

Engineering Applications of Computational Fluid Mechanics

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tcfm20>

Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression

Shahab S. Band, Essam Heggy, Sayed M. Bateni, Hojat Karami, Mobina Rabiee, Saeed Samadianfard, Kwok-Wing Chau & Amir Mosavi

To cite this article: Shahab S. Band, Essam Heggy, Sayed M. Bateni, Hojat Karami, Mobina Rabiee, Saeed Samadianfard, Kwok-Wing Chau & Amir Mosavi (2021) Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression, Engineering Applications of Computational Fluid Mechanics, 15:1, 1147-1158, DOI: [10.1080/19942060.2021.1944913](https://doi.org/10.1080/19942060.2021.1944913)

To link to this article: <https://doi.org/10.1080/19942060.2021.1944913>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 15 Jul 2021.



Submit your article to this journal [↗](#)



Article views: 490



View related articles [↗](#)



View Crossmark data [↗](#)

Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression

Shahab S. Band ^a, Essam Heggy ^{b,c}, Sayed M. Bateni ^d, Hojat Karami ^e, Mobina Rabiee^e, Saeed Samadianfard ^f, Kwok-Wing Chau ^g and Amir Mosavi ^{h,i,j}

^aFuture Technology Research Center, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan, People's Republic of China; ^bViterbi School of Engineering, University of Southern California, Los Angeles, CA, USA; ^cJet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA; ^dDepartment of Civil and Environmental Engineering and Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI, USA; ^eCivil Engineering Department, Semnan University, Semnan, Iran; ^fDepartment of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran; ^gDepartment of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China; ^hFaculty of Civil Engineering, Technische Universität Dresden, Dresden, Germany; ⁱJohn von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary; ^jDepartment of Informatics, J. Selye University, Komarno, Slovakia

ABSTRACT

Utilizing new approaches to accurately predict groundwater level (GWL) in arid regions is of vital importance. In this study, support vector regression (SVR), Gaussian process regression (GPR), and their combination with wavelet transformation (named wavelet-support vector regression (W-SVR) and wavelet-Gaussian process regression (W-GPR)) are used to forecast groundwater level in Semnan plain (arid area) for the next month. Three different wavelet transformations, namely Haar, db4, and Symlet, are tested. Four statistical metrics, namely root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and Nah-Sutcliffe efficiency (NS), are used to evaluate performance of different methods. The results reveal that SVR with RMSE of 0.04790 (m), MAPE of 0.00199%, R^2 of 0.99995, and NS of 0.99988 significantly outperforms GPR with RMSE of 0.55439 (m), MAPE of 0.04363%, R^2 of 0.99264, and NS of 0.98413. Besides, the hybrid W-GPR-1 model (i.e. GPR with Harr wavelet) remarkably improves the accuracy of GWL prediction compared to GPR. Finally, the hybrid W-SVR-3 model (i.e. SVR with Symlet) provides the best GWL prediction with RMSE, MAPE, R^2 , and NS of 0.01290 (m), 0.00079%, 0.99999, and 0.99999, respectively. Overall, the findings indicate that hybrid models can accurately predict GWL in arid regions.

ARTICLE HISTORY

Received 11 August 2020
Accepted 14 June 2021

KEYWORDS

Groundwater level prediction; hydrological model; Gaussian process regression; support vector; artificial intelligence; machine learning

Introduction

The sharp decline of groundwater level (GWL) in arid areas has attracted the attention of water resources engineers. By increasing the demand for water resources in industrial, residential, and agricultural sectors, forecasting the groundwater level has gained its particular importance in groundwater resources management (Nayak et al., 2006; Suryanarayana et al., 2014). Moreover, GWL predictions have been used to avoid the side effects of groundwater level decline including the loss of water pumping capacity from wells and ground subsidence (Moosavi et al., 2013; Shirmohammadi et al., 2013a).

Numerical methods are often computationally expensive and need many input variables in predicting GWL (Anderson & Woessner, 1992). Hence, in recent years, a number of studies used machine learning algorithms to predict GWL (Taormina et al., 2012; Porte et al., 2018; Lal & Datta, 2018; Su et al., 2020; Niu & Feng, 2021;

Alcalá García et al., 2021). By using air temperature, precipitation and GWL in preceding months, Chang et al. (2015) employed the feedforward neural network model to predict monthly GWL. Gholami et al. (2015) applied a multilayer perceptron (MLP) model to forecast GWL. Their results indicated that MLP model had a high capability for GWL forecasting. Rezaie-balf et al. (2017) utilized Multivariate Adaptive Regression Splines (MARS) and M5 Model Trees (MT) to simulate GWL fluctuation. Ghose et al. (2018) used radial basis neural network (RNN) to predict GWL. Rainfall, air temperature and humidity, runoff, and evapotranspiration were used as inputs in their model. Banadkooki, Ehteram, Ahmed, Teo, Fai, Afan, and El-Shafie (2020) attempted to predict GWL by hybrid radial basis neural network-whale algorithm (RNN-WA), hybrid MLP-WA, and genetic programming (GP) approaches. The results indicated that MLP-WA was more accurate than the

CONTACT Amir Mosavi  amir.mosavi@mailbox.tu-dresden.de; Shab S. Band  shamshirbands@yuntech.edu.tw

other two approaches for 3, 6, and 9 months ahead forecast scenarios.

Recently, Support Vector Regression (SVR) and Gaussian Process Regression (GPR) were utilized to predict different hydrologic variables such as water level in lakes and dams (Hipni et al., 2013; Lan, 2014), river flow forecasting (Esmailzadeh et al., 2017; Samadianfard et al., 2019a), global solar radiation (Samadianfard et al., 2019b), daily dew point temperature (Qasem et al., 2019), longitudinal dispersion coefficient in natural streams (Kargar et al., 2020), pan evaporation (Shabani et al., 2020) and standardized streamflow index (Shamshirband et al., 2020). SVR is a type of support vector machine (SVM) algorithm, which was introduced by Vapnik (1995). It belongs to supervised learning algorithms and can be used to solve regression problems (Cortes & Vapnik, 1995; Vapnik, 1999). SVR is able to achieve good prediction of variables of interest based on a small number of statistical samples. Compared to fuzzy logic and neural network approaches, SVR has a shorter training time, lower data dependence, and easier implementation. Gaussian Process Regression (GPR) is a Bayesian learning approach, which assumes the output probability distribution be Gaussian since it does not change the probability of inputs (O'Hagan, 2013; Sun et al., 2013). GPR is able to model complex relationships between inputs and outputs and cope with a broad range of simulation behaviors of alternative models (Karami et al., 2018; Liu & West, 2009; O'Hagan, 2013; Rasmussen & Nickisch, 2010). It deals with certain aspects of regression and prediction problems (Reis et al., 2005; Stedinger & Tasker, 1985; Vogel et al., 1999).

Using single techniques may be risky for water resources management. Hence, hydrologists usually take advantage of hybrid approaches to make better decisions for managing water resources. Various studies indicated that combining different techniques and developing hybrid methods could improve the prediction. Graf et al. (2019) applied 4 mother wavelets, namely Daubechies, Symlet, discrete Meyer and Haar, coupled with ANN to predict daily water temperature at a river in Poland. Zhu et al. (2020) applied a hybrid wavelet-multi layer perceptron approach to forecast daily surface water temperature at eight lakes in Polish. Campozano et al. (2020) investigated the capabilities of ANN to predict river discharge using wavelet analysis.

