Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression

Sujay Raghavendra. N and Paresh Chandra Deka

Cogent Engineering (2015), 2: 999414
CIVIL & ENVIRONMENTAL ENGINEERING | RESEARCH ARTICLE

Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression

Sujay Raghavendra. N* and Paresh Chandra Deka1

Abstract: This research demonstrates the state-of-the-art capability of Wavelet packet analysis in improving the forecasting efficiency of Support vector regression (SVR) through the development of a novel hybrid Wavelet packet–Support vector regression (WP–SVR) model for forecasting monthly groundwater level fluctuations observed in three shallow unconfined coastal aquifers. The Sequential Minimal Optimization Algorithm-based SVR model is also employed for comparative study with WP–SVR model. The input variables used for modeling were monthly time series of total rainfall, average temperature, mean tide level, and past groundwater level observations recorded during the period 1996–2006 at three observation wells located near Mangalore, India. The Radial Basis function is employed as a kernel function during SVR modeling. Model parameters are calibrated using the first seven years of data, and the remaining three years data are used for model validation using various input combinations. The performance of both the SVR and WP–SVR models is assessed using different statistical indices. From the comparative result analysis of the developed models, it can be seen that WP–SVR model outperforms the classic SVR model in predicting groundwater levels at all the three well locations (e.g. NRMSE_{WP–SVR} = 7.14, NRMSE_{SVR} = 12.27; NSE_{WP–SVR} = 0.91, NSE_{SVR} = 0.8 during the test phase with respect to

ABOUT THE AUTHOR
Paresh Chandra Deka, currently an associate professor in the Department of Applied Mechanics & Hydraulics, National Institute of Technology Karnataka, has over 20 years of experience in teaching undergraduate and postgraduate students. He earned his PhD from the prestigious IIT Guwahati. His current research interests are Hydrological Modeling, applications of Artificial Neural Network, Fuzzy logic, Genetic algorithm, Wavelet transformation, and Support Vector Machine in water resources system modeling. He has published over 21 journal articles and numerous national and international conference papers. He has successfully guided three PhD scholars and presently, seven research scholars are working under his esteemed guidance. He is a member of several profession bodies like Indian Society for Hydraulics, International Association of Hydraulic Research, and Indian Society for Technical Education, etc. He was visiting faculty at AIT Bangkok during the year 2012.

PUBLIC INTEREST STATEMENT
In general, Groundwater systems are characterized by non-stationary and nonlinear features. Modeling of these systems and predicting their future scenarios requires identification and capture of the rudimentary characteristics of the variables affecting them. The prolonged drought, population growth and emerging climate change impacts have increased demand for groundwater and sustainable management of groundwater resources is imperative to the agricultural, industrial, urban, rural and environmental viability. Such management requires not only a robust monitoring of groundwater levels, but also sufficient knowledge on future groundwater level trends. Hence a novel hybrid Wavelet packet–Support vector regression (WP–SVR) technique is proposed with a main objective to develop a reliable monthly groundwater level fluctuation forecasting system.
well location at Surathkal). Therefore, using the WP–SVR model is highly acceptable for modeling and forecasting of groundwater level fluctuations.

**Subjects:** Computation; Data Preparation & Mining; Civil, Environmental and Geotechnical Engineering; Water Engineering; Water Science

**Keywords:** groundwater systems; support vector machines; Wavelet packets; radial basis function (RBF); Wavelet packet–Support vector regression

1. **Introduction**

Groundwater is one of the major sources of supply for domestic, industrial, and agricultural purposes. In some areas, groundwater is the only dependable source of supply, while in some other regions, it is chosen because of its ready availability. The shallow water table depths have significant impacts on crop growth, vegetation development, and contaminant transport. Groundwater has many advantages over surface water for water supply. It is reliable in dry seasons or droughts because of the large storage (Hiscock, Rivett, & Davison, 2002). The consequences of aquifer depletion can lead to local water rationing, excessive reductions in yields, wells going dry or producing erratic ground water quality changes, changes in flow patterns of groundwater. Normal groundwater recharge to creeks and streams during low-flow periods could result in reduced supplies for surface water sources. To ensure the ecological sustainability of a watershed, the management of groundwater resources in conjunction with surface waters is very much substantial. So this necessitates the constant monitoring of the groundwater levels (Batu, 1998).

Assessment of groundwater levels from observation wells provides a principal source of information regarding the hydrological stresses acting over aquifers and how those stresses influence over groundwater recharge, storage, and discharge. It also allows water managers, engineers, and policy-makers to develop better strategies of groundwater management and protection. A better conception of the groundwater dynamics and underlying factors that affect groundwater levels paves way to investigate the requirement of agricultural, industrial, urban, and other demands and assess the benefits and costs of water conservation. The water levels if forecasted well in advance may help the administrators to plan for better utilization of groundwater resources (Healy & Cook, 2002).

