Proceedings of the 9ICEG
9th International Congress on Environmental Geotechnics
25-28 June, 2023 | Chania, Greece

https://doi.org/10.53243/ICEG2023-104

Regional-specific influence consideration for groundwater level prediction using machine learning-based approach

Incheol Kim[1], Jiyeong Lee[2], Qaisar Abbas[2], Jonghyeog Yoon[2], Donggun Nam[2], Jongwhan Eun[3], Junhwan Lee[4]

[1]*Postdoctoral associate, University of Nebraska-Lincoln, Lincoln, USA, email: ickim@unl.edu*

[2]*Ph.D student, Yonsei University, Seoul, Korea, representative email: jiyeong12@yonsei.ac.kr*

[3]*Associate professor, University of Nebraska-Lincoln, Lincoln, USA, email: jeun2@unl.edu*

[4]*Professor, Yonsei University, Seoul, Korea, email: junlee@yonsei.ac.kr*

ABSTRACT

Groundwater level (GWL) is one of the important climate change components, directly and indirectly related to geoenvironmental disasters such as ground subsidence, flood, and drought. For two decades, many data-based approaches, including machine learning (ML), have mainly focused on the optimization algorithm and computational process. In this study, a new methodology was proposed for considering the regional-specific infiltration characteristics of given study sites. For this, the time delay between the influencing factors and GWL of each study site was introduced to the ML process, based on the trend-fitting technology of the moving average. Two study sites in South Korea were selected and the performance of the proposed methodology was verified. In addition, it was found that the proper time lag consideration for the GWL prediction is quite important than the amount of learning data length. This indicates that regional-specific characteristics need to be taken into account for more effective and precise GWL prediction.

Keywords: Climate change, groundwater level, machine learning, time delay, moving average, prediction

1 INTRODUCTION

Groundwater level (GWL) is one of the important components influencing climate change-related geohazard, such as flooding, drought, and wildfire. From the geotechnical engineer's point of view addition, GWL fluctuation causes a change in the stress state within soils and might result in the stability and serviceability of the entire structure with both the decrease in the bearing capacity and unexpected settlement of the foundation structure (Morgan et al. 2010; Shahriar et al. 2015; Park et al. 2019; Xu et al. 2020; Rodrigues et al. 2021; Kim et al. 2022). Therefore, the investigation and estimation of GWL can be one of the essential steps for the sustainable design adaptative to climate change.

Several methodologies have been used to simulate and predict GWL fluctuation, such as analytical, numerical, statistical, and data-based approaches (Serrano and Workman 1998; Batelaan et al. 2003; Sahoo and Jha 2013; Kim and Lee 2018a, 2019; Rajaee et al. 2019; Kim and Lee 2022a). As a data-based approach, the artificial neural network (ANN), support vector machines (SVMs), and adaptive neuro-fuzzy inference system (ANFIS) have been the most frequently selected machine learning models for predicting GWL fluctuations (Rajaee et al. 2019; Vu et al. 2021). In this approach, the influencing components, such as precipitation, river stage, sea level, and groundwater pumping, are selected as inputs and they are trained with already given GWL time series data in the network structures. Accordingly, the quantity and quality of the given input datasets significantly affect the prediction performance of GWL (Rajaee et al. 2019; Kim and Lee 2022a and b).

The aspects of GWL fluctuation can be different depending on the local geological, geomorphic, and land use characteristics. For example, the response of GWL fluctuation due to rainfall within the silty soils should be slower than that within the sandy soils (Hoque et al. 2007; Kim et al. 2016). However, in the majority of cases, data-based approaches rarely consider regional-specific infiltrating characteristics for their prediction models (Kim and Lee 2022a).

In this study, a new methodology to consider the regional-specific characteristics of the data-based approaches was proposed based on the trend-fitting technology of the moving average. For this, a series of correlation analyses between influencing components and GWL was performed and GWL fluctuation was predicted using ANN as one of the representative machine-learning models. Two study sites in South Korea were selected and precipitation and river stage datasets were used as the main influencing components on GWL fluctuation. The prediction performance of the proposed and existing methodology was compared and discussed. Selected datasets, results, and discussion in this study were based on the results of Kim and Lee's previous research (2022a)

2 GROUNDWATER PREDICTION USING ANN

2.1 ANN Principle and Algorithm

ANN has been one of the models frequently selected for GWL prediction for about two decades. Generally, ANN consists of three layers, which are the input layer, the hidden layer, and the output layer. The input layer has some given or selected influence factors, the output layer is mainly the target variable, that is, GWL in this study, as shown in Figure 1. In the hidden layer, there are nodes connecting the given inputs and target output with multiple weight values. For obtaining the optimum weight values, the process of the feed-forward and backpropagation is repeated. The feedforward neural network (FNN) and Levenberg-Marquardt (LM) algorithm has been regarded as appropriate and effective for the prediction of GWL (Daliakopoulos et al. 2005; Nourani et al. 2008). The number of hidden layers and nodes is different depending on the cases. Therefore, the optimum numbers should be determined from some iterative testing process as well at the initial stage. Generally, the trial-and-error method or the genetic algorithm (GA) is selected for the prediction of GWL (Gerken et al. 2006; Sahoo and Jha 2013; Kim and Lee 2018b).

