



# 임계치 모형과 인공신경망 모형을 이용한 실시간 저수지 수위자료의 이상치 탐지

## Outlier Detection of Real-Time Reservoir Water Level Data Using Threshold Model and Artificial Neural Network Model

김마가<sup>a</sup> · 최진용<sup>b,†</sup> · 방재홍<sup>c</sup> · 이재주<sup>d</sup>

Kim, Maga · Choi, Jin-Yong · Bang, Jehong · Lee, Jaeju

### Abstract

Reservoir water level data identify the current water storage of the reservoir, and they are utilized as primary data for management and research of agricultural water. For the reservoir storage management, Korea Rural Community Corporation (KRC) installed water level stations at around 1,600 agricultural reservoirs and has been collecting the water level data every 10 minutes. However, various kinds of outliers due to noise and erroneous problems are frequently appearing because of environmental and physical causes. Therefore, it is necessary to detect outlier and improve the quality of reservoir water level data to utilize the water level data in purpose. This study was conducted to detect and classify outlier and normal data using two different models including the threshold model and the artificial neural network (ANN) model. The results were compared to evaluate the performance of the models. The threshold model identifies the outlier by setting the upper/lower bound of water level data and variation data and by setting bandwidth of water level data as a threshold of regarding erroneous water level. The ANN model was trained with prepared training dataset as normal data (T) and outlier (F), and the ANN model operated for identifying the outlier. The models are evaluated with reference data which were collected reservoir water level data in daily by KRC. The outlier detection performance of the threshold model was better than the ANN model, but ANN model showed better detection performance for not classifying normal data as outlier.

Keywords: Reservoir water level; outlier detection; threshold model; artificial neural network

### 1. 서론

수문 자료는 수자원을 이용하고 관리하기 위해 필요한 분석의 중요한 기초자료이며, 댐의 운영이나 홍수조절, 관개용수 관리 등 다양한 분야에 밀접하게 관련되어 있다. 특히 저수지 수위자료는 저수지 용량 곡선을 이용하여 저수용량을 산정하는 기준으로 사용되고 있으며 (Jeong and Kim, 2007), 저수지 운영 모형의 개발 (Shim et al., 1997; Jang et al., 2007), 저수지 수위 변화에 따른 제체 거동 연구 (Lee et al., 2014), 저수지 근처 지하수위의 변화에 따른 토양의 습윤화 (Kim et al., 2016), 땅밀림 산사태 발생 (Kim et al., 2017) 등 다양한

연구 분야에서 기초자료로 활용되고 있다.

최근에는 관측 장비와 컴퓨터의 발전으로 방대한 양의 수문 자료를 수집·축적할 수 있게 되었다. 한국에서는 10만 톤 이상의 저수용량을 가진 농업용 저수지에 대해 전국 약 1,600여개의 저수지 수위계측 장비가 설치되어 있다. 수위계측은 10분 단위 간격으로 이루어지며, 압력식 센서와 초음파식 센서를 사용하는 두 가지 방식이 있다. 그러나 압력식 센서의 경우 사통 내부로 토사물이 유입되거나 센서 부근에 퇴적된 토사 등으로 인해 이상치가 발생할 수 있으며, 초음파 센서의 경우 기온, 습도 등 환경 변화나 파랑, 센서 부근의 수초 등 장애물과 같은 물리적인 이유로 인한 이상치가 발생할 수 있다 (Bang et al., 2017). 효율적인 수자원의 관리와 연구를 위해 저수지 수위자료 등 수문자료의 신뢰성이 확보되어야 하며, 이를 위한 수문자료의 이상치 탐지 및 품질관리는 매우 중요하다.

수문 자료의 품질관리 및 보정을 위하여 지금까지 여러 연구들이 진행되어 왔다. Kilonsky and Caldwell (1991)은 고품질의 해수면 수위 데이터를 생산하기 위해 측정망을 구성하고 다양한 센서 측정 자료의 차이와 기준자료를 이용한 보정 방법을 분석하였다. Steiner et al. (1999)과 Harrison et al. (2000)은 레이더를 이용해 측정된 강우 자료를 보정하기 위해

<sup>a</sup> MS Student, Department of Rural Systems Engineering, Seoul National University

<sup>b</sup> Professor, Department of Rural Systems Engineering, Research Institute of Agriculture and Life Sciences, Seoul National University

<sup>c</sup> PhD Student, Department of Rural Systems Engineering, Seoul National University

<sup>d</sup> Rural Research Institute, Korea Rural Community Corporation

† Corresponding author

Tel.: +82-2-880-4583 Fax: +82-2-873-2087

E-mail: iamchoi@snu.ac.kr

Received: November 6, 2018

Revised: November 30, 2018

Accepted: December 6, 2018

실측 강우자료를 사용하였으며, 실측 강우자료의 품질에 따른 보정효과에 대해 분석하였다. Mourad and Bertrand-Krajewski (2002)는 농촌 지역의 장기 시계열 수문 자료에 대해 품질관리 과정을 구성하고 각 단계별로 품질등급을 매긴 후 이를 종합하여 품질관리를 수행하였다.

이상치를 탐지하기 위한 연구도 진행되어 왔는데 Gonzalez-Rouco et al. (1999)은 강우자료에 대해 통계적 방법을 이용하여 이상치를 탐지하고 기준자료를 이용한 균질성 보정을 수행하였다. Feng et al. (2004)은 강우자료를 포함한 기상자료에 대해 이상치 탐지 알고리즘을 개발하고 적용하였으며, 기준자료를 이용해 균질성을 검사하였다. Horsburgh et al. (2008)은 관측자료모형 (Observations Data Model, ODM)을 개발하여 자료의 품질에 따라 품질관리수준 (quality control level)을 표시하였다. Mounce et al. (2011)은 배수 시스템의 유량, 압력 자료에 대해 support vector machines을 이용하여 이상치를 탐지하는 연구를 수행하였다. Schneider et al. (2014)은 시계열 분포와 주변 관측소 자료를 활용해 월 강우량 자료에 대한 반자동 품질관리 방법을 제시하였다.

국내에서는 Kim et al. (2006)이 관계형 데이터베이스를 이용한 수문자료 관리시스템을 개발하였으며, 유량측정성과의 불확실도를 유량 측정 자료의 일관된 품질관리에 활용할 수 있도록 하였다. Kang et al. (2010)은 농업용수로에서 정확도가 높은 유량자료를 얻기 위해 수로 특성에 따른 계측기기 설치 기준을 정리하였으며, 주기적인 점검을 통해 측정의 불확실도를 저감하는 것으로 품질관리를 할 수 있다고 하였다. Park et al. (2015)은 수위자료의 이상치를 구분하고 보정할 수 있는 자료보정 알고리즘을 개발하고 지능형 수위 모니터링 플랫폼에 적용하였다. 수문자료의 품질관리 및 보정에 대한 연구는 대부분 강우자료를 대상으로 주변 관측지점의 측정 자료, 기준자료 등을 이용하여 보정하거나 같은 지점을 서로 다른 방법으로 측정한 후 이를 비교하여 품질관리 및 보정을 실시하고 있었으며, 단일 계측기로 측정하는 저수지 수위자료의 품질관리에 관한 연구는 미비하였다.

국내·외기관에서는 자동화 된 알고리즘과 수동 보정 방법을 결합하여 품질관리를 수행하고 있다. 국토해양부 한국건설기술연구원은 자기기록지 혹은 디지털 로거 방식으로 수집한 원시 수문자료에 대해 결측/이상치를 자동으로 점검하여 1차 품질등급을 부여한다. 이후 수동으로 자료의 결측과 이상치를 처리하여 2차 품질등급을 부여하며, 2차 품질등급이 확정된 자료는 최종 자료로 확정·발간된다. 한강홍수통제소는 수문자료 신뢰도 분석 시스템을 자체 개발하여 활용하고 있다. 시스템을 통해 관측소별 수위 및 유량 자료에 접속하고, 그래프를 이용하여 이상치 및 결측치를 판단한다. 보정은 선

형보간법을 이용한 자동보정방법을 사용한다. 미국 환경보호청 (Evisonmental Protction Agency, EPA)은 이상치검출 방법론으로 Extreme value test, Discordance test, Rosner's test, Walsh's test의 4 가지 방법을 추천하고 있다. Extreme value test, Discordance test는 표본의 수가 적은 경우 적용하는 방법론이고, 표본의 수가 많은 경우에는 Rosner's test 또는 Walsh's test를 적용하도록 하고 있다. 미국 지질조사국 (United Sates Geological Survey, USGS)에서는 수문자료 품질관리 시스템인 ADAPS (Automated Data Processing System)를 운영하고 있다. ADAPS는 수위자료를 수집하고 유량을 계산하며, USGS Bulletin No.17을 이론적 배경으로 하여 수집된 자료의 이상여부를 점검하고 보정 가능성을 판단한다 (Choi, 2018).

