Prediction of groundwater level variations in coastal aquifers with tide and rainfall effects using heuristic data driven models

Article *in* ISH Journal of Hydraulic Engineering · February 2020

DOI: 10.1080/09715010.2020.1729876

CITATIONS
7

READS
296

7 authors, including:

Jalal Shiri
University of Tabriz
146 PUBLICATIONS **4,741** CITATIONS
SEE PROFILE

Ozgur Kisi
University of Applied Sciences, Lübeck, Germany
543 PUBLICATIONS **21,992** CITATIONS
SEE PROFILE

Heesung Yoon
Korea Institute of Geoscience and Mineral Resources
34 PUBLICATIONS **879** CITATIONS
SEE PROFILE

Mohammad Hossein Kazemi
University of Tabriz
6 PUBLICATIONS **36** CITATIONS
SEE PROFILE

ISH Journal of Hydraulic Engineering

ISSN: 0971-5010 (Print) 2164-3040 (Online) Journal homepage: https://www.tandfonline.com/loi/tish20

Prediction of groundwater level variations in coastal aquifers with tide and rainfall effects using heuristic data driven models

Jalal Shiri, Ozgur Kisi, Heesung Yoon, Mohammad Hossein Kazemi, Naser Shiri, Mohammad Poorrajabali & Sepideh Karimi

To cite this article: Jalal Shiri, Ozgur Kisi, Heesung Yoon, Mohammad Hossein Kazemi, Naser Shiri, Mohammad Poorrajabali & Sepideh Karimi (2020): Prediction of groundwater level variations in coastal aquifers with tide and rainfall effects using heuristic data driven models, ISH Journal of Hydraulic Engineering, DOI: 10.1080/09715010.2020.1729876

To link to this article: https://doi.org/10.1080/09715010.2020.1729876

Published online: 27 Feb 2020.

ISH JOURNAL OF HYDRAULIC ENGINEERING
https://doi.org/10.1080/09715010.2020.1729876

Prediction of groundwater level variations in coastal aquifers with tide and rainfall effects using heuristic data driven models

Jalal Shiri[a], Ozgur Kisi[b], Heesung Yoon[c], Mohammad Hossein Kazemi[a], Naser Shiri[d], Mohammad Poorrajabali[a] and Sepideh Karimi[a]

[a]Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran; [b]Department of Civil Engineering, Ilia State University, Tbilisi, Georgia; [c]Groundwater Laboratory, Korea Institute of Geoscience and Mineral Resources (KIGAM), Daejeon, Korea; [d]Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

ABSTRACT

Correct simulation of groundwater level fluctuations is very important for optimum water resources management as it might affect the (subsurface) irrigation scheduling, groundwater exploitation studies, surface/groundwater interactions and identifying the aquifer characteristics. Such variations in coastal aquifers are more complicated due to the simultaneous effects of tide variations, especially in small time steps (e.g. hourly scales). The present study evaluated the capabilities of different heuristic data driven models for simulating groundwater table depth (GWD) fluctuations (with different lag times) in a coastal aquifer using the GWD, rainfall, and tide records. A k-fold testing cross-validation data assignment method was adopted here for defining the training and testing blocks. The obtained results showed that introducing rainfall and tide records along with the GWD data as models' inputs improved the models' performance accuracy, although accurate models could be obtained using only tide and rainfall data as external input parameters of models. Therefore, considering the overall performance of the models (which produced lower error values for all considered input combinations), all the applied models could simulate the GWD variations using external tide and rainfall data (without using GWD records).

ARTICLE HISTORY
Received 21 September 2019
Accepted 11 February 2020

KEYWORDS
Groundwater level; precipitation; soft computing; tide

1. Introduction

Groundwater, keeping out the polar ice caps and glaciers (Raghunath 2003), is a valuable and widely distributed water resource of the globe. Irrigated agriculture is the greatest user of ground/surface water resources accounting for about 80% of all consumptions. Moreover, groundwater is an important source of drinking, domestic, and industrial water requirements of the globe. Groundwater levels may experience different variations raised by differences between the recharge and discharge of groundwater, gaining/loosing waterway variations, tidal effects, urbanization, earthquake, land subsidence, and meteorological variables as well as global climatic changes (Todd and Mays 2005). Accurate prediction of such fluctuations is of crucial importance for an optimal management of water resources systems (Shiri et al. 2013; Barati 2018a, b; Zhang et al. 2018).

Substantial researches have been carried out so far for predicting groundwater level fluctuations, from which, the physics-based numerical approaches have been employed for characterizing the groundwater flow in aquifers. Despite their physical basis, a major downside of those models would be their dependency on substantial accurate data, which might not be available in some areas (Coppola et al. 2005). On the other hand, empirical models, utilizing Box and Jenkins (1976) and Hipel and Mc Leod (1994) models to simulate time series of groundwater table fluctuations, cannot produce accurate outcomes when the dynamical behavior of the hydrological system has considerable time variations (Bierkens 1998).

Existing literature proves that the soft computing (SC) methods have been successfully utilized for several water resources issues (Nassery et al. 2017; Alizadeh et al. 2017; Sadeghifar and Barati 2018). Alternative to empirical methods mentioned above, soft computing models have been applied for predicting groundwater level fluctuations. Among others, Rizzo and Dougherty (1994) applied neural networks (ANN) for the specification of the aquifer properties. Nayak et al. (2006) applied ANN for predicting groundwater level fluctuations in unconfined aquifers. Szidarovszky et al. (2007) proposed a combined ANN-numerical model to improve groundwater numerical model simulations. Also, Khalil et al. (2006), Gill et al. (2007) and Yoon et al. (2011) applied support vector machine (SVM) technique for predicting groundwater level variations. Shiri and Kisi (2011) compared gene expression programming (GEP) with neuro-fuzzy (NF) technique in predicting daily groundwater level fluctuations and introduced GEP as a promising tool in modeling groundwater table depth (GWD) fluctuations. Kisi and Shiri (2012) introduced a new hybrid wavelet-NF model for predicting short-term GWD fluctuations.

With minor exceptions, the reviewed literatures have used simple data assignment, which examines the performance accuracy of the applied model based on a single test set (for instance 30% of all patterns). This is a critical simplification, as the performance accuracy of the applied model for the rest of the available patterns cannot be assessed. Nonetheless, most of the studies have predicted groundwater levels at least for daily or longer prediction intervals, while hourly (and even shorter than hourly) prediction of GWD

CONTACT Jalal Shiri j_shiri2005@yahoo.com

variations are required for earthquake studies where the interrelationships between the GWD variations and dynamic behavior of the faults are studied. The present study aimed at assessing the boosted regression tree (BT), random forest (RF), multivariate adaptive regression spline (MARS), artificial neural networks (ANN), support vector machine (SVM) and gene expression programming (GEP) models in predicting hourly GWD values in a coastal aquifer, using the simultaneous records of the precipitation and tide heights, through a complete data scanning procedure (k-fold testing). In spite of the importance of hourly GWD fluctuations in earthquake studies, the selected time scale (hourly scale) is important for the National Groundwater Monitoring Network (NGMN) of Korea, where the abnormal measurement of the real-time monitoring system should be quickly found out.

2. Methodology

2.1. Boosted Regression Tree (BT)

BT includes an algorithm for involving tree-based approaches with boosting. It can be accepted as an improved regression-form related machine learning strategy (Freidman et al. 2000). Boosting improves the model accuracy by fitting new trees to the residual errors of the existing tree. During iterations, the existing tree does not change and the optimal (final) model is represented by the linear trees' combinations (Elith et al. 2008). BT uses multiple trees and copes with poor prediction accuracy as in single tree models. In this method, the numbers of trees are automatically optimized by an internal cross-validation procedure. The control parameters such as learning rate (that identifies the contribution of each tree to the final model), tree complexity (the node number of a single tree), and bag fraction, were obtained by trial-error procedure (França and Cabral 2015). Number of additive trees was selected as 150 based on trial and error. Different numbers of seeds for random number generator were evaluated and the optimum outputs were obtained for seed number = 1. More details about the BT method can be obtained from Freidman et al. (2000).