Wavelet transfer (WT) can analyze changes in time-series to provide the time scale of non-stationary signals (Adamowski, 2008; Adamowski & Chan, 2011; Karami et al., 2018; Kisi & Cimen, 2011; Samadianfard et al., 2018; Zhang & Dong, 2001; Zheng et al., 2000). In contrast to classical Fourier analysis (FA), which is a trigonometric polynomial, WT propagates functions

in waves. This form of translations and dilations is generated using a fixed function called mother wavelet (Zhu et al., 2020). It decomposes time series into sub-components and supplies more data at different resolutions, ultimately increasing the veracity of prediction models (Tikhmarine et al., 2019). WT has been widely used in water resources management studies (Karthika & Deka, 2015; Shafaei & Kisi, 2015; Shirmohammadi et al., 2013b). Also recently Rezaie-balf et al. (2017) coupled WT with MARS and MT to increase the accuracy of GWL forecast. Consequently, the goal of this study is to predict one-month ahead groundwater level in Semnan plain (Iran) by SVR, GPR, wavelet-support vector regression (W-SVR) and wavelet-Gaussian processes regression (W-GPR) approaches. Semnan plain has an arid climate with a mean annual rainfall of 140 (mm). Groundwater is the main source of water supply for residential and agricultural sectors in Semnan plain. However, its improper use has caused severe problems (e.g. sharp decline in groundwater level, land subsidence, and shortage of irrigation water) in Semnan province. Hence, the results of this study can help decision makers and stakeholders to manage groundwater resources more efficiently. Three different mother wavelets, namely Haar, db4, and Symlet, are tested in W-SVR and W-GPR approaches. Monthly air temperature, precipitation, relative humidity, and groundwater level for a period of 20 years (1996–2017) are used as inputs.

Materials and methods

Study area

Semnan plain is located in the east of Tehran and has an area of about 1207 km². The study area spans over the longitude of 53°3' E – 53°40' E, and latitude of 35°18' N – 35°43' N. The plain has an altitude of 874–1857 m above the sea level. It has an arid climate with an average annual precipitation of about 140 mm. More than 80% of precipitation in the Semnan plain evaporates. The remaining 20% is converted into the runoff and groundwater. The location of Semnan plain in Iran and its observation wells are shown in Figure 1.

Data

Various factors such as air temperature, humidity, precipitation, evapotranspiration, aquifer storage capacity, harvesting rate from wells, and recharging aquifers affect GWL (Adeleke et al., 2015). However, only a subset of these variables is available in Semnan plain. Based on literature (Banadkooki, Ehteram, Ahmed,

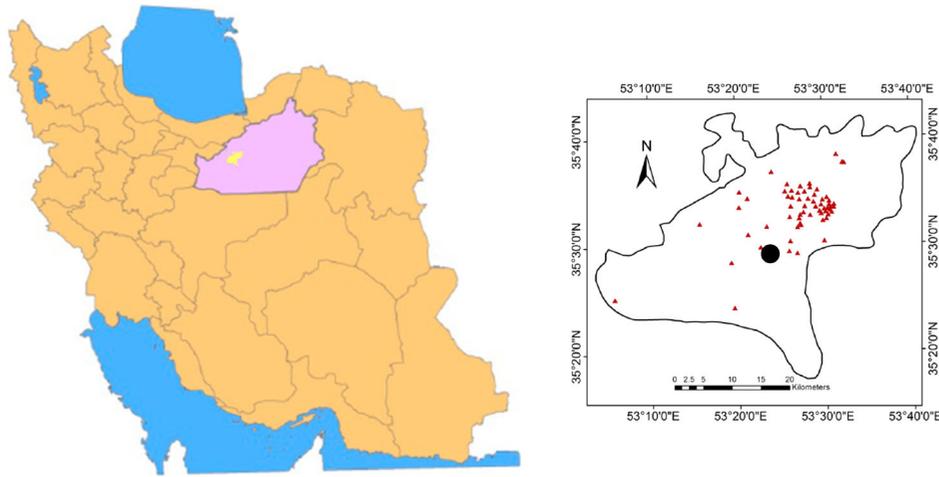


Figure 1. Location of Semnan plain (note: The red dots and black circle show the location of observation wells in Semnan plain and Merhamat Abad well, respectively.)

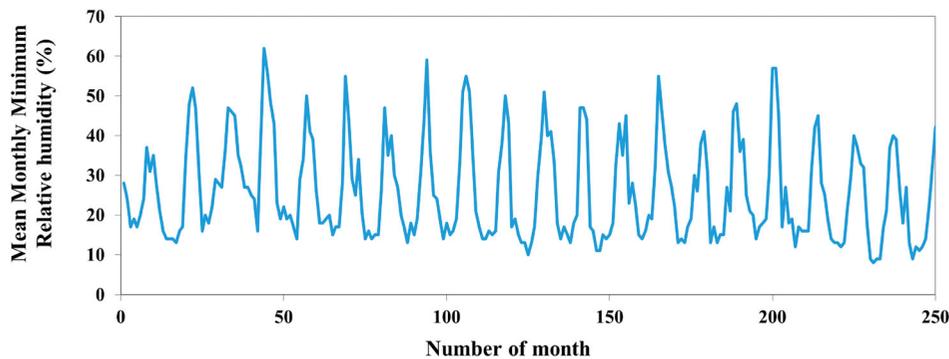


Figure 2. Time series of mean monthly minimum relative humidity (%) (1996–2017).

Teo, Fai, Afan, Sapitang, et al., 2020; Emamgholizadeh et al., 2014; Khoshand, 2021) and correlation analyses between GWL and available micrometeorological variables in Semnan plain, mean minimum air temperature (\bar{T}_{\min}), minimum relative humidity (\overline{RH}_{\min}), precipitation (P) and GWL in the current and previous three months are used as inputs. The micrometeorological data are provided by Semnan Meteorological Organization (<http://www.semnanweather.ir/>). Groundwater level data from the observation wells are provided by Semnan Provincial District Water Authority from March 1996 to February 2017 (<http://www.semnanweather.ir/>). The entire dataset is divided into training (March 1996–October 2012) and testing (November 2012–February 2017) datasets. Time series of \overline{RH}_{\min} , P , \bar{T}_{\min} and GWL in Semnan plain during 1996–2017 (20 years) are shown in Figures 2–5, respectively. Statistical metrics of the utilized variables are shown in Table 1.