In the past, several models have been proposed for forecasting groundwater levels using various stochastic, analytical, and soft computing techniques. In the recent times, the application of Artificial neural networks (ANN) and its hybrid models are found to be efficient in forecasting groundwater levels at different time scales (Chitsazan, Rahmani, & Neyamadpour, 2013; Daliakopoulos, Coulilbaly, & Tsanis, 2005; Jalalkamali, Sedghi, & Manshouri, 2011; Nayak, Rao, & Sudheer, 2006; Nourani, Mogaddam, & Nadiri, 2008; Shi & Zhu, 2009; Sreekanth et al., 2009; Sreenivasulu, Deka, & Nagaraj, 2012; Wanakule & Aly, 2005). ANN, due to its “black box” nature, immense computational burden, prone to overfitting, and the empirical nature of model development somehow paved way for hydrological modeling using Support Vector Machines (SVMs). The current standard SVM algorithm based on statistical learning theory proposed by Cortes and Vapnik (1995) is an approximation implementation of the method of structural risk minimization with a good generalization capability. SVMs have also been successfully applied in groundwater level estimation studies. Liu, Chang, and Zhang (2009) proposed least squares-support vector machine (LS-SVM) arithmetic based on chaos optimization for dynamic prediction of groundwater levels, taking into account of groundwater level dynamic series length and peak mutation characters. Their simulation results revealed that the developed LS-SVM model was very effective in reflecting the dynamic evolution process of groundwater levels in a well. Behzad, Asghari, and Coppola (2010) made an attempt with SVMs and ANNs for predicting transient groundwater levels in a complex groundwater system under variable pumping and weather conditions. They predicted water-level elevations at different time horizons i.e. daily, weekly, biweekly, monthly, and bimonthly. Here, particularly for longer prediction horizons, SVM outperformed ANN even when fewer data events were available for model development. Yoon, Jun, Hyun, Bae, and Lee (2011) demonstrated the efficacy of SVM model
in predicting short-term groundwater level fluctuations in a coastal aquifer giving significance to recharge from precipitation and tidal effect. The dilemma of salt water intrusion into coastal aquifer is accounted by considering the tidal effect. Sudheer, Shrivastava, Panigrahi, and Mathur (2011) proposed a hybrid Quantum behaved Particle Swarm Optimization (QPSO)-based SVM model for estimating the groundwater levels. QPSO function was adopted in their study in order to determine the optimal SVM parameters. Tapak, Rahmani, and Moghimbeigi (2014) developed a time series prediction model based on SVM using monthly peizometric groundwater level data. Their proposed SVM model outperformed the classic time series model in predicting groundwater levels. Raghavendra and Deka (2014a) investigated the potential of Sequential Minimal Optimization Algorithm-based Support Vector Regression (SMO-SVR) model for forecasting monthly groundwater level fluctuations observed in two shallow unconfined coastal aquifers. The relative performance of SMO-SVR models developed using three different kernel functions—Polynomial, RBF, and Pearson VII function-based universal kernel (PuK) was comparatively evaluated.

Discrete data pose a problem in SVR analysis. Even after the optimal choice of kernel and regularization parameter means, it would end up with all data being support vectors. The loss function of SVR doesn’t possess clear statistical interpretation; henceforth, expert knowledge of the problem cannot be efficiently constructed. For this, the representation and quality of data are of foremost importance. The knowledge discovery during training phase is more difficult if there is much of noisy and unreliable data. Consequently, this provides way for data pre-processing before using it in any regression models. This study investigates the applicability of the Wavelet packet transform (WPT) as a data pre-processing technique and is hybridized with SVR to forecast groundwater level fluctuations in coastal aquifers.

Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. Wavelet packet analysis can be used to denoise the highly non-linear, non-stationary time series (such as groundwater levels, rainfall, tidal levels, temperature) so that the new denoised time series can be used as inputs for an SVR model. Wavelet transformation has numerous applications in the field of hydrology and water resource engineering such as rainfall run-off modeling (Nourani, Kisi, & Komasi, 2011), forecasting of precipitation (Kisi & Shiri, 2011), daily pan evaporation prediction (Abghari, Ahmadi, Besharat, & Rezavehrinejad, 2012), urban water demand forecasting (Campisi-Pinto, Adamowski, & Oron, 2012). Wavelet packet-based SVM applications can be witnessed in various other domains of engineering (Hu, Zhu, & Ren, 2008; Manimala, Selvi, & Ahila, 2012; Subramanian & Henry, 2010; Tong, Song, Lin, & Zhao, 2006). Adamowski and Chan (2011) proposed a novel method of coupling discrete wavelet transforms (WA) and artificial neural networks (ANN) for groundwater level forecasting. Moosavi, Vafakhah, Shirmohammadi, and Behnia (2013) developed a Wavelet-ANFIS hybrid model for groundwater level forecasting with different prediction times. Ping, Qiang, and Xixia (2013) developed a combined model of chaos theory, wavelet and SVM to overcome the limitations, including challenges with determination of orders of non-linear models and low prediction accuracy in groundwater level forecasting. Suryanarayana, Sudheer, Mahmood, and Panigrahi (2014) emphasizes the application of integrated Wavelet analysis–Support vector regression (WA–SVR) model for predicting the groundwater level variations in three observation wells. They comparatively evaluate the WA–SVR model with ANN and Auto Regressive Integrated Moving Average models and found the superior performance of the WA–SVR model in predicting groundwater levels. Only few research works are carried out in groundwater level forecasting using simple discrete Wavelet analysis hybridized with SVM technique. No research has been published yet which explores, coupling Wavelet packet analysis with SVM for groundwater level forecasting with multiple hydrological input parameters. Hence, a novel hybrid Wavelet packet–Support vector regression (WP–SVR) technique is proposed with a main objective to develop a reliable monthly groundwater level fluctuation forecasting system. The performance of SVR models with multiple hydrological input scenarios is tested and the combination yielding higher efficiency is adopted in hybrid WP–SVR modeling. A comparative analysis of results obtained by WP–SVR and basic SVR is also presented.
2. Theory of Support vector regression

Consider a simple linear regression problem trained on data-set \( \mathbf{X} = \{ \mathbf{u}_i, \mathbf{v}_i; i = 1, \ldots, n \} \) with input vectors \( \mathbf{u}_i \) and linked targets \( \mathbf{v}_i \). In order to associate the inherited relations between the data-sets, a function \( g(\mathbf{u}) \) has to be formulated approximately, and thereby it can be used in future to infer the output “\( \mathbf{v} \)" for a new input data “\( \mathbf{u} \).”

Standard SVM regression employs a loss function \( L_i(\mathbf{v}, g(\mathbf{u})) \) which explains the deviation of the estimated function from the original one. Various forms of loss functions, namely, linear, quadratic, exponential, Huber’s loss function, etc. can be mined in the literature. The standard Vapnik’s \( \epsilon \) insensitive loss function is used in this context which is defined as:

\[
L_i(\mathbf{v}, g(\mathbf{u})) = \begin{cases} 
0 & \text{for } |\mathbf{v}_i - g(\mathbf{u}_i)| \leq \epsilon \\
|\mathbf{v}_i - g(\mathbf{u}_i)| - \epsilon & \text{otherwise} 
\end{cases}
\]  

(1)

The \( \epsilon \)-insensitive loss function, aid in determining \( g(\mathbf{u}) \) which can better approximate the actual output vector “\( \mathbf{v} \).” It also provides at most error tolerance “\( \epsilon \)” from the actual incurred targets “\( \mathbf{v}_i \)” for all training data as flat as possible. Consider the regression function defined by

\[
g(\mathbf{u}) = \mathbf{w} \cdot \mathbf{u} + \mathbf{b}
\]

(2)

where \( \mathbf{w} \in \mathbf{X} \), \( \mathbf{X} \) is the input space; \( \mathbf{b} \in \mathbb{R} \) is a bias term and \( (\mathbf{w} \cdot \mathbf{u}) \) is dot product of vectors \( \mathbf{w} \) and \( \mathbf{u} \). Flatness in Equation 2 refers to a smaller value of parameter vector \( \mathbf{w} \). By minimizing the norm \( ||\mathbf{w}||^2 \), flatness can be ascertained along with model complexity. Thus, regression problem can be stated as the following convex optimization problem:

\[
\min_{\mathbf{w}, \mathbf{b}, \xi, \xi^*} \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\
\text{subject to } (\mathbf{w} \cdot \mathbf{u}_i + \mathbf{b}) - \mathbf{v}_i \leq \epsilon + \xi_i \\
\xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \ldots, n
\]