2.2. Random Forests (RF)

RF is a bootstrap method based on classification and regression trees (CART). This method provides well predictions without overfitting in case of high dimensional input matrices (Breiman 2001). In RF, each tree is formed at first by randomly choosing some groups of input variables (features) to split on for each node, and, then, by computing the best split depended on the features in the calibration data. The obtained tree is grown utilizing the CART method (Breiman et al. 1984) to maximum size, without pruning. This subspace randomization procedure is blended by bagging (Breiman 1996; Biau 2012) to resample, by replacing, the calibration data each time a new individual tree is grown. Different numbers of trees were assessed to select the optimum random forest method and it was seen that the variations of errors were negligible for tree numbers beyond 150. Moreover, eight cycles gave the best outputs in calculating the mean error. The percentage decrease in training error was found to be as 5% by trial and error. Minimum child node size to stop and the maximum number of levels were chosen as 5, and 10, respectively, through a trial and error process. More details on RF theory can be gotten from Breiman (2001).

2.3. Multivariate Adaptive Regression Spline (MARS)

Multivariate adaptive regression spline performs a nonlinear regression by using piecewise linear segments (splines) between dependent and independent variables. MARS considers no specific assumption related to the underlying functional relationship between the input and output variables. This method involves forward and backward processes. In the forward process, input parameters' set is chosen and functions are added. This process, however, generally produces complex model resulting in overfitting. Such a model has poor prediction accuracy. To prevent this deficiency, the backward process is applied and unnecessary variables are eliminated among the previously selected set. Following functions are used in MARS which they project independent X variable to dependent Y variable, using a knot that identifies an inflection point along the inputs' inputs (Sharda et al. 2006; Adamowski et al. 2012):

$$Y = \max(0, X - c) \tag{1}$$

$$Y = \max(0, c - X) \tag{2}$$

where c is some determined threshold value. More details for understanding the MARS can be getting from Sharda et al. (2006). The optimum MARS architecture was obtained through a trial and error procedure. So, the values of the maximum number of basis functions, degree of interactions, penalty, and threshold were, respectively, chosen as 20,3,2,0.0002.

2.4. Artificial Neural Networks (ANNs)

ANNs involve consecutive layers comprising many artificial neurons that are linked from one layer to another. The network transforms the inputs to target output(s). The output of the input layer's neurons forms the hidden layer's inputs and the output of the hidden layer's neurons forms the output layer's inputs. In the neurons of hidden and output layers, some mathematical functions (transfer function) are used. The input layer's neurons are connected with the hidden layer's neurons while the output layer's neurons are only connected to the hidden layer's neurons. Thus, some non-linear relationships can be obtained between input and output variables through layers and transfer functions. ANNs are good candidate in predicting future values of possibly noisy multivariate time series based on past values (Adamowski et al. 2012; Shoaib et al. 2016). More detailed information about ANNs can be seen in, e.g. Haykin (1999).

Here, a three-layer feed-forward network was applied with different transfer functions in the hidden and output layers. One hundred networks were evaluated per training stage and the optimum transfer functions were selected. The minimum and maximum values of weight decay in the hidden layer were found as 0.0005 and 0.001. Moreover, it was found that the ANN model with Tanh and Exponential activation functions in hidden and output neurons comprising 25 neurons in the hidden layer produced the best outcomes.

2.5. Gene Expression Programming (GEP)

Gene expressing programming, proposed by Ferreira (2001), is one of the commonly used machine learning methods.

GEP has linear chromosomes of fixed length which they comprise multiple genes. Each chromosome encodes a smaller subprogram. Linear chromosomes provide the unconstrained operation of basic genetic operators, for example, mutation, transposition, and recombination. The most important advantages of GEP are (Ferreira 2001): (1) the chromosomes are basic elements: linear, compact, generally small, simple to manipulate genetically (replicate, mutate, recombine, etc.); (2) the expression trees are the expression of their respective chromosomes; they are elements whereupon selection acts and as indicated by fitness, they are chosen to reproduce with modification. GeneXpro program was utilized in this study for modeling groundwater depth fluctuations.

GEP implementation procedure was conducted following the suggestions of Shiri and Kisi (2011) for a similar subject. So, the root-relative square error (RRSE) was chosen as a fitness function. The functions set of GEP operation included the four basic operators (+,-,*,/) as well as ($\sqrt{}$, $\sqrt[3]{}$, ln(x), e^x, x^2, x^3). Length of head, h = 8, and three genes per chromosomes have been employed, which are commonly used values in the literature (e.g. Ferreira 2006). Finally, the sub trees were linked by addition linking function according to literature (e.g. Shiri and Kisi 2011).

2.6. Support Vector Machine (SVM)

SVM is a machine learning method based on statistical learning theory (Vapnik 1995). It is able to provide global optimization needed for improved generalization. The main advantages of utilizing the SVM model are getting high generalization accuracy, efficiency in computation, robustness for high dimensions, simplicity of application, its suitability for large dimensional objects and having a wide variety of applications (Nazari and Sanjayan 2015). This method was first proposed by Vapnik and more detailed information about its theory can be seen in, e.g. Vapnik (1995, 1998). Using a trial and error procedure, SVM capacity and epsilon constants were chosen as 10 and 0.15. Among the kernel functions, the radial basis function (Gamma = 0.20) gave the most accurate simulations of GWD. Further, the maximum iteration number was set as 1000 (iteratively) and the models were stopped at the error value of 0.005.

2.7. Study area and data used

Hourly groundwater table depth (GWD), precipitation (P), and tide levels (T) records from two observational wells in a coastal aquifer (located at the beach of the coastal town of Mukho in Donghae City) in Korea were used in the present study. There were totally 11 observation wells in the studied aquifer (OFs, BHs, and OBs) (Figure 1) but some of them have been washed by the waves. In the present study, the observational data from two wells, namely OB2 and BH5 were used. For OB2 well, hourly values from 23 May 2004 to 28 December 2004 (totally, 5274 patterns) and for BH5 well, hourly values from 1 April 2005 to 14 November 2006 (totally 14,222 patterns) were utilized for development and validation of the applied models. Table 1 sums up the statistics of the used data during the study period. X_{max}, X_{min}, X_{mean}, C_V, S_X, and C_{SX} in the table show the maximum, minimum, mean, coefficient of variation, standard deviation, and skewness coefficient of the data, respectively.

2.8. Data splitting and input parameters

As a predecessor of the present research, Yoon et al. (2011) applied ANN and SVM for a dynamic estimation of GWD in the same study area of the present paper. They utilized P, T, and GWD observations for the development and validation of ANN and SVM techniques. So, they generated five different input combinations of these parameters to feed their applied models. To be consistent, and based on the correlation analysis, the same input combinations were used in the present study as can be seen in Table 2. The cross-correlation analysis between the parameters showed the lag times of the both precipitation (P) and tide (T) records are 4 time steps (1 time step stands for 6 hr). So, the data with 4 lags (= 24 hr) were utilized in building the input matrix. The prediction horizons (lead times) for GWD records were considered as 1, 2, 4, 6, and 8 prediction intervals (each interval stands for 6 hr), similar to Yoon et al. (2011). Although hourly GWD predictions/analysis might not be of interest in agricultural issues, the selected prediction intervals (lead times) provided simulations of GWD variations for the following 1 (for interval 6) and 2 (for interval 8) days that would be of interest from the agricultural viewpoint.

An important step with building the soft computing models would be defining the training and testing portions of the all available patterns, with which, the models should be developed and validated. A common process would be the traditional two/three blocks mode, where the all available input-target patterns are divided into two/three portions, then the models are applied, tested and validated. However, a major drawback of such data partitioning methods would be their high dependency on the way the data have been divided, especially when there is an obvious time-dependency/time series trend among the studied variables (Shiri et al. 2014). So, the most robust k-fold testing data scanning procedure was adopted here for training and testing the applied techniques. In this way, the available patterns of each well were divided into 'k' parts and the models were trained using a part of patterns and tested using the remaining portion. The procedure was repeated till all the available patterns of each well were involved in training and testing stages. The smaller 'k' values would provide more accurate analysis of the error variations throughout the patterns, but the computational costs would not be assumable. Here, the minimum test set size was set as 1 day in both the wells. Given that the available records in OB2 wells were belonged to 220 days, a total 6600 training-testing process [220 days*6 heuristic models*5input combinations (see Table 2)] were established in this well. Similarly, in the BH5 well, a total of 17,760 training-testing procedures [592 days*6 heuristic models*5input combinations (see Table 2)] were carried out.