Support vector regression (SVR)

Support vector regression (SVR) is a nonparametric method, which was developed by Vapnik (1995). It has

been used as a machine learning tool for classification and regression (Mirarabi et al., 2019). The performance of SVR is highly dependent on its kernel. Consider a dataset $\{(x_i, d_i)\}_i^N$ where x_i is the input vector, d_i is the target value, and N is the number of data samples. The SVR function, $f(x)$, is given by,

$$f(x) = \omega \cdot \phi(x) + b \quad (1)$$

where ω is the weight vector, ϕ is the non-linear transfer function and b is the bias. To achieve a proper performance of $f(x)$, the regression problem is estimated by minimizing the structural risk function:

$$\text{minimize} : \frac{1}{2} \omega^2 + c \left(\sum_{i=1}^N (\xi_i + \xi_i^*) \right) \quad (2)$$

$$\text{Subject to} \begin{cases} \omega \cdot \phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ d_i - \omega \cdot \phi(x_i) + b_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, N \end{cases} \quad (3)$$

where ξ_i and ξ_i^* are the slack variables, c is the error penalty term, which denotes the trade-off of the risk term

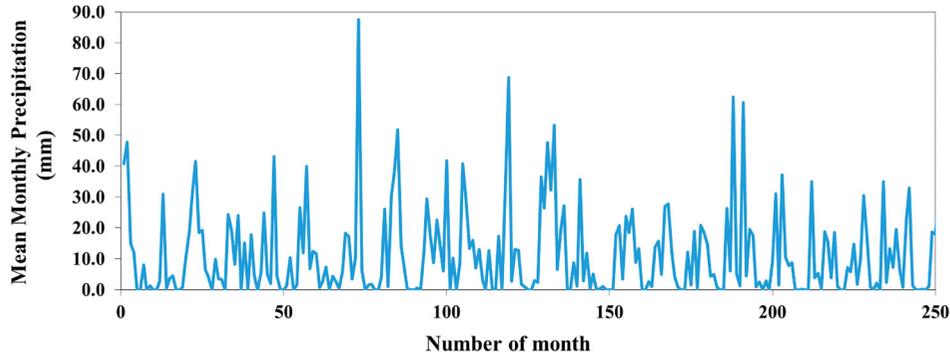


Figure 3. Time series of mean monthly precipitation (mm) (1996–2017).

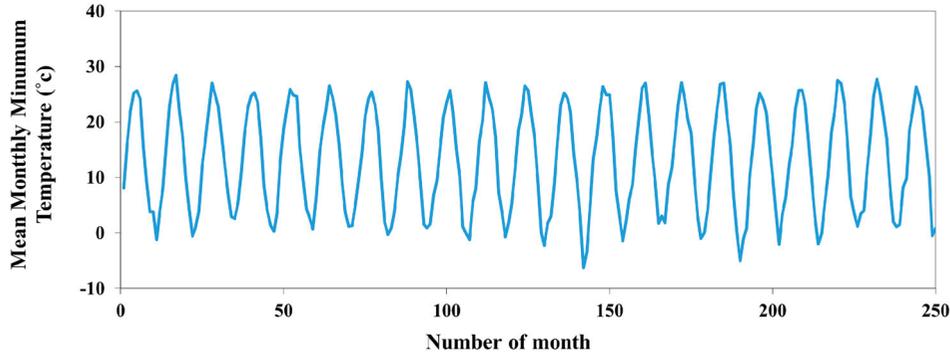


Figure 4. Time series of mean monthly minimum air temperature (°C) (1996–2017).

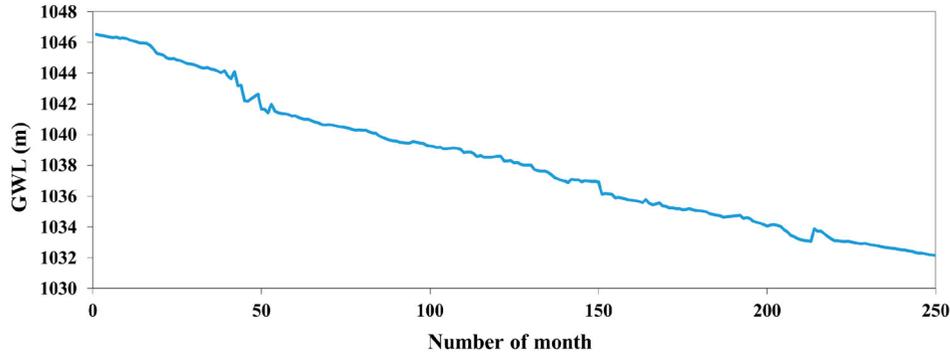


Figure 5. Time series of mean monthly groundwater level in Semnan plain (1996–2017).

Table 1. Statistical metrics of the utilized variables.

| Variable | Mean | Minimum | Maximum | Standard deviation |
|-----------------|---------|---------|---------|--------------------|
| P (mm) | 11.19 | 0.00 | 87.50 | 14.31 |
| RH_{\min} (%) | 25.90 | 8.00 | 62.00 | 12.49 |
| T_{\min} (°C) | 13.23 | -6.30 | 28.40 | 9.37 |
| GWL (m) | 1038.31 | 1032.08 | 1046.51 | 4.19 |

and regularization observed so that the deviation ε can be tolerated, and ε is the error tolerance (Suryanarayana et al., 2014). By making an appropriate change to the initial formula of the target function, we can turn it into a dual problem for quadratic programming. Quadratic programming is then used to solve the SVR problem:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot k(x_i, x) + b \quad (4)$$

where α_i and α_i^* are dual Lagrange multipliers, $k(x_i, x)$ is the kernel function, which is equal to $\phi(x_i)\phi(x)$. The above equation is used for both linear and non-linear regression analyses (Isazadeh et al., 2017). It is important to choose the right parameters to achieve a reliable predictive accuracy (Zounemat-Kermani et al., 2016). One of the widely used kernel function is the radial basic function, which is given below:

$$k(x_i, x) = \exp(\gamma x - x_i)^2 \quad (5)$$

where γ is the kernel width (Yu et al., 2018).

Gaussian process regression (GPR)

GPR is a non-parametric and stochastic approach that can solve nonlinear regression problems (Grbic et al., 2013). GPR learns the relationship between the predictor and target. According to linear regression, the response variable y can be defined as a function of p -dimensional variables $x = [x_1, \dots, x_p]^T$ parameterized by $w = [w_1, \dots, w_p]^T$ as shown in Equation (6) (Wang & Chen, 2015):

$$y_i = \sum_{d=1}^p w_d x_{id} + \varepsilon_i (i = 1, \dots, n) \quad (6)$$

where n is the number of data points, x_{id} is the d th covariate of x_i , and ε_i is the Gaussian noise with zero mean and variance σ_ε^2 (Wang & Chen, 2015).

