


Genetic programming based monthly groundwater level forecast models with uncertainty quantification

K. S. Kasiviswanathan^{1,3} · S. Saravanan²  · M. Balamurugan³ · K. Saravanan⁴

Received: 3 January 2016 / Accepted: 9 January 2016 / Published online: 29 January 2016
© Springer International Publishing Switzerland 2016

Abstract Modeling hydrogeologic processes facilitates in accurate prediction/forecasting of groundwater level variations. Still, the uncertainty in model prediction is a major concern that requires detailed investigation. There could be several factors which introduce uncertainty such as inherent assumption, various levels of model complexity and simplicity. In general, model inputs, parameters and structure are the major sources of uncertainty while quantifying model prediction uncertainty. In this study, a genetic programming (GP) based models have been employed for forecasting groundwater level variation along with prediction uncertainty quantification. Though various sources induce uncertainty in the model prediction, the input uncertainty quantification has received little attention. Hence, the input uncertainty has been considered for the analysis in this study. The proposed method is demonstrated using measured monthly values of rainfall and corresponding groundwater level data of Amarawathi basin, India. It is observed that the prediction along with uncertainty quantification improves the confidence level of models while making decisions, in particular for effective planning and management of groundwater resources.

Keywords Genetic programming · Groundwater · Prediction · Uncertainty

Introduction

Groundwater plays a vital role while estimating the overall water balance in hydrologic and hydrogeologic processes. The world's 30 % of fresh water exists in the form of ground water. Agriculture is a major activity in developing countries which solely depends on groundwater resources. In India, the contribution of groundwater in agricultural sector is about 50 % of the total irrigated area and nearly 80 % of total agricultural production depends on groundwater resources. However, during recent past, the groundwater resources face acute shortage due to over exploitation, urbanization and population growth. Also the excessive pumping of groundwater leads to several negative environmental consequences such as land subsidence, water quality issues, changes in flow patterns, sea water intrusion, etc. It may be noted that unlike the surface water, no clear guidelines or policies available for regulating the groundwater pumping. Hence, effective strategies should be developed for managing the groundwater, and also to maintain the equilibrium state.

Recent developments in computer modeling and sophisticated instruments that measure the groundwater level are useful tools which help administrators and policy makers for efficient utilization. Models, in general, are broadly classified into physics based and data driven approach. The physics based groundwater models are powerful tool while representing high spatial and temporal variability of aquifer parameters (Shiri and Kisi 2011). The rigorous model calibration procedure and large data required often discourages the physics based groundwater

✉ S. Saravanan
ssaravananirs@gmail.com

¹ Department of Civil Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

² Department of Civil Engineering, National Institute of Technology, Tiruchirappalli 620 015, India

³ Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600 036, India

⁴ School of Mechanical and Building Sciences, VIT University, Chennai 600 127, India

models. Some of the model parameters such as hydraulic conductivity, porosity and permeability are difficult to measure in field (Bierkens 1998; Knotters and Bierkens 2000), also which involves in cost and efforts. Consequently, there is considerable level of uncertainty associated with these parameters due to heterogeneity and complexity of the system. In such situations, the data driven modeling approaches produce reasonably accurate estimate of groundwater levels, further the results are comparable with physics based models. The main advantage of the data driven models are as follows: (i) do not require complete physics of the system, (ii) robustness, (iii) can be modeled with available observed input, output variables, (iv) can be generalized to an arbitrarily level of accuracy (Srivastav et al. 2007).

Variety of data driven models have been employed for groundwater level forecasting. In specific, the application of artificial neural network (ANN) models have produced promising results while forecasting the groundwater level (Dash et al. 2010; Ghose et al. 2010; Krishna et al. 2008; Nayak et al. 2006; Taormina et al. 2012). Also the comparison between different data driven models showed the relative improvement across the models (Yoon et al. 2011; Sreekanth et al. 2012). However, there is no clear evidence which indicates a single best model, which out performs under different condition. Hence the choice of particular model depends on the data used. In this study, the GP based models have been employed for forecasting groundwater level. In fact, GP has been reported in variety of hydrologic and hydro-geologic modeling such as: flood routing (Sivapragasam et al. 2008), evaporation (Shiri et al. 2013), ground water remediation (Aly and Peralta 1999), suspended sediment modeling (Kisi et al. 2012), stage discharge curve (Azamathulla et al. 2011), short-term water level fluctuations (Shiri and Kisi 2011) and rainfall-runoff model (Kisi et al. 2013). The detailed review of application of GP in water resources has been presented in ASCE Task Committee (2010). The potential advantage of using GP lies in optimizing the model structure and parameters simultaneously, whereas most of the other data driven models use predefined model structure for optimizing the model parameters. It may be noted that all these studies did not consider quantifying model prediction uncertainty rather focused on point prediction. The major reason could be the complex interaction between the modeled variables and high computation required. However, in recent times, the quantification of uncertainty has become a necessary exercise in modeling for improving the confidence and reliability of the developed models.

Hence, the focus of this study is to model groundwater level using GP approach along with quantification of prediction uncertainty. This is accomplished using a probabilistic framework which generates different realization of

model inputs for obtaining the variability of model output. The approach is demonstrated through data collected from three observation wells located in Amaravathi basin, India.

Study area and data description

Amaravathi River is one of the major tributaries of Cauvery River. It covers an area of 8280 km² which lies in between the latitudes 10°10'00"N and 11°02'10"N, the longitudes 77°03'24"E and 78°13'06"E (Fig. 1). The basin has four districts namely Coimbatore, Erode, Dindigul and Karur. The main stream originates at an altitude of 2600 m above mean sea level (MSL). From the total area of the basin, the forest and cultivated land area covers around 1100 and 3770 km² respectively. The remaining area comes under the category of barren and uncultivable lands. Paddy, sugarcane and groundnut are the major crops in this region.

The basin has three major soil types such as black soil, alluvial soil and calcareous soil. The area is underlined by crystalline rocks of Peninsular Gneissic Complex, comprising of hornblend-biotite-gneiss and charnokite. The basin has four distinct seasons that include South-West monsoon from June to September, North-East monsoon from October to December, the winter season from January to February and summer from March to May. The rainfall pattern in this region is highly heterogeneous, which receives rainfall due to South-East and North-West monsoons. The long term average annual rainfall is varying from 1300 to 500 mm.

