지하수관측망 자료를 활용하였다. 국가지하수관측망은 장기적이며 주기적으로 지하수위와 수질을 관측하고 이를 토대로 지하수의 부존 및 유동 특성과 배경 수질을 파악을 통하여 지하수자원의 효율적인 이용과 관리를 위한 합리적인 개발 계획과 보전계획을 수립할 목적으로 운영되고 있다. 1995년부터 설치되어 왔으며 2013년 기준 전국 348개소의 관측소가 운영되고 있다. 본 연구에서는 국가지하수관측망 지하수위 및 강우 시계열 자료를 수집·분석 하였고 이중 자료의 연속성이 뛰어나고 강우-지하수위 반응관계가 뚜렷한 충북보은보은(BE), 경기광명철산(CS), 강원홍천홍천(HC) 관측소의 2003년부터 2009년까지 지하수위 일변화 자료를 적용 대상으로 하였다. 강우 자료로 보은, 구로, 홍천 기상관측소의 일 단위 강우 자료를 이용하였다(Fig. 4). 강우와 지하수위의 교차상관 분석 결과 BE, CS, HC 관측소에 대해 각각 지연시간 2일에서 0.435, 0.589, 0.630의 최대 상관 계수를 보여주었다. 총 7년의 자료 가운데 2003년부터 2005년 자료를 모델 구성에 이용하고, 작성된 모델을 이용하여 2006년부터 2009년까지의 지하수위 변화를 예측하였다.

시계열 모델의 입력 구조를 설정하는데 있어 시행착오법, 통계적 방법 등 다양한 방법이 이용되지만 본 연구의 목적은 지하수위 변화 예측에 대한 TFNM, ANN, ANFIS 모델의 비교평가이므로 강우와 지하수위에 대해 각각 과거 3일까지의 정보를 이용하는 경우를 가정하였으며, 각 경우에 대해 각각 1에서 5의 lead time을 갖는 모형을 구성하고 예측 결과를 비교하였다.

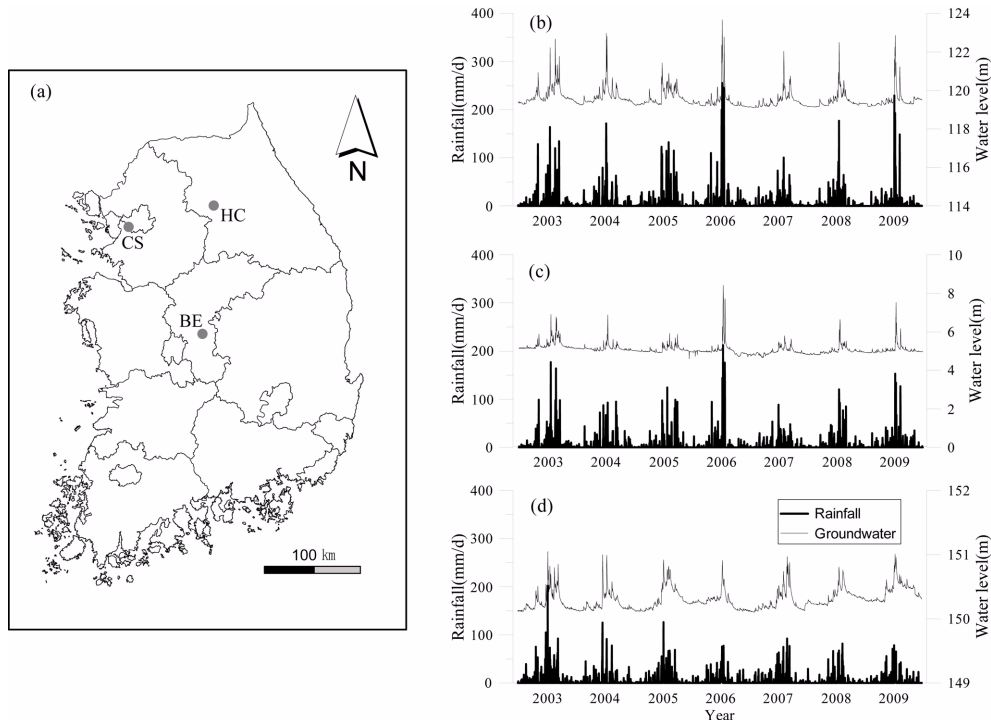


Fig. 4. Location of groundwater stations (a) and time series data of rainfall and groundwater level at (b) BE, (c) CS, and (d) HC stations.

