

# Forecasting Groundwater Levels using a Hybrid of Support Vector Regression and Particle Swarm Optimization

Saeed Mozaffari<sup>1</sup> · Saman Javadi<sup>1</sup> · Hamid Kardan Moghaddam<sup>2</sup> · Timothy O. Randhir<sup>3</sup>

Received: 9 September 2021 / Accepted: 15 March 2022 / Published online: 19 March 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

# Abstract

Forecasting the groundwater level is crucial to managing water resources supply sustainably. In this study, a simulation-optimization hybrid model was developed to forecast groundwater levels in aquifers. The model uses the PSO (Particle Swarm Optimization) algorithm to optimize SVR (Support Vector Regression) parameters to predict groundwater levels. The groundwater level of the Zanjan aquifer in Iran was forecasted and compared to the results of Bayesian and SVR models. In the first approach, the aquifers hydrograph was extracted using the Thiessen method, and then the time series of the hydrograph was used in training and testing the model. In the second approach, the time series data from each well was trained and tested separately. In other words, for 35 observation wells, 35 predictions were made. Aquifer's hydrograph was evaluated using the forecasted groundwater level in the wells. The results showed that the SVR-PSO hybrid model performed better than other models in terms of Root Mean Square Error (RMSE) and coefficient of determination  $(R^2)$  in both approaches. In the first approach, the SVR-PSO hybrid model forecasted the groundwater level in the next month with a training RMSE of 0.118 m and testing RMSE of 0.221 m. In the second approach, using the SVR-PSO hybrid model, the RMSE error was reduced in 88.57% of the wells compared to other models, and more reliable results were achieved. Based on the performance, the SVR-PSO hybrid model can be used as a tool for decision support and management of similar aquifers.

**Keywords** Bayesian networks · Groundwater level forecasting · Particle swarm optimization · Supportive vector regression · SVR-PSO hybrid model

Saman Javadi javadis@ut.ac.ir

<sup>&</sup>lt;sup>1</sup> Department of Water Engineering, College of Aburaihan, University of Tehran, Tehran, Iran

<sup>&</sup>lt;sup>2</sup> Department of Water Resources Study and Research, Water Research Institute, Tehran, Iran

<sup>&</sup>lt;sup>3</sup> Department of Environmental Conservation, College of Natural Sciences, University of Massachusetts, Amherst, USA

# **1** Introduction

The assessment of groundwater levels in aquifers plays an essential role in groundwater resource management, and changes in their levels could inform determination of groundwater volume in aquifers. In arid regions, where the surface water resources are limited (Akbarzadeh et al. 2016; Bajany et al. 2021), groundwater resources are seen as a reliable supply for various water demands (Ghafari et al. 2020). Forecasting groundwater level allows water managers to assess groundwater resources to balance supply and demand in water resources management (El Bilali et al. 2021; Sattari et al. 2018; Sheikhipour et al. 2018).

A time series of groundwater levels provides information for the sustainable management of groundwater. Groundwater levels could be forecasted by usingdata-driven methods or a conceptual approach. Recently, data-driven methods have been demonstrated to perform well in modeling groundwater levels (Adiat et al. 2020; Kouziokas et al. 2018; Mirzavand and Ghazavi 2015. Physical and numerical models, stochastic, analytical, and soft computational techniques are used to forecast groundwater levels. Numerical and Artificial Intelligence (AI) methods have also been widely used to simulate groundwater levels over time (Chitsazan et al. 2013; Jalalkamali et al. 2010; Sreenivasulu et al. 2012). Some studies have compared the performance of various groundwater forecasting models (Ankita et al. 2021; Mirarabi et al. 2019; Malekzadeh et al. 2019).

Support Vector Machines (SVM) was suggested to forecast groundwater levels as it has the advantage in reducing computational complexity, susceptibility to overfitting, and the experimental nature of other artificial intelligence methods (Sujay Raghavendra and Deka 2015). In optimizing the parameters for SVM, most of the studies use the trial and error approach. Behzad et al. (2009) observe that Support Vector Machines (SVM) show better performance and consistency in training than the Artificial Neural Networks (ANN) in predicting groundwater levels. However, they apply a time-consuming trial and error method to optimize SVM parameters. Mukherjee and Ramachandran (2018) forecasted groundwater levels using SVR, ANN, and linear regression models with satellite parameters for Terrestrial Water Storage (TWS) and meteorological variables based on the GRACE satellite data. A trial-and-error approach was also used in this study to optimize parameters in the SVR method. These studies show that the Support Vector Regression (SVR) enjoys superiority over other intelligent black-box models in groundwater forecasting (Guzman et al. 2019). Liu et al. (2021) studied using the SVM model, data absorption (DA), and trial and error method to optimize SVM parameters predicted groundwater level changes. The results show that adding Gravity Recovery and Climate Experiment GRACE data as a variable can improve the performance of SVM. The SVM-DA model performed relatively better than SVM on most stations.

El Bilali et al. (2021) compared four models of machine learning: Support Vector Regression, k- Nearest Neighbor (k-NN), Random Forest (RF), and Artificial Neural Network (ANN) in predicting the groundwater level of the Tanubart aquifer in Morocco. The SVR model showed the best forecast accuracy with the smallest forecast error in one of the piezometers. Rahbar et al. (2022) compared SVR, ANFIS, and ANN models for Daily Karst Spring Discharge Prediction in Chaharmahal and Bakhtiari province. The results showed that the SVR performed better than other models in all stages. However, all the studies have used the trial-and-error method for optimizing SVR parameters. In contrast, the generalized capability of SVR relies heavily upon the optimal values of three learning parameters including the penalty factor (C), the kernel parameter ( $\gamma$ ), and the permissible

error ( $\epsilon$ ). These parameters are interdependent and a change in one parameter affects the other linked parameters (Deka 2014). The optimization of SVR parameters through trial and error is challenging and is limited in effectively decreasing errors in groundwater forecasts. Thus, a more robust optimization that considers the interdependence of parameters of SVR is required for better forecasting of groundwater levels.