한편, 인공신경망 (Artificial Neural Network, ANN)은 인간의 뇌 구조를 컴퓨터로 구현하고자 개발된 개념으로 (Yoon et al., 2004), 시간적·공간적 변동이 심한 기상이나 수문 등 비선형 특성을 보이는 자연현상을 적절히 재현할 수 있는 모형화가 가능하다. 이러한 특징 덕분에 강우-유출해석, 강우 분포, 기온, 지하수 유동 등 수자원 및 환경 분야에서 적용과 응용이 활발하게 이루어지고 있다 (Choi and Kang, 2000). Kim (2000)은 낙동강 진동지점에서 신경망 모형을 이용하여 일유출량을 예측하고 다중회귀분석 모형과 결과를 비교하였고, Ahn et al. (2000b)은 보정천 유역의 호우 발생에 따른 하천 유출수문곡선을 학습하여 미학습 유역에 대한 적용성을 검토하였다. Kim and Salas (2000)는 홍수위와 유역평균강우량을 이용하여 홍수위를 예측하기 위한 신경망 모형을 제시하였으며, Ahn et al. (2000a)은 신경망 모형을 이용하여 강우 자료의 결측을 보완하고 강우 예측에 대해 연구하였다. Oh et al. (2002)은 하천수질예측을 위한 신경망 모형을 개발하였다. Yoon et al. (2004)은 신경망을 이용하여 남강댐의 홍수기 댐 유입량 예측을 위한 신경망 모형을 구축하였다. Kang and Lee (2008)는 한반도 영역을 대상으로 권역별 강수 발생확률을 신경망 모형을 통해 예측하였고, Oh et al. (2008)은 신경망을 이용하여 소양강댐 유역 12개 관측소를 대상으로 결측 수문자료 추정 및 실시간 자료 보정을 하였다. Kim et al. (2010)은 신경망을 이용하여 지상 우량계의 지점 강우 자료를 기준으로 레이더 강우 자료를 보정하였고, Ahn et al. (2012)은 눈 포장에서의 일별 지표유출량을 추정하기 위한 신경망 모형을 구성하였다. 신경망 모형을 수문자료에 적용한 연구는 회귀모델로 수문 사상을 예측하는 연구가 다수 있었고, 수문 자료의 품질관리 및 보정에 신경망을 적용한 연구도 있었으나 저수지 수위자료의 품질관리를 대상으로 한 연구는 미비하였다.

따라서 본 연구에서는 저수지 수위자료의 품질관리를 위한

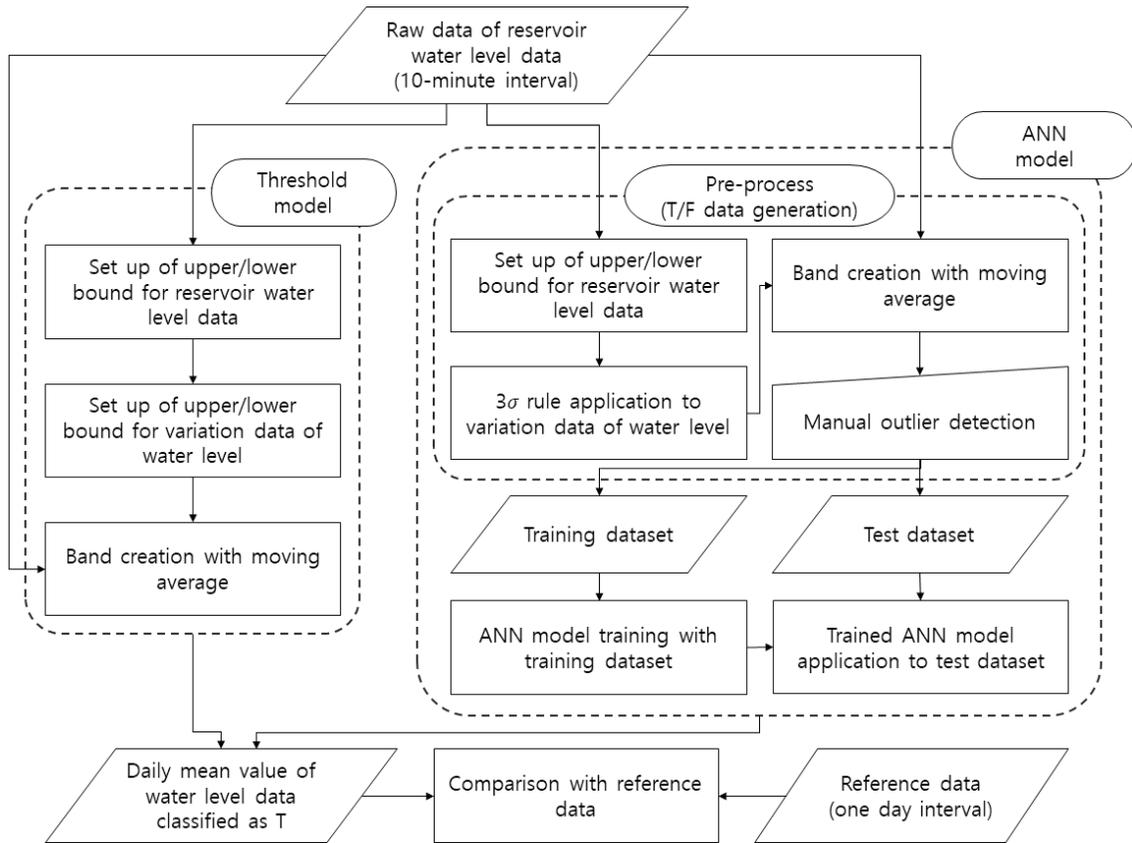


Fig. 1 Flow chart of the study

실시간 자동 이상치 탐지 방법을 구축하였으며, 개운 저수지의 수위자료를 대상으로 임계값을 이용한 이상치 탐지 모형과 인공지능망을 이용한 이상치 탐지 모형을 적용하여 성능을 분석하고 평가하고자 하였다.

## II. 재료 및 방법

본 연구에서는 임계치 모형과 인공지능망 모형을 이용하여 개운 저수지의 수위자료를 대상으로 이상치 탐지를 실시하였다. 개운 저수지의 수위자료는 한국농어촌공사에서 10분 단위로 측정된 자료로 2011. 01. 01. 0:00~2018. 06. 12. 11:40의 자료를 사용하였다. 기준 자료는 한국농어촌공사에서 제공하는 일 단위 수위자료를 사용하였으며, 획득한 기준자료의 기간은 2009. 08. 19.~2018. 05. 23.이다. 10분 단위 수위자료는 품질관리가 이루어지지 않은 원시 수위자료이며, 기준 자료는 관리자에 의해 수동으로 보정된 일 단위 수위자료이다. 모형의 성능 평가를 위해 모형 적용 결과의 일 평균 자료와 기준 자료를 비교해 산정한  $R^2$ ,  $MAE$ ,  $RMSE$ 를 이용하였으며, 일 평균 자료를 기준으로 모형 적용 결과 각 모형에서 이상치

로 분류되어 원시 수위자료에서 제외된 자료의 수를 비교하였다. Fig. 1은 본 연구의 흐름도를 나타낸 것이다.

### 1. 대상저수지 및 수위자료

본 연구는 강원도 홍천군 동면 개운리에 위치한 개운 저수지를 대상으로 수행하였다. 개운 저수지는 유역면적이 474 ha, 유효 저수량이 1,636 천 $m^3$ 에 해당하는 저수지로서 취수형식은 취수탑형이다. Table 1에 개운 저수지의 기본 제원을 나타내었다.

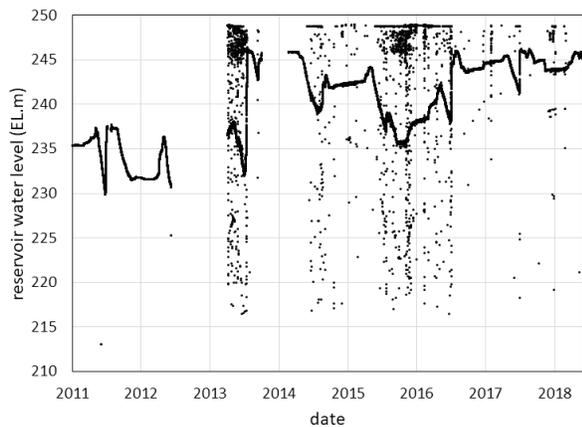
본 연구에서는 개운 저수지의 수위자료를 대상으로 이상치 탐지 모형을 적용하였고, 검증을 위해 한국농어촌공사에서 제공하는 일 단위 저수지 수위자료를 기준 자료로 사용하였다. 개운 저수지의 수위자료는 한국농어촌 공사에서 초음파식 센서를 이용해 10분 단위로 측정된 자료이며, 획득한 개운 저수지 수위자료의 측정 기간은 2011. 01. 01. 0:00~2018. 06. 12. 11:40이다. 기준자료는 센서를 통해 일 단위로 측정되며, 관리자에 의해 수동으로 보정된다. 관리자는 주관적인 판단에 의해 측정된 수위자료의 이상 여부를 검토하며, 이상이 있다고 판단될 경우 목측을 통해 수위자료를 수정하는 방식으로 측정된 수위자료를 보정한다. 획득한 기준 자료의 측정 기

**Table 1** Properties of the Gaeun reservoir

Dam property	Contents	Dam property	Contents
Dam body type	fill dam	Intake works	main facility
Dam body length (m)	224.0	Water-intake type	intake tower
Dam body height (m)	39.7	Total storage (m <sup>3</sup> )	1,649,375
Dead storage water level (EL,m)	222.50	Effective storage (m <sup>3</sup> )	1,636,475
Full water level (EL,m)	245.80	Benefited area (ha)	103
Flood water level (EL,m)	246.80	Basin area (ha)	474

**Table 2** Properties of water level data

Data	Measurement method	Measurement cycle	Measurement period
Sensor data	ultrasonic sensor	10 minute	2011. 01. 01. 0:00~2018. 06. 12. 11:40
Reference data	sensor/eye measurement	1 day	2009. 08. 19.~2018. 05. 23.