2.2 Influence Components on GWL Fluctuation

The main influencing factors on GWL fluctuation can be different depending on the geological, geomorphic, and land use characteristics of given study sites (Hoque et al. 2007; Kim et al. 2016; Kim and Lee 2018c). In most cases, however, these characteristics have been barely considered for the data-based GWL prediction. Meanwhile, Wilhite and Glantz (1985) and Guttman (1999) correlated GWL with precipitation by the time-series analysis [i.e. moving average (MA) of precipitation] to consider the regional infiltration characteristic of given study sites. Kim et al. (2016) conducted an influencing factor analysis on GWL fluctuation for urbanized and rural areas using the moving average (MA) of precipitation. Kim and Lee (2022a) introduced MA of precipitation as an input variable for GWL prediction by ANN and showed that MA of precipitation significantly enhanced the prediction performance of GWL.

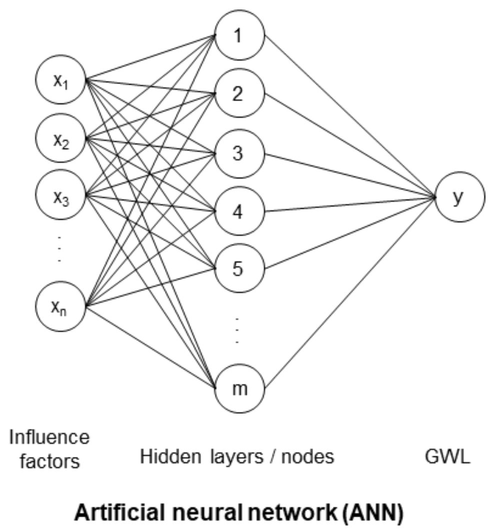
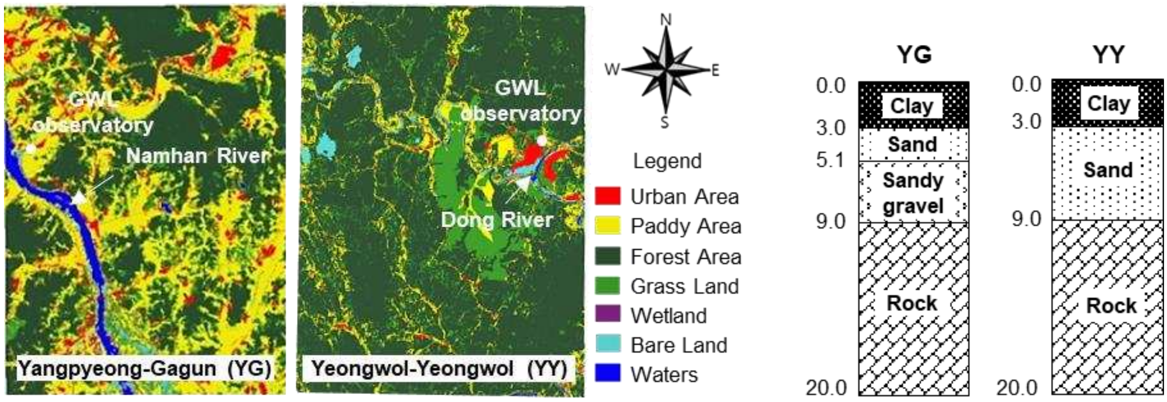