The basin experiences a maximum and minimum monthly mean temperature of 38.21 and 23.81 °C respectively. The average wind speed ranges between 16.03 and 0.2 kmph. The mean relative humidity is low in dry weather and high in the monsoon season. The sky is very cloudy during the monsoon season and is lightly clouded during non-monsoon season. The monthly climate data for the meteorological stations (i.e., K.Paramathy, Sundakampalayam, Uthamapalayam and Viralipatti) were obtained from the Public Works Department, Chennai.

The selected wells (K.Paramathy, Keeranur and Kuthiraiyar) for the study are illustrated in the Fig. 1 along with rain gauge and evaporation stations. The rationale behind in selecting these wells is geographically located in plain, highland and hilly regions.

Performance measures

Model prediction measures

The model prediction is evaluated with various statistical indices [i.e., correlation coefficient (CC), Nash–Sutcliffe

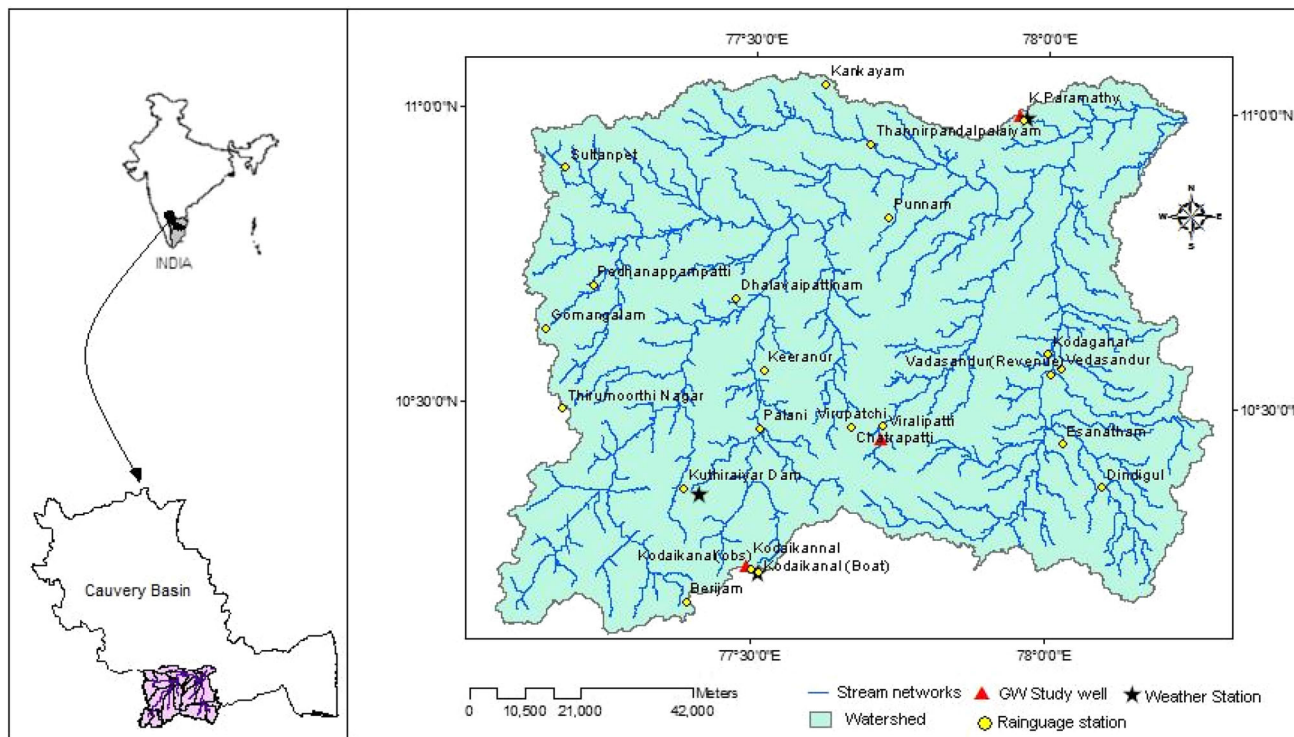


Fig. 1 Study area map of Amarawathi basin

efficiency (NE), root mean square (RMSE), and mean bias error (MBE) and defined as follows:

$$CC = \frac{\sum_{i=1}^n [(l_i^o - \bar{l}^o) \cdot (l_i^p - \bar{l}^p)]}{\sqrt{\sum_{i=1}^n (l_i^o - \bar{l}^o)^2 \cdot \sum_{i=1}^n (l_i^p - \bar{l}^p)^2}} \tag{1}$$

$$NE = \left\{ 1 - \frac{\sum_{i=1}^n (l_i^o - l_i^p)^2}{\sum_{i=1}^n (l_i^o - \bar{l}^o)^2} \right\} \times 100 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (l_i^o - l_i^p)^2} \tag{3}$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (l_i^o - l_i^p) \tag{4}$$

where, l_i^o and l_i^p are the observed and predicted water table values at time t, respectively; \bar{l}^o and \bar{l}^p are the mean of the observed and predicted water table values corresponding to n data points, n refers the number of degrees of freedom.

Model prediction uncertainty measures

The two uncertainty indices i.e., percentage of coverage (POC) and average width (AW) are used and is defined as follows.

$$POC = \left(\frac{1}{n} \sum_{i=1}^n c_i \right) \times 100 \tag{5}$$

$$AW = \frac{1}{n} \sum_{i=1}^n [l_i^{up} - l_i^{lp}] \tag{6}$$

where, l_i^{up} and l_i^{lp} are the upper and lower bound estimation of the i^{th} pattern; $c_i = 1$ if the observed values of target fall in the prediction band $[l_i^{up}, l_i^{lp}]$, otherwise $c_i = 0$.

