

Influence of the range of data on the performance of ANN- and SVM- based time series models for reproducing groundwater level observations

Influenza dell'intervallo di valori sulle prestazioni dei modelli di serie temporali fondati su ANN e SVM per la riproduzione di osservazioni di livelli piezometrici

Heesung Yoon, Yongcheol Kim, Soo-Hyoung Lee, Kyoochul Ha

Riassunto: In questo studio sono stati sviluppati dei modelli di serie temporali per lo studio delle fluttuazioni di livelli piezometrici utilizzando reti neurali artificiali (RNA) e macchine a vettori di supporto (MVS). Per stimare la sensitività dei modelli al rango di dati di ingresso nella costruzione del modello, test numerici sono stati condotti utilizzando misure orarie di dati piezometrici in un acquifero costiero presso l'isola di Jeju in Corea del Sud. Le prestazioni dei due modelli è simile ed accettabile quando l'intervallo dei valori di ingresso è compreso nel set di dati utilizzato per la costruzione del modello. Quando invece l'intervallo dei valori dei dati è superiore a questa soglia, entrambi i modelli mostrano risultati predittivi anomali: oscillazioni dei valori nel caso del RNA e valori costanti nel caso dei MVS. Il risultato del test numerico indica che è necessario ottenere diversi tipi di variabili di ingresso ed uscita ed utilizzarli nella fase di costruzione di modelli predittivi basati su serie temporali.

Abstract: *In the present study, we designed time series models for predicting groundwater level fluctuations using an artificial neural network (ANN) and a support vector machine (SVM). To estimate the model sensitivity to the range of data set for the model building, numerical tests were conducted using hourly measured groundwater level data at a coastal aquifer of Jeju Island in South Korea. The model performance of the two models is similar and acceptable when the range of input variable lies within the data set for the model building. However, when the range of input variables is beyond it, both the models showed abnormal prediction results: an oscillation for the ANN model and a constant value for SVM. The result of the numerical tests indicates that it is necessary to obtain various types of input and output variables and assign them to the model building process for the success of design time series models of groundwater level prediction.*

Keywords: *time series models, groundwater level, artificial neural network, support vector machine.*

Parole chiave: modelli di serie temporali, livello piezometrico, reti neurali artificiali, macchine a vettori di supporto.

Yongcheol KIM ✉

Korea Institute of Geoscience and Mineral Resources,
124 Gwahang-no, Yuseong-gu, Daejeon, Republic of Korea
yckim@kigam.re.kr

Heesung YOON

Soo-Hyung LEE

Korea Institute of Geoscience and Mineral Resources,
124 Gwahang-no, Yuseong-gu, Daejeon, Republic of Korea

Ricevuto/Received: 04 March 2019-Acettato/Accepted: 15 March 2019

Publicato online/Published online: 29 March 2019

This is an open access article under the CC BY-NC-ND license:
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

© Associazione Acque Sotterranee 2019

Introduction

For a sustainable use of groundwater resources, it is necessary to monitor and predict groundwater level fluctuations (Masetti et al. 2016). As for the groundwater level prediction, typically, a physics-based numerical model has been developed and widely applied to the real field problems. As the quantity and quality of monitoring networks for the management of water resources are enlarged and improved, researches on development of time series models using measurement data have been increased. Recently, it has been reported that machine learning methods are effective and useful to solve non-linear problems and its application research to the field of earth and environment science is increasing (Maier and Dandy 2000; Nayak et al. 2006; Gill et al. 2007; Dogan et al. 2009; Yoon et al. 2016). In the present study, we designed time series models of groundwater level prediction using an artificial neural network and a support vector machine and applied them to observed groundwater level data at Handong area in the Jeju Island, South Korea. The accuracy of groundwater level prediction and the model sensitivity to data set of the model building process were examined through numerical tests.