The output value $y = [y_1, \dots, y_n]^T$, as a linear function of w and ε_i , is also Gaussian with zero mean and covariance matrix C . The Gaussian process can be defined as $p(y) = G(0, C)$, where G denotes the Gaussian distribution. The equation for C is given by Wang and Chen (2015):

$$C_{ij} = C(x_i, x_j) = \sigma_w^2 x_i^T x_j + \delta_{ij} \sigma_\varepsilon^2 \quad (7)$$

where δ_{ij} is equal to 1 for $i = j$, and 0 for $i \neq j$. Equation (8) presents the Gaussian variance (Wang & Chen, 2015):

$$\sigma_*^2 = C(x_*, x_*) - k_*^T C^{-1} k_* \quad (8)$$

where x refers to the input values and $k_* = [C(x_*, x_1), \dots, C(x_*, x_n)]^T$. Accordingly, the log-likelihood of the training data can be defined by Equation (9) to take hyper-parameters in the presence of maximum likelihood computation (Wang & Chen, 2015):

$$L = \left(-\frac{1}{2}\right) \log(\det C) - \frac{1}{2} y^T C^{-1} y - \frac{n}{2} \log(2\pi) \quad (9)$$

Equation (10) presents the derivative of the log-likelihood with respect to each hyper parameter (Wang & Chen, 2015):

$$\frac{\partial L}{\partial \theta} = \left(-\frac{1}{2}\right) \text{tr} \left(C^{-1} \frac{\partial C}{\partial \theta} \right) + \left(\frac{1}{2}\right) y^T C^{-1} \left(\frac{\partial C}{\partial \theta} \right) C^{-1} y \quad (10)$$

Wavelet transform (WT)

The wavelet transform (WT) has been successful in the field of signal processing (Starck & Murtagh, 2001). In general, WT replaces local oscillating base functions (called wavelets) with infinitely oscillating base functions of Fourier transform. In fact, wavelets are displaced versions of a stem wavelet with a value of $\psi(t)$ (Selesnick

et al., 2005). These wavelets, by selecting and combining with a low-pass scale function with a value of $\varphi(t)$, can form a balanced base shape for signals of a one-dimensional time series (1-D). Accordingly, any finite energy analog signal $x(t)$ can be separated in terms of wavelets and scale functions by Equation (11) (Selesnick et al., 2005):

$$x(t) = \sum_{n=-\infty}^{\infty} c(n) \vartheta(t-n) + \sum_{j=0}^{\infty} \sum_{n=-\infty}^{\infty} d(j, n) 2^{\left(\frac{j}{2}\right)} \psi(2^j t - n) \quad (11)$$

where $\psi(t)$ is the mother wavelet function, j is the scale factor, and n is the time shift. It should be noted that in WT, the basic functions are translations of a function named the mother wavelet. Besides, the scaling coefficients $c(n)$ and wavelet coefficients $d(j, n)$ are computed via the inner products as shown in Equations (12) and (13) (Selesnick et al., 2005):

$$c(n) = \int_{-\infty}^{+\infty} x(t) \vartheta(t-n) dt \quad (12)$$

$$d(j, n) = 2^{\left(\frac{j}{2}\right)} \int_{-\infty}^{+\infty} x(t) \psi(2^j t - n) dt \quad (13)$$

$c(n)$ and $d(j, n)$ provide a time–frequency analysis of the signal by measuring its frequency content (controlled by the scale factor j) at different times (controlled by the time shift n).

This study uses three different mother wavelets, namely Haar, db4, and Symlet, to compare their performance to improve groundwater level modeling when applying hybrid W-SVR and W-GPR models.

Performance indicators

Root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and Nah-Sutcliffe efficiency (NS) are used to measure the accuracy of predictions. The expression for each of these metrics is shown below (Faizollahzadeh Ardabili et al., 2018):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (14)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|(x_i - y_i)|}{y_i} \times 100 \quad (15)$$

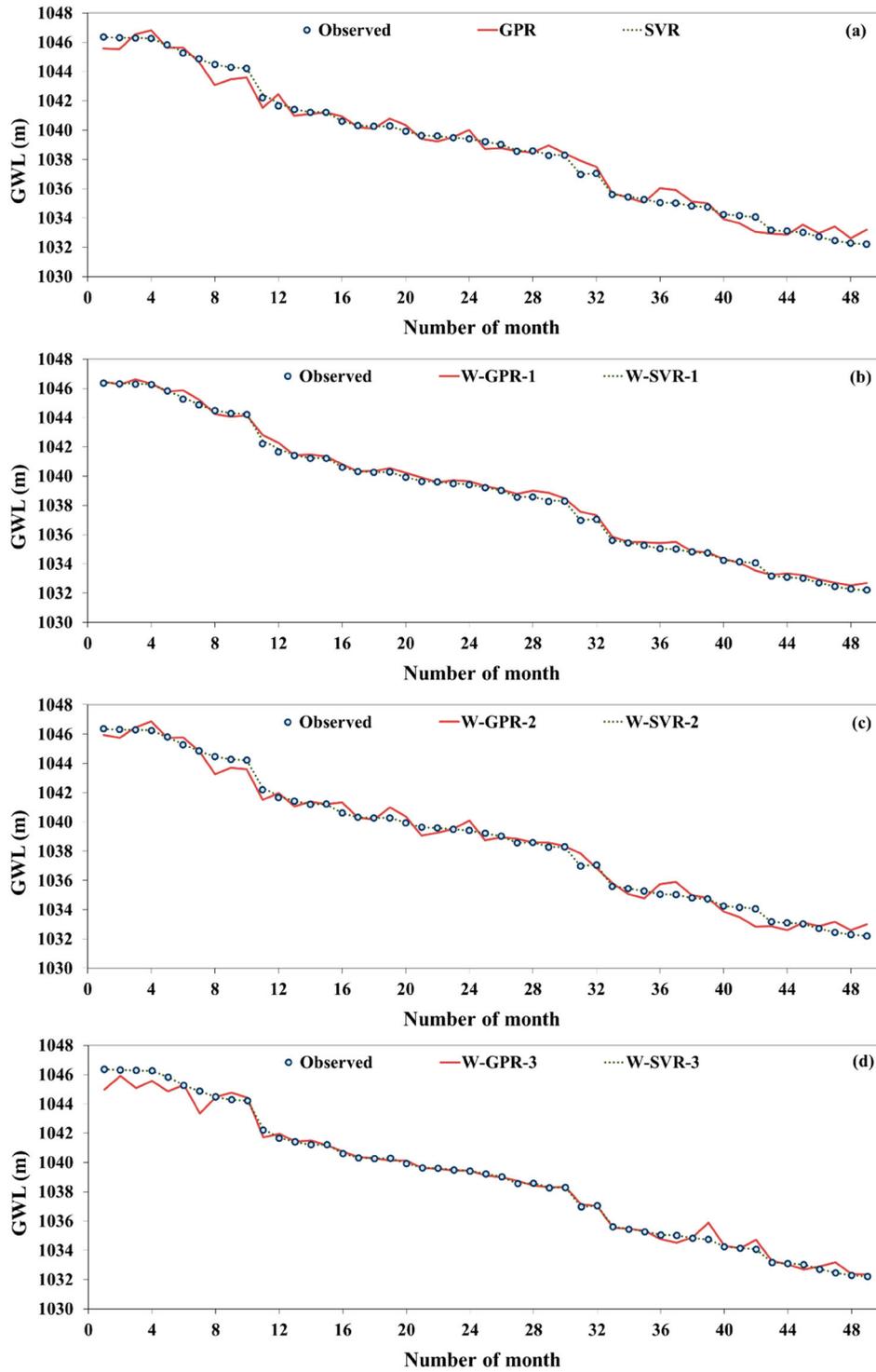


Figure 6. Time series of monthly GWL observations and predictions from (a) GPR and SVR, (b) W-GPR-1 and W-SVR-1, (c) W-GPR-2 and W-SVR-2, and (d) W-GPR-3 and W-SVR-3 during 2012–2017.

$$NS = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{y})^2} \quad -\infty \ll NS \ll 1 \quad \ll R^2 \ll 1 \quad (17)$$

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad 0 \quad (16)$$

where x_i and y_i are the i th observed and predicted groundwater levels, respectively. \bar{x} , \bar{y} are the average of observed and predicted groundwater level data, respectively, and n is the number of total data.