Methodology

Genetic programming

Genetic programming (GP) is an evolutionary algorithm based on Darwinian theories of natural selection and survival to approximate the equation, in symbolic form, that best describes how the output relates to the input variables (Sivapragasam et al. 2008). The algorithm initializes with a randomly generated arithmetic operators (+, −, ×, ÷) and mathematical functions (sin, cos, tan, exp, log, sigmoid) from population. Based on understanding of the physical process, the combination of these operators and functions along with inputs form a set of equations in each generation/iteration. The individuals in other words genes consist of operators (i.e., +, / and ×) and terminals (3, 4, x and z)

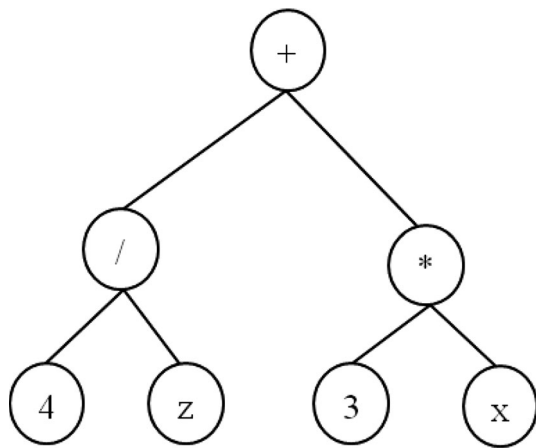


Fig. 2 A typical individual that returns $((4/z) + (3 \times y))$

are illustrated in Fig. 2. The example may be interpreted as $((4/z) + (3 \times y))$.

Each set of potential solutions is then subjected to an evolutionary process called ‘fitness’ (a measure of how well

they solve the problem). The individual programs that best fit the data are selected from the initial population. This paper used mean square error (*MSE*) as a measure of fitness function. GP attempts to combine the elements of existing solutions in order to create a new solution, with some of the features of each “parent”. Based on this, the generated model equations that best fit are selected to exchange part of the information between them to produce optimal model equations through ‘crossover’ and ‘mutation’ processes which mimics the natural world reproduction system. The GP equation which less fitted the data are discarded. This evolution process is repeated over successive generations and is driven towards finding symbolic expressions describing the data, which can be interpreted to derive knowledge about the process. The flow chart describes the GP model evolution at different stages (Fig. 3). More details on GP can be obtained from (Koza 1992; Babovic and Keijzer 2000; Khu et al. 2001). This study used DTREG software (Phillip 2012) to develop GP based groundwater fluctuation model.

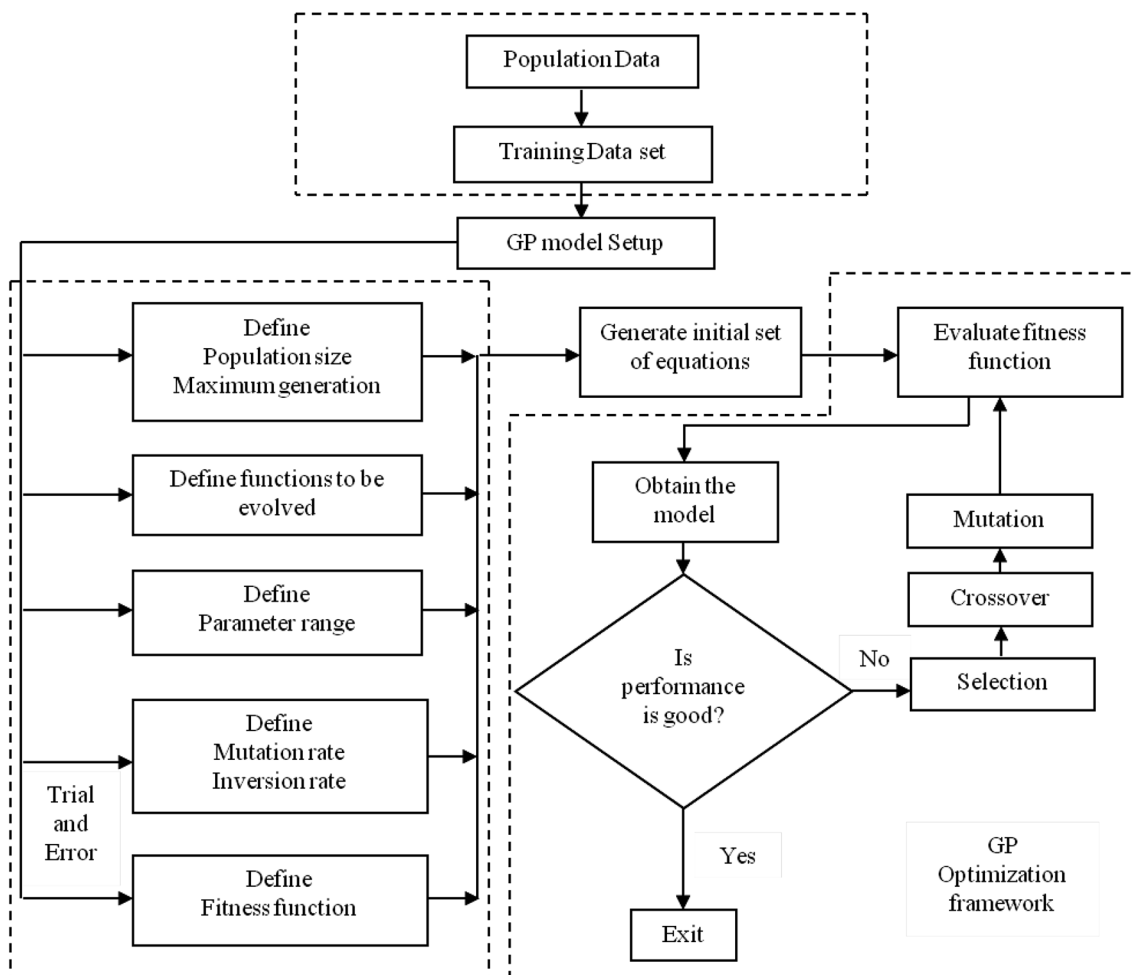


Fig. 3 Flowchart of proposed methodology

Input selection

There is a growing interest in modeling the relationship between the physical variables with variety of test procedure to account the non-linearity. However, identifying the significant input variables generally require a priori knowledge about the system to be modeled (Campolo et al. 1999; Thirumalaiah and Deo 2000). In fact most of the occasion, the relationship between the variables is not clearly known a priori, and hence often an analytical technique, such as auto-, cross-correlation is employed (Sajikumar and Thandaveswara 1999; Sudheer et al. 2002; Kasiviswanathan and Sudheer 2013). However, the cross-, auto- correlation is only being able to detect the linear dependence between the variables, while the modeled relationship may be highly nonlinear. Nonetheless, the cross-, auto- correlation methods represent the most popular analytical techniques for selecting appropriate inputs (Bowden et al. 2004a, b).