Materials and methods

Study site and observed data

The study site is HD0 groundwater observatory at Handong area in the Jeju Island of the South Korea (Fig. 1). The distance between the HD0 and north-eastern coast of the Jeju Island is 2.1 km. The depth of the well and depth to the water table from the ground is 120m and about 41m, respectively. The Jeju Island, where the most of land surface is composed of highly permeable material, has no perennial stream (Won et al. 2006). Groundwater is a main water resource, thus, the sustainable use and systematic management of groundwater is an essential issue in the Jeju Island.



Fig. 1 - Location of the study site.
Fig. 1 - Ubicazione del sito studiato.

For the development of prediction model of groundwater level, we used hourly measured time series data of groundwater level at HD0 observatory, rainfall and tide level from 30th December, 2016 to 10th January, 2017 (Fig. 2). Results of cross correlation analyses between rainfall, tide level and groundwater level showed that the groundwater level fluctuation of the study site is highly influenced by the tidal level. Three rainfall events exist during the measurement period; however, the correlation between rainfall and groundwater level is insignificant (Fig. 3).

Artificial neural network

The present study employed an artificial neural network (ANN) and a support vector machine (SVM) to construct time series models of groundwater level prediction. The ANN is a mathematical structure which contains interconnected artificial neurons that mimics the way of information processing of biological neural networks. Generally, the ANN is composed of three layers: input, hidden and output layer, and each layer contains a certain number of nodes (Fig. 4a). A node in a layer is connected with nodes of next layer by weights and has bias values. The goal of ANN model building is to find optimized weights and biases for the given data set through repeated updating based on estimation errors, which is called training or model building process (Hagan 1996; Yoon et al. 2011). In this study, a back propagation algorithm (Rumelhart et al. 1986) is used to train a feedforward neural network which is the most common ANN structure.

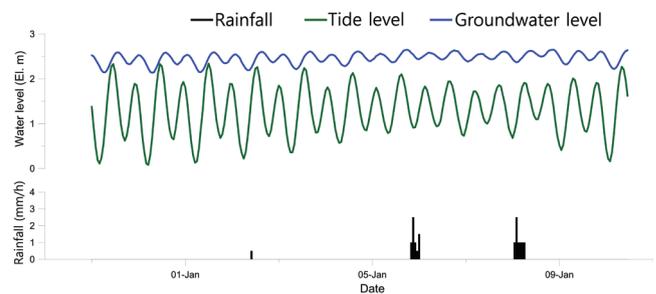


Fig. 2 - Observed time series data of groundwater level, tide level, and rainfall.

Fig. 2 - Osservazioni delle serie temporali dei livelli piezometrici e di marea e pioggia.

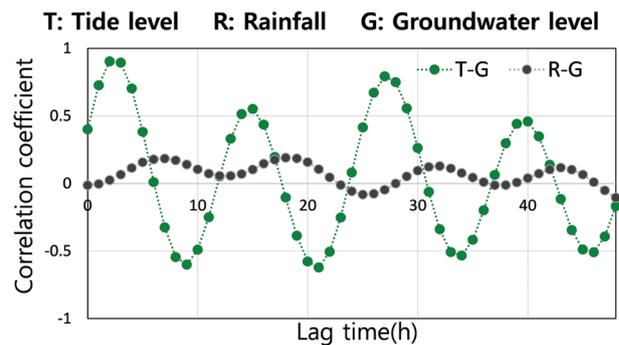


Fig. 3 - Results of cross correlation analyses between groundwater level, tide level and rainfall.

Fig. 3 - Risultati della correlazione incrociata fra livelli piezometrici e di marea e pioggia.

Support vector machine

The SVM, suggested by Vapnik (1995), is based on a structural risk minimization rather than an empirical risk minimization of the ANN, which can enhance generalization ability of the model and make the model find a global minimum. The goal of SVM model building is to find support vectors weights that are proved to be adequate for constructing model structure. An output value is estimated by inner product between input and support vectors transformed by a kernel function (Fig. 4b). In the present study a radial basis function is employed as the kernel function and a sequential minimum optimization algorithm (Platt 1999) is used to select support vectors and weights the SVM structure.