One of the most powerful tools for representing a complex system is Bayesian Networks (BN). Since the BN can address uncertainty in the relationship between forecasting variables, its application in forecasting hydrologic processes has been growing recently (Aguilera et al. 2011). Yang et al. (2015) successfully implementated a Bayesian approach in water resources and environmental research. Farmani et al. (2009) used an integrated approach based on a Bayesian networks model and an Evolutionary Multi-Objective Optimization (EMO) to manage groundwater pollution in Copenhagen, Denmark. They observe that the method assists water managers in assessing the cost and benefit consequences of alternative measures and determining the best decisions under uncertain conditions. Moghaddam et al. (2019) forecasted groundwater level in the Birjand plain using BN, groundwater simulation model (MODFLOW), and ANN using 13 piezometer observations over 12 years. The results show that BN models produce better results compared to ANN and simulation models (MODFLOW). The BN model was also used to qualitatively predict an aquifer's parameters (Ammar et al. 2009; Hantush and Chaudhary 2013). These studies show a growing number of applications of BN in forecasting groundwater.

Aquifers are important in supplying water for multiple needs, and the assessment of groundwater levels is critical to using and adequately managing water resources. The Support Vector Machine (SVM), Bayesian Networks (BN), and Support Vector Regression (SVR) have been identified as appropriate and efficient tools for forecasting groundwater levels, according to recent studies. The penalty factor (C), the kernel parameter ( $\gamma$ ), and the permissible error ( $\epsilon$ ) are three learning parameters on which the critical capability of SVR firmly depends on these interdependently-allied parameters (Deka 2014). Therefore, the PSO algorithm was used to optimize SVR parameters for the first time in groundwater level prediction in this study. Moreover, RBF was used as a kernel performance due to its accuracy, efficiency, and excellent performance compared to the other kernel functions. The main objective of this study is to develop the SVR-PSO hybrid model and compare it with recent data-driven models (SVR and BN Models). Furthermore, the developed models for groundwater level forecast were examined through well-oriented and aquiferoriented approaches.

### 2 Material and Methods

### 2.1 Method

In this study, all scenarios of modeling were performed under two approaches (aquiferoriented and well-oriented). In the first approach (aquifer-oriented), the total volume of the aquifer is assessed by applying the Thiessen method (using a weighted average of the water level at 35 observation wells). Then, the total volume of the aquifer was forecasted in various modeling scenarios in this approach. In the second approach (well-oriented), these 35 observation wells were separately modeled (trained and tested). Then the aquifer hydrograph was derived using trained and tested values of the water level of the wells. This hydrograph was compared to the observed aquifer's hydrograph, and the forecasting accuracy was examined. The second approach is distributed forecasting of groundwater levels in which changes in groundwater levels could be assessed in distributed wells. The SVR and SVR-PSO were programmed in MATLAB (2018). In both approaches, the groundwater level was forecasted through the three models (SVR, BN, and SVR-PSO), and their results were compared.

The results of the models were evaluated using monthly data from 35 observation wells in Zanjan plain, Iran, from 2004 to 2017. Data from 129 months were used for the training period (2004 to 2014) and 23 months for the testing period (2015 to 2017). The input variables of forecasting were precipitation, exploitation of groundwater (Discharge), temperature, evaporation, and groundwater level in the current month, and output was the groundwater level model in the next month. Different scenarios were used to assess the impacts of various inputs in simulation. For all models, five different scenarios were evaluated using input data. In the first scenario, all 5 input parameters were used for prediction. In the second scenario the evaporation parameter was removed due to it's less importance than other parameters. In the third scenario, both evaporation and temperature were eliminated. In the next scenario, two parameters of discharge and groundwater level in this month were used for forecasting, and in the last scenario, the only groundwater level in the month was used for forecasting (Table 1). After modeling and conducting a verification test, the results obtained from the best scenarios in each model were compared. Finally, the results of the approaches were compared and analyzed. Figure 1 displays the models' procedure, the number of scenarios, and methods.

#### 2.1.1 Support Vector Regression (SVR) Model

The Support Vector Machine (SVM) is a supervised learning model developed by Vapnik in 1995 (Vapnik 2013). This model has been broadly used for classification and regression purposes (Hosseini and Mahjouri 2014; Mirarabi et al. 2019; Panahi et al. 2020). The SVM was first developed for classification and later expanded for regression analysis (Safavi and Esmikhani 2013). The objective of the Support Vector Regression (SVR) is to train the function F(x) for the training patterns x in which the trained values fall into the thickness curve ( $\varepsilon$ ) to have the least error in the test (Smola and Schölkopf 2004). The minimization of structural risk in SVR is superior to conventional empirical risk minimization used in neural networks or classical statistical methods (Dai et al.

MODEL	Scenario	Input Data in Model							
		GWL	Discharge	Precipitation	Temperature	Evaporation			
SVR-PSO SVR BN	1	~	~	✓	✓	✓			
	2	✓	✓	✓	✓				
	3	✓	$\checkmark$	✓					
	4	✓	$\checkmark$						
	5	$\checkmark$							

Table 1 Different scenarios of input parameters to SVR-PSO, SVR, and BN models



Fig. 1 Flowchart of the research procedure

2012). Moreover, it cannot fall in local minimum error like neural networks because SVR can achieve the desired results with a small amount of data. As a matter of fact, adding unnecessary variables would create a more complex model than is required. Moreover, the complex model is susceptible to overfitting of training data. Therefore, SVR is a more authentic approach for regression with little-size datasets (Karimipour et al. 2019; Wu et al. 2008).