## 2.6. 오차 지표

본 연구에서는 각 모형에 대한 적합성 및 예측 능력을 비교하고 평가하기 위해 평균 제곱 오차 제곱근(RMSE: root mean squared error), 상관계수(CORR: correlation coefficient), 평균 상대 백분율 오차(MARE: mean absolute relative error)의 세 가지 오차 지표를 이용하였다. RMSE 및 CORR은 시계열 모델의 예측력을 평가하는 데 있어 대표적으로 이용되는 오차 지표이다. 이 외에 서로 다른 범위를 가지는 자료에 대한 오차 지표로 평균 절대 백분율 오차(MAPE: mean absolute percentage error)가 사용되지만 본 연구에서는 지하수 심도가 아닌 지하수위 자료를 이용하였기 때문에 MAPE를 이용하는 것이 효율적이지 않다고 판단된다. 따라서 본 연구에서는 오차 절대값을 해당 관측 시계열의 최대 변동폭으로 나눈 값을 이용하는 MARE를 정의하고 이를 활용하였다. 각 오차 지표의 수학적 표현은 다음의 식과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (obs_i - est_i)^2} \quad (10)$$

$$CORR = \frac{\frac{1}{n} \sum_{i=1}^n (obs_i - \overline{obs_i})(est_i - \overline{est_i})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (obs_i - \overline{obs_i})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (est_i - \overline{est_i})^2}} \quad (11)$$

$$MARE = \frac{1}{n} \left( \sum_{i=1}^n \frac{|obs_i - est_i|}{obs_{max} - obs_{min}} \right) \times 100 \quad (12)$$

여기서  $n$ 은 자료의 수,  $obs$ 는 관측값,  $est$ 는 예측값,  $\overline{obs}$ 와  $\overline{est}$ 는 각각 관측값과 예측값의 평균값,  $obs_{max} - obs_{min}$ 은 관측값의 최대 변동폭을 의미한다. BE, CS, HC 관측소의 최대 변동폭은 각각 0.95, 3.82, 4.55 m이다.

## 3. 연구결과 및 토의

### 3.1. Lead time 1 예측 결과

각 관측소에 대한 Lead time 1에서의 모델들의 지하수위 예측 결과 전체적인 오차 지표를 살펴보면, RMSE는 0.025 m~0.269 m로, CORR은 0.748~0.988, MARE는 0.87%~11.7%의 분포를 보여준다(Table 1). 모델 별 오차 지표를 비교해 보면 모든 오차 지표에서 TFNM 보다 ANN 및 ANFIS가 비교적 높은 예측력을 보여주는 것을 알 수 있다. 이와 같은 결과는 강우-지하수위 사이에 비선형적인 관계가 존재할 수 있고 이를 학습하는데 있어 선형적 관계를 가정하는 TFNM보다 비선형성을 고려할 수 있는 ANN 및 ANFIS가 보다 효과적이기 때문인 것으로 판단된다. MARE 값을 통해 관측소 별 오차를 비교해보면 모든 모델에 대해 BE 관측소에 대한 오차가 가

**Table 1.** Model performance of the three models with lead time 1

Station	TFNM			ANN			ANFIS		
	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)
BE	0.137	0.748	11.7	0.025	0.988	1.29	0.043	0.962	1.91
CS	0.192	0.896	4.02	0.102	0.939	0.94	0.153	0.851	1.23
HC	0.269	0.900	4.38	0.189	0.938	1.29	0.135	0.950	0.87

**Table 2.** Model performance with lead times at BE station

Lead time	TFNM			ANN			ANFIS		
	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)
1	0.137	0.748	11.7	0.025	0.988	1.29	0.043	0.962	1.91
2	0.139	0.722	11.8	0.045	0.958	2.37	0.052	0.942	2.48
3	0.141	0.700	11.9	0.054	0.938	3.08	0.061	0.918	3.16
4	0.143	0.677	12.0	0.061	0.921	3.61	0.067	0.898	3.80
5	0.145	0.655	12.1	0.068	0.901	4.26	0.072	0.882	4.11