Figure 1. *Conceptual structure for artificial neural network (ANN)*

3 ANALYSIS OF INFLUENCE COMPONENTS ON GWL

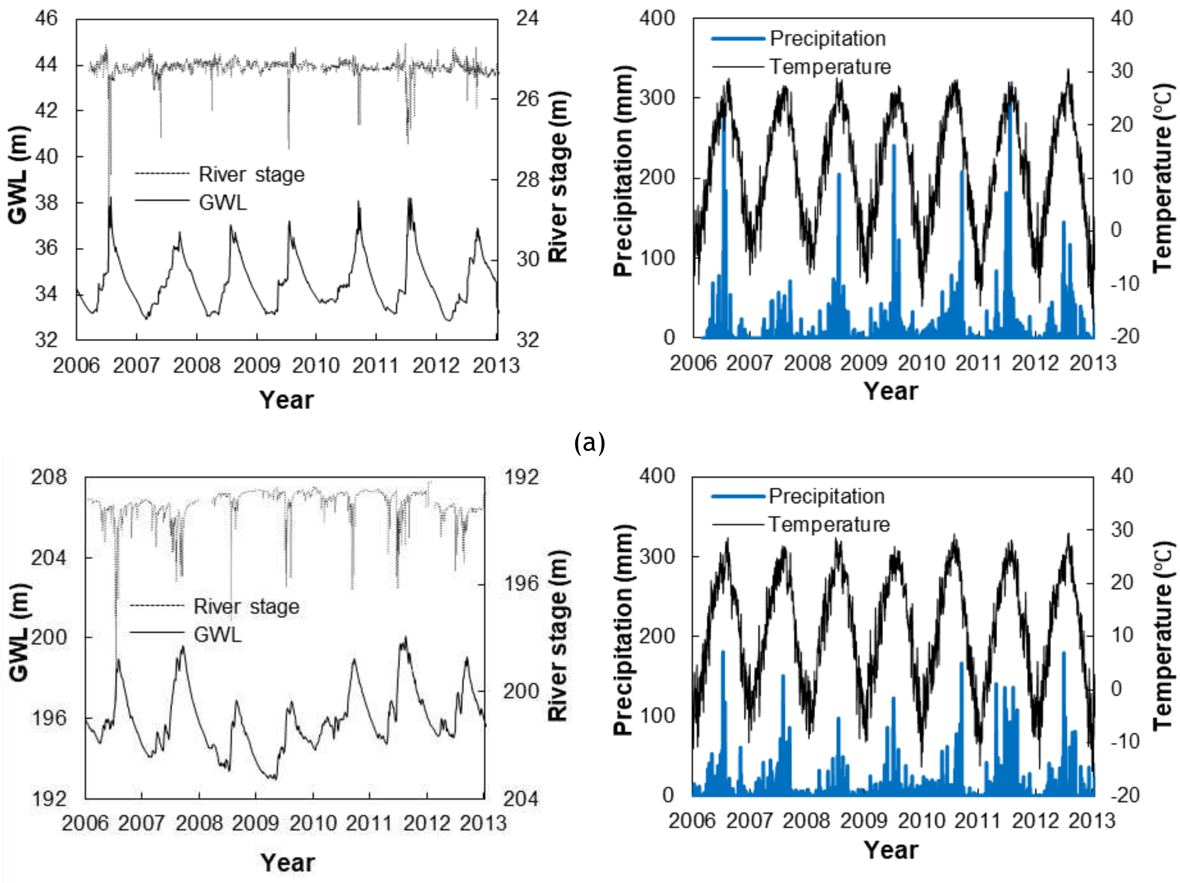
3.1 Study Site

In this study, two study sites in Korea were selected and introduced to the influencing factor analysis and GWL prediction by ANN. The selected study sites are Yangpyeong-Gagun (YG) and Yeongwol-Yeongwol (YY) and Figure 2(a) shows the location and the ground surface condition of the study sites. Both study sites are in rural areas and close to a river. There are clay layers from the surface layer to 3 m depth and below the layer sand or sandy gravel layer exists at both study sites, as shown in Figure2(b). Figure 3 shows the 7-year measured GWL, river stage (RS), precipitation, and temperature datasets, which were used for the influencing factor analysis and GWL prediction by ANN in this study. The datasets show strong seasonal period change and especially precipitation is quite large in the summer season.



(a) (b)

Figure 2. *Study site information: (a) location and surface conditions and (b) soil profile.*

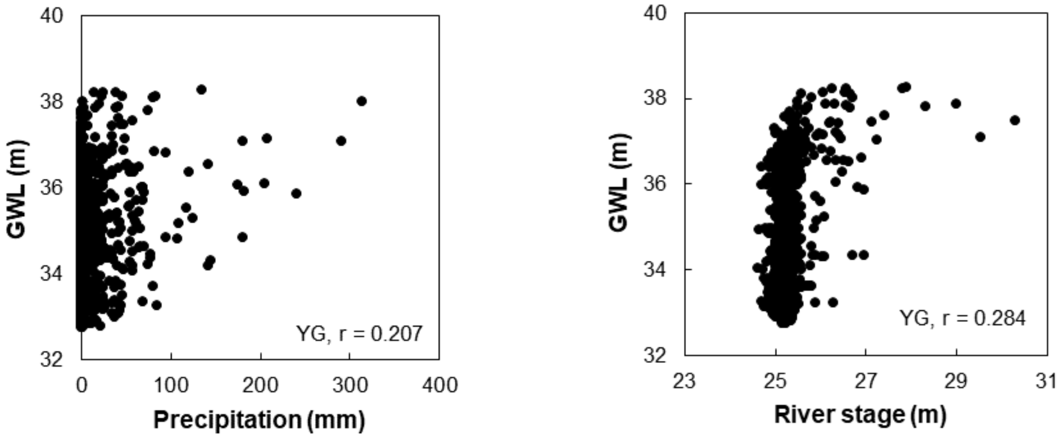


(b)