Two statistical indices, namely the root-mean-square error (*RMSE*) and the scatter index (*SI*) were used for assessing the performance accuracy of the obtained models as:

$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(GWD_{iM} - GWD_{io})^2}}{\overline{GWD}} \quad (1)$$

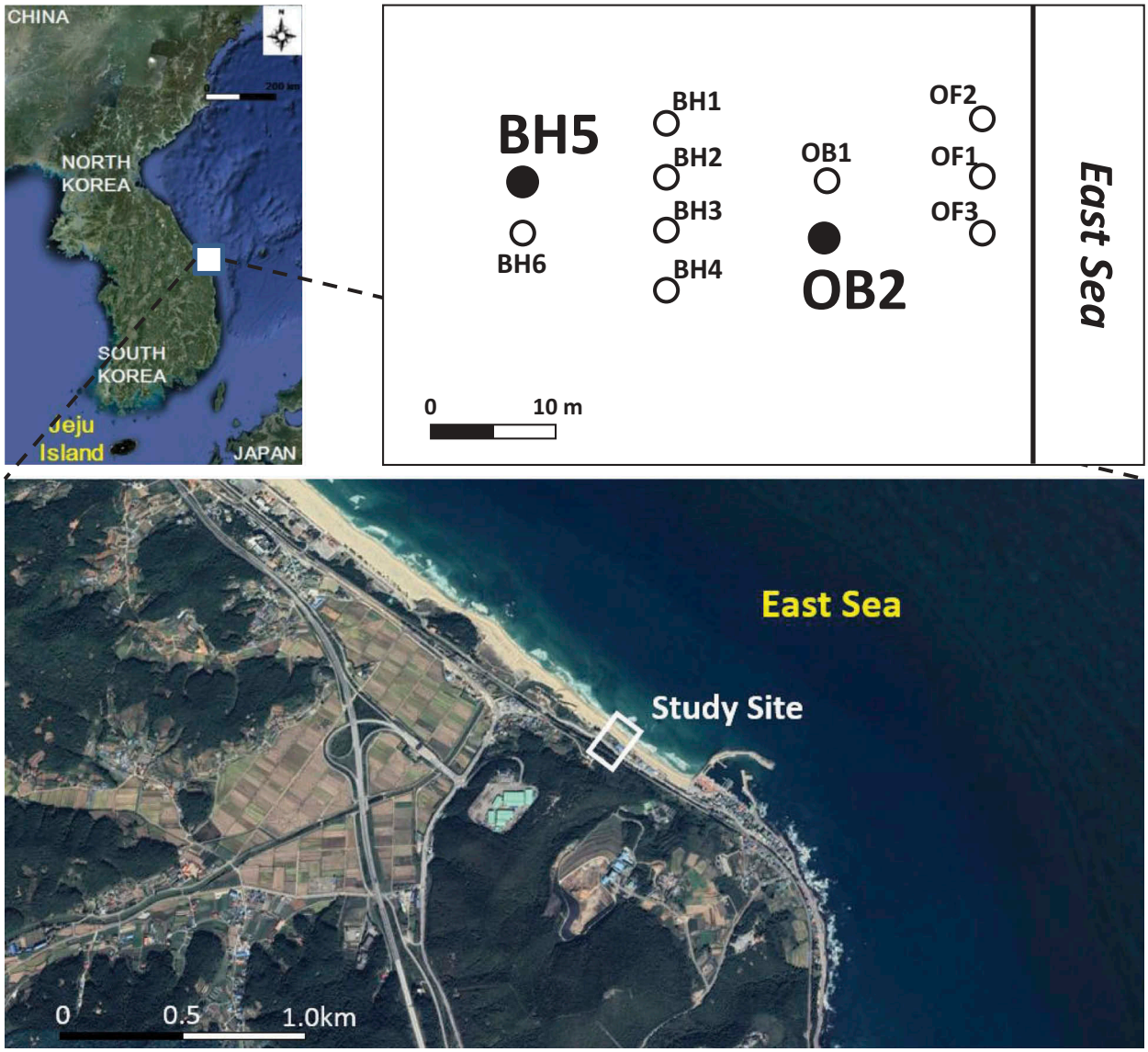

Figure 1. Localization of the studied points.

Table 1. Statistical parameters of the used data set during the study period.

		X_{max}	X_{min}	X_{mean}	S_X	C_V	C_{SX}
OB2 well	GWD(m)	3.91	2.03	2.46	0.24	0.09	1.84
	P (mm)	53.5	0	0.23	1.56	6.71	15.19
	T (m)	0.56	−0.32	0.08	0.12	1.51	−0.008
BH5 well	GWD(m)	4.85	2.66	3.03	0.23	0.08	2.19
	P (mm)	51	0	0.21	1.40	6.53	14.07
	T (m)	3.9	−4.9	0.44	1.18	2.63	−0.19

Table 2. Input combinations used to feed the applied models.

	Used variable per combination		
Input combinations	P	T	GWD
Input1	■	■	
Input2			■
Input3	■		■
Input4		■	■
Input5	■	■	■

$$SI = \frac{RMSE}{\overline{GWD}} \tag{2}$$

where GWD_{iM} and GWD_{io} denote the predicted and observed GWD values at the i^{th} time step, respectively, $\overline{GWD}$ is the mean of observed values, and N stands for the number of time steps. *RMSE* illustrates the average error values by giving more weight to large errors and can take on values from 0 (perfect fit) to ∞ (the worst fit). The dimensionless *RMSE* (called *SI* here) can present a suitable insight for comparing the performances of various models, as it eliminates the effect of the magnitude range of the target parameter. So, using *SI* it is possible to compare the models that have been used for simulating a unique target parameter in different situations (with different ranges of magnitudes). Instead of taking an average from the performance indicators from the all available days, a global estimation vector of the complete period was made for each well, and utilized for calculating the performance indicators.

3. Results and discussions

3.1. General statements

Table 3 sums up the global statistical indicators of the BT, RF, MARS, ANN, SVM, and GEP models for each studied well. These values have been calculated by taking an average from the indicators of the applied lead times for each model and input combination. As can be seen from the table, the models showed similar performance accuracy for BH5 and

Table 3. Global error statistics of the applied models.

	Input1		Input2		Input3		Input4		Input5	
	RMSE(m)	*SI*	*RMSE*(m)	*SI*	*RMSE*(m)	*SI*	*RMSE*(m)	*SI*	*RMSE*(m)	*SI*
OB2 well										
BT	0.194	0.079	0.149	0.060	0.150	0.061	0.146	0.059	0.152	0.062
RF	0.198	0.080	0.152	0.062	0.147	0.060	0.146	0.059	0.145	0.059
MARS	0.205	0.083	0.158	0.064	0.155	0.063	0.153	0.062	0.151	0.061
ANN	0.261	0.106	0.244	0.099	0.156	0.063	0.151	0.062	0.156	0.063
SVM	0.212	0.086	0.174	0.070	0.160	0.065	0.163	0.066	0.163	0.066
GEP	0.210	0.075	0.165	0.057	0.159	0.055	0.158	0.054	0.156	0.053
	RMSE	*SI*	*RMSE*	*SI*	*RMSE*	*SI*	*RMSE*	*SI*	*RMSE*	*SI*
BH5 well										
BT	0.211	0.070	0.148	0.049	0.144	0.048	0.144	0.047	0.128	0.047
RF	0.213	0.070	0.147	0.048	0.142	0.047	0.142	0.047	0.141	0.046
MARS	0.213	0.070	0.144	0.048	0.140	0.046	0.142	0.047	0.140	0.046
ANN	0.214	0.070	0.159	0.053	0.145	0.048	0.146	0.048	0.148	0.049
SVM	0.232	0.076	0.156	0.051	0.167	0.055	0.155	0.051	0.149	0.049
GEP	0.225	0.066	0.159	0.044	0.154	0.043	0.154	0.043	0.154	0.042

OB2 wells for all adopted input combinations, although some slight differences between *SI* values might be detected for all cases.