The correlation coefficient (R^2) indicates the degree of correlation between the observed and predicted values. The closer it is to 1, the more the observed and predicted data are correlated. RMSE and MAPE denote the error of predictions. The closer they are to 0, the more accurate the prediction is. The closer NS is to 1, the more accurate the prediction is.

Results and discussion

In this study, the hybrid Haar-, db4-, and Symlet-SVR models are named W-SVR-1, W-SVR-2, and W-SVR-3, respectively. Similarly, the Haar, db4, and Symlet mother wavelets combined with GPR are called W-GPR-1, W-GPR-2, and W-GPR-3, respectively. The results of SVR, GPR, W-SVR, and W-GPR models are compared in Figure 6 and Table 2. As can be seen, SVR significantly outperforms GPR. RMSE and MAPE of GWL predictions from SVR are 91.3% and 95.4% less than those of GPR (Figure 6(a)). Besides, all three hybrid W-SVR models perform better than their corresponding hybrid W-GPR approaches (Figure 6(b–d)). GWL predictions from all models can capture the continuous decline in GWL observations. W-GPR-1, as the most accurate W-GPR model, provides GWL predictions with lower RMSE and MAPE values (0.29406 (m) and 0.02274%) and higher R^2 and NS values (0.99866 and 0.99554) in comparison with standalone GPR with RMSE, MAPE, R^2 and NS values of 0.55439, 0.04363, 0.99264 and 0.98413, respectively. Besides, W-GPR-1 reduces RMSE and MAPE of standalone GPR by 47.0% and 47.9%, respectively. The statistical metrics indicate that W-SVR-3 yield the most precise GWL predictions with RMSE, MAPE, R^2 and NS values of 0.01290 (m), 0.00079%, 0.99999 and 0.99999, respectively. W-SVR-3 reduces RMSE and MAPE of SVR by 73.1% and 60.3%, respectively. The overall comparison of GPR, SVR, W-GPR, and W-SVR models illustrates that hybrid models have higher capabilities in predicting GWL compared to standalone GPR and SVR models. The findings indicates that Haar and Symlet mother wavelets have higher positive effects on improving error metrics of standalone GPR and SVR models, respectively. Thus, W-GPR-1 and W-SVR-3 may be recommended as the most accurate models for GWL prediction.

Scatterplots of observed and predicted GWL values from different models are shown in Figure 7. As can be seen, the predictions from GPR and W-GPR fall around the 45-degree line (left column in Figure 7), implying that these models can reliably predict GWL. GWL predictions from SVR and W-SVR mostly fall on the 1:1 line, denoting these models can forecast GWL very precisely. Overall, a comparison of scatterplots in the left and

Table 2. Statistical metrics of different approaches.

| Approach | Mother wavelet | Statistical metrics | | | |
|----------|----------------|---------------------|----------|---------|----------|
| | | R^2 | RMSE (m) | NS | MAPE (%) |
| GPR | – | 0.99264 | 0.55439 | 0.98413 | 0.04363 |
| SVR | – | 0.99995 | 0.04790 | 0.99988 | 0.00199 |
| W-GPR-1 | Haar | 0.99866 | 0.29406 | 0.99554 | 0.02274 |
| W-GPR-2 | db4 | 0.98621 | 0.51029 | 0.98619 | 0.03969 |
| W-GPR-3 | Symlet | 0.99051 | 0.47297 | 0.98814 | 0.02738 |
| W-SVR-1 | Haar | 0.99996 | 0.04759 | 0.99988 | 0.00244 |
| W-SVR-2 | db4 | 0.99996 | 0.02861 | 0.99996 | 0.00133 |
| W-SVR-3 | Symlet | 0.99999 | 0.01290 | 0.99999 | 0.00079 |

right columns shows that SVR-based models are superior to GPR-based ones. Besides, among the three tested mother wavelets, Haar and Symlet have the best efficacy when combined with GPR and SVR, respectively. Comparing the results of GPR and SVR (first row) with those of W-GPR and W-SVR (second–fourth rows) illustrates that WT, as a data preprocessing technique, is able to map the relationship between GWL and its influential variables more robustly. This happens because WT decomposes complex time series by catching information on different levels, which results in a more effective structure for machine learning methods (Adamowski & Sun, 2010; Nourani et al., 2009; Zhou et al., 2020). In fact, the decomposed input variables give GPR and SVR models more flexibility in simulating the relationship between input and output variables. As a result, WT shows a significant contribution in explaining the hidden relationship between meteorological variables and GWL in Semnan plain.

Taylor diagram is designed to graphically indicate which of several prediction models are more realistic. This diagram is based on the geometric linkage between the correlation coefficient and standard deviation. It can be presented in two forms: semicircle (showing negative and positive correlation) and quadrant (indicating positive correlation only). In both forms, the values of the correlation coefficients are expressed on the arc of the circle. The graph evaluation method is tailored such that any model whose location is in the graph closer to the reference point is more accurate and therefore more appropriate (Taylor, 2001).

The impact of different mother wavelets on the accuracy of groundwater level predictions has not been fully investigated in the literature. The Taylor diagram in Figure 8 shows a comparison of all models used in this research. The results indicate that applying different mother wavelets, namely Haar, db4, Symlet, for developing hybrid W-SVR and W-GPR models can provide different performances. The green point (observed on the horizontal axis) indicates the position of the reference point according to standard deviation. The pink and orange points (W-SVR-3 and W-SVR-2) are closer to the