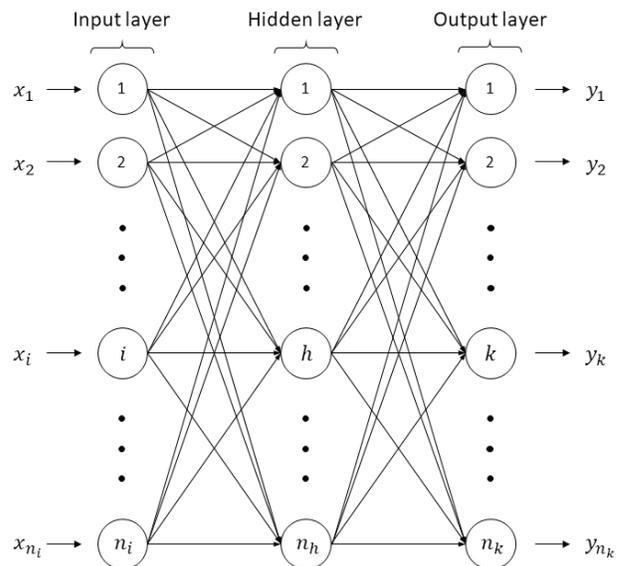
**Fig. 2** 10-minute interval raw water level data of the Gaeun reservoir

간은 2009. 08. 19.~2018. 05. 23.이다. Table 2는 본 연구에서 사용한 수위자료의 기본 정보를 정리한 것이다.

Fig. 2는 개운 저수지의 10분 단위 수위자료를 그래프로 도시한 것이다. 측정 기간 내에서 spike noise가 다수 발생하였으며, 측정값은 0~23542 EL.m 범위에서 나타났다. 그래프 개형을 도시하기 위해 y축의 범위를 210~250 EL.m로 제한하였다.

## 2. 인공신경망 (Artificial Neural Network, ANN)

인공신경망 모형은 뉴런과 시냅스로 이루어져 있는 인간의 뇌 구조를 단순화하여 만든 연산모델로 단순한 연산자들의 결합으로 구성되어 있다. 인공신경망 모형은 문제를 해결하기 위한 직접적인 지식이나 방법을 설정하지 않아도 주어진 자료를 통해 문제를 해결할 수 있으며 (Yeo et al., 2010), 복잡한 비선형 및 다차원 입출력 관계를 모델링하기에 적절한 모형으로 주로 회귀, 분류의 문제를 해결하기 위해 사용된다.



**Fig. 3** General structure of multi-layer perceptron

인공신경망 모형에서 가장 일반적으로 적용되는 구조는 여러 개의 층으로 구성되는 다층 퍼셉트론 (Multi-Layer Perceptron, MLP)이다 (Seo et al., 2017). Fig. 3은 각 층의 노드 수가  $n_i$ ,  $n_h$ ,  $n_k$ 인 입력층, 은닉층, 출력층을 가지는 다층 퍼셉트론의 일반적인 구조를 도식화 한 것이다.

세 개의 층과  $J$ 개의 은닉노드를 가지는 다층 퍼셉트론은 식 (1)과 같은 연산을 수행한다.

$$\begin{aligned}
 o(\mathbf{x}) &= f\left(w_0 + \sum_{j=1}^J w_j \cdot f\left(w_{0j} + \sum_{i=1}^n w_{ij} x_i\right)\right) \\
 &= f\left(w_0 + \sum_{j=1}^J w_j \cdot f\left(w_{0j} + \mathbf{w}_j^T \mathbf{x}\right)\right)
 \end{aligned} \tag{1}$$

여기서  $\mathbf{x} = \{x_i, i = 1, 2, \dots, n\}$ 는 입력벡터,  $f$ 는 활성화 함수,  $\mathbf{o}$ 는 출력벡터,  $w_0$ 는 출력노드에 대한 편향,  $w_j$ 는 가중치,  $\mathbf{w}_j = (w_{1j}, \dots, w_{nj})$ 는 가중치 벡터,  $w_{oj}$ 는  $j$ 번째 은닉노드에 대한 편향이다 (Günther and Fritsch, 2010).

활성화 함수 (activation function)는 각 노드에서 입력신호의 총합을 출력신호로 변환하는 함수로 이때 출력값은 다음 노드에서 입력값으로 사용된다. 본 연구에서는 은닉층의 활성화 함수로 일반적으로 많이 사용하는 시그모이드 함수 (sigmoid function)를 적용하였다. 역전파알고리즘 (back-propagation)은 전방향 (feed-forward)의 계산을 통하여 출력층의 오차를 계산하고, 후방계산을 통하여 오차의 미분을 은닉층으로 역전파시켜 각 층 사이의 가중치를 조절하고 오차를 줄여나가는 방법으로, 본 연구에서는 역전파알고리즘으로 경사하강법 (gradient descent method)을 적용하여 인공신경망 모형의 학습을 진행하였다. 출력층의 오차를 계산하는 손실 함수로는 교차 엔트로피 오차 함수 (cross entropy cost function)를 사용하였고, 출력층의 활성화함수로 소프트맥스 함수 (softmax function)를 사용하였다. 소프트맥스 함수는  $K$ 개의 다른 사건에 대해 확률분포를 계산하는 함수로 출력 확률 범위가 0에서 1이며 모든 사건  $K$ 개의 확률의 합은 1이므로 분류의 문제를 해결하기에 적합하다. 본 연구에서는 소프트맥스 함수를 이용하여 저수지 수위자료의 이상치를 분류하였다. 시그모이드 함수, 교차 엔트로피 오차 함수와 소프트맥스 함수의 기본 형태는 각각 식 (2), (3), (4)와 같다.

$$h(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

$$E = - \sum_k t_k \log y_k \quad (3)$$

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (4)$$

여기서  $h(x)$ 는 시그모이드 함수,  $E$ 는 교차 엔트로피 오차,  $\sigma$ 는 소프트맥스 함수이고,  $t_k$ 는 정답 레이블,  $y_k$ 는 인공신경망 모형의 출력값,  $k$ 는 데이터의 차원 수이며,  $\mathbf{z} = (z_1, z_2, \dots, z_j)$ 는 출력층의 입력 벡터이다.

### 3. 이상치 탐지 모형

본 연구에서는 저수지 수위자료의 실시간 이상치 탐지를 위해 수위자료와 수위변화량 자료에 임계치를 설정한 임계치

모형과 원시 수위자료의 이상치를 분류하여 생성한 학습 자료를 통해 학습시킨 인공신경망 모형을 저수지 수위자료의 이상치 탐지에 적용하였다.

#### 가. 임계치 모형

임계치 모형은 수위자료의 상하한 설정, 수위변화량의 상한 설정, 이동평균을 이용한 band 형성의 3 단계로 구성하였고 각 단계에서 기준을 벗어나는 수위자료를 이상치로 분류하였다.

##### 1) 수위자료의 상하한 설정 (1 단계)

수위자료의 상하한 설정 단계에서는 저수지 제원을 기준으로 수위자료의 상하한을 설정하고, 이를 벗어나는 측정값을 이상치로 분류하였다. 수위자료의 상하한은 홍수위를 상한, 사수위를 하한으로 설정하였으며, 수위자료가 이상치로 분류된 경우에는 해당 시점의 수위자료를 이전 시점의 수위자료로 대체하여 다음 단계에 적용하였다.

##### 2) 수위변화량 자료의 상한 설정 (2 단계)

수위변화량 자료의 상한 설정 단계에서는 수위변화량이 설정한 상한을 벗어나는 경우에 해당하는 시점의 수위자료를 이상치로 분류하였다. 수위변화량은 이전 시점 측정값과 해당 시점 측정값의 차로 계산하였으며, 수위변화량 자료의 상한은 만수위와 사수위의 차 ( $h$ )의 일정 비율 ( $\alpha$ )로 결정하였다.  $\alpha$ 는 0.1, 0.2, 0.3, 0.4, 0.5를 적용하여 그래프 개형이나 기준자료와 비교하여 계산한 통계치를 이용해 성능이 뛰어난 것으로 선정하였다. 수위자료가 이상치로 분류된 경우에는 해당 시점의 수위자료를 이전 시점의 수위자료로 대체하여 다음 단계에 적용하였다.

##### 3) 이동평균을 이용한 band 형성 (3 단계)

이동평균을 이용한 band 형성 단계에서는 2 단계를 적용한 수위자료의 이동평균으로 계산한 기준값으로부터 일정 폭을 가지는 band를 형성하고, 최종적으로 원시자료가 band를 벗어나는 경우를 이상치로 분류하였다. 임계치 모형의 적용 결과는 원시 수위자료에서 이상치로 분류된 수위자료와 결측 자료를 제외하여 얻었다. 이동평균은 해당 시점을 포함한 10개 수위자료를 사용하였고, band의 폭은 0.05 m를 사용하였다. band의 폭은 0.01~0.5 m를 시행착오법으로 적용하고 그래프 개형을 비교하여 이상치 탐지 성능이 뛰어난 0.05 m로 결정하였다.

#### 나. 인공신경망 모형

인공신경망 모형은 정상 수위자료를 T, 이상치와 결측 자

료를 F로 구분한 학습 자료 (T/F dataset)를 생성하여 신경망 모형을 학습시킨 모형으로, 검증기간에 대해 모형을 적용하여 수위자료의 이상치를 분류하였다. 학습 자료는 수위자료의 상하한 설정,  $3\sigma$  규칙 적용, 이동평균을 이용한 band 형성, 수동 이상치 탐지의 4 단계를 통해 수위자료를 정상 수위자료와 이상치로 분류하여 생성하였다. 학습 자료는 훈련 데이터 (training dataset)와 시험 데이터 (test dataset)로 구분하여 훈련 데이터를 통해 인공신경망 모형을 학습시키고 시험 데이터를 통해 모형의 성능을 평가하였다. 인공신경망 모형의 은닉노드 수, 학습률, 입력 자료는 시행착오법을 통해 결정하였다.