From Table 1, the statistics of the applied patterns were similar for both the wells, even the OB2 time series presented lower scatters (in terms of the skewness, standard deviation, etc.) than those of BH5. Therefore, one might seek the reason for such differences in the total number of patterns used to establish the models. As the robust k-fold testing data management scenario was adopted here, all the available patterns have been involved in training and testing stages. However, a model might be sensitive to the total number of patterns that have been incorporated in the training phase so that the model can learn using more patterns. Again, Table 3 shows that the GEP produced the lowest *SI* values in both the wells that were because it has lower sensitivity to the number of inputs/patterns as it belongs to the information content in the used data. This includes data gathering in different conditions; much like one would do employing a design of experiment approach. Therefore, GEP could be employed using any number of data, provided the information content in the data set is sufficient to evolve a generalized formulation (Deschaine 2014; Shiri 2017). In the contrast, it seems the rest of the applied models has been sensitive to the total number of the patterns used to feed them, as they have given better results for BH5 (with higher amount of data records) than the OB2 (with lower available patterns). Nevertheless, this might be evaluated through the information saturation content to find the suitable amount of patterns that should be utilized for training these models. Accordingly, a further modeling process was carried out in BH5 using the same number of total patterns of OB2 (220 days) for all the applied models and input combinations. Analyses of the results (not presented here) confirmed the superiority of the models in BH5 over the OB2, again. Hence, such differences cannot be explained using the available information about the water table fluctuations in each well and it may need further analysis based on aquifer dynamic characteristics that were not available to the authors.

Further, given that the depths of groundwater table for OB2 and BH5 are 2.46 m and 3.03 m, respectively, both wells have been influenced by the meteorological factors (e.g. air temperature and rainfall) and tidal effects to the same extent. Nevertheless, as the GWD in BH5 is slightly greater than OB2, the required time for passing the infiltrated rainfall through the porous media to join the groundwater would be somewhat longer there. Analysis of the simultaneous variations of T, P, and GWD (Figure 2) showed that the trend of variations of all three parameters is close in both the wells, albeit there were some discrepancies between the occurrence of extreme points in time series of these parameters. Moreover, the linear correlation values between GWD, T, and P were higher in OB2 than the BH5: 0.169, 0.461 and 0.159 between P-T, T-GWD, and P-T, respectively, for OB2 and 0.082, 0.264 and 0.154 between P-T, T-GWD, and P-T, respectively, for BH5. However, the models showed better performances in BH5 than OB2 that shows the better capability of the models in hindering nonlinear processes like GWD prediction. Nonetheless, the performance accuracy of the models relied on P and T (Input1) was lower than the rest of the inputs. This might be foreshadowed because these models have been constructed using only P and T data without information on GWD variations. For the other applied input combinations, Input 5 comprising P, T, and GWD values as inputs gave the best results, as could be expected. However, differences between mean *SI* values of this input combination with Inputs 2–4 (and even input 1) were negligible, so, it could be stated that the models excluding GWD from input matrix could predict GWD variations in different lead times accurately.

Figure 3 illustrates the cumulative distribution functions (CDF) of observed and predicted GWDs in both the wells for GEP models. From the figures, it can be seen that the deviations from the CDF of the observed GWD are increased by increasing the prediction intervals (lead times) for both the wells. Comparing between two wells, these deviations were more obvious in OB2 than BH5 well.

Nevertheless, one of the advantages of GEP over the other models is in giving mathematical expressions of the studied phenomenon. Table 4 presents the mathematical expressions of the optimal GEP models for the both wells.

3.2. Local assessment of the models

Figures 4 and 5 present the *SI* values of the applied models for different input combinations and lead times (prediction intervals) in both the wells. Analyzing the figures, some closures could be highlighted as follows. First, it is seen that the *SI* values fluctuations between the models are low for both the wells and all input combinations ($\Delta SI<0.1$). Second, there were obvious variations of *SI* values for different input combinations for both the wells. The first input combination

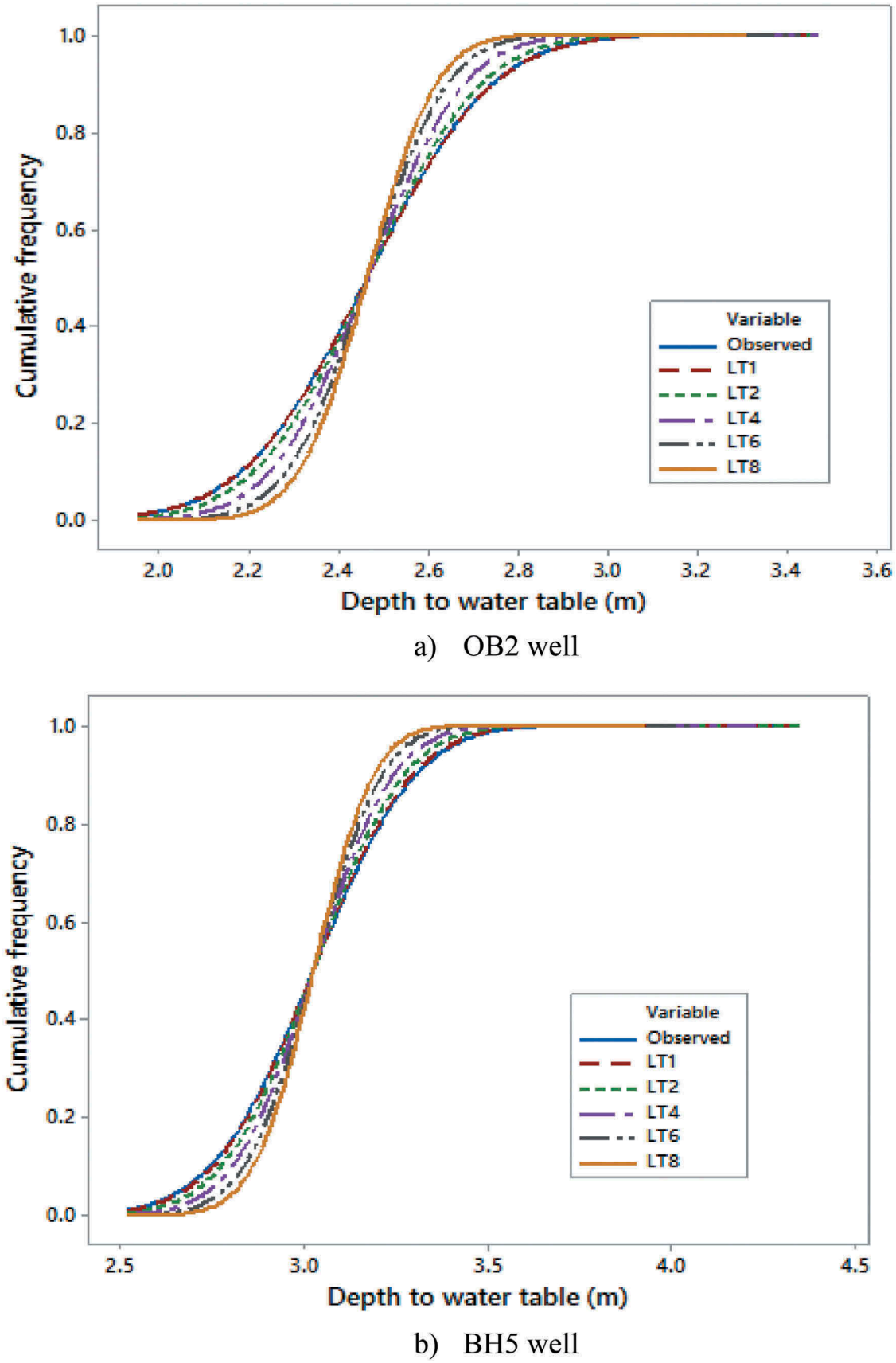

Figure 2. Cumulative distribution functions of observed and predicted GWDs with different lead times for the GEP model.