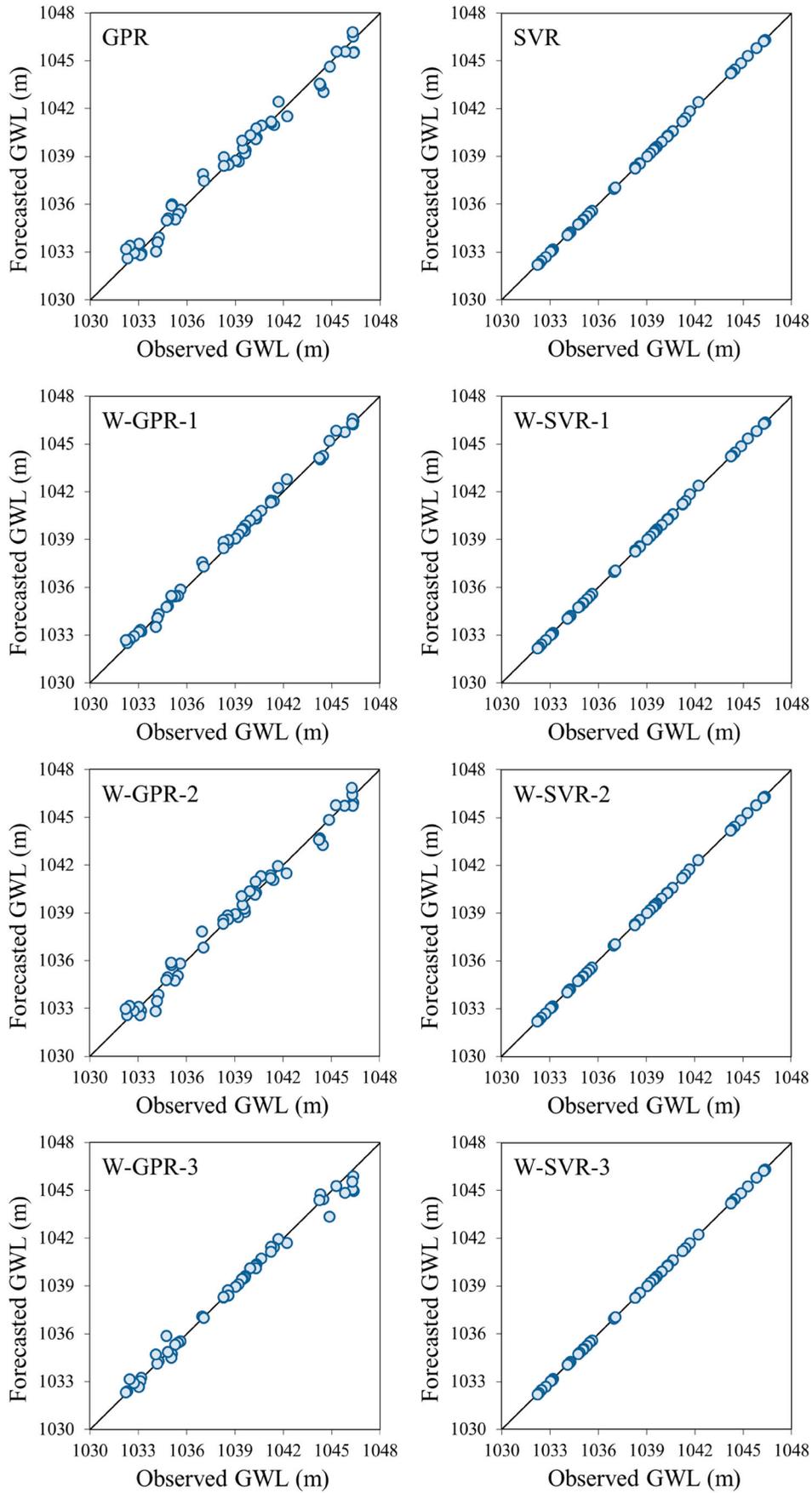


Figure 7. Scatterplots of monthly observed and predicted values from GPR and SVR (1st row), (b) W-GPR-1 and W-SVR-1 (2nd row), W-GPR-2 and W-SVR-2 (3rd row), and W-GPR-3 and W-SVR-3 (4th row) during 2012–2017.

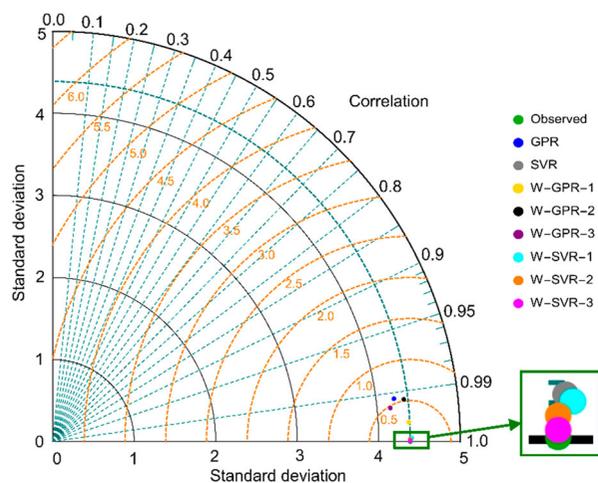


Figure 8. Taylor diagram for the utilized models.

reference point and provide better results. The blue point (GPR) falls outside the semicircle and is far from the reference point. However, by adding wavelet analysis to this model, the results are improved and the yellow point (W-GPR-1) is closer to the observed point. Yet, at the end, the best model is W-SVR-3 with the Symlet mother wavelet. These results show that the effectiveness of different mother wavelets should be explored for predicting GWL. These findings may be used by local authorities for monitoring and managing groundwater level decline of Semnan plain.

Conclusion

Groundwater level (GWL) prediction is of vital importance for residential, agricultural and industrial sectors, especially in arid regions. In this study, support vector regression (SVR), Gaussian process regression (GPR), hybrid wavelet-SVR (W-SVR), and hybrid wavelet-GPR (W-GPR) methods with three different mother wavelets (i.e. Haar, db4, Symlet) are used to predict GWL in Semnan plain, which is an arid region. Performance of these approaches is compared based on the root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and Nah-Sutcliffe efficiency (NS) metrics. Minimum air temperature, minimum relative humidity, precipitation and groundwater level of three preceding and current months are employed as inputs to predict the one-month ahead GWL. All approaches can reliably capture the continuous decline in GWL during the testing period (2012–2017). However, SVR can predict GWL more accurately than GPR. RMSE and MAPE of GWL predictions from SVR are respectively 0.04790 (m) and 0.00199%, which are 91.3% and 95.4% less than those of GPR. Besides, adding any of the three mother wavelets to GPR and SVR improves

their performance. It is found that Harr and Symlet mother wavelets are the most effective ones when combined with GRP and SVR, respectively. The hybrid W-SVR-3 model (that uses Symlet mother wavelet) reduces RMSE and MAPE of GWL predictions by 73.1% and 60.3% compared to SVR. Similarly, the hybrid W-GPR-1 model (that used Harr wavelet) decreases RMSE and MAPE of predicted GWL values by 46.9% and 47.9% relative to GPR. Overall, the results show that both GPR and SVR and specifically their combination with the wavelet analysis can accurately predict GWL, ultimately allowing water resources planners to create effective measures for conserving groundwater resources in Semnan plain.

Acknowledgements

Open Access Funding by the Publication Fund of the TU Dresden.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Shahab S. Band <http://orcid.org/0000-0001-6109-1311>

Essam Heggy <http://orcid.org/0000-0001-7476-2735>

Sayed M. Bateni <http://orcid.org/0000-0002-7134-0067>

Hojat Karami <http://orcid.org/0000-0002-2017-7204>

Saeed Samadianfard <http://orcid.org/0000-0002-6876-7182>

Kwok-Wing Chau <http://orcid.org/0000-0001-6457-161X>

Amir Mosavi <http://orcid.org/0000-0003-4842-0613>

References

- Adamowski, J. F. (2008). Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology*, 353(3), 247–266. <https://doi.org/10.1016/j.jhydrol.2008.02.013>
- Adamowski, J., & Chan, H. F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology*, 407(1–4), 28–40. <https://doi.org/10.1016/j.jhydrol.2011.06.013>
- Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390(1–2), 85–91. <https://doi.org/10.1016/j.jhydrol.2010.06.033>
- Adeleke, O. O., Makinde, V., Eruola, A. O., Dada, O. F., Ojo, A. O., & Aluko, T. J. (2015). Estimation of groundwater recharges using empirical formulae in Odeda Local Government Area, Ogun State, Nigeria. *Challenges*, 6(2), 271–281. <https://doi.org/10.3390/challe6020271>
- Alcalá García, F. J., Martínez Pagan, P., Catarina Paz, M., Navarro Bernal, M., Pérez Cuevas, J., Domingo Poveda, F. (2021). Combining of MASW and GPR imaging and hydro-geological surveys for the groundwater resource evaluation