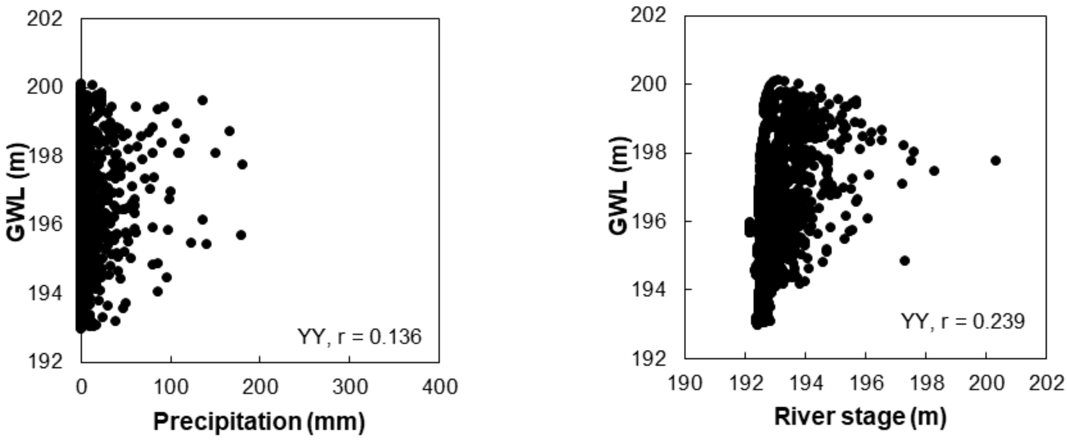
Figure 3. *Observed GWL, RS, precipitation, and temperature datasets at (a) YG and (b) YY.*

3.2 Correlation Analysis: Precipitation and River Stage (RS)

The influencing degree of precipitation and RS on GWL was analyzed using the 7-year measured datasets of Figure 3. Figure 4 shows the general correlation analysis results for YG and YY. The coefficient of correlation, r, at YY was 0.207 and 0.284 with precipitation and RS, respectively [Figure 4(a)] and it indicates that the correlation of GWL with precipitation and RS is quite small. For YY, a similar tendency was observed as shown in Figure 4(b). It is interesting that the coefficients of correlation were so small that it indicates almost no correlation between GWL and precipitation although similar seasonal period changes were observed in Figure 3. The possible reason for this might be due to the time delay effect between precipitation and GWL fluctuation. The low permeable clay soil of the ground surface might delay the rainwater infiltration into GWL. However, the time delay effect was not considered in the general correlation analysis. The other possible reason might be the difference in data continuity between GWL and precipitation. For example, the time series data of GWL is continuous with time, but the precipitation data is intermittent.



(a)



(b)

Figure 4. *Correlation between GWL with precipitation and RS at (a) YG and (b) YY.*

3.3 Correlation Analysis: Moving Average (MA) of Precipitation

MA is one of the time-series models and its trend-fitting method makes the original time-series data be smooth. Some academic reports revealed the concept of MA is useful to correlate GWL with precipitation considering the time delay effect and the different data continuity (Wilhite and Glantz 1985; Guttman 1999; Kim et al. 2016). In this study, the trend-fitting of MA for precipitation data (MA of precipitation) was introduced for additional correlation analysis between GWL and precipitation and the arithmetical expression is following:

$$MA_{t,m} = \frac{Pre_t + Pre_{t-1} + \cdots + Pre_{t-(m-1)}}{m} = \frac{1}{m}\sum_{i=0}^{m-1} Pre_{t-1} \quad (1)$$

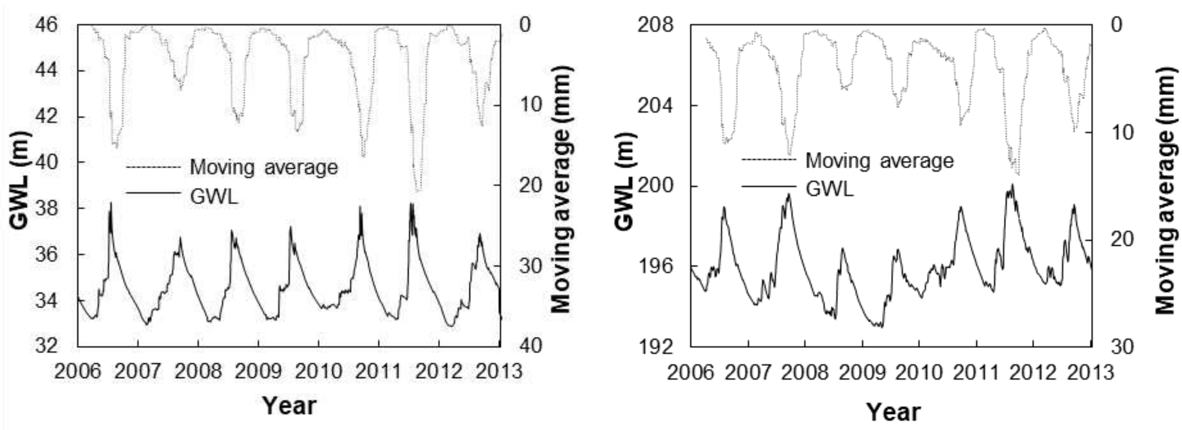
where $MA_{t,m}$ is MA of precipitation value for *m* days on day *t*. *Pre* is precipitation. Note that MA of precipitation is different depending on the considered number of days (*m*). For example, the longer time delay between GWL and precipitation corresponds with the larger value of *m*.