(comprising tide and rainfall as inputs) has produced the worst simulations while the rest of the combinations have given more accurate results. Next, as could be expected, the first lead time has given the best simulations for all cases, since this lead time has had the lowest prediction interval distance from the corresponding input variables. However, as can be seen from the figures, the general values of *SI* are lower than (or around) 0.1 which shows the good abilities of the applied models for simulating GWD fluctuations in the studied locations. Nevertheless, analysis of the *SI* range variations in both the wells showed that the variation range is higher for OB2 than the BH5 which shows the higher *SI*

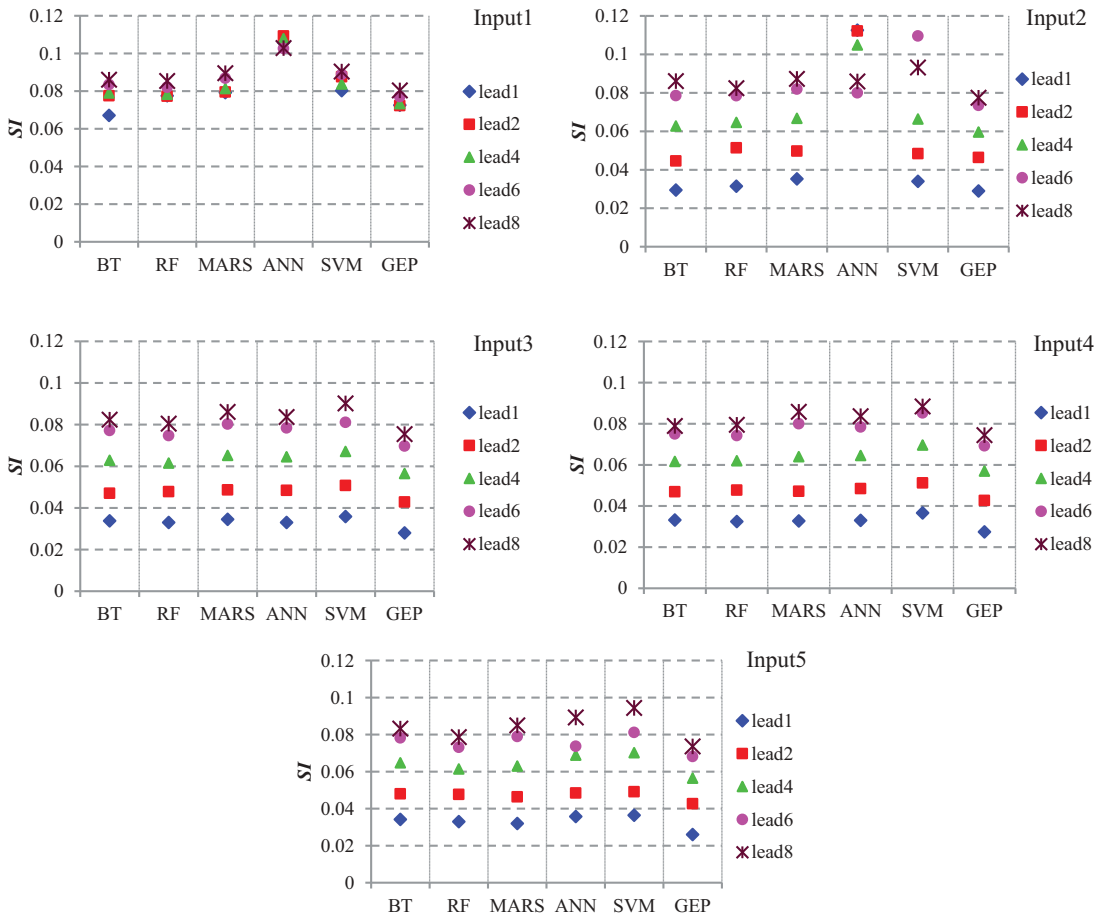

Figure 3. *SI* values of the applied models for different input combinations and lead times: OB2 well.

Table 4. GEP expressions of the optimal models.

model	Expressions
BH5	
Rainfall-based	$GW_{t+1} = \{Arctg[Arctg(GW_{t-1} - GWt)(2.428 - P_{t-3})]\}^3 + 0.321(GW_{t-1} - GW_{t-2}) + GW_t$
Tide-based	$GW_{t+1} = Arctg[Sin(Arctg(T_{t-1}))] + Arctg[Exp(Arctg(GW_t))] + 2GW_{t-2} + GW_t - 6.47$
OB2	
Rainfall-based	$GW_{t+1} = Arctg[(0.246 + T_{t-1}) + GW_{t-1} + GW_{t-2} + T_{t-2} - 1.1032] + GW_t + 0.172$
Tide-based	$GW_{t+1} = P_{t-1} Sin \frac{Cos(GW_t).P_{t-1}+P_{t-2}}{GW_{t-2}-8.1628} + GW_t - P_{t-2} + 1.364P_{t-2}$

differences between the adopted lead times in this well. Finally, GEP showed the best simulations among the applied models in both the locations, whereas ANN and SVM gave the worst predictions (specifically in OB2).

3.3. Temporal assessment of the models

Figure 6 presents the temporal variations of the *SI* values for the GEP models during a sample monthly period (30 test days) for both the wells. *SI* values have shown considerable variations for different days of the considered test periods in both the wells, although the maximum values of this index were around 0.1 for all cases which demonstrated the high capability of the applied models in simulating GWL variations. With some exceptions, the minimum variations of *SI* values were belonged to the first lead time while the lead6 and lead8 prediction intervals presented the highest variations for this index. Such variations in the statistical indicator magnitudes might highlight the need for a complete data scanning procedure (e.g. k-fold testing) when the heuristic models were applied for modeling different phenomena, as advised by Shiri et al. (2015). These variations might be attributed to the temporal variations of the factors influencing the GWD fluctuations during a day (24 h) or a month (30 days). A sample variation of tide and GWL parameters during a day for OB2 well has been illustrated in Figure 7, which shows the direct partial influence of tide variations on GWD. While GWD magnitudes were monotonously fluctuated for some hours during a day, obvious variations can be detected for a part of day, especially when the tide magnitudes started to reduce. After peak value, the tide magnitude is reducing and after about 3 h the GWL is increasing (see the tide values after 13:00 h and GWL after 16:00 h). This might show the effect of tide rise (due to the loading factor at the boundary) on rising GWD but with a small time delay.

Figure 8 shows the observed and simulated hourly GWD values of GEP-based models (fed with input5) during sample test periods (from 01-May2005 to 31-May2005 for BH5; and from 1 June 2004 to 30 June 2004 for OB2). From the figure, it is seen that the model has simulated the GWD fluctuations accurately, although there are some discrepancies between the observed and simulated values for both the wells. It was