장 큰 것을 알 수 있다. 이에 대한 원인으로 강우 및 지하수 관측소 사이의 거리의 차이를 생각해 볼 수 있다. 그러나 BE, CS, HC 지하수 관측소와 기상 관측소와의 거리는 각각 0.5, 0.9, 1.4 km로 모두 가깝게 위치하고 있고 BE 관측소의 경우가 가장 가까워서 기상관측소까지의 거리가 원인이라고 판단할 수 없다. 다만 항공 사진을 통해 지하수 관측소 주변을 살펴보면 CS 및 HC와 달리 BE 관측소 주변에는 논밭이 다수 존재하고 있어 지하수 위 자료에 지하수 양수에 의한 인위적 영향이 포함되어 있을 가능성이 있다. 양수정과 관측정이 가까이 위치하고 있고 단일 양수정인 경우 양수의 영향이 자료에 뚜렷하게 나타나지만 양수정과 거리가 가깝지 않고 다중 양수정의 영향이 복잡적으로 존재한다면 자료에서 눈으로 양수 영향 관찰할 수 없더라도 강우-지하수위 상관성을 저해할 수 있다고 판단된다. 또한 본 연구에서는 강우 입력만을 고려하고 있어서 이러한 인위적인 영향이 고려되기 어려운 상황이기 때문에 BE 관측소 예측 오차가 비교적 크게 나타났을 가능성이 있다. 이 외에 TFNM과 ANN 모델은 CS 관측소에서 가장 좋은 예측 결과를, ANFIS 모델은 HC 관측소에서 가장 좋은 예측 결과를 보여주었다(Table 1).

### 3.2. Lead time 변화에 따른 예측 결과

Lead time에 따른 각 모델의 반응을 관찰하기 위하여 Lead time을 1부터 5까지 증가시키면서 모델을 작성하고 각 모델 및 관측소 별 예측 결과를 비교평가 하였다. Table 2는 BE 관측소에서의 Lead time에 따른 각 모델 별 오차지표를 보여준다. 모든 모델들이 Lead time이 증

가할수록 오차가 증가하는 것을 볼 수 있다. TFNM의 경우 예측 오차가 가장 크게 나타났으나 Lead time 증가에 따른 오차 증가 폭은 비교적 크지 않았다. RMSE를 기준으로 보면 ANN의 예측 오차가 ANFIS보다 약간 낮은 것으로 나타나지만 전체적으로 큰 차이를 보이지 않았다. ANN과 ANFIS는 Lead time에 따른 예측 오차 및 오차 증가 폭이 유사하게 변화하는 것을 알 수 있다. Fig. 5a에서 BE 관측소에서 Lead time 1, 3, 5에 대한 각 모델의 예측 결과를 도시하였다. TFNM의 예측 결과 지하수위 기저가 변화하는 구간에서 오차가 크게 발생하는 것을 알 수 있다. 또한 Lead time이 증가할수록 지하수위 침투 부분에서 과소 예측이 심해지는 것을 관찰할 수 있다. 관측값과 예측값의 1:1 도시 결과를 보면 분포의 폭이 비교적 넓고 대체로 과소 예측되는 경향이 있음을 보여준다. ANN 모델의 예측 결과의 경우 TFNM 모델보다 예측력이 크게 증가하지만 역시 Lead time이 증가할수록 침투 부분에서의 과소 예측 현상이 심해지는 것을 볼 수 있다(Fig 5b). 관측값과 예측값의 1:1 도시 결과 TFNM 모델에 비해 분포 폭이 크게 줄지만 과소 예측되는 경향이 있음을 알 수 있다. ANFIS 모델의 예측 결과 전체적으로 ANN 모델 결과와 유사하지만 침투 부분에서 큰 폭으로 벗어나는 예측 오차들이 발생하는 것을 볼 수 있다. 이는 모델의 학습 단계에서 이용된 자료군의 범위를 넘어선 자료에 대해 ANFIS가 적절하게 예측하지 못하고 있음을 나타낸다. 본 연구에서는 ANFIS의 소속함수를 설정할 때 입력 가능 자료의 범위를 모델 학습단계에서 보정함으로써 이러한 문제를 최대한 해결하려고 하였으나 소