Given the dataset  $(x_n.y_n)$ , where  $x_n$  and  $y_n$  denoting independent and dependent variables respectively, and n = 1,2,3, ..., N, in which N is considered to be the total number of input and output data pairs, the linear regression function can be written as follows (Elbisy 2015):

$$f(x) = \sum_{n=1}^{N} w_n \varphi_n(x) + b \tag{1}$$

The value of  $\varphi_n(x)$  in Eq. (1) displays the features of input functions and  $w_n$  and b are coefficients determined by minimizing Eq. (2) using Eq. (3) as constraints (Elbisy 2015):

$$P(f(x)) = \frac{C}{N} \sum_{n=1}^{N} l_{\varepsilon} (Y_n f(x_n)) + \frac{\|w\|^2}{2}$$
(2)

$$l_{\varepsilon}(Y_{n}f(x_{n})) = \begin{cases} 0if |Y_{n}f(x_{n})| \le \varepsilon \\ |Y_{n}f(x_{n})| - \varepsilon otherwise \end{cases}$$
(3)

In Eq. (3),  $l_{\varepsilon}(Y_n, f(x_n))$  is the insensitive loss function, and C is the regulator constant defined by the user, which determines the exchange curve between the smoothing model and experimental risk. The  $\varepsilon$  is permissible tolerance between observed and computed values. The  $\frac{||w||^2}{2}$  is a smoothing component. By introducing the slack variables  $\xi$  and  $\xi^*$  in Eq. (3), the general form of optimization equation is formulated as follows (Elbisy 2015):

$$\min\left\{\varphi(w.b.\xi.\xi^*) = \frac{\|w\|^2}{2} + C\sum_{n=1}^N (\xi + \xi^*)\right\} s.t : \begin{cases} Y_i - w.\varphi(x) - b \le \varepsilon + \xi_i.\xi_i \ge 0\\ w.\varphi(x) - Y_i + b \le \varepsilon + \xi_i^*.\xi_i^* \ge 0 \end{cases}$$
(4)

### Deringer

To solve the above optimization model, a Lagrangian form of equations can be used. Based on the Lagrangian form, the regression equation can be written as Eq. (5) (Elbisy 2015).

$$f(x) = \sum_{n=1}^{N} (a_i \cdot a_i^*) + K(x_i \cdot x) + bs.t : 0 \le a_i^* \le C0 \le a_i \le C$$
(5)

In which  $K(x_i,x)$  is the kernel equation that can be used to solve the problem in N-dimensional space (Eq. (6)) (Elbisy 2015; Suryanarayana et al. 2014).

$$K(x_i.x) = \exp\left(\frac{\parallel x_i - x \parallel^2}{2\gamma^2}\right) \qquad RBF$$
  

$$K(x_i.x) = x_i.x \qquad \text{linear}$$
  

$$K(x_i.x) = (1 + x_i.x)^q \qquad \text{polynomial}$$
(6)

The radial basis function (RBF) kernel was used in this study since Several studies reported that the RBF kernel equation produces the best results compared to other existing kernels such as linear, polynomial, and sigmoid (Al-Fugara et al. 2020; Jin et al. 2021; Rajaee et al. 2019). Three numbers assigned to parameters  $\varepsilon \ \mathcal{S} \ C$ .  $\gamma$  obtained by trial and error method and 27 states were evaluated as the total permutations for parameters $\varepsilon$ (0.1.0.01.0.001).C(10.100.1000). $\gamma$ (1.0.1.0.01). The best model corresponds to the least RMSE error in the test.

#### 2.1.2 The Development of (SVR-PSO) Model

Although the SVR model is superior to artificial intelligence and statistic models, the parameters of the model require optimization to have desired results (Xiong and Xu 2006). In this study, the PSO optimization algorithm was applied to minimize RMSE in the test phase and determine the desired SVR parameters, and the SVR-PSO hybrid model was developed (Fig. 2). The performance of the SVR-PSO hybrid model was evaluated using RMSE.

The Particle Swarm Optimization (PSO) algorithm is an evolutionary computation technique (Kennedy and Eberhart 1995). This technique is based on the flock of birds (considered to be particles), with each bird searching for least distance in the search-space solutions, seeking food considered to be the best solution. Firstly, particles are positioned in the search space of the N-dimensional problem. The objective function in each particle's current position could be estimated by each of these particles, demonstrating a potential solution. Then, particles repeatedly fly in the swarm to search for the best solution. The pbest and gbest are the best positions obtained by the particle and swarm, respectively. The next position of each particle is updated by gbest and pbest, seeking the most desired performance (Poli et al. 2007). This algorithm is commonly used in optimization problems due to its simplicity and excellent capability (Li et al. 2018; Patil



Fig. 2 Flowchart of GWL forecast using PSO-SVR hybrid model

et al. 2019; Shourian et al. 2008). The procedure of forecasting groundwater levels using SVR-PSO is as follows:

- 1. Setting the first value for iteration i=0, and dividing the dataset (152 data) into two groups; training and testing (129 samples for the training period and 23 samples for the testing period).
- Normalizing the entire datasets (training and testing data) at the interval of [0,1] using Eq. (7) to enhance the ability of the model for identification of inputs and outputs relationship (Zounemat-Kermani et al. 2016):

$$X_{ik} = \frac{x_{ik} - X_{ik}^{min}}{X_{ik}^{max} - X_{ik}^{min}}$$
(7)

where  $X_{ik}$ ,  $x_{ik}$ ,  $X_{ik}^{max}$  and  $X_{ik}^{min}$  are the normalized value, the main value, the maximum and minimum values of variable k in the dataset, respectively.