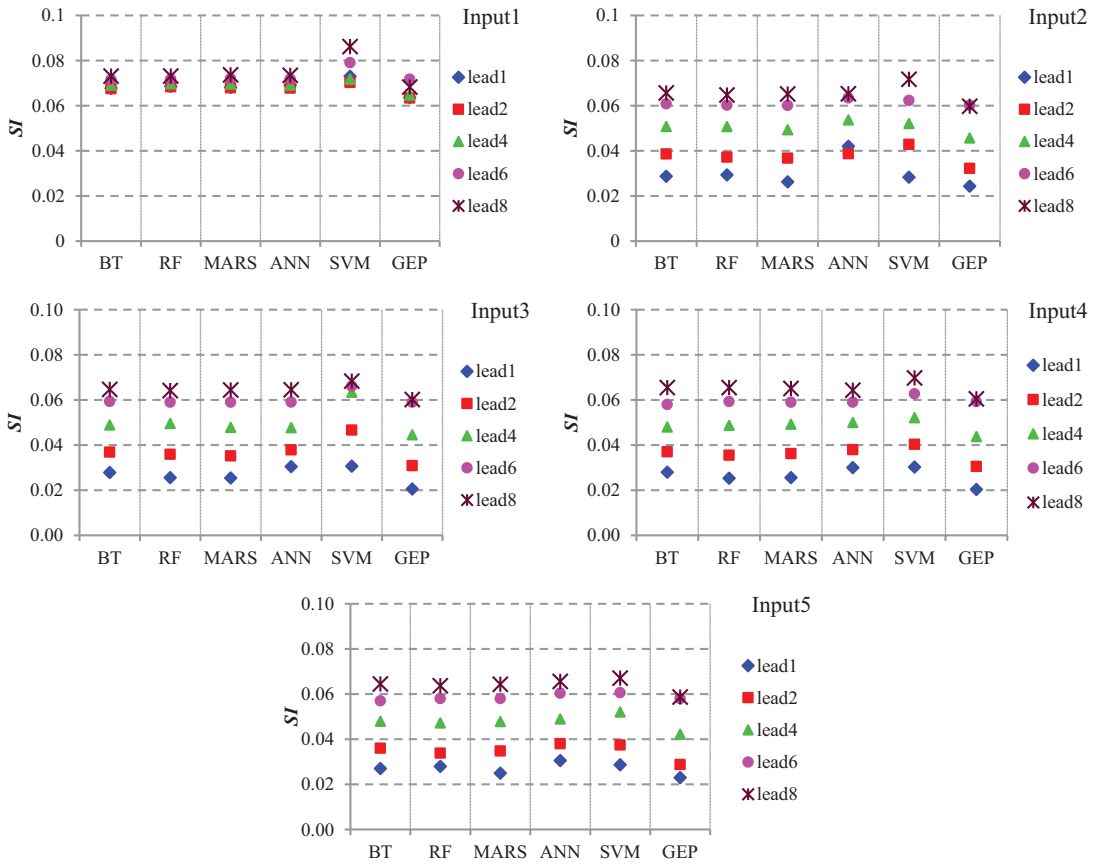

Figure 4. *SI* values of the applied models for different input combinations and lead times: BH5 well.

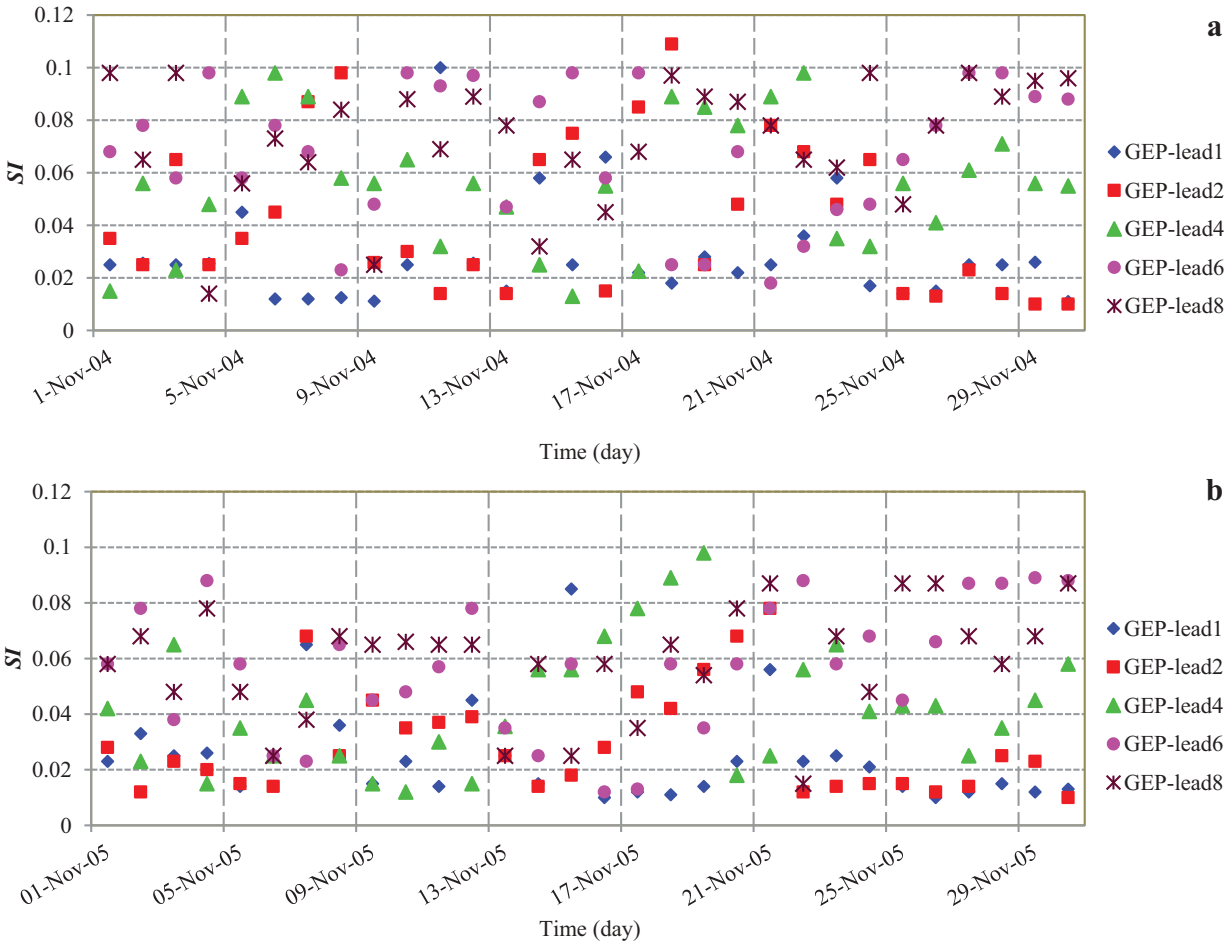

Figure 5. Temporal variations of the *SI* values during a sample test period (1 month) for: a) OB2, b) BH5.

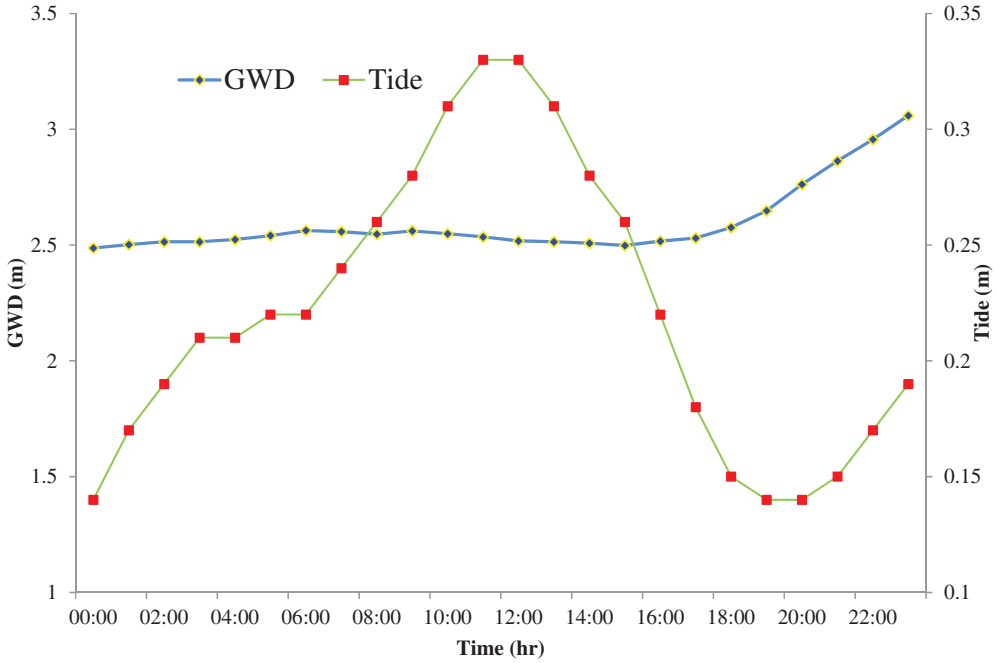

Figure 6. Inverse effect of tidal variations on GWD variations for OB2 (1 November 2004).