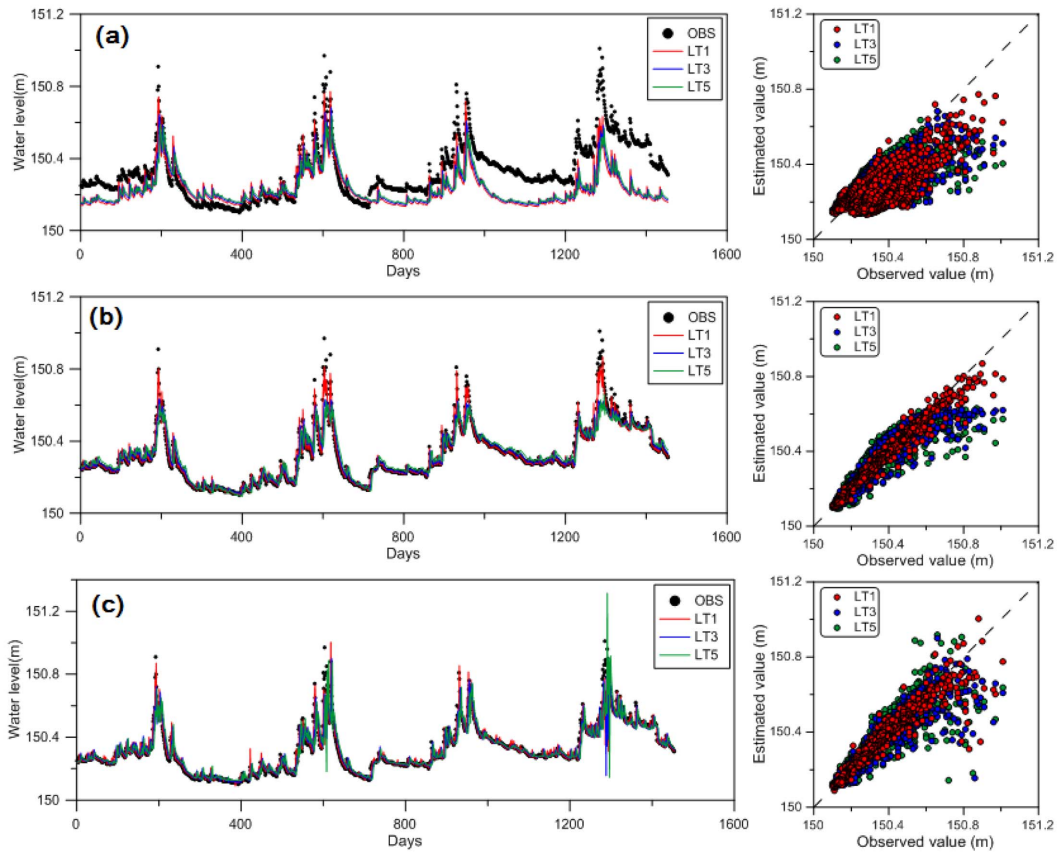


Fig. 5. Results of groundwater level prediction at BE station: (a) TFNM; (b) ANN; (c) ANFIS.

Table 3. Model performance with lead times at CS station

Lead time	TFNM			ANN			ANFIS		
	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)
1	0.192	0.896	4.02	0.102	0.939	0.94	0.153	0.851	1.23
2	0.235	0.752	4.50	0.173	0.803	1.64	0.174	0.808	1.68
3	0.252	0.675	4.72	0.195	0.741	1.91	0.196	0.735	1.86
4	0.257	0.650	4.78	0.209	0.699	2.32	0.226	0.627	2.15
5	0.266	0.606	4.92	0.219	0.658	2.38	0.230	0.609	2.34

Table 4. Model performance with lead times at HC station

Lead time	TFNM			ANN			ANFIS		
	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)
1	0.269	0.900	4.38	0.189	0.938	1.29	0.135	0.950	0.875
2	0.351	0.757	5.28	0.222	0.858	2.04	0.244	0.823	1.89
3	0.379	0.683	5.72	0.257	0.800	2.50	0.245	0.800	2.50
4	0.399	0.625	6.00	0.285	0.750	2.80	0.326	0.669	2.91
5	0.405	0.596	6.11	0.306	0.702	3.14	0.339	0.628	3.15