3. Randomly determining the position and velocity of each particle and also learning parameters (inertia weight and the maximum number of iterations) in PSO.

- 4. Computing the SVR value function (RMSE Test) for each particle in its current position.
- Computing the current performance value for each particle and evaluating its improvement compared to prior pbest (update pbest, otherwise pbest remains unchanged).
- Comparing new pbest values with gbest value, update the gbest if the new pbest presenting a better value than gbest value, otherwise gbest remains unchanged.
- 7. Computing and updating the velocity and position of the particle, respectively.
- 8. Adding a number to iteration (i=i+1).
- 9. If the iterative termination criterion is not satisfactory, go to step 5; otherwise, gbest presents the best-optimized parameter of SVR (Hasanipanah et al. 2017).

Modeling by SVR-PSO hybrid model was similar to SVR model except that the parameters  $\varepsilon$  and *C*. $\gamma$  were optimized employing the PSO algorithm. The range of optimization for each particle should be determined to optimize in the PSO algorithm and obtain the optimal point. The number of particles and the maximum number of necessary iterations in the SVR-PSO hybrid model were 100 and 1000, respectively. Stopping occurs when there is no significant difference in the rate of optimization between two consecutive iterations. This value is considered by the program at 10^-6. But if the optimization does not reach the desired point, the optimization should be removed from the program after a certain number of iterations to avoid the infinite loop. Most optimizations in this article reached a desired point at around100 repetitions, to ensure this value 1000 repetitions were considered.. Ranges of parameters are:  $\varepsilon$ —(0 to 1);  $\gamma$ —(0.0001 to 1), and C—(1 to 1000).

#### 2.1.3 Bayesian Network

Judea Pearl first introduced the BN model in 1988 (Pearl 1988). The foundation of this model is based on the Bayes rule presented by Thomas Bayes in the eighteenth century (Aguilera et al. 2011; Uusitalo 2007). The advantage of this model is its application risk and uncertainty analyses compared to the other data-driven models, which merely give the forecasted values (Aguilera et al. 2011). Laplace developed this theory and the probabilistic logic was determined based on this theory (Farmani et al. 2009). If E and F are assumed events which  $P(E) \neq 0$  and  $P(F) \neq 0$ , then:

$$P(E|F) = \frac{P(E)P(F|E)}{P(F)}$$
(8)

The BN model provides forward and backward computation for analysis. The effect of each of the input variables on the outputs of the model could be determined by forecasting the targeted variable applying the status of the input variables in combination with having forecasted variable status (Aguilera et al. 2011; Uusitalo 2007). The BN model comprises a series of interconnected nodes that examine both occurrence and non-occurrence for each process. The joint probability distribution of n events, including  $E_1.E_2....E_n$  where  $P(E) \neq 0$  for  $1 \le i \le n$  obtained from Eq. (9) (Roozbahani et al. 2018):

$$P(E_i|F) = \frac{P(E_i)P(F|E_i)}{P(FP(E_1)P(F|E_1) + P(E_2)P(F|E_2) + \dots + P(E_n)P(F|E_n))}$$
(9)

The GeNie software was used in this study for modeling, training, and validation. The Path Condition (PC) and Necessary Path Condition (NPC) algorithms are the most commonly used in training due to their simplicity.

### 2.1.4 Performance Criteria for Model

The performance of the developed models is evaluated using statistical indicators such as coefficient of determination ( $R^2$ ), Root Mean Square Error (RMSE) (Nossent and Bauwens 2012). RMSE and  $R^2$  used not only to evaluate the accuracy of the models but also to compare them in this study as given in Eqs. (10)–(11) (Krause et al. 2005; Wunsch et al. 2018).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2}$$
(10)

$$R^{2} = \left(\frac{\sum_{i=1}^{n} (P_{i} - \overline{P})(O_{i} - \overline{O})}{\sqrt{\sum_{i=1}^{n} (P_{i} - \overline{P})^{2}} \sqrt{\sum_{i=1}^{n} (O_{i} - \overline{O})^{2}}}\right)^{2}$$
(11)

where  $P_i$  indicates the forecasted values,  $O_i$  is the observed values,  $\overline{P}$  indicates mean predicated values,  $\overline{O}$  indicates the mean observed values. The RMSE shows the difference between the observed and the predicted value. The RMSE increases from zero to large positive values as the difference between predicted and observed values increases. A relatively low amount of RMSE and a high amount of  $R^2$  (up to one) indicate an efficient model (Rezaie-Balf et al. 2017).

#### 2.2 Study Area

The aquifer studied is located in Zanjan plain of Iran, at coordinates 48° to 48°60′ E and 36°20′ to 37°N, with an area of approximately 2154 km<sup>2</sup>. This aquifer is one of the most important water resources for agricultural, industrial, and drinking purposes due to its proper water supply and lack of reliable surface water resources in the region. The annual water consumption is 490 MCM, in which 400 MCM of this amount is supplied from the aquifer.

The geographical location of the Zanjan plain and observation wells are illustrated in Fig. 3. This aquifer is in critical situation facing a 0.54-m annual drop due to overexploitation and recent droughts. The general slope in this plain surface is declining from southeast to northwest. Groundwater levels are between 1,540 and 1,800, dropping from east to west. The average annual temperature in central areas is 11 and in high land is 8 degrees °C. The maximum precipitation of 600 mm occurs in the southwest, and the minimum precipitation of 250 mm is in the northwest.