Figure 5 shows MA of precipitation for YG and YY calculated using the 7-year measured precipitation data and Eq. (1). The values of *m* were 82 and 105 for YG and YY, respectively, which were obtained from the best correlation between GWL and MA of precipitation. It is shown from Figure 5 that MA of precipitation is the clear seasonal period fluctuations with time and the peak values for MA of precipitation matches well with GWL for both study sites. Figure 6 shows the results of the correlation analysis between GWL and MA of precipitation. The coefficient of correlation, r, were 0.839 and 0.862 for YG and YY. The correlation of MA of precipitation was significantly high compared to that of precipitation (Figure 4). This indicates that the time delay between GWL and precipitation was the important factor for the correlation analysis in YG and YY. That might be because the clay soil of the ground surface, which generally has a low permeable characteristic, caused the time delay of rainwater infiltration to GWL.

4 PREDICTION OF GWL USING ANN

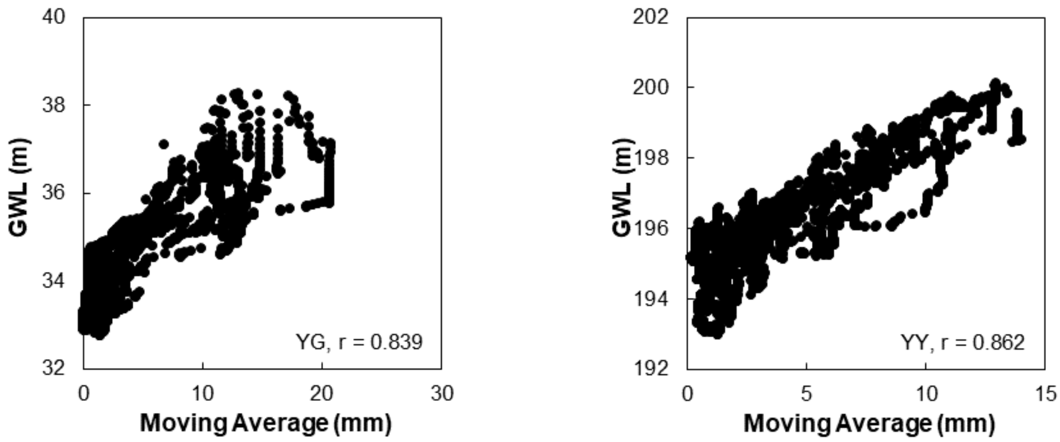
4.1 ANN Structure and Input Types

Additionally, the prediction performance of GWL for ANN was investigated with the introduction of MA of precipitation as an input, which includes the regional time-delay information between GWL and precipitation as a sort of geological, geomorphic, and land use characteristics.



(a) (b)

Figure 5. *Moving average (MA) of precipitation and GWL with time for (a) YG and (b) YY.*



(a) (b)

Figure 6. *Correlation between GWL and MA of precipitation for (a) YG and (b) YY.*

In this study, the basic ANN structure was used as shown in Figure 1 and the feedforward neural network (FNN) and Levenberg-Marquardt (LM) algorithm were used for the training algorithm. The numbers of hidden layers and nodes were determined by the trial-and-error method, and they were 3 and 4 for YH and 1 and 5 for YY, respectively. Two input types were prepared for GWL prediction. The first input type was precipitation, RS, and temperature and the second input type was precipitation, RS, and temperature, MA of precipitation. Note that temperature is frequently used to consider seasonal effects although it might be not a direct influencing factor on GWL fluctuation. The used data length was 6-year for training and 1-year for testing of Figures 3 and 5.

4.2 Results and Discussion

Figure 7 shows the measured and predicted GWL with time at YG and YY. For YG without MA of precipitation in Figure 7(a), the coefficient of correlation and the root mean square error (RMSE) were 0.570 and 0.96 m, respectively, indicating the predicted fluctuation does not match well with the measured one. For YG with MA of precipitation in Figure 7(b), the coefficient of correlation and RMSE were 0.976 and 0.36 m, respectively. It indicates that the predicted fluctuation trend was improved, and the mean error between the measured and predicted GWL was smaller. Similar results were observed at YY as shown in Figures 7(c) and (d). In both study sites, especially, the value and time for the peak GWL with MA of precipitation were predicted better than those without MA of precipitation. Figure 8 shows the compared results between the measured and predicted GWL for YG and YY, more clearly showing the enhanced prediction performance with MA of precipitation. As a result, it is found that at the study sites MA of precipitation as the input for ANN contributed to the improvement of the prediction performance of GWL. That might be because MA of precipitation includes the time delay information between GWL and precipitation and influenced the value of weights in the hidden layer and nodes as an influencing input for the data training process of ANN.