속함수에 다양한 종류에 대한 고찰이 더 필요할 것으로 판단된다. 관측값과 ANFIS 모델 예측값의 1:1 도시 결

과 앞서 언급한 바와 같이 일부 큰 폭으로 벗어나는 값들이 존재하고 전체적으로 과소 평가되는 경향은 줄어들

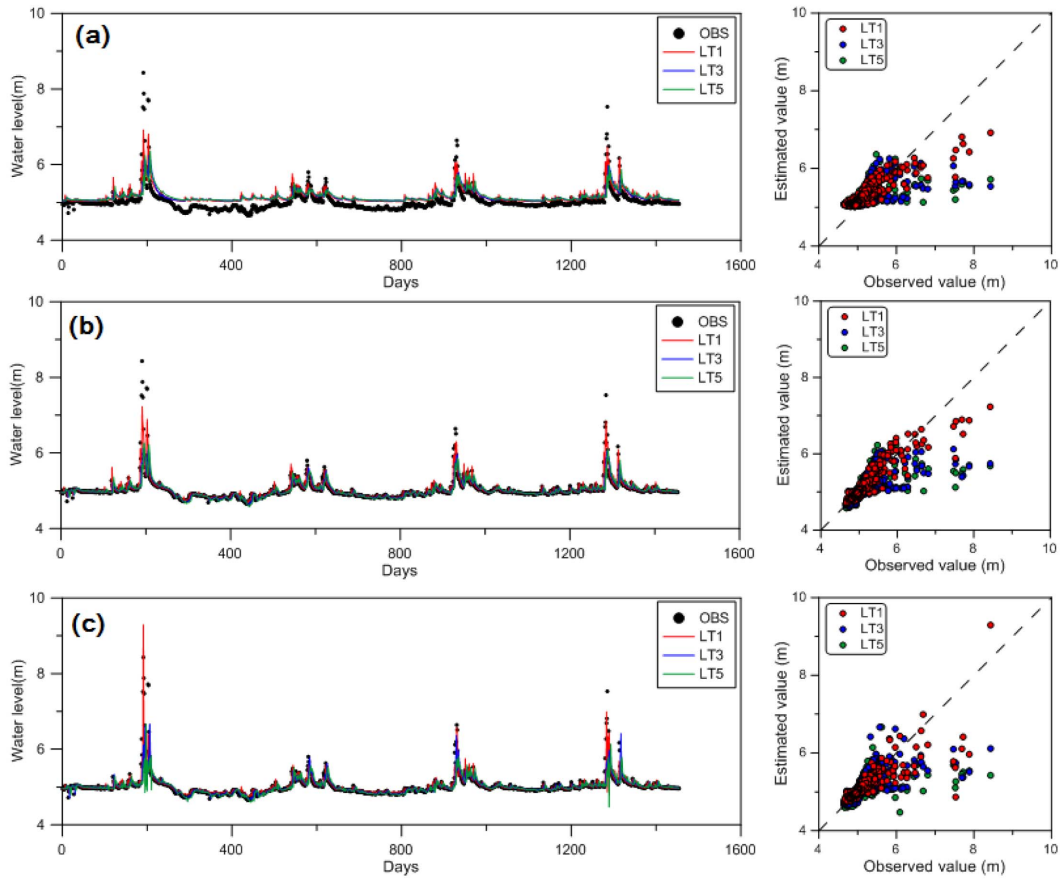


Fig. 6. Results of groundwater level prediction at CS station: (a) TFNM; (b) ANN; (c) ANFIS.

Table 5. Sensitivity analysis for the input structure at BE station

Model	TFNM			ANN			ANFIS		
	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)	RMSE (m)	CORR	MARE (%)
P1G1	0.135	0.748	11.2	0.026	0.988	1.36	0.026	0.987	1.29
P1G2	0.135	0.750	11.2	0.027	0.988	1.42	0.027	0.986	1.38
P1G3	0.135	0.749	11.3	0.026	0.987	1.39	0.031	0.980	1.43
P2G1	0.136	0.746	11.5	0.025	0.989	1.24	0.041	0.965	2.11
P2G2	0.137	0.748	11.7	0.027	0.988	1.36	0.030	0.981	1.41
P2G3	0.137	0.749	11.7	0.026	0.988	1.45	0.037	0.970	1.54
P3G1	0.136	0.746	11.6	0.025	0.989	1.38	0.032	0.980	1.33
P3G2	0.136	0.748	11.7	0.027	0.988	1.51	0.025	0.988	1.26
P3G3	0.137	0.748	11.7	0.025	0.988	1.29	0.043	0.962	1.91

것을 볼 수 있다. 이러한 결과는 CS 및 HC 관측소에서  
도 유사하게 관찰되었다(Tables 3, 4, Figs. 6, 7).

관측소 별 Lead time에 따른 MARE 값의 변화를 살  
펴보면 TFNM 모델의 오차가 ANN 및 ANFIS 모델보다  
큰 것을 알 수 있다(Fig. 8). TFNM과 ANN 및 ANFIS  
모델의 MARE 차이는 BE 관측소에서 가장 크게 나타났

다(Fig. 8a). 이는 앞서 언급한바와 같이 교차상관분석 결  
과 최대 상관계수에서의 지연시간은 동일하나 BE 관측소  
의 강우-지하수위 자료의 선형적인 상관성이 비교적 낮고  
다른 관측소에 비해 강우 외의 요인이 지하수위 변화에  
영향을 미쳤을 가능성이 높아 TFNM 모델의 적합성이 떨  
어지기 때문인 것으로 판단된다. 전체적으로 각 모델에 대