Fig. 3 The geographic location of the Zanjan aquifer

## 3 Results and Discussion

A lack of a simple relationship between input and output data is one of the error sources in data-driven models. Thus, various combination input variables (different scenarios) should be examined to find the optimal variable for the data-driven models. To have a better groundwater level forecast and determine the optimal input variables, five different scenarios of input variables combination were tested for the BN model. Precipitation, discharge, temperature, evaporation, and groundwater level (GWL) in the current month are five forecasters used to forecast the groundwater level in the next month (NMGWL). Figure 4 shows the BN structure. Five different scenarios of the input data were examined in the forecasting by SVR, and SVR-PSO hybrid models are shown in Table 1.

### 3.1 Comparison of the Models in the First Approach (Aquifer-Oriented)

Various scenarios were assessed for the three models in each approach, and the superior scenario was selected based on the least RMSE value. Figure 5 displays the results of the forecasted groundwater hydrograph for the best scenario of the three models for the aquifer-oriented approach in the testing period. The results showed that in the first approach, the error in the SVR model was higher than the BN model. However, SVR optimization parameters and developing SVR-PSO led to a better result, and thus the best outcome was achieved using this model among the three models.

The results of the models for the first approach are presented in Table 2. Due to an excellent coefficient of determination  $(R^2)$  in most models, the RMSE test is used as the



Figure. 4. BN structure for the groundwater level forecast

criterion for comparison. Table 2 shows that the models were appropriately performed in the training period, and thus examined in the testing period. The second scenario in the SVR-PSO hybrid model using four parameters, precipitation, discharge, temperature, and groundwater level in a month, showed the least RMSE compared to the other scenarios in all three models.

Overall, the hybrid SVR-PSO hybrid model can accurately forecast the groundwater level in the aquifer, and the coefficient of determination for all scenarios is higher than



Fig. 5 Comparison of observed and forecasted groundwater level during the test period for aquifer-oriented approach

MODEL		Training		Testing		Optimized parameters SVR		
		RMSE(m)	R <sup>2</sup>	RMSE(m)	R <sup>2</sup>	C	Y	3
SVR-PSO	1	0.186	0.994	0.228	0.871	504	0.0005	0.2710
	2	0.188	0.994	0.221	0.873	541	0.0007	0.2817
	3	0.185	0.994	0.255	0.839	727	0.0005	0.2845
	4	0.185	0.994	0.256	0.836	974	0.0006	0.0001
	5	0.367	0.981	0.360	0.674	962	0.0671	0.6121
SVR	1	0.153	0.996	0.326	0.867	1000	0.01	0.01
	2	0.158	0.996	0.324	0.862	1000	0.01	0.1
	3	0.167	0.995	0.318	0.816	100	0.01	0.1
	4	0.174	0.995	0.351	0.818	1000	0.01	0.1
	5	0.337	0.983	0.365	0.673	100	0.1	0.001
BN	1	0.199	0.994	0.265	0.845	-	-	-
	2	0.213	0.994	0.273	0.844	-	-	-
	3	0.222	0.994	0.287	0.835	-	-	-
	4	0.251	0.995	0.316	0.994	-	-	-
	5	0.333	0.992	0.387	0.836	-	-	-

 Table 2
 Comparison of the performance criteria of SVR-PSO, SVR, and BN models in the aquifer-oriented approach and optimized parameters SVR

0.8, except scenario 5. The value of coefficients signifies a good correlation between the observed and forecasted groundwater levels. Scenario 2 of the SVR-PSO hybrid model produced the best results having  $R^2 = 0.873$  and RMSE=0.221 m amongst all models and developed scenarios for the aquifer.

### 3.2 Comparison of the Models in the Second Approach (Well-Oriented)

After modeling and optimizing its parameters, each well played a key role in deriving the aquifer's hydrograph in the second approach (well-oriented). In other words, all models were first trained and tested using all observation wells in each scenario, then the particular values of  $C.\gamma$  and  $\varepsilon$  found for each well in the study area. Finally, the hydrograph for the aquifer was derived by the Thiessen method using the forecasted groundwater level.

Table 3 shows the results of the well-oriented approach. The coefficient of determination ( $R^2$ ) was used for the training period in all different methods and scenarios, i.e., its value was greater than 0.9. The coefficient of determination ( $R^2$ ) in the testing period in all SVR and SVR-PSO hybrid models except scenario 5 was desired. Scenario 2 of the SVR-PSO hybrid model demonstrated the best training and testing RMSE values, which were 0.2024 and 0.2194 m, respectively. Moreover, the BN model displayed the worst results in the well-oriented approach.