Numerical tests of time series models

Numerical tests were conducted to estimate the sensitivity of ANN- and SVM-based time series models to the range of data set allocated to training stage.

Test 1

In Test 1, a front part of the time series data of 150 hours is allocated to the model training data set. Input variables of the model include past measurement data of rainfall, tide and groundwater level. In the training data of Test 1 there exists one small rainfall event, however, in the rest of the

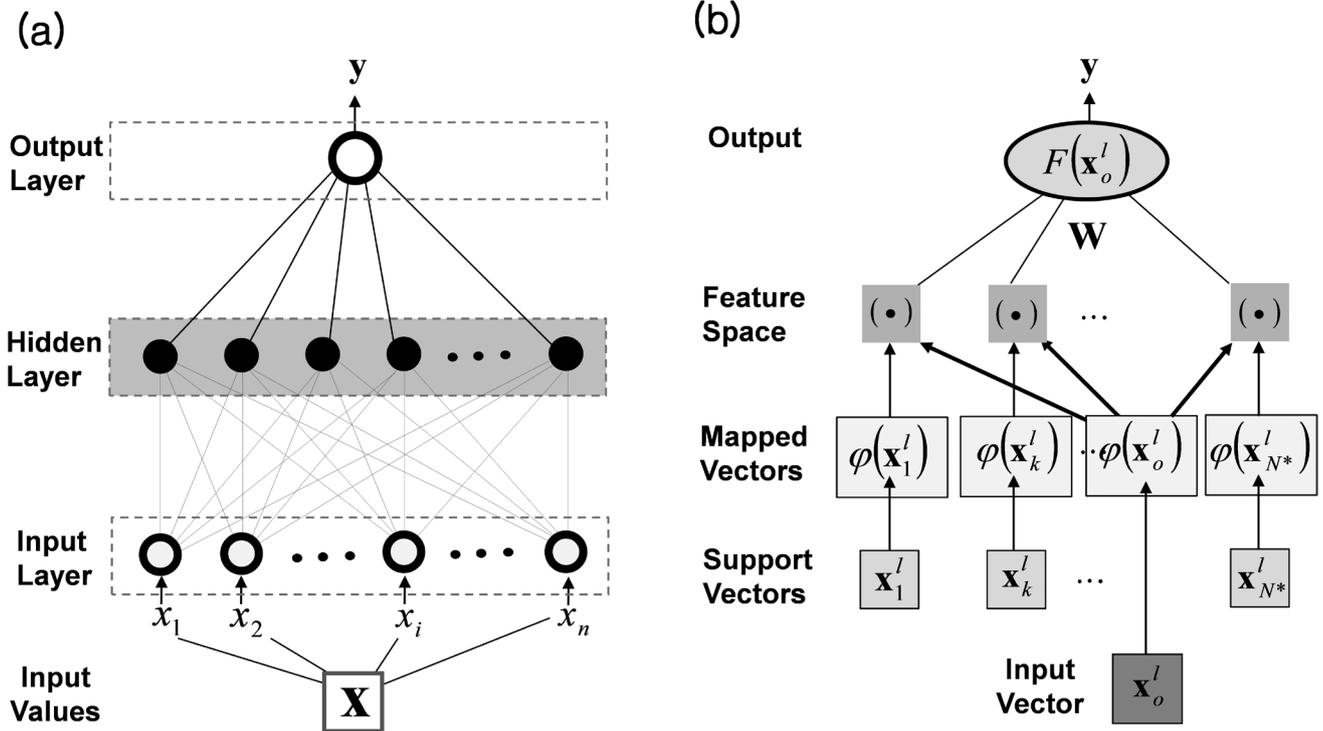


Fig. 4 - Schematic diagrams of ANN(a) and SVM(b) structures.
 Fig. 4 - Diagramma schematico delle strutture di RNA e MVS.

training data set, that is a validation data set, there exists two relatively big rainfall events.