The model proposed in this study used monthly rainfall and groundwater level information of selected three wells over 30 years of period to forecast future groundwater level. The data were collected between January, 1980 and December, 2009. Out of the total available data (30 years), first 21 years of data (252 sets) was used for model calibration and the remaining was used for model validation. Initially the potential input variables corresponding to different lag time periods are identified through statistical analysis suggested by (Sudheer et al. 2002). In such analysis, various antecedent values at different lagged time model inputs (i.e., rainfall) were analyzed for their influence on model output (i.e., groundwater level). The quantitative analysis of cross correlation function between inputs and output was employed to determine appropriate inputs for the model to be trained. Similarly, an autocorrelation function (ACF) suggests the most influencing antecedent groundwater level in producing the better autoregressive process with model output. Based on this analysis, the modeler can select appropriate inputs which would result the models of parsimonious and subsequently eliminates the input variables which has less influence. With this procedure, the input variables are identified to develop GP model for the wells selected and is presented in Table 1.

It is clear that different well has different potential lagged values of rainfall and depth to water table. The statistically obtained lagged values of depth to water table values are consistent across the wells for the modeling. It is interesting to note that while selecting the lagged values of rainfall, the immediate and long term precedent values were identified which has good correlation with output variable of the model. This information reinforces that the groundwater table fluctuation is subjected to both immediate and long term response of rainfall.

Uncertainty quantification

The calibration of the GP model is carried out with measured values of inputs–output identified through correlation analysis. While these models are used to predict future variables for the unseen input variables, it might be possible that the models may not be able to produce better results because of inherent uncertainty. Consequently, the point prediction is always questionable, which in turn lacks explaining the model robustness and limitations. Therefore, the quantification of uncertainty in the form of prediction interval is a necessary exercise in modeling which helps in improving the confidence of the developed model. There are different methods of uncertainty analysis is reported in literature (Shrestha and Solomatine 2006; Montanari 2007).

In recent decades, the quantification of input uncertainty has gained significant attention among researchers, since the input variables are one among the major source for causing the variability in model predictions. In general, the model validation is carried out with calibrated model of $f(.)$ with deterministic structure, parameter and input data and it is denoted as follows:

$$y_i = f(x_i, p) \tag{7}$$

where, x denotes model inputs, p denotes model parameters, y refers model output at any instance ‘ i ’.

In this study, a simplified approach is proposed to account the input uncertainty of GP model. In such approach, the errors (e_i) are probabilistically sampled for quantifying the prediction uncertainty of GP models. Therefore the Eq. (7) changes into Eq. (8) which has errors multiplied with inputs.

Table 1 Identified input variables for developing GP models

Well	Climatic input	Initially identified model inputs
K.Paramathy	Rainfall	$l_t = f(R_{t-14}, R_{t-13}, R_{t-12}, R_{t-2}, R_{t-1}, l_{t-2}, l_{t-1})$
Keeranur	Rainfall	$l_t = f(R_{t-14}, R_{t-13}, R_{t-3}, R_{t-2}, R_{t-1}, l_{t-2}, l_{t-1})$
Kuthiraiyar	Rainfall	$l_t = f(R_{t-14}, R_{t-13}, R_{t-12}, R_{t-2}, R_{t-1}, l_{t-3}, l_{t-2}, l_{t-1})$

$$y_i = f(e_i \times x_i, p) \tag{8}$$

It is known that the error is normally distributed with zero mean and with specified values of standard deviation. In this framework, the assumption is that the error follows log-normal distribution with zero mean and pre-fixed standard deviation values, since we require positive values of errors to be multiplied. The error forced input variables (i.e., sampled error is multiplied with measured values of rainfall and water level values) are then simulated through the calibrated GP models for quantifying the prediction uncertainty. It is to be noted that input errors associated with rainfall and water levels may not be same and therefore, the independent error values are sampled from distinct log normal distribution.

Results and discussion

GP model development

Based on the input selection, the patterns of input–output vector corresponding to calibration and validation period is separated. In which the calibration data is used for training the GP model. The selection of suitable model setting parameters is listed in Table 2. This generally requires extensive trails of different combination of parameters for better predicting the model output. After setting these parameters, the GP creates initial sets of number of equations (i.e., which is equal to population value). These equations are evaluated with fitness function such as ‘MSE’; subsequently the performance of the model is evaluated. Based on the fitness criteria, models with good

Table 2 The values of parameters used in GP model evolution

Parameter	Value
Population size	50
Generation	2000
Crossover rate	0.96
Mutation rate	0.04
Fitness function	Mean square error

Table 3 Trained GP models for predicting groundwater fluctuation

Well	Climatic input	GP equation
K.Paramathy	Rainfall	$l_t = (-13.16 * e^{(R_{t-1}-6.20)*l_{t-2}}) + (-15.76 * e^{(R_{t-2}-7.12)*l_{t-2}}) + l_{t-1} + \frac{0.05}{l_{t-2}}$
Keeranur	Rainfall	$l_t = (l_{t-1} + 0.12) * \left(\frac{0.95}{\sqrt{e^{0.30 * R_{t-1}}}}\right)$
Kuthiraiyar	Rainfall	$l_t = (-0.08 * l_{t-2} * R_{t-1}) + \frac{0.004 * R_{t-1}}{R_{t-1}-1.005} + 0.08 + l_{t-2}$

performance are preserved using the process of cross over and mutation. This process continues for the number of generation and the final model is obtained if there is no recognized level of improvement in model performance (Fig. 3). The GP equations trained for each well are listed in Table 3.

It is noted that the finally evolved GP equations consist different levels of complexity as well simplifications. Hence, the GP model might contain various combinations of short and long term lagged values of rainfall information. It clearly indicates that GP model explicitly considers the nonlinearity associated with the variables of interest and not necessary that all the input variables identified through cross-, auto- correlation methods should be included.