1) 학습 자료 생성

가) 수위자료의 상하한 설정 (1 단계)

수위자료의 상하한 설정 단계에서는 저수지 제원을 기준으로 수위자료의 상하한을 설정하고, 이를 벗어나는 측정값을 이상치로 분류하였다. 수위자료의 상하한은 홍수위를 상한, 사수위를 하한으로 설정하였으며, 수위자료가 이상치로 분류된 경우에는 해당 시점의 수위자료를 제외하고 다음 단계에 적용하였다.

나)  $3\sigma$  규칙 적용 (2 단계)

$3\sigma$  규칙 적용 단계에서는 해당 시점 수위자료에서 이전 시점 수위자료를 뺀 수위변화자료를 대상으로  $3\sigma$  규칙을 적용하였다. 이때 결측 자료에 대해서는 수위변화를 계산하지 않고 이상치로 분류하였다.  $3\sigma$  규칙에 따르면 정규 분포에서 평균 양쪽으로  $3\sigma$  이내 범위에 변량의 대부분의 값이 포함되며, 따라서  $3\sigma$  규칙에 의해 수위변화자료에서 평균 양쪽으로  $3\sigma$ 를 벗어나는 경우에 해당하는 시점의 수위자료를 이상치로 분류하였다. 수위자료의 경우 정규 분포를 따르지 않기 때문에  $3\sigma$  규칙을 적용하기 부적절하며, 수위변화자료의 경우 평균 주변으로 변량이 밀집되어 있는 분포를 가지므로  $3\sigma$  규칙을 적용하기에 적절하다고 판단하였다. 수위자료가 이상치로 분류된 경우에는 해당 시점의 수위자료를 제외하고 다음 단계에 적용하였다.

다) 이동평균을 이용한 band 형성 (3 단계)

이동평균을 이용한 band 형성 단계에서는 2 단계를 적용한 수위자료의 이동평균으로 계산한 기준값으로부터 일정 폭을 가지는 band를 형성하고, 원시 수위자료가 band를 벗어나는 경우를 이상치로 분류하였다. 이동평균은 해당 시점을 포함한 10개 수위자료를 사용하였고, band의 폭은 0.05 m를 사용하였다. band의 폭은 0.01 ~ 0.5 m를 시행착오법으로 적용하고

그래프 개형을 비교하여 이상치 탐지 성능이 뛰어난 0.05 m로 결정하였다.

라) 수동 이상치 탐지 (4 단계)

수동 이상치 탐지 단계에서는 3 단계까지 적용한 이후 남아있는 spike noise를 그래프 개형을 통해 수동으로 확인하고 이상치로 분류하였다. 또한 수위가 급격히 변하는 구간에서 이동평균이 수위자료의 변화를 즉각적으로 반영하지 못해서 정상 수위자료가 이상치로 분류된 경우를 찾아 정상 수위자료로 재분류하였다. 이 경우에 연속된 정상 수위자료가 이상치로 처리되므로 3 단계 적용 결과에서 연속하여 이상치가 나타난 구간에 대해 수동으로 확인 과정을 거쳤다. 최종적으로 학습 자료는 정상 수위자료를 1, 이상치로 분류된 수위자료와 결측 자료를 0으로 구분하여 얻었다.

Fig. 4는 인공신경망의 학습 자료에서 이상치로 분류된 수위자료를 제외하고 그래프로 도시한 것이다. 그래프에서 이상치가 대부분 제외된 것으로 보아 학습 자료에서 이상치의 분류가 적절히 이루어진 것을 확인할 수 있으며, 학습 자료를 정답으로 보고 인공신경망 모형을 학습시켰다. 최종적으로 인공신경망 모형의 적용 결과는 인공신경망 모형에 의해 이상치로 분류된 수위자료를 원시 수위자료에서 제외하여 얻었다.

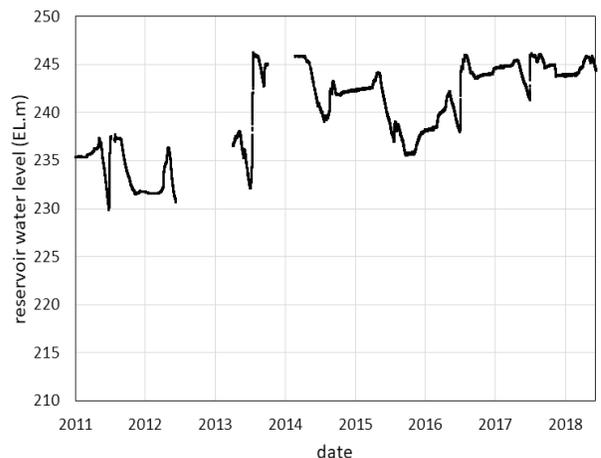


Fig. 4 T dataset after data pre-process

2) 인공신경망 모형 구성

본 연구에서 적용한 인공신경망 모형은 입력층 (input layer), 은닉층 (hidden layer), 출력층 (output layer)의 3개 층으로 구성하였다. 은닉층의 활성화 함수로는 일반적으로 많이 사용하는 것으로 알려진 시그모이드 함수를 적용하였으며, 출력층의 활성화 함수로 소프트맥스 함수를 적용하였고 손실

**Table 3** Input data of the artificial neural network model for outlier detection

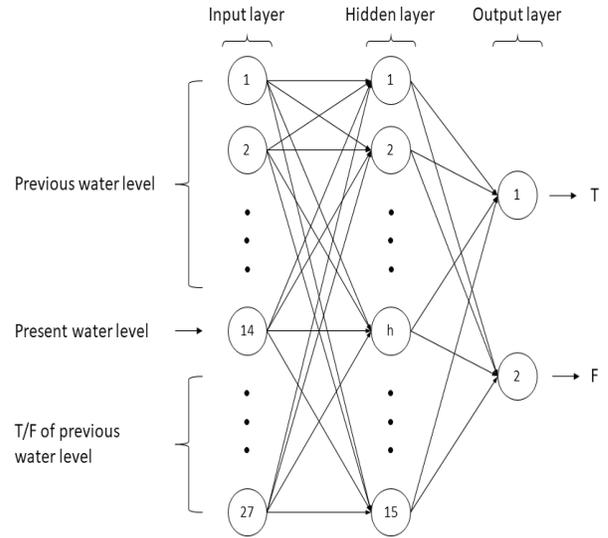
Data type	Contents
water level ( $wl_t$ )	$wl_t$
variation ( $v_t$ )	$v_t = wl_t - wl_{t-1}$
reservoir regularization ( $rr_t$ )	$rr_t = \frac{wl_t - dl}{fl - dl}$
limited regularization ( $lr_t$ )	$lr_t = \begin{cases} 1 & (\text{if } rr_t > 1) \\ rr_t & (\text{if } 0 < rr_t < 1) \\ -1 & (\text{if } rr_t < 0) \end{cases}$
regularization ( $r_t$ )	$r_t = \frac{wl_t - \min(wl)}{\max(wl) - \min(wl)}$
variation regularization ( $vr_t$ )	$vr_t = \frac{v_t - \min(v)}{\max(v) - \min(v)}$

\* $wl$ : measured reservoir water level,  $v$ : variation of reservoir water level,  $rr$ : regularized reservoir water level with dead water level ( $dl$ ) and flood water level ( $fl$ ),  $lr$ : regularized reservoir water level with dead water level ( $dl$ ) and flood water level ( $fl$ ) within range of -1 to 1,  $r$ : regularized reservoir water level,  $vr$ : regularized variation of reservoir water level

함수로 교차 엔트로피 오차 함수를 적용하였다.

입력 자료는 전처리 방식에 따라 6가지 자료를 적용하고, 성능이 가장 뛰어난 자료를 선정하였다. 이상치 여부를 판단하고자 하는 시점에 해당하는 입력 자료와 해당 시점 이전의 입력 자료 및 학습 자료의 이상치 여부를 적용하였으며, 입력 자료는 수위자료 (water level,  $wl$ ), 수위변화자료 (variation,  $v$ ), 저수지 제원을 이용한 일반화 자료 (reservoir regularization,  $rr$ ), 제한된 일반화 자료 (limited regularization,  $lr$ ), 일반화 자료 (regularization,  $r$ ), 수위변화자료의 일반화 자료 (variation regularization,  $vr$ )를 사용하였다. 학습 자료의 이상치 여부는 학습 자료 생성 결과를 기준으로 정상 수위자료를 1, 이상치와 결측 자료를 0으로 구분하였다. 각 입력 자료에 대한 전처리 방법에 대한 설명을 Table 3에 정리하였다.

입력층의 노드는 입력 자료의 수로 결정되며, 입력 자료의 수는 사용하는 이전 시점 자료의 개수를 바꿔가며 시행착오법을 통해 결정하였다. 마찬가지로 방법으로 은닉층의 노드와 학습률에 대해서 시행착오법을 적용하여 은닉 노드 수와 학습률을 결정하였다. 시행착오법을 수행한 결과 13개의 이전 시점 자료 및 이상치 여부를 입력 자료로 사용하였고 은닉 노드는 15개, 학습률은 0.005를 사용하였다. Fig. 5는 본 연구에서 사용한 인공신경망 모형의 모식도이다.



**Fig. 5** The structure of the artificial neural network model for outlier detection

모형의 검증을 위해 학습 자료를 시계열로 나열한 후 기준 시점 이전을 훈련 데이터, 이후를 시험 데이터로 사용하였다. 저수지 수위자료의 경우 이상치의 시간적 분포가 고르지 않기 때문에 전체 기간의 전반부 70%를 훈련 데이터, 나머지 30%를 시험 데이터로 구분하면 이상치가 한쪽에 편중된다. 따라서 본 연구에서는 이상치로 분류된 수위자료를 대상으로 이상치의 70%가 훈련 데이터에 포함되도록 하는 시점을 기준 시점으로 하여 훈련 데이터와 시험 데이터를 구분하였다. 또한 학습 시 측정 시점에 따른 영향을 제거하기 위해 훈련 데이터는 임의 순서로 재배치한 후 학습을 진행하였다.