For the ANN model design, three model parameters should be determined: number of nodes in the hidden layer (HN), learning rate (LR), and momentum (MM). In the present study, the model parameters were selected through a grid search method in the range of HN: [2, 5, 10, 15, 20, 25], LR: [0.001, 0.005, 0.01, 0.015, 0.02, 0.025], MM: [0.0, 0.1, 0.3, 0.5, 0.7, 0.9]. The SVM model design also requires the selection of three model parameters: C, ϵ and σ . And they were selected as the same method in the range of C: [0.5, 1.0, 3.0, 5.0, 7.0, 10.0], ϵ : [0.01, 0.05, 0.1, 0.11, 0.12, 0.13], σ : [0.5, 1.0, 1.5, 2.0, 2.5, 3.0]. The selected model parameters of the ANN and SVM are shown in Table1.

Tab. 1 - The selected ANN and SVM model parameters in the Test 1.

Tab. 1 - Parametri selezionati per i modelli RNA e MVS nel Test 1.

ANN			SVM		
HN	LR	MM	C	ϵ	σ
15	0.001	0.5	7.0	0.05	3.0

Root mean squared error (RMSE) between observed and predicted groundwater level values is calculated for the ANN and SVM time series models (Tab. 2). Figure 5 shows the results of time series modeling of the groundwater level fluctuations. The groundwater level prediction results for the training data set show that proper models have been made. However, the prediction for the validation data set shows

abnormal values: highly oscillatory values of the ANN model and constant values of the SVM in a specific period of time where the two rainfall events occurred. The amount of the rainfall in the validation data set is about five times higher than the rainfall in the training.

Tab. 2 - RMSE values of groundwater level prediction in Test 1.

Tab. 2 - Radice dell'errore quadratico medio dei valori piezometrici nel Test 1.

Error	ANN		SVM	
	Model Building	Model Validation	Model Building	Model Validation
RMSE (m)	0.015	0.113	0.018	0.109

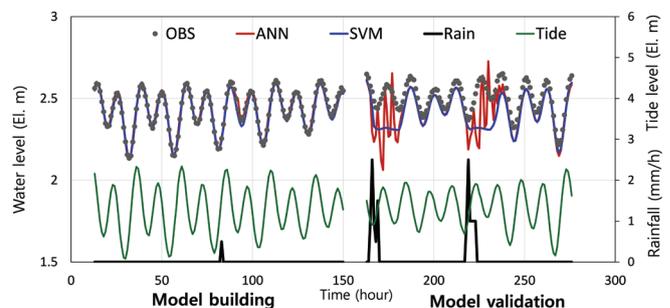


Fig. 5 - Results of groundwater level prediction in Test 1.

Fig. 5 - Risultati delle previsioni dei livelli piezometrici nel Test 1.

Test 2

In Test 2, the rear part of the time series data of 100 hours with the two relatively high rainfall events is allocated to the model training data set. The selected model parameters are described in Tab. 3. The RMSE values and prediction results of the Test 2 are shown in Tab. 4 and Fig. 6, respectively.

Tab. 3 - The selected ANN and SVM model parameters in the Test 2.

Tab. 3 - Parametri selezionati per i modelli RNA e MVS nel Test 2.

ANN			SVM		
HN	LR	MM	C	ϵ	σ
15	0.001	0.0	10.0	0.05	2.5

Tab. 4 - RMSE values of groundwater level prediction in Test 2.

Tab. 4 - Radice dell'errore quadratico medio dei valori piezometrici nel Test 2.