The GP model developed for K.Paramavathy well has more number of parameters and variables compared to other wells. This indicates the complex nature and high non-linearity. Keeranur well has only l_{t-1} and R_{t-1} information as input variables. It can be interpreted as high nonlinearity of these variables, which has strong correlation with groundwater level fluctuations. Other probable reason could be ascribed as this well is located in high elevation which causes immediate response through lagged variables of R_{t-1} and l_{t-1} . It is interesting to note that the final GP model did not consist l_{t-1} variable in the case of Kuthiraiyar well. However the presence of l_{t-2} indicates that in hilly areas a slight delayed response of water table depth has more influence with future predicted values of groundwater level.

Efficacy of model performance

The GP model performance is evaluated using various statistical indices such as mean biased error, root mean square error (RMSE), Nash–Sutcliffe efficiency (NE) and the coefficient of correlation between the measured and computed groundwater level values. The performance indices values are presented in Table 4. The results indicate that process has been modeled with reasonable accuracy across selected wells. Though a slight improved model performance is observed during calibration, performances

Table 4 Performance of GP model across different wells

Well	Calibration				Validation			
	CC	NE (%)	RMSE (m)	MBE (m)	CC	NE (%)	RMSE (m)	MBE (m)
K.Paramathy	0.88	76.45	0.67	0.11	0.80	61.88	1.23	0.26
Keeranur	0.85	71.05	1.75	0.23	0.81	64.25	2.35	0.27
Kuthiraiyar	0.90	80.43	1.56	0.06	0.83	66.41	2.43	-0.05

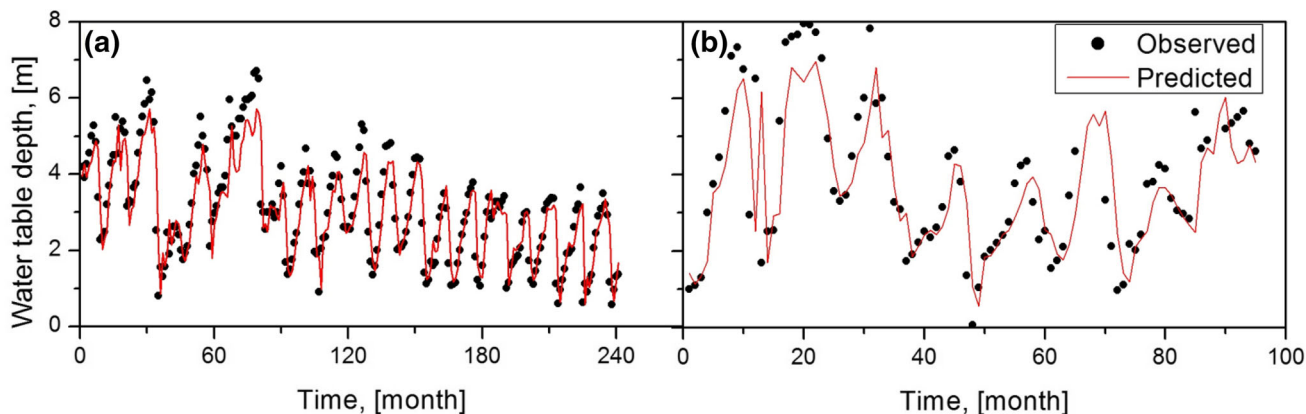


Fig. 4 Observed and predicted values of water table depth for K.Paramathy well **a** Calibration and **b** Validation

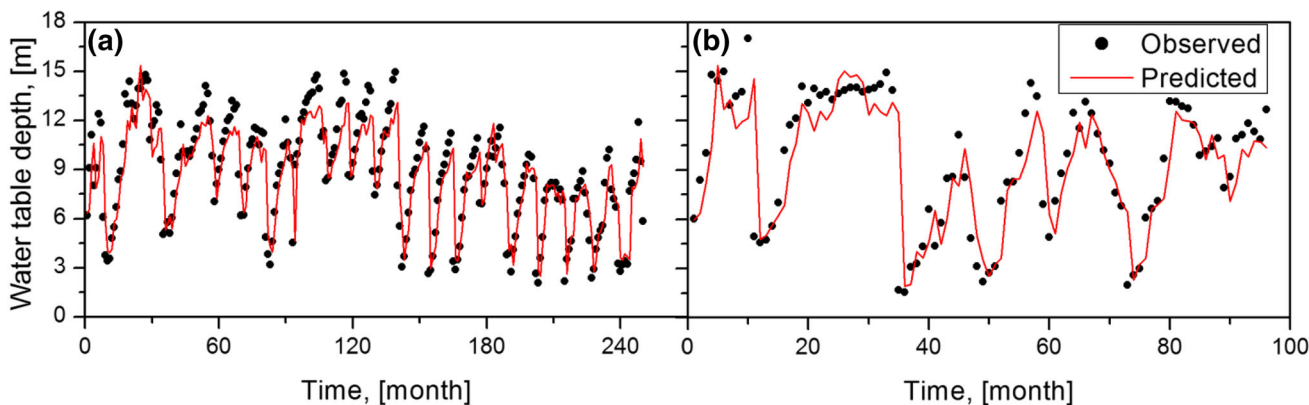


Fig. 5 Observed and predicted values water table depth for Keeranur well **a** Calibration and **b** Validation

across calibration and validation periods are consistent across the wells selected for the study.