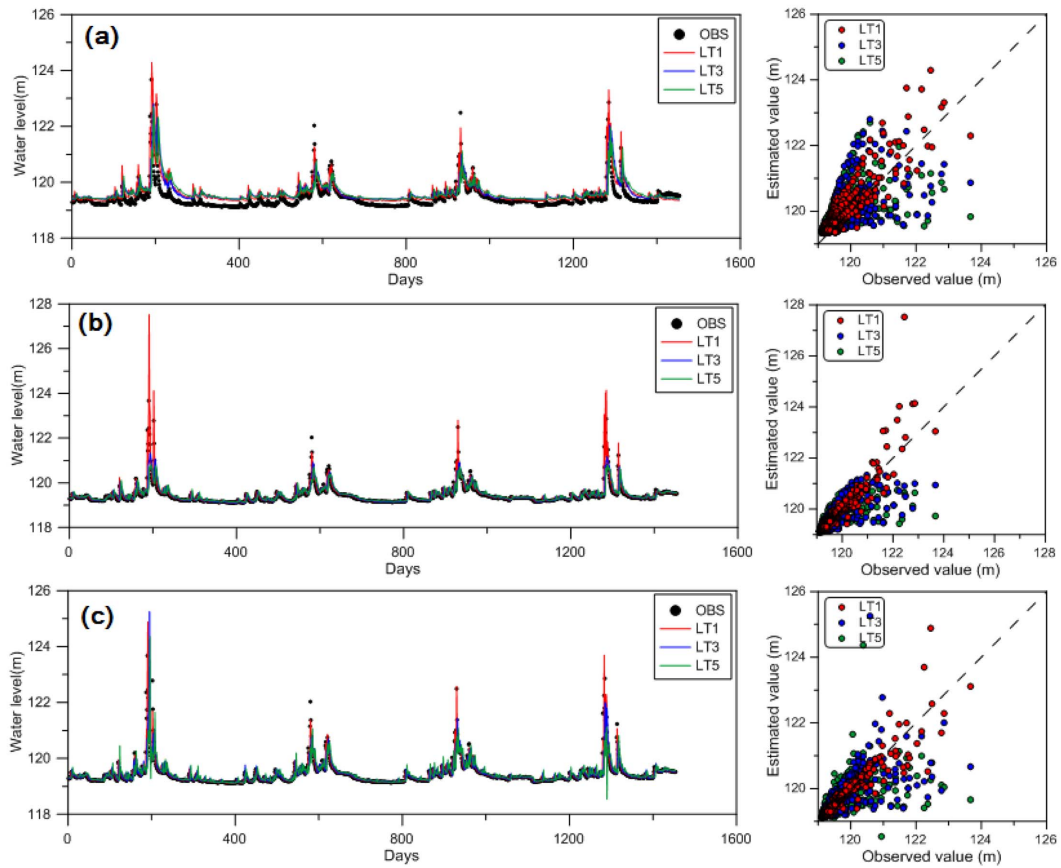


Fig. 7. Results of groundwater level prediction at HC station: (a) TFNM; (b) ANN; (c) ANFIS.

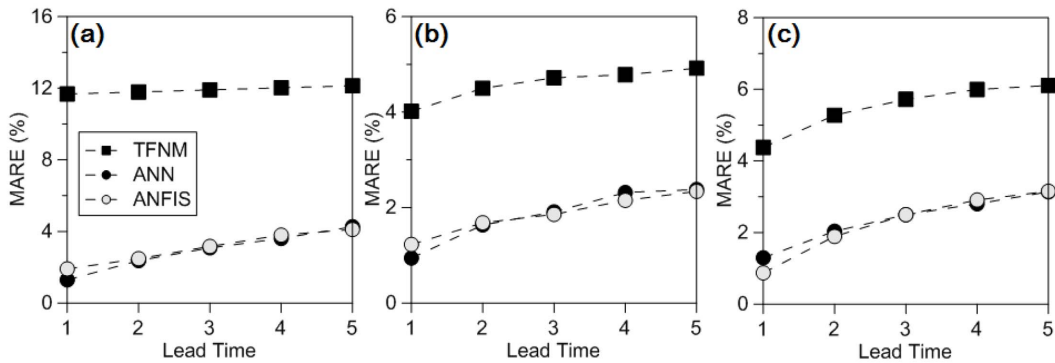


Fig. 8. MARE values of the three models with lead times at (a) BE, (b) CS, and (c) HC stations.

해 Lead time이 증가함에 따라 MARE값이 증가하는 것을 볼 수 있다. ANN 및 ANFIS 모델 간의 MARE는 큰 차이가 없어 본 연구 자료에 대해서는 우위를 평가하기 어려웠다.