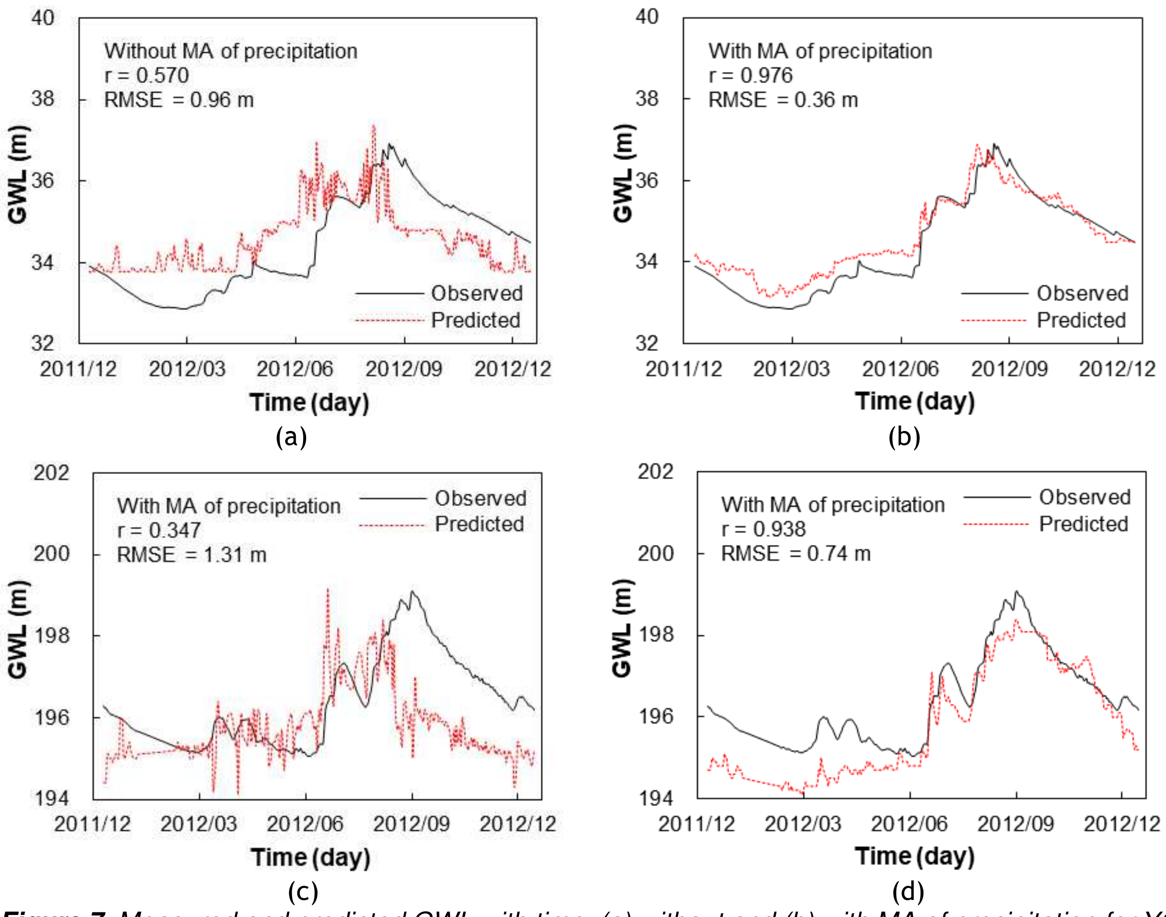


Figure 7. *Measured and predicted GWL with time: (a) without and (b) with MA of precipitation for YG and (c) without and (d) with MA of precipitation for YY.*