In the case of K.Paramathy well, the GP model has NE of 76.45 % with CC of 0.88 during calibration and the validation simulation produced NE of 61.88 % with CC of 0.79. The model generalization is good in calibration and validation period. The positive values of MBE indicates the under estimation of model prediction both in calibration, validation phase. The RMSE obtained in validation is 1.23 m against the mean value of water level fluctuation 4.02 m in validation shows less residual variance against total variance. The model prediction (i.e., groundwater level fluctuation) is plotted with observed values for

calibration and validation periods and it is illustrated in Fig. 4. It is also observed that the model in some points fails to capture sharp fluctuation of water level. In the case of Keeranur well, the model’s NE during calibration and validation is 71.05 and 64.25 % respectively. The MBE indices show similar performance as obtained for K.Paramavathy well. Figure 5 illustrates the model predictions of Keeranur well during calibration and validation periods. The RMSE value during validation is 2.35 m whereas the mean water level fluctuation is 9.29 m which shows better model generalization. While comparing performance across different wells using NE values, the GP model for Kuthiraiyar well has produced better results than other two

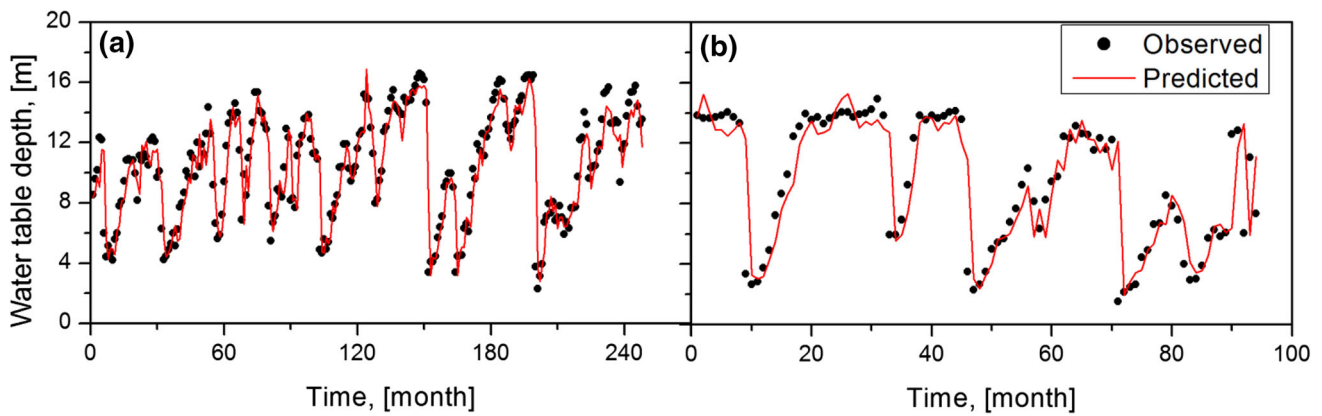


Fig. 6 Observed and predicted values water table depth for Kuthiraiyar well a Calibration and b Validation

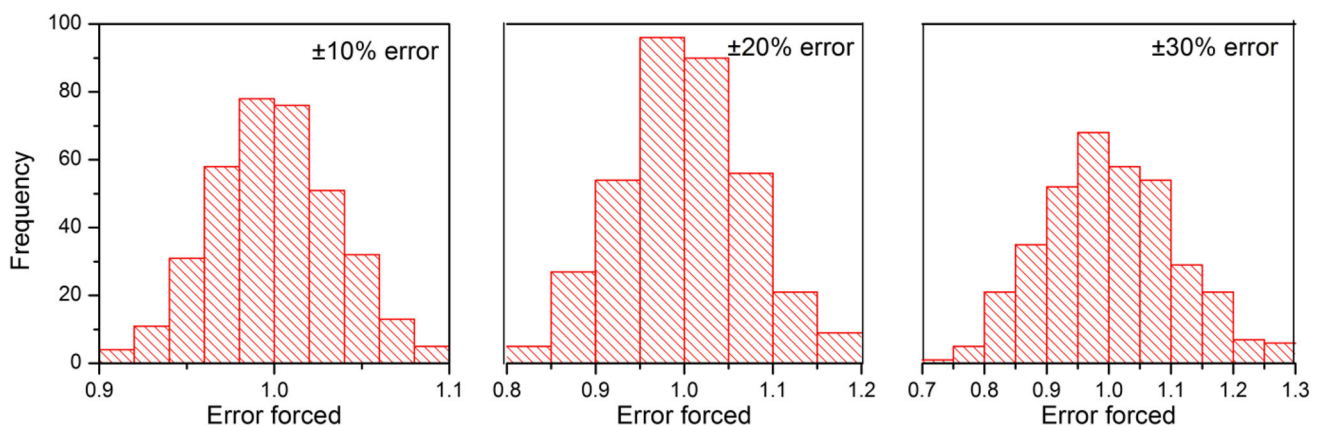


Fig. 7 Histogram of errors sampled from the lognormal distribution

Table 5 Quantitative estimation of prediction interval

Error forced (%)	K.Paramathy		Keeranur		Kuthiraiyar	
	AW (m)	POC (%)	AW (m)	POC (%)	AW (m)	POC (%)
±10	0.52	20.21	3.32	60.00	2.12	50.53
±20	0.98	40.43	7.00	81.05	4.30	78.94
±30	1.55	47.87	11.65	87.37	6.58	89.47

wells. It has NE value of 80.43 % in calibration and 66.41 % in validation. The negative MBE value (i.e., -0.0484 m) indicates model over prediction during validation. The observed and predicted values are plotted in Fig. 6 during calibration and validation period. Overall, the model prediction is reasonably good across different water level fluctuations.

Prediction uncertainty of GP models

Based on the methodology presented, the input data are perturbed with forced errors ranges (from ±10 to ±30 %) that are sampled from log-normal distribution (Kasiviswanathan

2013). For brevity, the error sampled from histogram corresponds to ±10, ±20 %, and ±30 % are presented in Fig. 7. The error forced inputs are validated through developed GP model for constructing the prediction interval of the model output. The prediction uncertainty for the validation data are estimated using indices AW and POC for the selected wells. The results are presented in Table 5. It is expected that increasing the input error increases the predictive uncertainty at the output. Hence, the AW value increases across all the wells when the error increases from ±10 to ±30 %. It is obvious that increasing the average width of prediction band would tend to contain more observed points, consequently which increases the percentage of coverage values. It is also