Different scenarios were assessed for the models in a well-oriented approach, and the ideal scenario was selected as a scenario having the least RMSE value among all scenarios of each model. Having compared observed and forecasted groundwater levels in a well-oriented approach (Fig. 6), the superiority of the SVR model over BN was observed. Moreover, forecasted groundwater levels are close to observed values in SVR-PSO hybrid models and forecasted value using the SVR model is better than the BN model. Generally, the

Table 3Comparison of the performance criteria of SVR- PSO, SVR, and BN models in the well-oriented approach	MODEL		Training		Testing	
				<b>R</b> <sup>2</sup>	RMSE(m)	R <sup>2</sup>
	SVR-PSO	1	0.1974	0.9938	0.2316	0.8929
		2	0.2024	0.9935	0.2194	0.8997
		3	0.2060	0.9935	0.2256	0.8961
		4	0.2173	0.9930	0.2237	0.8974
		5	0.3226	0.9837	0.3678	0.6695
	SVR	1	0.1940	0.9940	0.3275	0.8667
		2	0.1929	0.9940	0.2822	0.8987
		3	0.2010	0.9936	0.2846	0.8890
		4	0.2081	0.9931	0.2709	0.8924
		5	0.3182	0.9838	0.3753	0.6720
	BN	1	0.5536	0.9935	0.2194	0.8997
		2	0.6092	0.9938	0.2316	0.8929
		3	0.6214	0.9935	0.2256	0.8961
		4	0.9964	0.6370	0.9365	0.7407
		5	0.9920	0.8571	0.8210	0.9484

SVR and SVR-PSO provide more accurate predictions for the groundwater level in the study case. On the other hand, the BN model estimates groundwater levels higher than actual values. Additionally, the best values were obtained by applying the SVR-PSO hybrid model to forecast groundwater level.

Table 4 displays the averaged performance criteria for 35 observation wells for the superior scenarios of the models in a well-oriented approach. The BN method shows the worst results indicating that the performance of the model was not desired for the wells. The accuracy of BN for observation wells declines for wells in which groundwater levels temporally fluctuate (Kardan and Roozbahani 2016; Moghaddam et al. 2019).

The superior scenario of the models in a well-oriented approach was selected, and the percentage of best performance criteria for 35 observation wells are compared in Table 5.



Fig. 6 Observed and forecasted groundwater levels in test period for well-oriented approach

Table 4The average of SVR-PSO, SVR, and BN performance	Model	Training		Testing		
criteria for 35 observation wells in the present study		Average RMSE(m)	Average R <sup>2</sup>	Average RMSE(m)	Average R <sup>2</sup>	
	SVR-PSO	0.535	0.9545	0.397	0.8299	
	SVR	0.511	0.9584	0.479	0.8172	
	BN	0.874	0.9506	0.889	0.7868	

The superiority of models in the second approach (well-oriented) is defined by the percentage of the best performance criteria for each model in groundwater forecasts. Moreover, the RMSE test values were less in 88.75% of cases of SVR-PSO forecasts compared to the other two models (Table 5). Generally, applying the SVR model shows an accurately proper performance; however, overfitting errors could occur in this model. Thus, using the PSO optimization algorithm and optimizing SVR parameters assist forecast with excellent performance.

Groundwater level forecast is a critical element in groundwater resource consumption. Determining an appropriate approach with the least inaccuracy in groundwater level forecasting could support better water management. Each well-oriented and aquifer-oriented method enjoys its advantages and disadvantages, which could be used depending on the forecasting purpose. The very low computational complexity is considered one of the benefits of the aquifer-oriented approach compared to the well-oriented method. Moreover, high computations, particularly in large aquifers with numerous observation wells, are a drawback for the well-oriented approach. Reasonably faster modeling and acceptable results obtained in the aquifer-oriented approach is due to not having severe fluctuations in the aquifer's averaged groundwater level. However, only the general state of the aquifer could be obtained in the aquifer-oriented approach demonstrating its serious disadvantages. Thus, changes in groundwater level cannot be acquired in a distributed manner as obtained in a well-oriented approach. Having compared observed and simulated groundwater levels using aquifer-oriented and well-oriented approaches in the Bayesian model, the results of the aquifer-oriented approach were more accurate than a well-oriented approach. Nonetheless, the correlation and accuracy of the two approaches in the SVR-PSO hybrid model were similar, and scenario 2 in this model yields the best results. Therefore, the first approach is perceived as a proper solution for more precise decision-making and water resources management in the region and similar aquifers.

Table 5         Percentage of the	Model	Training		Testing	
and BN models in 35 observed wells		RMSE	$R^2$	RMSE	<b>R</b> <sup>2</sup>
	SVR-PSO	14	8.57	88.57	57
	SVR	71.43	51.43	5.71	20
	BN	14.29	40.00	5.71	22.86

## 4 Conclusion

Considering the importance and sensitivity of groundwater levels in water resource management, the accuracy of groundwater level forecasting is critical. Therefore, in this study, the efficiency of various forecasting models is assessed to determine a more accurate groundwater level forecaster. Three BN, SVR, SVR-PSO hybrid models were developed in this study, considering different scenarios for two well-oriented and aquifer-oriented approaches. Monthly groundwater levels in a vital aquifer from 2015 to 2017 (23 months) were forecasted to examine the developed models.

The SVR-PSO hybrid model presented the best performance compared to the developed models. Moreover, the highest correlation was observed between observed and forecasted groundwater levels in the SVR-PSO hybrid model. The results showed that SVR is better at modeling groundwater levels than the BN model. The use of the PSO optimization algorithm hastens finding the optimal parameters of the SVR model, leading to an increase in the speed of modeling and an improvement of the results. The analysis of the results showed that the SVR-PSO hybrid model in both well-oriented and aquifer-oriented approaches could produce better results than the other two models. In the aquifer-oriented approach, the SVR-PSO hybrid model could forecast groundwater level in the next month with RMSE training and testing of 0.188 and 0.266 m, respectively. However, the SVR-PSO hybrid model showed a better performance even in the well-oriented approach than the other two models. Thus, the proposed SVR-PSO hybrid model works better in both well-oriented and aquifer-oriented approaches.