#### 4. 모형 검증

모형의 성능을 평가하기 위한 기준 자료로 한국농어촌공사에서 제공하는 일 단위 저수지 수위자료를 사용하였다. 한국농어촌공사에서 제공하는 일 단위 수위자료는 센서를 통해 일 단위로 측정되며, 관리자에 의해 수동으로 보정된 자료이다. 관리자는 센서를 통해 자동으로 입력된 값을 확인한 후 이상이 있다고 판단되는 경우 목측을 통해 자료를 보정한다. 한국농어촌공사에서 제공하는 일 단위 수위자료의 보정은 관리자의 주관적인 판단과 수동 보정에 의존하고 있으나 그럼에도 지속적으로 보정이 이루어지고 있는 자료이기 때문에 기준 자료로 사용하기 적절하다고 판단하였다. 모형의 성능 평가를 위한 검증 기간은 획득한 기준 자료의 기간과 인공신경망 모형에서 시험 데이터에 해당하는 기간을 고려하여 2013. 12. 05.~2018. 05. 23.으로 선정하였다. 모형의 성능 평가는 그래프 개형이나  $R^2$ ,  $MAE$ ,  $RMSE$ 와 같은 통계치를

비교하여 수행하였으며, 통계치는 모형 적용 결과의 일 평균 자료와 기준 자료를 비교하여 산정하였다.

### III. 결과 및 고찰

#### 1. 임계치 모형 적용

##### 가. 수위변화량 자료의 상한 설정

임계치 모형에서는 수위변화량 자료의 상한을 설정하기 위해  $\alpha$ 에 따른 이상치 탐지 결과를 비교하고 그에 따라  $\alpha$ 를 결정하였다. Fig. 6은  $\alpha$ 에 따른 이상치 탐지 결과를 그래프로 도시한 것이다. 그래프에서 확인할 수 있듯이  $\alpha$ 가 0.3, 0.4,

0.5인 경우는 이상치 탐지 성능이 현저히 낮으며,  $\alpha$ 가 0.2인 경우도 0.1인 경우에 비해 이상치 탐지 성능이 떨어진다. 이는 수위자료 측정 간격이 10분으로 매우 짧아서 물리적으로 허용 가능한 수위변화량 값이 작기 때문이다. 따라서  $\alpha$ 가 커져서 수위변화량 자료의 상한이 커질 경우 이상치로 판단되는 자료가 정상 수위자료로 분류될 가능성이 크다. 자료의 측정 간격이 길어질수록 허용 가능한 수위변화량 자료의 상한도 더 커질 것으로 사료되며, 본 연구에서는 그래프 개형 비교 결과에 따라  $\alpha$ 값을 0.1로 적용하는 것이 적절하다고 판단하여 임계치 모형에  $\alpha = 0.1$ 을 적용하였다.

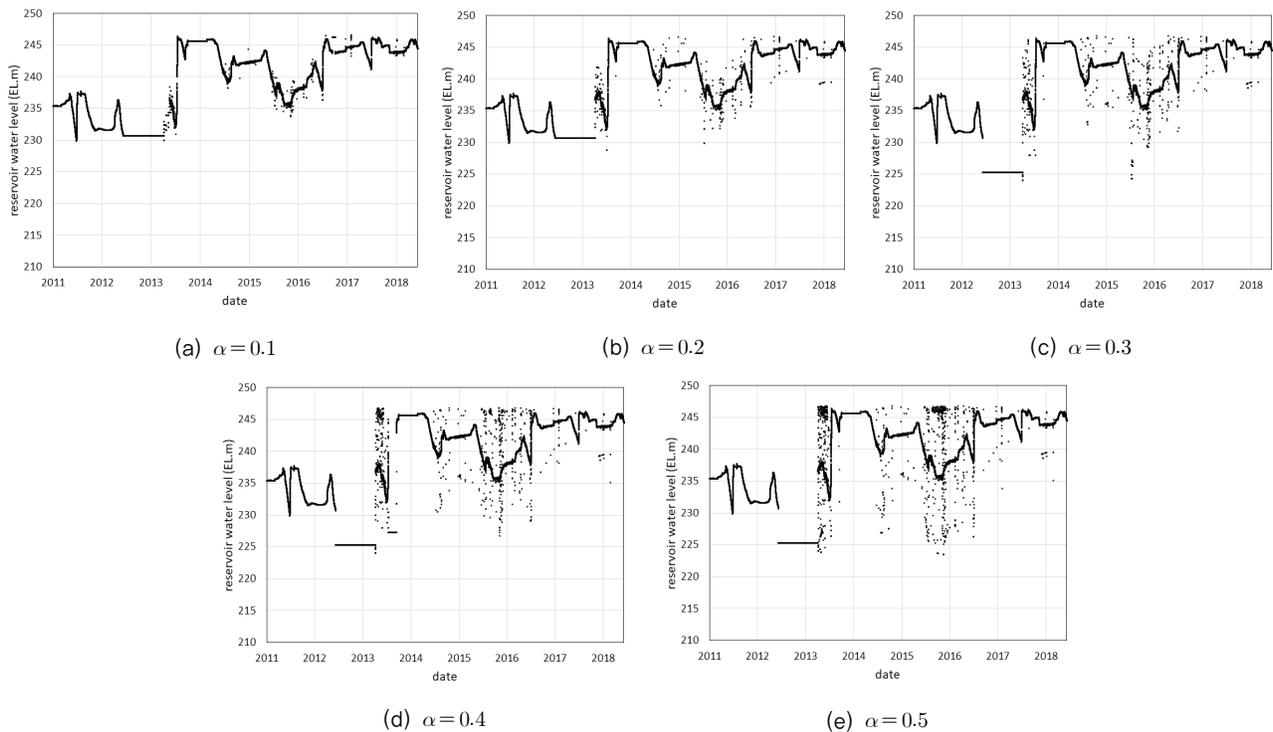


Fig. 6 The results of the threshold model according to value of  $\alpha$

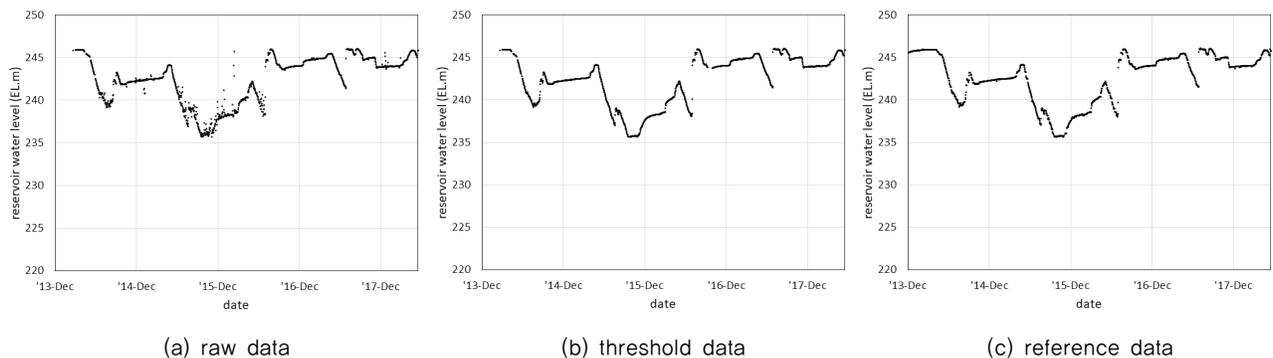


Fig. 7 Daily mean value of (a) raw data, (b) threshold data, (c) reference data after threshold model application

나. 임계치 모형 적용 결과

Fig. 7은 검증기간에 대해 원시 수위자료의 일 평균 자료 (raw data)와 임계치 모형 적용 결과의 일 평균 자료 (threshold data), 기준 자료 (reference data)를 그래프로 도시한 것이다. 평균을 하게 되면 이상치가 보정되는 효과가 있음에도 원시 수위자료의 일 평균 자료 그래프에서 이상치가 남아있는 것을 확인 할 수 있다. 임계치 모형 적용 결과의 일 평균 자료는 기준 자료와 거의 유사한 그래프 개형을 보여 이상치 탐지가 적절히 이루어진 것을 확인할 수 있었다. 기준 자료의 경우 측정 간격이 하루 단위이고 담당자의 주관적인 판단과 수동 보정에 의존하고 있기 때문에 그래프가 다소 매끄럽지 않은 부분이 있지만, 임계치 모형 적용 결과의 일 평균 자료 그래프는 기준 자료의 그래프보다 매끄럽게 나타났다. 이는 임계치 모형의 경우 10분 단위 수위자료를 대상으로 적용하여 기준 자료보다 자료의 밀도가 더 높기 때문인 것으로 사료된다.

2. 인공신경망 모형 적용

가. 입력 자료 선정

Table 4는 시험 데이터에 해당하는 기간에 대해 입력 자료에 따른 인공신경망 모형의 적용 결과와 학습 자료의 이상치 여부를 비교하여 오차를 정리한 것이다. T는 정상 수위자료, F는 이상치 또는 결측 자료를 의미하며 인공신경망 모형은 학습 자료의 T/F 구분을 정답으로 여기고 학습을 수행하였다. 해당 기간의 학습 자료는 T가 210,511개, F가 27,107개로 총 237,618개로 나타났다.