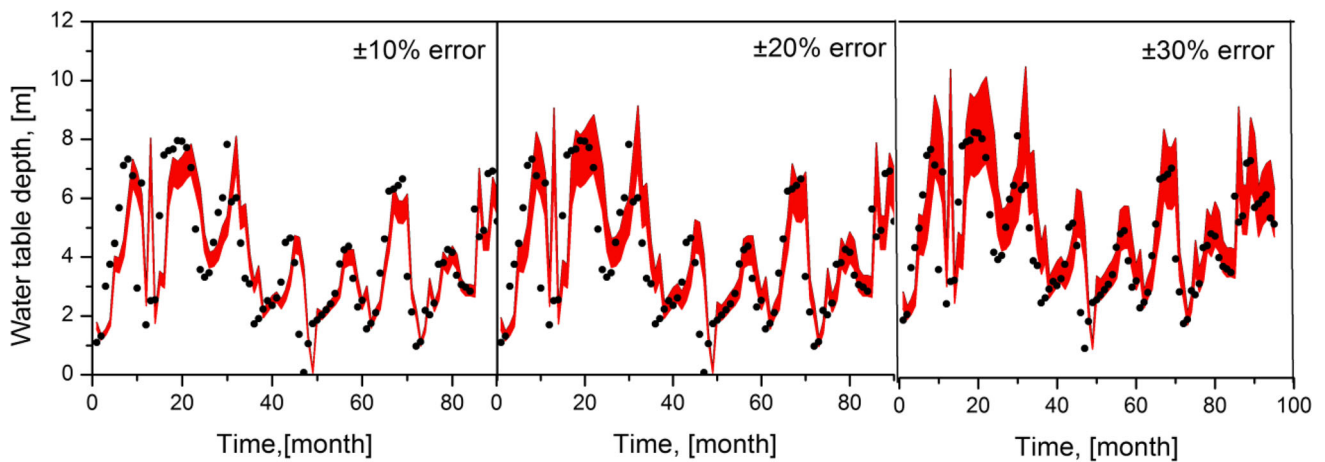


Fig. 8 The prediction interval of GP model for K.Paramathy well during validation (The dotted points denote observed values)

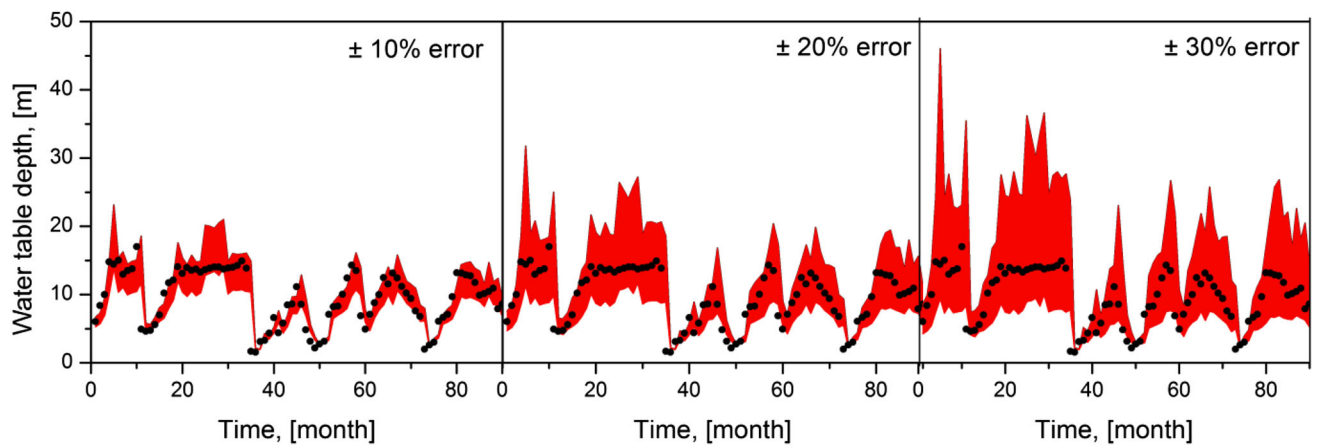


Fig. 9 The prediction interval of GP model for Keeranur well during validation (The dotted points denote observed values)

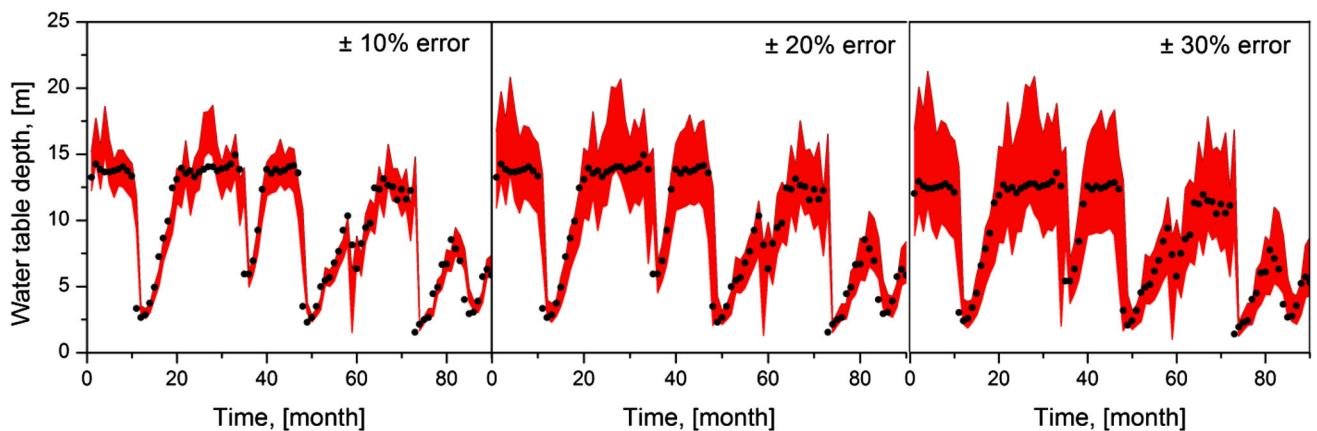


Fig. 10 The prediction interval of GP model for Kuthiraiyar well during validation (The dotted points denote observed values)

noted that K.Paramavathy well has a less POC even with considerable increase in its average width (Fig. 8). However, the models developed for Keeranur and Kuthiraiyar has

significant POC variations while there is increase in input error (Figs. 9, 10). The reason could be the developed GP model for K. Paramavathy well may not be able to capture the

variance of the underlined non-linear process; consequently the model covers less number of observed values. This information would be useful while analyzing the model behavior under uncertain condition. The increased POC and reduced AW indicates the robustness of the model, where the model can be considered as more reliable.