- in a coastal urban area in southern Spain. *Applied Sciences*, 11(7), 3154. <https://doi.org/10.3390/app11073154>
- Anderson, M. P., & Woessner, W. W. (1992). *Applied groundwater modeling: Simulation of flow and advective transport*. Gulf Professional Publishing.
- Banadkooki, F. B., Ehteram, M., Ahmed, A. N., Teo, F. Y., Fai, C. M., Afan, H. A., & El-Shafie, A. (2020). Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. *Natural Resources Research*, 29(5), 3233–3252. <https://doi.org/10.1007/s11053-020-09634-2>
- Banadkooki, F. B., Ehteram, M., Ahmed, A. N., Teo, F. Y., Fai, C. M., Afan, H. A., Sapitang, M., & El-Shafie, A. (2020). Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. *Natural Resources Research*, 29, 3233–3252. <https://doi.org/10.1007/s11053-020-09634-2>
- Campozano, L., Mendoza, D., Mosquera, G., Palacio-Baus, K., Céleri, R., & Crespo, P. (2020). Wavelet analyses of neural networks based river discharge decomposition. *Journal of Hydrological Processes*, 34(11), 2302–2312. <https://doi.org/10.1002/hyp.13726>
- Chang, J., Wang, G., & Mao, T. (2015). Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. *Journal of Hydrology*, 529, 1211–1220. <https://doi.org/10.1016/j.jhydrol.2015.09.038>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Emamgholizadeh, S., Moslemi, K. H., & Karami, H. (2014). Prediction the groundwater level of Bastam Plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). *Water Resources Management*, 28(15), 5433–5446. <https://doi.org/10.1007/s11269-014-0810-0>
- Esmailzadeh, B., Sattari, M. T., & Samadianfard, S. (2017). Performance evaluation of ANNs and an M5 model tree in Sattarkhan Reservoir inflow prediction. *ISH Journal of Hydraulic Engineering*, 23(3), 283–292. <https://doi.org/10.1080/09715010.2017.1308277>
- Faizollahzadeh Ardabili, S., Najafi, B., Shamshirband, S., Bidgoli, B. M., Deo, R. C., Chau, K.-W. (2018). Computational intelligence approach for modeling hydrogen production: A review. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 438–451. <https://doi.org/10.1080/19942060.2018.1452296>
- Gholami, V. C. K. W., Chau, K. W., Fadaee, F., Torkaman, J., & Ghaffari, A. (2015). Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers. *Journal of Hydrology*, 529, 1060–1069. <https://doi.org/10.1016/j.jhydrol.2015.09.028>
- Ghose, D., Das, U., & Roy, P. (2018). Modeling response of runoff and evapotranspiration for predicting water table depth in arid region using dynamic recurrent neural network. *Groundwater for Sustainable Development*, 6, 263–269. <https://doi.org/10.1016/j.gsd.2018.01.007>
- Graf, R., Zhu, S., & Sivakumar, B. (2019). Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *Journal of Hydrology*, 578. <https://doi.org/10.1016/j.jhydrol.2019.124115>
- Grbic, R., Kurtagić, D., & Slišković, D. (2013). Stream water temperature prediction based on Gaussian process regression. *Expert Systems with Applications*, 40(18), 7404–7414. <https://doi.org/10.1016/j.eswa.2013.06.077>
- Hipni, A., El-shafie, A., Najah, A., Othman, A. K., Husain, A., & Mukhlisin, M. (2013). Daily forecasting of dam water levels: Comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). *Water Resources Management*, 27(10), 3803–3823. <https://doi.org/10.1007/s11269-013-0382-4>
- Isazadeh, M., Biazar, S. M., & Ashrafzadeh, A. (2017). Support vector machines and feed-forward neural networks for spatial modeling of groundwater qualitative parameters. *Environmental Earth Sciences*, 76(17), 610. <https://doi.org/10.1007/s12665-017-6938-5>
- Karami, H., Karimi, S., Bonakdari, H., & Shamshirband, S. (2018). Predicting discharge coefficient of triangular labyrinth weir using extreme learning machine, artificial neural network and genetic programming. *Neural Computing and Applications*, 29(11), 983–989. <https://doi.org/10.1007/s00521-016-2588-x>
- Kargar, K., Samadianfard, S., Parsa, J., Nabipour, N., Shamshirband, S., Mosavi, A., & Chau, K. W. (2020). Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 311–322. <https://doi.org/10.1080/19942060.2020.1712260>
- Karthika, B. S., & Deka, P. C. (2015). Prediction of air temperature by hybridized model (wavelet-ANFIS) using wavelet decomposed data. *Aquatic Procedia*, 4, 1155–1161. <https://doi.org/10.1016/j.aqpro.2015.02.147>
- Khoshand, A. (2021). Application of artificial intelligence in groundwater ecosystem protection: A case study of Semnan/Sorkheh plain, Iran. *Environment, Development and Sustainability*, 1–15. <https://doi.org/10.1007/s10668-021-01361-9>
- Kisi, O., & Cimen, M. (2011). A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*, 399(1), 132–140. <https://doi.org/10.1016/j.jhydrol.2010.12.041>
- Lal, A., & Datta, B. (2018). Genetic programming and gaussian process regression models for groundwater salinity prediction: Machine learning for sustainable water resources management. *IEEE Conference on Technologies for Sustainability (SusTech)*, 1–7.
- Lan, Y. (2014). Forecasting performance of support vector machine for the Poyang Lake's water level. *Water Science & Technology*, 70(9), 1488–1495. <https://doi.org/10.2166/wst.2014.396>
- Liu, F., & West, M. (2009). A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, 4(2), 393–411. <https://doi.org/10.1214/09-BA415>
- Mirarabi, A., Nassery, H. R., Nakhaei, M., Adamowski, J., Akbarzadeh, A. H., & Alijani, F. (2019). Evaluation of data-driven models (SVR and ANN) for groundwater-level prediction in confined and unconfined systems. *Environmental Earth Sciences*, 78(15), 489. <https://doi.org/10.1007/s12665-019-8474-y>
- Moosavi, V., Vafakhah, M., Shirmohammadi, B., & Behnia, N. (2013). A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. *Water Resources Management*, 27(5), 1301–1321. <https://doi.org/10.1007/s11269-012-0239-2>