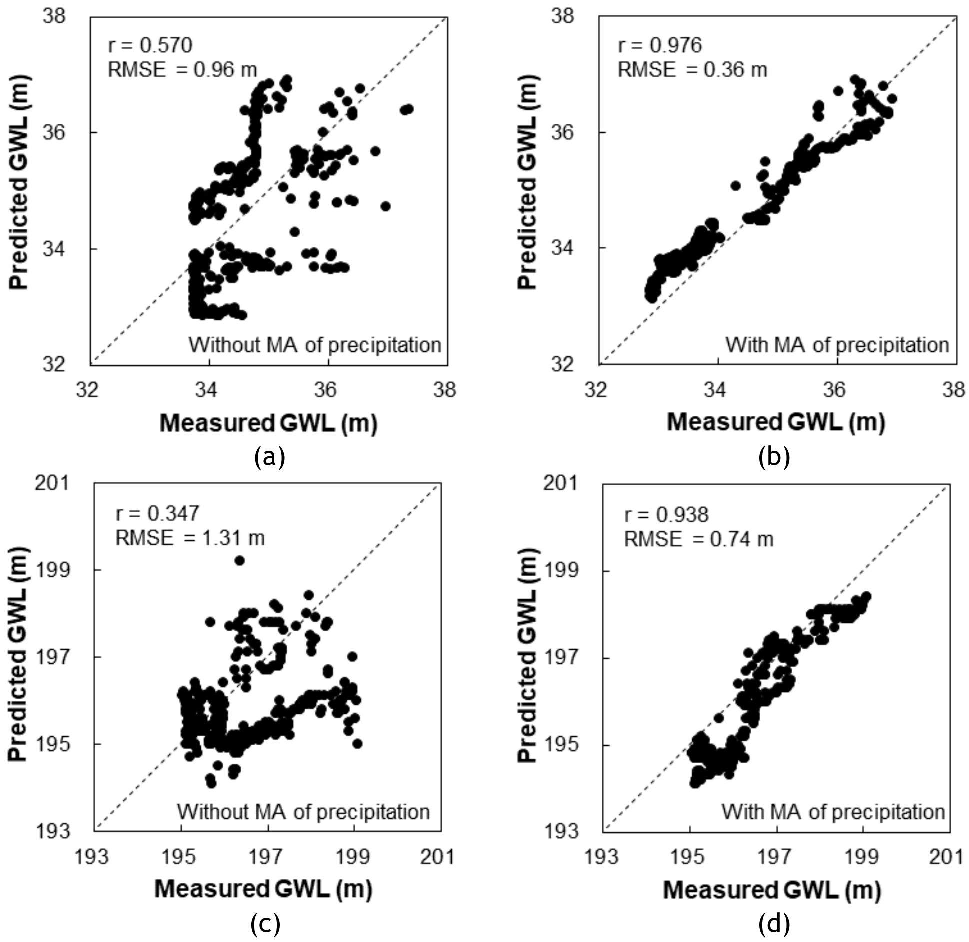


Figure 8. *Compared results between measured and predicted GWL: (a) without and (b) with MA of precipitation for YG and (c) without and (d) with MA of precipitation for YY.*

5 SUMMARY AND CONCLUSION

GWL is one of the important climate change components, which can influence the stability and serviceability of geotechnical infrastructures. ANN has been frequently selected to predict GWL fluctuation as the data-based approach. In this study, the moving average (MA) of precipitation was introduced as one of the inputs to consider the effect of the time delay between GWL and precipitation. From a series of correlation analyses, it was revealed that MA of precipitation has a stronger correlation with GWL than that of precipitation at the given study sites. In addition, MA of precipitation as input of ANN contributed to enhancing the prediction performance for GWL. That might be because MA of precipitation includes a sort of information about geological, geomorphic, and land use characteristics of the local regions. In this study, only two study sites were addressed. This finding needs to be vitrificated by conducting the analysis for various study sites.

6 ACKNOWLEDGEMENTS

This study was partially funded by the Collaboration Initiatives Global Impact Grant from the University of Nebraska Foundation (Project ID: 21000) and by the Basic Science Research Program through the National Research Foundation of Korea and Korea Agency for Infrastructure Technology Advancement with grants funded by the government of Korea (Nos. 2020R1A2C201196613 and 22SMIP-A158708-03).

REFERENCES

Batelaan, O., De Smedt, F., & Triest, L. (2003). Regional groundwater discharge: Phreatophyte mapping, groundwater modelling and impact analysis of land-use change. Journal of Hydrology, 275, 86–108. https://doi.org/10.1016/S0022-1694(03)00018-0

Daliakopoulos, I.N., Coulibaly, P., & Tsanis, I.K. (2005). Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 309, 229-240. https://doi.org/10.1016/j.jhydrol.2004.12.001

Gerken, W.C., Purvis, L.K., & Butera, R.J. (2006). Genetic algorithm for optimization and specification of a neuron model. Neurocomputing, 69, 1039-1042. DOI: 10.1109/IEMBS.2005.1615421

Guttman, N.B. (1999). Accepting the standardized precipitation index: A calculation algorithm. Journal of the American Water Resources Association, 35(2), 311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Hoque, M.A., Hoque, M.M., & Ahmed, K.M. (2007). Declining groundwater level and aquifer dewatering in Dhaka metropolitan area, Bangladesh: Causes and quantification. Hydrogeology Journal, 15, 1523–1534. https://doi.org/10.1007/s10040-007-0226-5

Kim, I., Park, D., Kyung, D., Kim, G., Kim, S., & Lee. J. (2016). Comparative influences of precipitation and river stage on groundwater levels in near-river areas. Sustainability, 8(1), 1–16. https://doi.org/10.3390/su8010001