The aquifer's hydrograph obtained by the SVR-PSO hybrid model displays that this model has a great ability to forecast groundwater levels. The high correlation and low RMSE value indicate that the SVR-PSO hybrid model could be reliably applied for the groundwater forecast. Therefore, forecasting groundwater levels using the SVR-PSO hybrid model in the first approach (aquifer-oriented) can be applied as a forecasting tool for decision support systems used by managers and water resource stakeholders to address water scarcity in similar aquifers.

**Contributions** Saeed Mozaffari: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing-Original Draft. Saman Javadi: Writing-Review & Editing, Formal analysis, Supervision. Hamid Kardan Moghaddam: Formal analysis, Visualization, Investigation. Timothy O. Randhir: Writing-Review & Editing.

Funding The authors did not receive support from any organization for the submitted work.

Availability of Data and Materials Authors have restrictions on sharing data.

**Code Availability** All analyses were made by MATLAB(R2018b).

# Declarations

Ethics Approval Relevant research content in this study was in accordance with the ethical standards of the institutional and national research committee.

Consent to Participate All of the authors consent to participate in the relevant research content in this paper.

**Consent for Publication** All of the authors consent to publish the paper, and it has not been published previously nor is it being considered by any other peer-reviewed journal.

Competing Interests/Conflicts of Interest The authors declare that they have no competing interests.

### References

- Adiat K, Ajayi O, Akinlalu A, Tijani I (2020) Prediction of groundwater level in basement complex terrain using artificial neural network: a case of Ijebu-Jesa, southwestern Nigeria. Appl Water Sci 10:8
- Aguilera P, Fernández A, Fernández R, Rumí R, Salmerón A (2011) Bayesian networks in environmental modelling. Environ Model Softw 26:1376–1388
- Akbarzadeh F, Hasanpour H, Emamgholizadeh S (2016) Groundwater level prediction of Shahrood Plain using RBF neural networks. J Watershed Manag Res 7
- Al-Fugara Ak, Ahmadlou M, Shatnawi R, AlAyyash S, Al-Adamat R, Al-Shabeeb AA-R, Soni S (2020) Novel hybrid models combining meta-heuristic algorithms with support vector regression (SVR) for groundwater potential mapping. Geocarto Int 1–20
- Ankita P, Dadhich Rohit, Goyal Pran N, Dadhich (2021) Assessment and prediction of groundwater using Geospatial and ANN modeling. Water Resour Manage 35(9):2879–2893. https://doi.org/10.1007/ s11269-021-02874-8
- Ammar K, McKee M, Kaluarachchi J (2009) Bayesian method for groundwater quality monitoring network analysis. J Water Resour Plan Manag 137:51–61
- Bajany DM, Zhang L, Xu Y, Xia X (2021) Optimisation Approach toward Water Management and Energy Security in Arid/semiarid Regions. Environ Process 8:1455–1480
- Behzad M, Asghari K, Coppola EA Jr (2009) Comparative study of SVMs and ANNs in aquifer water level prediction. J Comput Civ Eng 24:408–413
- Chitsazan M, Rahmani G, Neyamadpour A (2013) Groundwater level simulation using artificial neural network: a case study from Aghili plain, urban area of Gotvand, south-west. Iran Geopersia 3:35–46
- Dai H, Zhang H, Wang W, Xue G (2012) Structural reliability assessment by local approximation of limit state functions using adaptive Markov chain simulation and support vector regression. Comput Aided Civ Inf Eng 27:676–686
- Deka PC (2014) Support Vector Machine Applications in the Field of Hydrology: a Review Applied Soft Computing 19:372–386
- El Bilali A, Taleb A, Brouziyne Y (2021) Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region. J Afr Earth Sci 181:104244
- Elbisy MS (2015) Support vector machine and regression analysis to predict the field hydraulic conductivity of sandy soil. KSCE J Civ Eng 19:2307–2316
- Farmani R, Henriksen HJ, Savic D (2009) An evolutionary Bayesian belief network methodology for optimum management of groundwater contamination. Environ Model Softw 24:303–310
- Ghafari S, Banihabib ME, Javadi S (2020) A framework to assess the impact of a hydraulic removing system of contaminate infiltration from a river into an aquifer (case study: Semnan aquifer). Groundw Sustain Dev 10:100301
- Guzman SM, Paz JO, Tagert MLM, Mercer AE (2019) Evaluation of Seasonally Classified Inputs for the Prediction of Daily Groundwater Levels: NARX Networks Vs Support Vector Machines. Environ Model Assess 24:223–234
- Hantush MM, Chaudhary A (2013) Bayesian framework for water quality model uncertainty estimation and risk management. J Hydrol Eng 19:04014015
- Hasanipanah M, Shahnazar A, Amnieh HB, Armaghani DJ (2017) Prediction of air-overpressure caused by mine blasting using a new hybrid PSO–SVR model. Engineering with Computers 33:23–31
- Hosseini SM, Mahjouri N (2014) Developing a fuzzy neural network-based support vector regression (FNN-SVR) for regionalizing nitrate concentration in groundwater. Environ Monit Assess 186:3685–3699
- Jalalkamali A, Sedghi H, Manshouri M (2010) Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman plain. Iran J Hydroinformatics 13:867–876
- Jin J et al. (2021) Support vector regression for high-resolution beach surface moisture estimation from terrestrial LiDAR intensity data. Int J Appl Earth Obs Geoinf 102:102458
- Kardan MH, Roozbahani A (2016) Evaluation of Bayesian networks model in monthly groundwater level prediction (Case study: Birjand aquifer). Water Resour Manage 5
- Karimipour A, Bagherzadeh SA, Taghipour A, Abdollahi A, Safaei MR (2019) A novel nonlinear regression model of SVR as a substitute for ANN to predict conductivity of MWCNT-CuO/water hybrid nanofluid based on empirical data. Physica A 521:89–97