- Nayak, P., SatyajiRao, Y., & Sudheer, K. (2006). Groundwater level forecasting in a shallow aquifer using artificial neural network. *Water Resources Management*, 20(1), 77–90. <https://doi.org/10.1007/s11269-006-4007-z>
- Niu, W. J., & Feng, Z. K. (2021). Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society*, 64. <https://doi.org/10.1016/j.scs.2020.102562>
- Nourani, V., Alami, M. T., & Aminfar, M. H. (2009). A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence*, 22(3), 466–472. <https://doi.org/10.1016/j.engappai.2008.09.003>
- O'Hagan, A. (2013). Polynomial chaos: A tutorial and critique from a statistician's perspective. *SIAM/ASA Journal on Uncertainty Quantification*, 20, 1–20. http://courses.washington.edu/me333afe/Polynomial-chaos_OHagan.pdf
- Porte, P., Isaac, R. K., Mahilang, K. K. S., Sonboier, K., & Minj, P. (2018). Groundwater level prediction using artificial neural network model. *International Journal of Current Microbiology and Applied Sciences*, 7(2), 2947–2954. <https://doi.org/10.20546/ijcmas.2018.702.358>
- Qasem, S. N., Samadianfard, S., Sadri Nahand, H., Mosavi, A., Shamshirband, S., & Chau, K. W. (2019). Estimating daily dew point temperature using machine learning algorithms. *Water*, 11(3), 582. <https://doi.org/10.3390/w11030582>
- Rasmussen, C. E., & Nickisch, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *The Journal of Machine Learning Research*, 11, 3011–3015. https://www.jmlr.org/papers/volume11/rasmussen10a/rasmussen10a.pdf?_hstc=200028081.1bb630f9cde2cb5f07430159d50a3c91.1524182400081.1524182400082.1524182400083.1_hssc=200028081.1524182400084_hsfp=1773666937
- Reis, D. S., Stedinger, J. R., & Martins, E. S. (2005). Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation. *Water Resource Research*, 41(10), W10419. <https://doi.org/10.1029/2004WR003445>
- Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of Hydrology*, 553, 356–373. <https://doi.org/10.1016/j.jhydrol.2017.08.006>
- Samadianfard, S., Asadi, E., Jarhan, S., Kazemi, H., Kheshtgar, S., Kisi, O., Sajjadi, S., & Abdul Manaf, A. (2018). Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. *Soil Tillage Research*, 175, 37–50. <https://doi.org/10.1016/j.still.2017.08.012>
- Samadianfard, S., Jarhan, S., Salwana, E., Mosavi, A., Shamshirband, S., & Akib, S. (2019a). Support vector regression integrated with fruit fly optimization algorithm for river flow forecasting in Lake Urmia Basin. *Water*, 11(9), 1934. <https://doi.org/10.3390/w11091934>
- Samadianfard, S., Majnooni-Heris, A., Qasem, S. N., Kisi, O., Shamshirband, S., & Chau, K. W. (2019b). Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate. *Engineering Applications of Computational Fluid Mechanics*, 13(1), 142–157. <https://doi.org/10.1080/19942060.2018.1560364>
- Selesnick, I. W., Baraniuk, R. G., Kingsbury, N. C. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6), 123–151. <https://doi.org/10.1109/MSP.2005.1550194>
- Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, 11(66), 1–17. <https://doi.org/10.3390/atmos11010066>
- Shafaei, M., & Kisi, O. (2015). Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. *Water Resources Management*, 30(1), 79–97. <https://doi.org/10.1007/s11269-015-1147-z>
- Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi, A., Nabipour, N., & Chau, K. W. (2020). Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 339–350. <https://doi.org/10.1080/19942060.2020.1715844>
- Shirmohammadi, B., Moradi, H., Moosavi, V., Semiromi, M. T., & Zeinali, A. (2013b). Forecasting of meteorological drought using Wavelet ANFIS hybrid model for different time steps (case study: Southeastern part of east Azerbaijan province, Iran). *Natural Hazards*, 69(1), 389–402. <https://doi.org/10.1007/s11069-013-0716-9>
- Shirmohammadi, B., Vafakhah, M., Moosavi, V., & Moghadamnia, A. (2013a). Application of several data-driven techniques for predicting groundwater level. *Water Resources Management*, 27(2), 419–432. <https://doi.org/10.1007/s11269-012-0194-y>
- Starck, J.-K., & Murtagh, F. (2001). Astronomical image and signal processing: Looking at noise, information and scale. *IEEE Signal Processing Magazine*, 18(2), 30–40. <https://doi.org/10.1109/79.916319>
- Stedinger, J. R., & Tasker, G. D. (1985). Regional hydrologic analysis: 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, 21(9), 1421–1432. <https://doi.org/10.1029/WR021i009p01421>
- Su, Z., Wu, J., He, X., & Elumalai, V. (2020). Temporal changes of groundwater quality within the groundwater depression cone and prediction of confined groundwater salinity using Grey Markov model in Yinchuan area of northwest China. *Exposure and Health*, 12(3), 447–468. <https://doi.org/10.1007/s12403-020-00355-8>
- Sun, A. Y., Wang, D., & Xu, X. (2013). Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology*, 511, 72–81. <https://doi.org/10.1016/j.jhydrol.2014.01.023>
- Suryanarayana, C., Sudheer, C., Mahmood, V., & Panigrahi, B. K. (2014). An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. *Neurocomputing*, 145, 324–335. <https://doi.org/10.1016/j.neucom.2014.05.026>
- Taormina, R., Chau, K.-W., & Sethi, R. (2012). Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Engineering Applications of Artificial Intelligence*, 25(8), 1670–1676. <https://doi.org/10.1016/j.engappai.2012.02.009>
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical*

- Research: Atmospheres*, 106(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Tikhamarine, Y., Souag-Gamane, D., & Kisi, O. (2019). A new intelligent method for monthly streamflow prediction: Hybrid wavelet support vector regression based on grey wolf optimizer (WSVR–GWO). *Arabian Journal of Geosciences*, 12(17), 540. <https://doi.org/10.1007/s12517-019-4697-1>
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag.
- Vapnik, V. (1999). *The nature of statistical theory*. Springer Science & Business Media.
- Vogel, R. M., Wilson, I., & Daly, C. (1999). Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, 125(3), 148–157. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1999\)125:3\(148\)](https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148))
- Wang, B., & Chen, T. (2015). Gaussian process regression with multiple response variables. *Chemometrics and Intelligent Laboratory Systems*, 142, 159–165. <https://doi.org/10.1016/j.chemolab.2015.01.016>
- Yu, H., Wen, X., Feng, Q., Deo, C., Si, R., & Wu, M. (2018). Comparative study of hybrid wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, northwest China. *Water Resources Management*, 32(1), 301–323. <https://doi.org/10.1007/s11269-017-1811-6>
- Zhang, B.-L., & Dong, Z.-Y. (2001). An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research*, 59(2), 121–129. [https://doi.org/10.1016/S0378-7796\(01\)00138-9](https://doi.org/10.1016/S0378-7796(01)00138-9)
- Zheng, T., Girgis, A. A., & Makram, E. B. (2000). Hybrid wavelet-Kalman filter method for load forecasting. *Electric Power Systems Research*, 54(1), 11–17. [https://doi.org/10.1016/S0378-7796\(99\)00063-2](https://doi.org/10.1016/S0378-7796(99)00063-2)
- Zhou, F., Liu, B., & Duan, K. (2020). Coupling wavelet transform and artificial neural network for forecasting estuarine salinity. *Journal of Hydrology*, 588, 125127. <https://doi.org/10.1016/j.jhydrol.2020.125127>
- Zhu, S., Ptak, M., MundherYaseen, Z., Dai, J., & Sivakumar, B. (2020). Forecasting surface water temperature in lakes: A comparison of approaches. *Journal of Hydrology*, 585. <https://doi.org/10.1016/j.jhydrol.2020.124809>
- Zounemat-Kermani, M., Kişi, O., Adamowski, J., & Ramezani Charmahineh, A. (2016). Evaluation of data driven models for river suspended sediment concentration modeling. *Journal of Hydrology*, 535, 457–472. <https://doi.org/10.1016/j.jhydrol.2016.02.012>