### 3.3. 모델 입력 구조에 따른 민감도 분석

본 연구에서는 모델 간 예측 결과 비교를 위해 강우

및 지하수위 자료에 대해 과거 3일까지의 관측값을 입력 성분으로 이용하는 입력 구조(P3G3)를 공통적으로 이용하였다. 입력 구조에 대한 각 모델들의 민감도를 분석하기 위해 BE 관측소를 대상으로 강우 및 지하수위 자료에 대해 1일부터 3일까지의 입력 성분 조합으로 9개의 입력 구조를 지닌 모델들을 구성하고 예측 결과를 비교하였다 (Table 5). 분석 결과 TFNM 및 ANN 모델은 입력 구조

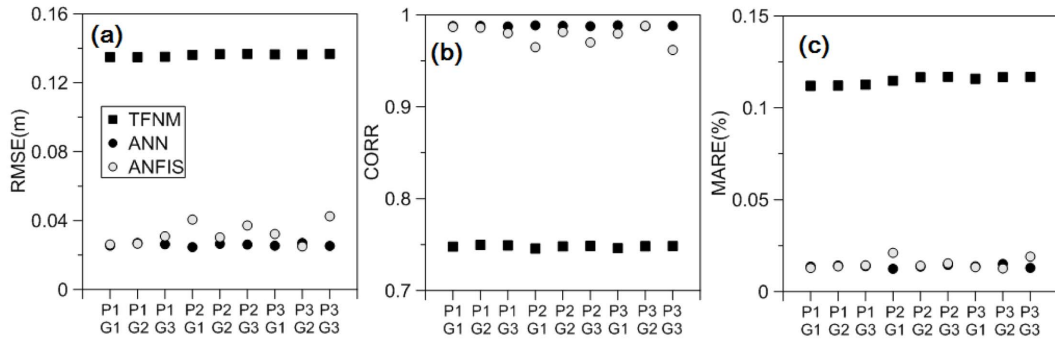


Fig. 9. Result of sensitivity analysis for the input structure at BE station: (a) RMSE, (b) CORR, (c) MARE.

에 따라 예측 오차가 큰 변화를 보이지 않았다(Fig. 9). ANFIS 모델의 경우 비교적 높은 민감도를 보였고 RMSE를 기준으로 P3G2 입력 구조에서 가장 낮은 오차를 보여주었다. 본 연구의 민감도 분석을 통해 ANFIS 모델의 경우 입력 시계열의 종류 결정 및 각 시계열 별 과거 관측 자료 사용 정도 결정 등 입력 구조를 설정하는데 보다 주의를 기울여야 한다는 것을 알 수 있다.

#### 4. 요약 및 결론

본 연구에서는 강우에 의한 지하수위 변화 예측에 대해 직접 예측(direct prediction) 방법을 이용한 시계열 모델의 성능을 비교 평가 하였다. 시계열 예측 모델로 대표적인 선형 시계열 모델인 TFNM과 비선형 학습 알고리즘에 기반한 ANN 및 ANFIS를 고려하였고 충북보은보은, 경기광명철산, 강원홍천홍천 국가지하수관측소 지하수위 자료 및 기상관측소 강우 자료를 이용하였다.

지하수위 예측 결과 TFNM 모델에 비해 ANN 및 ANFIS 모델의 예측 오차가 더 작은 것으로 나타났다. ANN 및 ANFIS의 예측 결과는 대체로 유사한 것으로 평가되었다. Lead time이 증가할수록 지하수위 침투 부분에서 과소 예측되는 경향을 보여주며 예측 오차가 증가하였다. ANFIS의 경우 과소 예측되는 경향이 적지만 학습 구간을 벗어나는 예측에 대해 상하로 크게 벗어나는 현상이 발생할 수 있다는 것을 알 수 있었다. 이후 강우-지하수위 변화 예측에 있어 적합한 ANFIS 모델 파라미터 설정과 관련된 추가적인 연구가 수행되어야 할 것이다.

관측소 별 예측 결과를 살펴보면 강우-지하수위 간 선형 상관성이 낮고 주변에 농업활동이 예상되는 충북보은보은 관측소의 예측 오차가 MARE 기준으로 특히 TFNM 모델에 대해 가장 높았다. 입력 구조에 대한 시계열 모델의 민감도 분석 결과 TFNM 및 ANN에 비해

ANFIS 모델의 민감도가 큰 것으로 조사되었다. ANFIS 모델 적용시 입력 시계열의 선정 및 과거 관측 자료 이용 정도 등 입력 구조의 설정에 주의를 기울여야 할 것으로 판단된다.

시계열 예측 방법 중 본 연구에서 이용된 직접 예측 방법은 지하수 자동 관측소의 이상 자료 판별 및 결측 자료 보완 연구 등에 적용될 수 있다. 본 연구에서 이용된 시계열 모델, 접근 방법 및 연구 결과는 이후 지하수위 관측망의 효율적인 운영과 지하수 자원 관리에 활용될 수 있을 것으로 기대한다.