- Kennedy J, Eberhart R (1995) Particle swarm optimization (PSO). In: Proc. IEEE International Conference on Neural Networks, Perth, Australia, pp 1942–1948
- Kouziokas GN, Chatzigeorgiou A, Perakis K (2018) Multilayer feed forward models in groundwater level forecasting using meteorological data in public management. Water Resour Manage 32:5041–5052
- Krause P, Boyle D, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment
- Li Y, He L, Peng B, Fan K, Tong L (2018) Remote sensing inversion of water quality parameters in longquan lake based on PSO-SVR algorithm. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE, pp 9268–9271
- Liu D, Mishra AK, Yu Z, Lü H, Li Y (2021) Support vector machine and data assimilation framework for Groundwater Level Forecasting using GRACE satellite data. J Hydrol 603:126929
- Malekzadeh M, Kardar S, Saeb K, Shabanlou S, Taghavi L (2019) A novel approach for prediction of monthly ground water level using a hybrid wavelet and non-tuned self-adaptive machine learning model. Water Resour Manage 33:1609–1628
- MATLAB P (2018) 9.5.0.944444 (R2018b) Natick, Massachusetts: The MathWorks Inc
- Mirarabi A, Nassery H, Nakhaei M, Adamowski J, Akbarzadeh A, Alijani F (2019) Evaluation of datadriven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems. Environ Earth Sci 78:489
- Mirzavand M, Ghazavi R (2015) A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. Water Resour Manag 29:1315–1328
- Moghaddam HK, Moghaddam HK, Kivi ZR, Bahreinimotlagh M, Alizadeh MJ (2019) Developing comparative mathematic models, BN and ANN for forecasting of groundwater levels. Groundw Sustain Dev 9:100237
- Mukherjee A, Ramachandran P (2018) Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: Analysis of comparative performances of SVR, ANN and LRM. Journal of Hydrology 558:647–658
- Nossent J, Bauwens W (2012) Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol'sensitivity analysis of a hydrological model. In: EGU General Assembly Conference Abstracts p 237
- Panahi M, Sadhasivam N, Pourghasemi HR, Rezaie F, Lee S (2020) Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). J Hydrol 588:125033
- Patil MB, Naidu MN, Vasan A, Varma MR (2019) Water Distribution System Design Using Multi-Objective Particle Swarm Optimisation arXiv preprint arXiv:190306127
- Pearl J (1988) Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers San Mateo, Representation & Reasoning
- Poli R, Kennedy J, Blackwell T (2007) Particle Swarm Optimization Swarm Intelligence 1:33-57
- Rahbar A, Mirarabi A, Nakhaei M, Talkhabi M, Jamali M (2022) A comparative analysis of data-driven models (SVR, ANFIS, and ANNs) for daily karst spring discharge prediction. Water Resour Manag. https://doi.org/10.1007/s11269-021-03041-9
- Rajaee T, Ebrahimi H, Nourani V (2019) A review of the artificial intelligence methods in groundwater level modeling. J Hydrol
- Rezaie-Balf M, Zahmatkesh Z, Kim S (2017) Soft computing techniques for rainfall-runoff simulation: local non-parametric paradigm vs. model classification methods. Water Resour Manag 31:3843–3865
- Roozbahani A, Ebrahimi E, Banihabib ME (2018) A framework for ground water management based on bayesian network and MCDM techniques. Water Resour Manag 32:4985–5005
- Safavi HR, Esmikhani M (2013) Conjunctive use of surface water and groundwater: application of support vector machines (SVMs) and genetic algorithms. Water Resour Manag 27:2623–2644
- Sattari MT, Mirabbasi R, Sushab RS, Abraham J (2018) Prediction of groundwater level in Ardebil plain using support vector regression and M5 tree model Groundwater 56:636–646
- Sheikhipour B, Javadi S, Banihabib ME (2018) A hybrid multiple criteria decision-making model for the sustainable management of aquifers. Environ Earth Sci 77:712
- Shourian M, Mousavi S, Tahershamsi A (2008) Basin-wide water resources planning by integrating PSO algorithm and MODSIM. Water Resour Manag 22:1347–1366

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222

- Sreenivasulu D, Deka PC, Nagaraj G (2012) Investigation of the effects of meteorological parameters on groundwater level using ANN Artificial Intelligent Systems and Machine. Learning 4:39–44
- Sujay Raghavendra N, Deka PC (2015) Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression Cogent Engineering 2:999414

- Suryanarayana C, Sudheer C, Mahammood V, Panigrahi BK (2014) An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam. India Neurocomputing 145:324–335
- Uusitalo L (2007) Advantages and challenges of Bayesian networks in environmental modelling. Ecol Model 203:312–318
- Vapnik V (2013) The nature of statistical learning theory. Springer science & business media
- Wu C, Chau KW, Li YS (2008) River Stage Prediction Based on a Distributed Support Vector Regression. J Hydrol 358:96–111
- Wunsch A, Liesch T, Broda S (2018) Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). J Hydrol 567:743–758
- Xiong W-L, Xu B-G (2006) Study on optimization of SVR parameters selection based on PSO. J Sysem Simul 18:2442–2445
- Yang L, Zhao X, Peng S, Zhou G (2015) Integration of Bayesian analysis for eutrophication prediction and assessment in a landscape lake. Environ Monit Assess 187:4169
- Zounemat-Kermani M, Kişi Ö, Adamowski J, Ramezani-Charmahineh A (2016) Evaluation of data driven models for river suspended sediment concentration modeling. J Hydrol 535:457–472

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.