전체 오차 (total error)는 전체 자료에 대한 잘못 분류한 자료의 비율을 의미하며, 누락오차 (omission error)는 원래 자료에 대한 모형이 잘못 분류한 자료의 비율을 의미한다. 예를 들어 F의 누락오차는 학습 자료의 F에 대한 모형이 T로 잘못 분류한 자료의 비율이다. 수행오차 (performance error)는 모형이 분류한 자료에 대한 잘못 분류한 자료의 비율을 의미하며, 예를 들어 F의 수행오차는 모형이 F로 분류한 자료 중에 포함

된 학습 자료의 T의 비율이다.

Table 4에서 F의 누락오차 (omission error of F)에 따르면 수위변화자료 (*v*)를 입력 자료로 적용한 인공신경망 모형의 성능이 가장 뛰어나며, 이는 학습 자료의 F 중 가장 많은 수를 이상치로 분류한 모형이라는 의미이다. F의 수행오차 (performance error of F)에 따르면 수위자료 (*wl*)를 입력 자료로 적용한 인공신경망 모형의 성능이 가장 뛰어나며, 이는 모형이 분류한 F에 학습 자료의 T가 가장 적게 포함되었다는 의미이다. 전체 오차 (total error)에 따르면 수위변화자료 (*v*)를 적용한 인공신경망 모형의 성능이 가장 뛰어난 것으로 나타났다으며, 이는 전체 자료에 대해 모형이 잘못 분류한 자료의 수가 가장 적은 것을 의미한다.

모형 적용 결과는 이상치로 분류된 자료를 제외하고 표시하게 되는데 학습 자료의 T는 210,511개, F는 27,107개이므로 전체 T/F 자료 수를 고려해 봤을 때 F의 수행오차 (performance error of F)에서 의미하는 이상치로 분류되어 제외된 T가 전체 T에 미치는 영향보다 F의 누락오차 (omission error of F)에서 의미하는 이상치로 분류되지 않고 남아있는 F가 전체 F에 미치는 영향이 더 크다. 따라서 본 연구에서는 전체 오차 (total error), F의 누락오차 (omission error of F)에서 성능이 뛰어난 수위변화자료를 입력 자료로 선정하였다.

나. 인공신경망 모형 적용 결과

Fig. 8은 검증 기간에 대해 원시 수위자료의 일 평균 자료 (raw data), 학습 자료의 일 평균 자료 (target data), 인공신경망 모형 적용 결과의 일 평균 자료 (ANN data)와 기준 자료 (reference data)를 그래프로 도시한 것이다. 학습 자료의 일 평균 자료는 기준 자료와 비슷한 그래프 개형을 보이고 있어 이상치가 적절히 탐지된 것을 확인할 수 있다. 인공신경망 모형 적용 결과의 일 평균 자료는 이상치가 어느 정도 제외되어 원시 수위자료의 일 평균 자료에 비해 개선된 것을 확인할 수 있으나 학습 자료의 일 평균 자료의 수준보다는 성능이 떨어지는 것으로 나타났다. 이는 Table 4에서 확인할 수 있는 것처럼 학습 자료와 비교하였을 때 F의 누락오차 15.43%가 일 자료 평균 그래프에 반영되었기 때문이다.

Table 4 The errors by input data in application of the artificial neural network model for outlier detection

Data type	Total error (%)	Omission error of F (%)	Performance error of F (%)
<i>wl</i>	2.67	23.45	0.00
<i>v</i>	2.29	15.43	5.22
<i>rr</i>	2.51	18.32	4.30
<i>lr</i>	2.53	19.11	3.63
<i>r</i>	3.47	21.89	9.81
<i>vr</i>	3.49	21.71	10.16

3. 임계치 모형과 인공신경망 모형 성능 비교

Table 5는 검증기간에 대해 원시 수위자료의 일 평균 자료 (raw data), 임계치 모형 적용 결과의 일 평균 자료 (threshold data), 학습 자료의 일 평균 자료 (target data), 인공신경망 모형 적용 결과의 일 평균 자료 (ANN data)를 기준 자료와 비교한 통계치를 정리한 것이다.

원시 수위자료의 일 평균 자료는  $R^2$ 가 0.982로 상당히 높

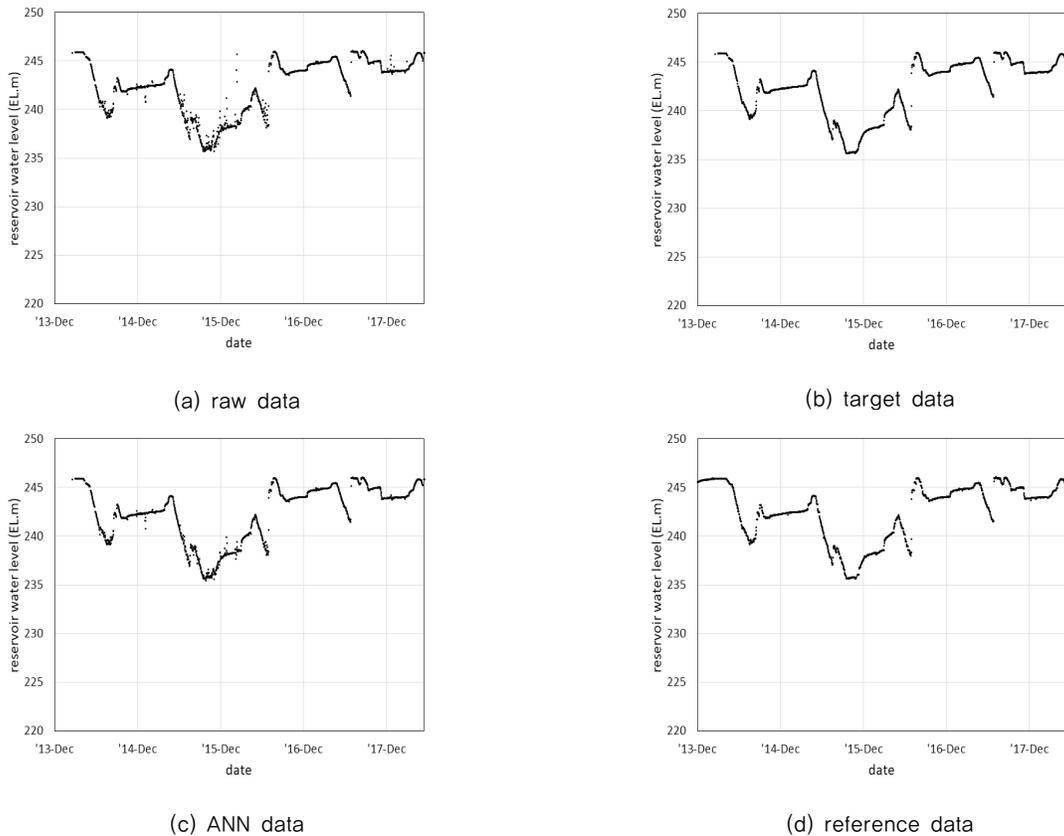


Fig. 8 Daily mean value of (a) raw data, (b) target data, (c) ANN data, (d) reference data after ANN model application

았는데, 이는 원시 수위자료의 경우 정상 수위자료가 이상치보다 많으므로 일 평균이 정상 수위자료로 수렴하는 경향이 있기 때문이다. 임계치 모형 적용 결과의 일 평균 자료, 학습 자료의 일 평균 자료, 인공지능경망 모형 적용 결과의 일 평균 자료는  $R^2$ 가 각각 1.000, 0.999, 0.997로 원시 수위자료에 비해 소폭 증가한 것으로 나타났다.

원시 수위자료의 일 평균 자료에서  $MAE$ ,  $RMSE$ 는 각각 0.126 m, 0.396 m였다. 임계치 모형 적용 결과의 일 평균 자료는  $MAE$ ,  $RMSE$ 가 각각 0.034 m, 0.063 m로 가장 가장 낮아 임계치 모형의 성능이 가장 뛰어난 것으로 나타났다. 인공지능경망 모형을 적용 결과 일 평균 자료는  $MAE$ ,  $RMSE$ 가 각각 0.065 m, 0.149 m로 나타나 원시 수위자료의 일 평균 자료에 비해 품질이 개선된 결과를 보였으나, 학습 자료의 일 평균 자료의  $MAE$ ,  $RMSE$ 인 0.034 m, 0.067 m에는 미치지 못하는 것으로 나타났다.

Fig. 9는 원시 수위자료의 일 평균 자료, 임계치 모형 적용 결과의 일 평균 자료, 학습 자료의 일 평균 자료, 인공지능경망 모형 적용 결과의 일 평균 자료와 기준 자료를 각각 비교한 산점도를 도시한 것이다. (a)에서 원시 수위자료의 일 평균 자료의 spike noise로 인해 기준 자료보다 크게 측정된 이상치

가 다수 발생한 것을 확인할 수 있으며, (b)와 (d)에서 모형의 적용 결과 이상치가 제거되어 저수지 수위자료에 비해 개선된 것을 확인할 수 있다. (b)와 (d)의 산점도를 비교해 보면 임계치 모형이 인공지능경망 모형에 비해 성능이 더 뛰어난 것으로 나타났으며, (c)에서 인공지능경망 모형 학습 자료가 임계치 모형과 비슷한 수준으로 이상치를 탐지한 것을 확인할 수 있다.