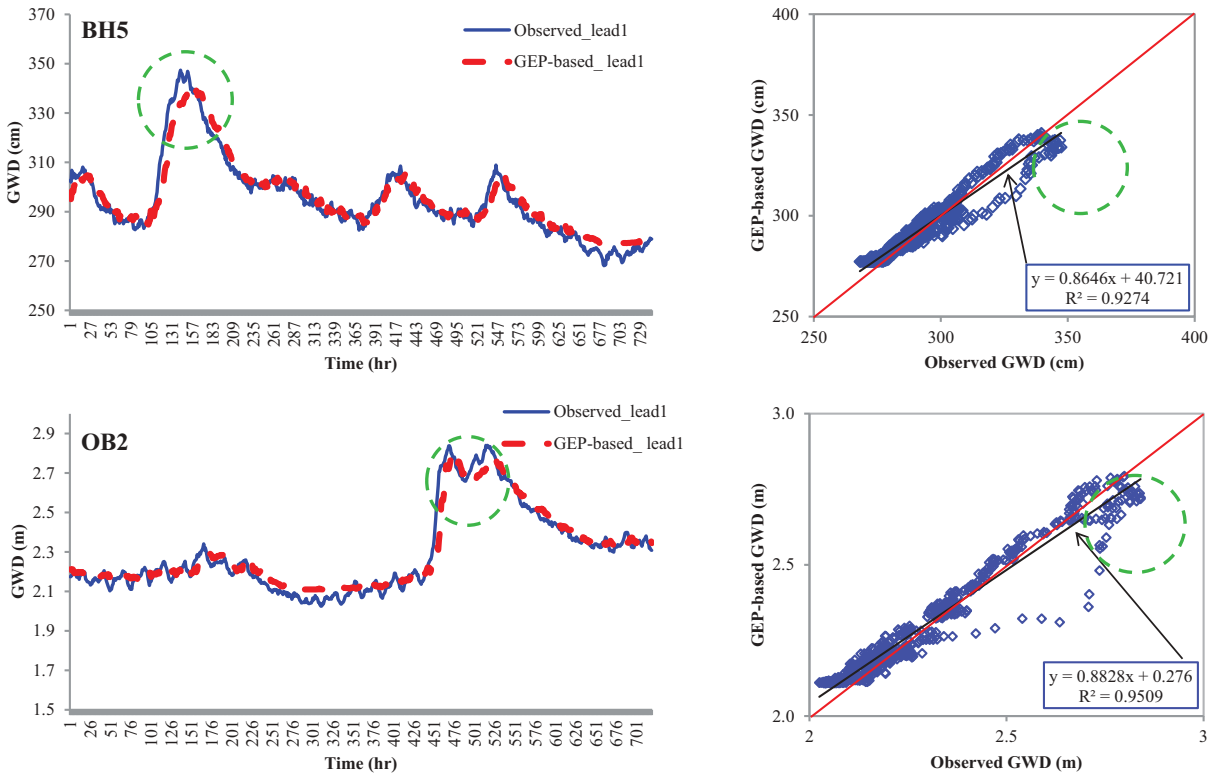

Figure 7. Observed vs. simulated GWD values for the input combination 5 (GEP5) during sample test periods in BH5 (May 2005) and OB2 (June 2004).

observed that the inconsistencies for all cases (all test periods) tend to underestimate the peak GWD values. The corresponded underestimations of the GWD time series for the presented cases have been circulated in the scatterplots in Figure 7. This might be due to the difficulties to produce the peak values which had higher distance with the mean of the observed GWD data (see Cv values of Table 1).

Overall, the comparison of several input combinations clearly indicated that importing previous GWD as input slightly increases models' accuracies and GWD of both wells could be successfully estimated with only hourly precipitation and tide data as inputs. Models showed better accuracy in BH5 well compared to OB2 even though there are high linear correlations between the inputs and output data of the OB2 well. This indicated the high nonlinearity or complexity of the investigated phenomenon. Among the applied models, the GEP model generally provided the best accuracy in estimating GWD in both wells closely followed by the BT, RF and MARS models while the ANN and SVM had the worst results. The optimal GEP model with five inputs tends to underestimate peak GWD

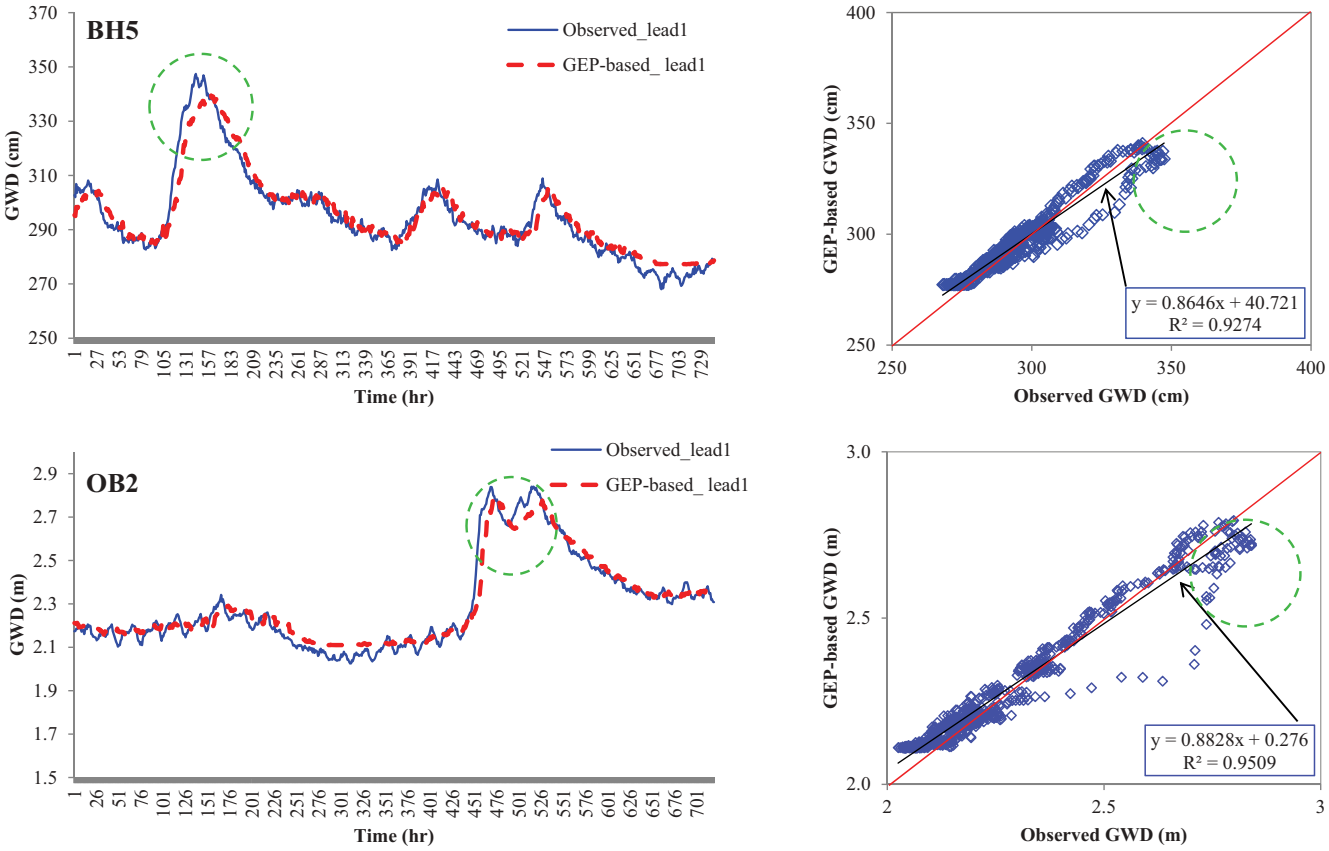

Figure 8. Observed vs. simulated GWD values for the input combination 5 (GEP5) during sample test periods in BH5 (May2005) and OB2 (June 2004).

values in both wells. The main reason caused this underestimation might be the low number of peak values in the whole data and the applied model cannot adequately learn the process.

4. Conclusions

The study compared the six different heuristic methods, BT, RF, MARS, ANN, SVM, and GEP, in simulating groundwater level depth (GWD) fluctuations of coastal aquifers. Hourly GWDs, precipitation, and tide levels from two observational wells (OB2 and BH5) located in the coastal town of Mukho in Donghae City, Korea were utilized for model development. In the applications, various training-test blocks were employed based on the cross-validation data assignment (CVDA) procedure. It was seen that the accuracies of the applied models were fluctuated with respect to different training-test data set and this indicated the necessity of the CVDA process. The models provided more accurate results for BH5 compared to OB2 well. All the models employed accurately simulated the GWD variations with respect to *RMSE* and *SI* by utilizing tide and rainfall data as inputs. GEP model, in general, gave the best estimates in the both wells and its accuracy was closely followed by the BT, RF and MARS models. The ANN and SVM generally provided the worst GWD estimates. It can be concluded that the GWD of both wells could be accurately estimated by the heuristic methods (e.g. GEP) using precipitation and tide level data. The present study utilized data from two coastal wells for assessing the impacts of tide and rainfall parameters on GWD variations. Further studies would be needed using more sites and other heuristic models to verify these conclusions in other regions with different aquifer–sea interactions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Korea Ministry of Environment as "[GAIA Geo-Advanced Innovative Action Project 2016000530004]".