#### 사 사

본 연구는 국토교통부가 출연하고 국토교통과학기술진흥원에서 위탁시행한 물관리연구사업(11기술혁신C05)에 의한 '수변지하수활용고도화'연구단의 연구비 지원에 의해 수행되었습니다.

#### References

Box, G.E.P. and Jenkins, G.M., 1976, Time Series Analysis-Forecasting and Control, Holden-Day, San Francisco, California, USA, 575 p.

Coulibaly, P., Anctil, F., Aravena, R., and Bobee, B., 2001, Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.*, **37**(4), 885-896.

French, M.N., Krajewski, W.F., and Cuykendall, R.R., 1992, Rainfall forecasting in space and time using a neural network. *J. Hydrol.*, **137**(1-4), 1-31.

Herrera, L.J., Pomares, H., Rojas, I., Guillen, A., Prieto, A., and Valenzuela, O., 2007, Recursive prediction for long term time series forecasting using advanced models. *Neurocomputing*, **70**(16-18), 2870-2880.

- Hipel, K.W., Mcleod, A.I., and Lennox, W. C., 1977, Advances in Box-Jenkins modeling: 1. Model construction, *Water Resour. Res.*, **13**(3), 567-575.
- Hong, Y.S.T. and White, P.A., 2009, Hydrological modeling using a dynamic neuro-fuzzy system with on-line and local learning algorithm, *Adv. Water Resour.*, **32**, 110-119.
- Hu, T.S., Lam, K.C., and NG, S.T., 2005, A modified neural network for improving river flow prediction, *Hydrol. Sci. J.*, **50**(2), 299-318.
- Jang, J., 1993, ANFIS: Adaptive-network-based fuzzy inference system, *IEEE T. on syst. man cyb.*, **23**(3), 665-685.
- Ji, Y., Hao, J., Reyhani, N., and Lendasse, A., 2005, Direct and recursive prediction of time series using mutual information selection, *Lect. notes comp. sc.*, **3512**, 1010-1017.
- Karunanithi, N., Grenney, W.J., Whitley, D., and Bovee, K., 1994, Neural networks for river flow prediction, *J. of Comp. Civil Eng.*, **8**(2) 201-220.
- Kisi, O. and Shiri, J., 2012, Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations, *Hydrol. Res.*, **43**(3), 286-300.
- Knotters, M. and Bierkens, M.F.P., 2000, Physical basis of time series models for water table depths, *Water Resour. Res.*, **36**(1), 181-188.
- Maier, H.R. and Dandy, G.C., 2000, Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, *Environ. Modell. Softw.*, **15**, 101-124.
- Mohanty, S., Jha, M.K., Kumar, A., and Sudheer, K.P., 2010, Artificial neural network modeling for groundwater level forecasting in a river island of eastern India, *Water Resour. Manag.*, **24**(9), 1845-1865.
- Rai, S.N. and Singh, R.N., 1995, Two-dimensional modelling of water table fluctuation in response to localized transient recharge, *J. Hydrol.*, **167**, 167-174.
- Rumelhart, D.E., McClelland, J.L., and The PDP Research Group, 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, Massachusetts, USA, 516 p.
- Tankersley, C.D., Graham, W.D., and Hatfield, K., 1993, Comparison of univariate and transfer function models of groundwater fluctuations, *Water Resour. Res.*, **29**, 3517-3533.
- van Geer, F.C. and Zuur, A.F., 1997, An extension of Box-Jenkins transfer/noise models for spatial interpolation of groundwater head series, *J. Hydrol.*, **192**, 65-80.
- Yi, M.J., Kim, G.B., Sohn, Y.C., Lee, J.Y., and Lee, K.K., 2004, Time series analyses for the groundwater level in the National Groundwater Monitoring Network, *J. Geol. Soc. Korea*, **50**(2), 293-307.
- Yi, M.J. and Lee, K.K., 2004, Transfer function-noise modeling of irregularly observed groundwater heads using precipitation data, *J. Hydrol.*, **288**, 272-287.
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., and Lee, K.-K., 2011, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, *J. Hydrol.*, **396**(1-2), 128-138.
- Yoon, H., Kim, Y., Ha, K., and Kim, G.B., 2013, Application of groundwater-level prediction models using data-based learning algorithms to National Groundwater Monitoring Network data, *J. Eng. Geol.*, **23**(2), 137-147.
- Zealand, C.M., Burn, D.H., and Simonovic, S.P., 1999, Short-term streamflow forecasting using artificial neural networks, *J. Hydrol.*, **214**(1-4), 32-48.