일 평균 자료를 산정할 때 이상치로 분류된 자료는 제외하고 산정하므로 해당 일의 측정 자료가 모두 이상치로 분류된 경우 일 평균 자료에서도 해당 일의 수위자료는 제외된 것으로 나타나게 된다. 이상치로 분류되어 원시 수위자료에서 제외된 자료의 일 수를 비교하기 위해 원시 수위자료의 일 평균

Table 5 The statistical parameters ( $R^2$ ,  $MAE$ ,  $RMSE$ ) compared to reference data

Statistical parameters	Raw data	Threshold data	Target data	ANN data
$R^2$	0.982	1.000	0.999	0.997
$MAE$ (m)	0.126	0.034	0.034	0.065
$RMSE$ (m)	0.396	0.063	0.067	0.149

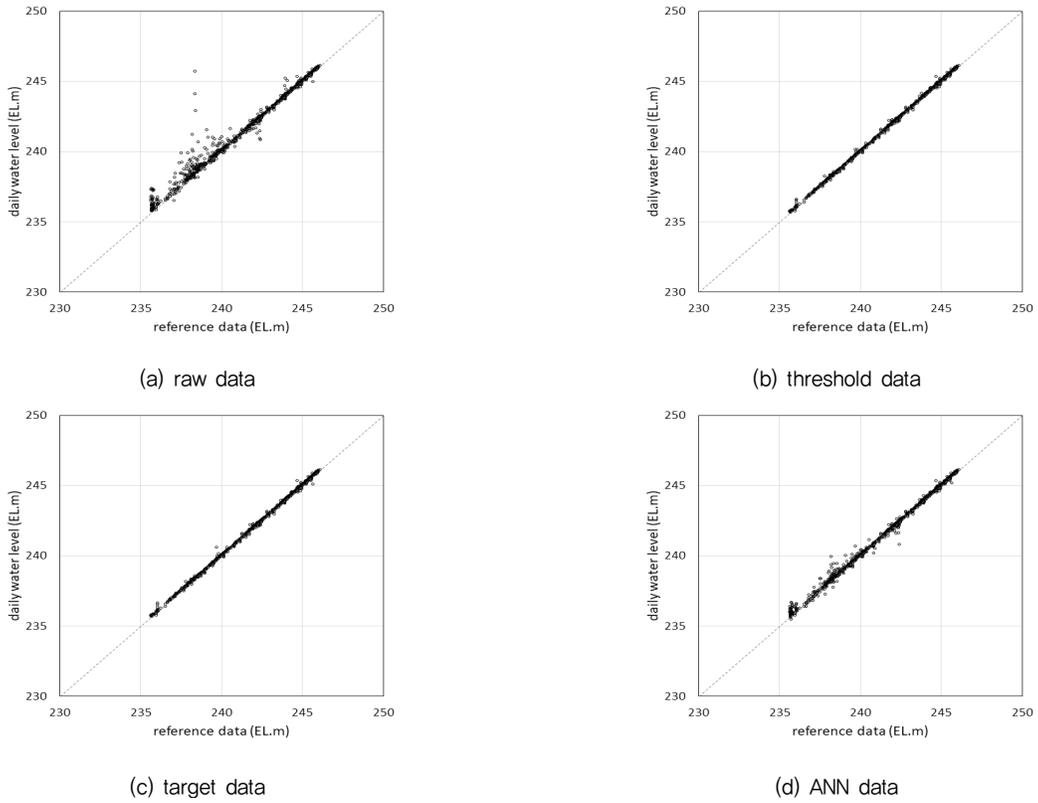


Fig. 9 The scatter plot of (a) raw data, (b) threshold data, (c) target data, (d) ANN data compared with reference data

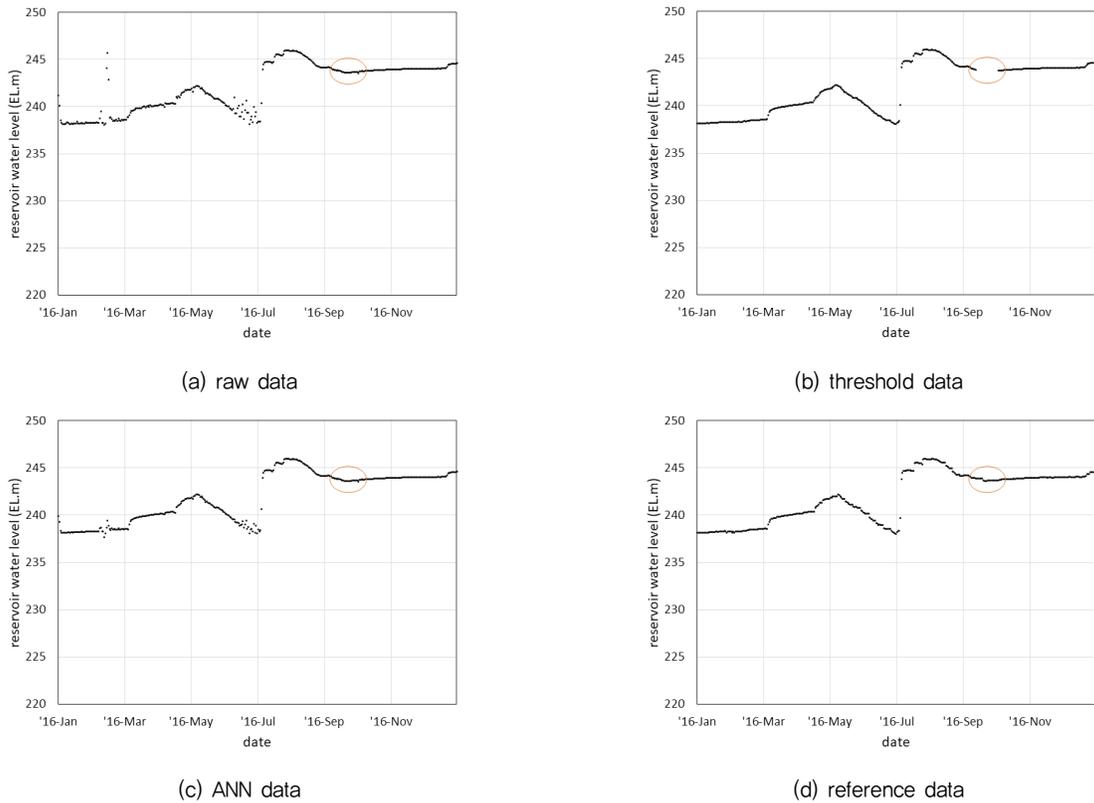


Fig. 10 Daily mean value of (a) raw data, (b) threshold data, (c) ANN data, (d) reference data (2016. 01. 01.~2016. 12. 31.)

자료, 임계치 모형 적용 결과의 일 평균 자료, 인공신경망 모형 적용 결과의 일 평균 자료를 기준으로 제외된 자료의 일 수를 산정하였다. 원시 수위자료의 일 평균 자료와 인공신경망 모형 적용 결과의 일 평균 자료에서 나타난 제외된 자료의 일 수는 109일로 동일하며 제외된 시점 또한 동일하다. 반면 임계치 모형 적용 결과의 일 평균 자료에서 나타난 제외된 자료의 일 수는 129일로 원시 수위자료의 일 평균 자료보다 20일이 더 많았는데 이는 임계치 모형의 경우 2016. 09. 13. ~ 2016. 10. 12.의 자료를 이상치로 분류하여 추가적으로 제외했기 때문이다. Fig. 10은 2016. 01. 01~2016. 12. 31.의 원시 수위자료, 임계치 모형 적용 결과, 인공신경망 모형 적용 결과의 일 평균 자료를 그래프로 도시한 것으로 임계치 모형에서 정상 수위자료 임에도 이상치로 분류하여 추가적으로 제외한 자료를 확인할 수 있다.

즉, 임계치 모형은 인공신경망 모형에 비해 정상 수위자료를 더 많이 이상치로 분류하였다. 통계치로 비교해보면 임계치 모형이 인공신경망 모형보다 성능이 뛰어나지만 F의 수행 오차는 더 크게 나타났다. 임계치 모형의 적용 결과를 인공신경망의 학습 자료와 비교해 보았을 때 F의 수행 오차는 13.40%로 이는 인공신경망 모형에서 F의 수행 오차인 5.22%의 약 2.6배이다.

#### IV. 결론

본 연구에서는 한국농어촌공사에서 10분 단위로 측정된 개운 저수지 수위자료를 대상으로 임계치를 이용한 이상치 탐지 모형과 인공신경망을 이용한 이상치 탐지 모형을 적용하고 분석하였다. 모형의 적용 결과를 분석하고 평가하기 위해서 한국농어촌공사에서 제공하는 일 단위 수위자료를 기준자료로 이용하였으며, 모형이 이상치로 분류한 자료와 결측 자료를 제외하여 모형 적용 결과를 얻었다. 기준 자료와 원시 수위자료의 측정 간격이 다르기 때문에 모형 적용 결과의 일 평균을 통해 자료의 간격을 동일하게 한 후 기준 자료와 비교하였다.

원시 수위자료는  $R^2$ , MAE, RMSE가 각각 0.982, 0.126 m, 0.396 m였다. 임계치 모형 적용 결과  $R^2$ , MAE, RMSE는 각각 1.000, 0.034 m, 0.067 m였고, 인공신경망 모형 적용 결과  $R^2$ , MAE, RMSE는 각각 0.997, 0.065 m, 0.149 m로 원시 수위자료와 비교했을 때 이상치가 적절히 탐지되어 품질이 개선된 것으로 나타났으며 임계치 모형이 인공신경망 모형보다 더 뛰어난 성능을 보였다.

그러나 일 평균 자료를 기준으로 산정한 제외된 자료의 일 수는 임계치 모형에서 129일, 원시 수위자료와 인공신경

망 모형에서 109일로 임계치 모형에서 20일 가량 더 많았으며, F의 수행오차 또한 인공신경망 모형에 비해 더 큰 것으로 나타났다. 이는 임계치 모형이 인공신경망 모형에 비해 정상 수위자료를 이상치로 잘못 분류한 수가 더 많다는 것을 의미한다.