References

Adamowski, J., Chan, H.F., Prasher, S.O., and Sharda, V.N. (2012). "Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data." *J. Hydroinf.*, 14(3), 731–744. doi:10.2166/hydro.2011.044.

Alizadeh, M.J., Shahheydari, H., Kavianpour, M.R., Shamloo, H., and Barati, R. (2017). "Prediction of longitudinal dispersion coefficient in natural rivers using a cluster-based Bayesian network." *Environ. Earth Sci.*, 76(2), 86. doi:10.1007/s12665-016-6379-6.

Barati, R. (2018a). "Discussion of "Modeling water table depth using adaptive Neuro-Fuzzy inference system by Umesh Kumar Das, Parthajit Roy and Dillip Kumar Ghose (2017)." *ISH J. Hydraulic Eng.* Online- In press, 1–4. doi:10.1080/09715010.2018.1488224.

Barati, R. (2018b). "Discussion of "Study of the spatial distribution of groundwater quality using soft computing and geostatistical models by Saman Maroufpoor, Ahmad Fakheri-Fard and Jalal Shiri (2017)." *ISH J. Hydraulic Eng.*, 122–125.

Biau, G. (2012). "Analysis of a random forests model." *J. Mach. Learn. Res.*, 13(2012), 1063–1095.

Bierkens, M.F.P. (1998). "Modeling water table fluctuations by means of a stochastic differential equation." *Water Resour. Res.*, 34(10), 2485–24499. doi:10.1029/98WR02298.

Box, G.E.P., and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*, Holden day, Boca Raton, FL.

Breiman, L. (1996). "Bagging predictors." *Mach. Learn.*, 24, 123–140. doi:10.1007/BF00058655.

Breiman, L. (2001). "Random forests." *Mach. Learn.*, 45, 5–32. doi:10.1023/A:1010933404324.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and regression trees*, Chapman & Hall, New York.

Coppola, E., Rana, A., Poulton, M., Szidarovszky, F., and Uhl, V. (2005). "A neural network model for predicting aquifer water level elevations." *Ground Water*, 43(2), 231–241. doi:10.1111/gwat.2005.43.issue-2.

Deschaine, L.M. (2014). *Decision support for complex planning challenges: Combining expert systems, engineering-oriented modeling, machine learning, information theory, and optimization technology*, Chalmers University of Technology, Sweden, 233.

Elith, J., Leathwick, J.R., and Hastie, T. (2008). "A working guide to boosted regression trees." *J. Anim. Ecol.*, 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Ferreira, C. (2001). "Gene expression programming: A new adaptive algorithm for solving problems." *Complex Syst.*, 13(2), 87–129.

Ferreira, C., 2006. Gene expression programming: Mathematical Modeling by an artificial intelligence. Springer, Berling, Heidelberg New York, 478pp.

França, S., and Cabral, H.N. (2015). "Predicting fish species richness in estuaries: Which modelling technique to use?" *Environ. Modell Software*, 66, 17–26. doi:10.1016/j.envsoft.2014.12.010.

Freidman, J., Hastie, T., and Tibshirani, R. (2000). "Additive logistic regression: A statistical view of boosting." *Annal. Stat.*, 28(2), 337–407. doi:10.1214/aos/1016218223.

Gill, M.K., Asefa, T., Kaheil, Y., and McKee, M. (2007). "Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique." *Water Resour. Res.*, 43(7), W07416. doi:10.1029/2006WR005298.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*, Macmillan College Publishing Co., New York.

Hipel, K.W., and Mc Leod, A.I. (1994). *Time series modeling of water resources and environmental systems* Dev, water sci Vol. 45, Elsevier Science, New York.

Khalil, A.F., McKee, M., Kemblowski, M., Asefa, T., and Bastidas, L. (2006). "Multiobjective analysis of chaotic dynamic systems with sparse learning machines." *Adv. Water Resour.*, 29, 72–88. doi:10.1016/j.advwatres.2005.05.011.

Kisi, O., and Shiri, J. (2012). "Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations." *Hydrol. Res.*, 43 (3), 286–300. doi:10.2166/nh.2012.104b.

Nassery, H.R., Adinehvand, R., Salavitabar, A., and Barati, R. (2017). "Water management using system dynamics modeling in semi-arid regions." *Civil Eng. J.*, 3(9), 766–778. doi:10.21859/cej-030913.

Nayak, P.C., Satyaji Rao, Y.R., and Sudheer, K.P. (2006). "Groundwater level forecasting in a shallow aquifer using artificial neural network approach." *Water Resour. Manage.*, 20, 77–90. doi:10.1007/s11269-006-4007-z.

Nazari, A., and Sanjayan, J.G. (2015). "Modelling of compressive strength of geopolymer paste, mortar and concrete by optimized support vector machine." *Ceram. Intl.*, 41(2015), 12164–12177. doi:10.1016/j.ceramint.2015.06.037.

Raghunath, H.M., 2003. Groundwater. New age International Publishers. 2nd edition. New Delhi, India. 563 pp.

Rizzo, D.M., and Dougherty, D.E. (1994). "Characterization of aquifer properties using artificial neural networks: Neural Kriging." *Water Resour. Res.*, 30(2), 483–497. doi:10.1029/93WR02477.

Sadeghifar, T., and Barati, R. (2018). "Estimation of alongshore sediment transport rate using adaptive Neuro-fuzzy inference system – A case study application for the southern shorelines of the Caspian Sea." *J. Soft Comput. Civil Eng.*, 2(4), 72–85.

Sharda, V., Prasher, S.O., Patel, R.M., Ojavasi, P.R., and Prakash, C. (2006). "Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques." *Agric. Water Manage.*, 83, 233–242. doi:10.1016/j.agwat.2006.01.003.

Shiri, J. (2017). "Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran." *Agric. Water Manage.*, 188, 101–114. doi:10.1016/j.agwat.2017.04.009.

Shiri, J., and Kisi, O. (2011). "Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations." *Comput. Geosci.*, 37(10), 1692–1701. doi:10.1016/j.cageo.2010.11.010.

Shiri, J., Kisi, O., Yoon, H., Lee, K.K., and Nazemi, A.H. (2013). "Predicting groundwater level fluctuations with meteorological effect implications: A comparative study among soft computing techniques." *Comput. Geosci.*, 56, 32–44. doi:10.1016/j.cageo.2013.01.007.

Shiri, J., Marti, P., and Singh, V.P. (2014). "Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning." *Hydrol. Process*, 28 (3), 1215–1225. doi:10.1002/hyp.v28.3.

Shiri, J., Sadraddini, A.A., Nazemi, A.H., Marti, P., Fakheri Fard, A., Kisi, O., and Landeras, G. (2015). "Independent testing for assessing the calibration of the Hargreaves–Samani equation: New heuristic alternatives for Iran." *Comput. Electron. Agric.*, 117, 70–80. doi:10.1016/j.compag.2015.07.010.

Shoaib, M., Shamseldin, A.Y., Melville, B.W., and Khan, M.M. (2016). "A comparison between wavelet based static and dynamic neural network approaches for runoff prediction." *J. Hydrol.*, 535, 211–225. doi:10.1016/j.jhydrol.2016.01.076.

Szidarovszky, F., Coppola, E., Long, J., Hall, A., and Poulton, M. (2007). "A hybrid artifiucial neural network-numerical model for groundwater problems." *Ground Water*, 45(5), 590–600. doi:10.1111/j.1745-6584.2007.00330.x.

Todd, D.K., and Mays, L.W. (2005). *Groundwater hydrology, third revision*, John Wiley and Sons Inc., New Jersy, USA, 636.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer, New York.

Vapnik, V. (1998). *Statistical learning theory*, Wiley, New York.

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. doi:10.1016/j.jhydrol.2010.11.002.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J. (2018). "Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas." *J. Hydrol.*, 561, 918–929. doi:10.1016/j.jhydrol.2018.04.065.