Error	ANN		SVM	
	Model Building	Model Validation	Model Building	Model Validation
RMSE (m)	0.043	0.017	0.062	0.018

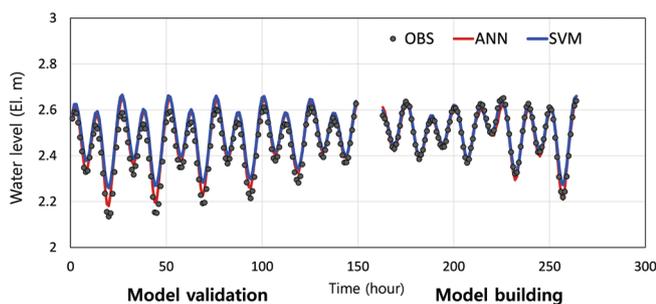


Fig. 6 - Results of groundwater level prediction in Test 2.

Fig. 6 - Risultati delle predizioni dei livelli piezometrici nel Test 2.

The prediction results of the Test 2 show that no abnormal errors occur for the validation data set as well as the training. Generally, it is reported that machine learning method is useful tool to solving nonlinear problems but especially for the interpolation. The result of the numerical tests indicates that the ANN- and SVM-based time series models can be effective for predicting groundwater level change, however, their extrapolation ability is limited for the given data of this study. For the development of time series models based on machine learning methods such as ANN and SVM, it is important and necessary to obtain various types of data and include it to model building process.

Summary and Conclusions

In the present study time series models for predicting groundwater level fluctuations were developed using ANN and SVM and applied them to groundwater level data at HD0 groundwater observatory in northeastern part of the

Jeju Island, South Korea. Numerical tests were conducted to estimate the model sensitivity to training data set as well as the prediction accuracy. The result of numerical tests shows that for input values within the range of training data set the prediction accuracy of the ANN and SVM models are similarly high, however, for input values beyond the range of training data set the ANN model estimated abnormal oscillatory values and the SVM constant values. This result indicates that for the development of an accurate and stable time series models based on machine learning methods such as the ANN and the SVM the allocation of the training data set with wide range and various types of observations. For the sustainable use of groundwater resource in coastal aquifers it is necessary to monitor and predict the groundwater level fluctuations. It is expected that the developed time series models and the results of numerical tests will be practically used for the management of groundwater resources of South Korea.

Acknowledgment: This research was supported by the Basic Research Project (19-3411) of the Korea Institute of Geoscience and Mineral Resources and the Korea Ministry of Environment (MOE) as "Demand Responsive Water Supply Service Program (RE20191097)".

REFERENCES

- Dogan E, Sengorur B, Koklu R (2009) Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J Environ Manage*; 90(2): 1229-35.
- Gill MK, Asefa T, Kaheil Y, McKee M (2007) Effect of missing data on performance of learning algorithms for hydrologic predictions: implications to an imputation technique. *Water Resour Res*; 43: W07416. doi:10.1029/2006WR005298.
- Hagan MT, Demuth HB, Beale M (1996) *Neural Network Design*. PWS Publishing Company, Boston, Massachusetts;
- Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ Modell Software*; 15: pp 101–24.
- Masetti M, Pedretti D, Sorichetta A, Stevenazzi S, Bacci F (2016) Impact of a Storm-Water Infiltration Basin on the Recharge Dynamics in a Highly Permeable Aquifer. *Water Resour Manag* 30(1): 149–165. DOI: 10.1007/s11269-015-1151-3.
- Nayak PC, Satyaji Rao YR, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resour Manage*; 20: 77–90.
- Platt JC (1999) Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges CJC, Smolar AJ, eds. *Advances in Kernel Methods-Support Vector Learning*, Cambridge, MA; MIT Press; pp 41-65.
- Rumelhart DE, McClelland JL (1986) The PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.
- Vapnik VN (1995) *The Nature of Statistical Learning Theory*. New York, NY: Springer-Verlag.
- Won JH, Lee JY, Kim JW (2006) Groundwater occurrence on Jeju Island, Korea. *Hydrogeol J*; 14:532-547.
- Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydro*; 396(1–2): 128–38.
- Yoon H, Hyun Y, Ha K, Lee KK, Kim GB (2016) A method to improve the stability and accuracy of ANN and SVM based time series models for long-term groundwater level predictions. *Comput Geosci* 2016; 90: 144-55.