인공신경망 모형의 경우 학습 자료의 영향을 크게 받으므로 학습 자료 생성 과정을 더욱 엄밀히 하여 신뢰도 높은 학습 자료를 구축한다면 모형의 성능을 개선할 수 있을 것으로 기대된다. 또한 인공신경망 모형은 입력 자료에 따라 결과가 달라지므로 선행 강우와 기상 자료 등을 입력 자료에 추가적으로 활용하여 성능의 변화를 분석하고 모형을 개선시킬 수 있을 것으로 사료된다.

임계치 모형이 인공신경망 모형에 비해 이상치 탐지 성능은 더 뛰어났으나 F의 수행오차에서는 인공신경망 모형의 성능이 더 뛰어났다. 그러나 저수지 수위자료에서 T의 수가 F수에 비해 약 7.8배 많기 때문에 수행오차가 미치는 영향이 크지는 않으므로 인공신경망 모형에 비해 임계치 모형의 성능이 더 우수하다고 사료되며, 학습 자료와 입력 자료의 개선을 통해 인공신경망 모형의 성능을 향상시킬 수 있을 것이라 기대된다.

모형을 이용한 이상치 탐지 방법을 저수지 수위자료에 적용하면 현재 사용하고 있는 수동 보정 방법에 비해 시간과 노동력을 크게 절감할 수 있으며, 저수지 수위자료의 신뢰성을 높여 현장에서 활용도를 크게 증가시킬 수 있을 것으로 사료된다.

#### 감사의 글

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2017R1E1A1A01077413).

#### REFERENCES

1. Ahn, J. H., M. S. Kang, I. H. Song, K. D. Lee, J. H. Song, and J. R. Jang, 2012. Estimation of surface runoff from paddy plots using an artificial neural network. *Journal of the Korean Society of Agricultural Engineers* **54(4)**: 66-71 (in Korean). doi:10.5389/KSAE.2012.54.4.065.
2. Ahn, S. J., I. S. Yeon, and K. I. Kim, 2000a. Rainfall forecasting Using neural network. *Journal of The Korean Society of Civil Engineers* **20(5B)**: 711-722 (in Korean).
3. Ahn, S. J., K. W. Jun, and K. I. Kim, 2000b. Forecasting

- of runoff hydrograph using neural network algorithms. *Journal of Korean Water Resources Association* **33(4)**: 505-515 (in Korean).
4. Bang, J. H., Y. H. Lee, S. Y. Jeong, and J. Y. Choi, 2017. A study of outlier detection on time series of water-level in agricultural reservoir. In *Proceedings of the Korean Society of Agricultural Engineers Conference* 60 (in Korean).
  5. Choi, J. K., and M. S. Kang, 2000. Application of neural network to water resources. *Journal of Korean National Committee on Irrigation and Drainage* **7(2)**: 248-258 (in Korean).
  6. Choi, J. Y., 2018. Development of quality control methods for water level data and irrigation water supply estimation, 4-10. Korea Rural Community Corporation.
  7. Feng, S., Q. Hu, and W. Qian, 2004. Quality control of daily meteorological data in China, 195-2000: a new dataset. *International Journal of Climatology* **24**: 853-870. doi:10.1002/joc.1047.
  8. González-Rouco, J. F., J. L. Jiménez, V. Quesada, and F. Valero, 1999. Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate* **14**: 964-978. doi:10.1175/1520-0442(2001)014<0964:QCAHOP>2.0.CO;2.
  9. Günther, F., and S. Fritsch, 2010. Neuralnet: training of neural networks. *The R Journal* **2(1)**: 30-38.
  10. Harrison, D. L., S. J. Driscoll, and M. Kitchen, 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Applications* **7(2)**: 135-144. doi:10.1017/S1350482700001468.
  11. Horsburgh, J. S., D. G. Tarboton, D. R. Maidment, and I. Zaslavsky, 2008. A relational model for environmental and water resources data. *Water Resources Research* **44(5)**. doi:10.1029/2007WR006392.
  12. Jang, S. H., J. Y. Yoon, S. D. Kim, and Y. N. Yoon, 2007. An establishment of operation and management system for flood control and conservation in reservoir with date: I. establishment of real-time inflow prediction model using recorded rainfall data. *Journal of The Korean Society of Civil Engineers* **27(2B)**: 133-140 (in Korean).
  13. Jeong, G. H., and T. W. Kim, 2007. Comparing water distribution model with reservoir and water system management. *Water and Future* **40(10)**: 38-43 (in Korean).
  14. Kang, B. S., and B. K. Lee, 2008. Predicting probability of precipitation using artificial neural network and mesoscale numerical weather prediction. *Journal of The Korean Society of Civil Engineers* **28(5B)**: 485-493 (in Korean).
  15. Kang, M. G., H. S. Jeong, and J. T. Kim, 2010. Efficient management of agricultural canal systems through quality management of water level and water quantity data. *Rural Resources* **52(2)**: 87-96 (in Korean).
  16. Kilonsky, B. J., and P. Caldwell, 1991. In the pursuit of high-quality sea level data. In *Oceans 91 Proceedings* (2): 669-675. doi:10.1109/OCEANS.1991.627921.
  17. Kim, C. H., J. A. Ryu, D. G. Kim, and G. B. Kim, 2016. Analysis of the effects of drainage systems in wetlands based on changes in groundwater level, soil moisture content, and water quality. *The Journal of Engineering Geology* **26(2)**: 251-260 (in Korean). doi:10.9720/kseg.2016.2.251.
  18. Kim, H. G., M. I. Kim, M. S. Lee, Y. S. Park, and J. H. Kwak, 2017. Correlation of deep landside occurrence and variation of groundwater level. *Journal of The Korea Society of Forest Engineering and technology* **15(1)**: 1-12 (in Korean).
  19. Kim, H. K., S. M. Kim, and S. W. Park, 2006. Development of hydrologic data management system based on relational database. *Journal of Korean Water Resources Association* **39(10)**: 855-866 (in Korean).
  20. Kim, S. J., Y. S. Kwon, K. H. Lee, and H. S. Kim, 2010. Radar rainfall adjustment by artificial neural network and runoff analysis. *Journal of The Korean Society of Civil Engineers* **30(2B)**: 159-167 (in Korean).
  21. Kim, S. W., 2000. A study on the forecasting of daily streamflow using the multilayer neural networks model. *Journal of Korean Water Resources Association* **33(5)**: 537-550 (in Korean).
  22. Kim, S. W., and J. D. Salas, 2000. The flood water stage prediction based on neural networks method in stream gauge station. *Journal of Korean Water Resources Association* **33(2)**: 247-262 (in Korean).
  23. Lee, C. W., Y. S. Mang, and Y. S. Kim, 2014. Behavior of fill dam subjected to continuous water level change and overflow. *Journal of the Korean Geo-Environmental Society* **15(6)**: 41-48 (in Korean). doi:10.14481/jkges.2014.15.6.41.
  24. Mounce, S. R., R. B. Mounce, and J. B. Boxall, 2011. Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics* **13(4)**: 672-686. doi:10.2166/hydro.2010.144.
  25. Mourad, M., and J. L. Bertrand-Krajewski, 2002. A method for automatic validation of long time series of data in urban hydrology. *Water Science & Technology* **45(4-5)**: 263-270. doi:10.2166/wst.2002.0601.

26. Oh, C. R., S. C. Park, H. M. Lee, and Y. P. Pyo, 2002. A forecasting of water quality in the Youngsan river using neural network. *Journal of The Korean Society of Civil Engineers* **22(3B)**: 372-382 (in Korean).
27. Oh, J. W., J. H. Park, and Y. K. Kim, 2008. Missing hydrological data estimation using neural network and real time data reconciliation. *Journal of Korean Water Resources Association* (**10**): 1059-1065 (in Korean). doi:10.3741/JKWRA.2008.41.10.1059.
28. Park, J. H., M. S. Kang, J. H. Song, and S. M. Jun, 2015. Design and implementation of IoT-based intelligent platform for water level monitoring. *Journal of the Korean Society of Rural Planning* **21(4)**: 177-186 (in Korean). doi:10.7851/ksrp.2015.21.4.177.
29. Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data its role in quantifying the global water cycle. *Theoretical and Applied Climatology* **115(1-2)**: 15-40. doi:10.1007/s00704-013-0860-x.
30. Seo, Y. M., E. H. Choi, and W. K. Yeo, 2017. Reservoir water level forecasting using machine learning models. *Journal of the Korean Society of Agricultural Engineers* **59(3)**: 97-110 (in Korean). doi:10.5389/KSAE.2017.59.3.097.
31. Shim, S. B., S. K. Kim, R. H. Park, and D. K. Hoh, 1997. Decision support system for real-time reservoir operation during flood period. *Journal of Korean Water Resources Association* **30(5)**: 431-439 (in Korean).
32. Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden, 1999. Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resources Research* **35(8)**: 2487-2503. doi:10.1029/1999WR900142.
33. Yeo, W. K., Y. M. Seo, S. Y. Lee, and H. K. Ji, 2010. Study on water stage prediction using hybrid model of artificial neural network and genetic algorithm. *Journal of Korean Water Resources Association* **43(8)**: 721-731 (in Korean). doi:10.3741/JKWRA.2010.43.8.721.
34. Yoon, K. H., B. C. Seo, and H. S. Shin, 2004. Dam inflow forecasting for short term flood based on neural networks in Nakdong river basin. *Journal of Korean Water Resources Association* **37(1)**: 67-75 (in Korean).