Kim, I. & Lee. J. (2018a). Prediction model for spatial and temporal variation of groundwater level based on river stage. Journal of Hydrologic Engineering, 23(6), 06018002-1-8. doi.org/10.1061/(ASCE)HE.1943-5584.0001658

Kim, I. & Lee. J. (2018b). The effect of seasonal input on predicting groundwater level using artificial neural network. Ecology and Resilient Infrastructure, 5(3), 125–133. (in Korean). https://doi.org/10.17820/eri.2018.5.3.125

Kim, I. & Lee. J. (2018c). Influencing factor analysis on groundwater level fluctuation near river. Ecology and Resilient Infrastructure, 5(2), 72–81. (in Korean). https://doi.org/10.17820/eri.2018.5.2.072

Kim, I. & Lee. J. (2019). Proposed correlation model for groundwater level prediction based on river stage considering changes in hydrological and geological conditions. Journal of Hydrologic Engineering, 24(10), 04019042-1-14. doi.org/10.1061/(ASCE)HE.1943-5584.0001849

Kim, I. & Lee. J. (2022a). Performance analysis of ANN prediction for groundwater level considering regional-specific influence components. Groundwater, 60(3), 344-361. https://doi.org/10.1111/gwat.13156

Kim, I. & Lee. J. (2022b). Enhanced data utilization approach to improve the prediction performance of groundwater level using semianalytical and data process models. Journal of Hydrologic Engineering, 27(10), 04022021. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002201

Kim, I., Kim, G., & Lee. J. (2022). Centrifuge investigation of groundwater-induced footing settlement in sand considering rising and falling GWL phases. J. Geotech. and Geoenviron., 148(11), 04022104. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002920

Morgan, A. B., Sanjay, K. S., & Sivakugan, N. (2010). An experimental study on the additional settlement of footings resting on granular soils by water table rise. Soils Found., 50(2), 319–324. http://dx.doi.org/10.3208/sandf.50.319

Nourani, V., Mogaddam, A. A., & Nadiri, A. O. (2008). An ANN-based model for spatiotemporal groundwater level forecasting. Hydrological Processes, 22, 5054-5066. doi.org/10.1002/hyp.7129

Nourani, V. & Mousavi, S. (2016). Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method. Journal of Hydrology, 536, 10-25. doi.org/10.1016/j.jhydrol.2016.02.030

Oh, Y. Y., Hamm, S. Y., Yoon, H., & Kim, G. B. (2016). Analytical and statistical approach for evaluating the effects of a river barrage on river-aquifer interactions. Hydrological Processes, 30, 3932-3948. https://doi.org/10.1002/hyp.10920

Park, D., Kim, I., Kim, G., & Lee, J. (2019). Effect of groundwater fluctuation on load carrying performance of shallow foundation. Geomech. Eng., 18(6), 575-584. https://doi.org/10.12989/gae.2019.18.6.575

Rajaee, T, Ebrahimi, H., & Nourani. V. (2019). A review of the artificial intelligence methods in groundwater level modeling. Journal of Hydrology, 572, 336-351. doi.org/10.1016/j.jhydrol.2018.12.037

Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 model three approaches for groundwater level forecasting. Journal of Hydrology, 553, 356-373. doi.org/10.1016/j.jhydrol.2017.08.006

Rodrigues, R. A., Soares, F. V. P., & Sanchez, M. (2021). Settlement of footings on compacted and natural collapsible soils upon loading and soaking. J. Geotech. and Geoenviron., 147(4), 04021010. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002479

Sahoo, S. & Jha, M. K. (2013). Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeology Journal, 21, 1865-1887. doi: 10.1007/s10040-013-1029-5

Serrano, S. E. & Workman. S. R. (1998). Modeling transient stream/aquifer interaction with the non-linear Boussinesq equation and its analytical solution. Journal of Hydrology, 206, 145-255. doi.org/10.1016/S0022-1694(98)00111-5

Shahriar, M. A., Sivakugan, N., Das, B. M., Urquhart, A., & Tapiolas, M. (2015). Water table correction factors for settlements of shallow foundations in granular soils. Int. J. Geomech., 15(1), 06014015. https://doi.org/10.1061/(ASCE)GM.1943-5622.0000391

Tapoglou, E., Karatzas, G. P., Trichakis, I. C., & Varouchakis, E. A. (2014). A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. Journal of Hydrology, 519, 3193-3203. doi.org/10.1016/j.jhydrol.2014.10.040

Wilhite, D.A. & Glantz, M.H. (1985). Understanding the drought phenomenon: The role of definitions. Water International, 10, 110–120. https://doi.org/10.1080/02508068508686328

Xu, Y., Leung, C. F., Yu, J., & Chen, W. (2020). Centrifuge model study on settlement of strip footing subject to rising water table in loess. Can. Geotech. J., 57, 992-1005. https://doi.org/10.1139/cgj-2018-0337