Remarks

The presented results clearly illustrate that quantification of uncertainty helps in improving the confidence of the models as compared to the single point forecast values of groundwater level fluctuations. While developing GP models couple of concerns have arisen.

1. Can GP model be developed with rainfall, evapotranspiration information together in case of groundwater level fluctuation? In doing so, we observed that the finally evolved equation hides the information of evapotranspiration and does not present in GP equation. In addition, the performance of model is also reduced. From the region map it is clear that the wells are located near streams and interactions between groundwater/surface water in losing and/or gaining streams would affect the groundwater level to great extent.
2. Can the information of river stage as another input variable improve the model prediction? The answer is yes. However in this study, we don't have the information of river stage data near the wells. Therefore, further analysis with more climatic and hydrologic variables would require for better understanding the actual process involved.

Conclusion

The following specific conclusions are listed based on the present study

1. GP based model captures the non-linearity of groundwater level fluctuations reasonably good with hydrogeology and meteorological observed data without requiring explicit knowledge of physics of the system.
2. The model prediction at 1 month lead time depends on short term monthly lagged values of rainfall and previous water levels rather long term lagged values. This indirectly indicates the high porosity and hydraulic conductivity values of filed under study.
3. The uncertainty analysis with forced error in the input shows the model output variability which helps in improving the model confidence in real case applications.

4. It is also noted that this study considers only error in inputs however quantification of parameter and model structure uncertainty further improves the quality of prediction band which is in turn offer better understanding of the process being modeled.
5. The results from GP based groundwater level models suggest that this can be directly incorporated while developing scenarios of pumping strategies and hence for consumptive use of any basin.
6. In the case of real time operation of well, these GP models can be coupled with optimizer for efficient decision making in advance with specified levels of uncertainty.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal studies This paper does not contain any studies with human participants or animals performed by any of the authors.

Informal consent For this type of study formal consent is not required.

References

- Aly AH, Peralta RC (1999) Optimal design of aquifer cleanup systems under uncertainty using a neural network and Genetic algorithm. *Water Resour Res* 35(8):2523–2532
- ASCE Task Committee (2010) State of the art for genetic algorithms and beyond in water resources planning and management on evolutionary computation in environmental and water resources Engineering. *J Water Resour Plan Manag* 136:412–432
- Azamathulla HM, Mohammad H, Aminuddin A, Ghani Leow CS, Chang CK, Zakaria NA (2011) Gene-expression programming for the development of a stage-discharge curve of the Pahang river. *Water Resour Manag* 25:2901–2916
- Babovic V, Keijzer M (2000) Rainfall runoff modeling based on genetic programming. *Nord Hydrol* 33:331–346
- Bierkens MFP (1998) Modeling water table fluctuations by means of a stochastic differential equation. *Water Resour Res* 34:2485–2499
- Bowden GJ, Dandy GC, Maier HR (2004a) Input determination for neural network models in water resources applications. Part 1—background and methodology. *J Hydrol* 301(1–4):75–92
- Bowden GJ, Dandy GC, Maier HR (2004b) Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *J Hydrol* 301(1–4):93–107
- Campolo M, Soldati A, Andreussi P (1999) Forecasting river flow rate during low flow periods using neural networks. *Water Resour Res* 35(11):3547–3552
- Dash NB, Panda SN, Remesan R, Sahoo N (2010) Hybrid neural modeling for groundwater level prediction. *Neural Comput Appl* 19:1251–1263
- Ghose DK, Panda SS, Swain PC (2010) Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. *J Hydrol* 394:296–304

- Kasiviswanathan KS (2013) Uncertainty analysis of neural network based hydrologic models. Unpublished PhD thesis, IIT Madras
- Kasiviswanathan KS, Sudheer KP (2013) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-012-0600-2
- Khu ST, Liong SY, Babovic V, Madsen H, Muttill N (2001) Genetic programming and its application in real-time runoff forecasting. *J Am Water Resour Assoc* 37:439–451
- Kisi O, Hossein Zadeh, Dalir A, Cimen M, Shiri J (2012) Suspended sediment modeling using genetic programming and soft computing techniques. *J Hydrol* 450–451:48–58
- Kisi O, Shiri J, Tombul M (2013) Modeling rainfall-runoff process using soft computing techniques. *Comput Geosci* 51:108–117
- Knotters M, Bierkens MFP (2000) Physical basis of time series models for water table depths. *Water Resour Res* 36:181–188
- Koza JR (1992) Genetic programming: on the programming of computers by natural selection. MIT Press, Cambridge
- Krishna B, Satyajirao YR, Vijaya T (2008) Modeling groundwater levels in an urban coastal aquifer using artificial neural networks. *Hydrol Process* 22:1180–1188
- Montanari A (2007) What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrol Process* 21:841–845
- Nayak PC, Satyajirao YR, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resour Manag* 20:77–90
- Phillip HS (2012) DTREG predictive modeling software user manual
- Sajikumar N, Thandaveswara BS (1999) A non-linear rainfall-runoff model using an artificial neural network. *J Hydrol* 216:32–35
- Shiri J, Kisi O (2011) Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Comput Geosci* 37:1692–1701
- Shiri J, Marti P, Singh VP (2013) Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning. *Hydrol Process*. doi:10.1002/hyp.9669
- Shrestha DL, Solomatine DP (2006) Machine learning approaches for estimation of prediction interval for the model output. *Neural Netw* 19:225–235
- Sivapragasam C, Maheswaran R, Venkatesh V (2008) Genetic programming approach for flood routing in natural channels. *Hydrol Process* 22:623–628
- Sreekanth PD, Sreedevi PD, Ahmed S, Geethanjali N (2012) Comparison of FFNN and ANFIS models for estimating groundwater level. *Environ Earth Sci* 62(6):1301–1310
- Srivastav RK, Sudheer KP, Chaubey I (2007) A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resour Res* 43:W10407. doi:10.1029/2006WR005352
- Sudheer KP, Gosain AK, Ramasastri KS (2002) A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol Process* 16(6):1325–1330
- Taormina R, Chau KW, Sethi R (2012) Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng Appl Artif Intell* 25:1670–1676
- Thirumalaiah K, Deo MC (2000) Hydrological forecasting using neural networks. *J Hydrol Eng* 5(2):180–189
- Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396:128–138