(3)

where \( \xi_i \) and \( \xi_i^* \) are slack variables introduced to evaluate the deviation of training samples outside \( \epsilon \)-insensitive zone. The trade-off between the flatness of \( g \) and the quantity up to which deviations greater than \( \epsilon \) are tolerated is depicted by \( C > 0 \). Whenever a training error occurs, \( C \) is a positive constant influencing the degree of penalizing. Underfitting and overfitting of training data can be prevented by minimizing the regularization term \( \frac{1}{2} ||\mathbf{w}||^2 \) along with the training error term \( C \sum_{i=1}^{n} (\xi_i - \xi_i^*) \) in Equation 3. The minimization problem in Equation 3 represents the primal objective function. A dual set of variables, \( \alpha_i \) and \( \alpha_i^* \), is introduced for the corresponding constraints and the problem is dealt by constructing a Lagrange function from the primal objective function. Optimality conditions are utilized at the saddle points of a Lagrange function, steering to the formulation of the dual optimization problem:

\[
\max_{\gamma, \pi} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \pi_i)(\alpha_j - \pi_j)(\mathbf{u}_i \cdot \mathbf{u}_j) - \epsilon \sum_{i=1}^{n} (\pi_i + \pi_i^*) + \sum_{i=1}^{n} \gamma_i (\alpha_i - \pi_i) \\
\pi_i, \pi_i^* \geq 0, \quad i = 1, 2, \ldots, n \quad \alpha_i, \pi_i \geq 0, \quad i = 1, 2, \ldots, n
\]

(4)

After determining Lagrange multipliers, \( \alpha_i \) and \( \alpha_i^* \), the parameter vectors \( \mathbf{w} \) and \( \mathbf{b} \) can be evaluated under Karush–Kuhn–Tucker (KKT) complementarity conditions (Fletcher, 2000) which are not discussed here. Therefore, the prediction is a linear regression function that can be expressed as:

\[
g(\mathbf{u}) = \sum_{i=1}^{n} (\alpha_i - \pi_i)(\mathbf{u}_i \cdot \mathbf{u}) + \mathbf{b}
\]

(5)
Thus, SVM regression expansion is derived; where “w” is represented as a linear combination of the training patterns “v_i” and “b” can be found using primary constraints. For \(|g(u_i)| \geq \varepsilon\), Lagrange multipliers may be non-zero for all the samples inside the \(\varepsilon\)-tube and these remaining coefficients are labeled as support vectors.

In order to craft SVM regression to deal with non-linear cases, pre-processing of training patterns “u_i” has to done by mapping the input space \(\chi\) into some feature space \(\mathfrak{F}\) using non-linear hypothesis function \(\varphi = \chi \rightarrow \mathfrak{F}\) and then applied to the standard support vector algorithm. Let \(u_i\) be mapped into the feature space by non-linear function \(\varphi(u)\) and hence, the decision function is given by:

\[
g(w, b) = w \cdot \varphi(u) + b
\]  

(6)

This non-linear regression problem can be expressed as the following optimization problem:

\[
\begin{align*}
\min_{w, b, u_i, \xi_i^+, \xi_i^-} & \quad \frac{1}{2} \|w]\|^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-) \\
\text{subject to} & \quad v_i - (w \cdot \varphi(u_i) + b) \leq \varepsilon + \xi_i^+ \\
& \quad (w \cdot \varphi(u_i) + b) - v_i \leq \varepsilon + \xi_i^- \\
& \quad \xi_i^+ \xi_i^- \geq 0, \quad i = 1, 2, \ldots, n
\end{align*}
\]  

(7)

where \(w\) is the weight vector of coefficients, \(\xi_i^+\) and \(\xi_i^-\) are the upper and lower training bounds of the region where the errors less than \(\varepsilon\) are ignored and \(b\) is a constant. The index \(i\) labels the “n” training cases. The \(v \in \pm 1\) is the class label and \(u_i\) is the independent variable.

Then, the dual form of the non-linear SVR can be expressed as:

\[
\begin{align*}
\max_{\alpha, \alpha_i} & \quad -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i) (\alpha_j - \alpha_j) \left( \varphi(u_i) \cdot \varphi(u_j) \right) - \varepsilon \sum_{i=1}^{n} \alpha_i (u_i, \overline{u_i}) + \sum_{i=1}^{n} v_i (\alpha_i - \overline{\alpha_i}) \\
\text{subject to} & \quad 0 \leq \alpha_i \leq C \quad i = 1, 2, \ldots, n \\
& \quad 0 \leq \overline{\alpha_i} \leq C \quad i = 1, 2, \ldots, n
\end{align*}
\]  

(8)

The “kernel trick” \(K(u_i, u_j) = \langle \varphi(u_i), \varphi(u_j) \rangle\) is used for computations in input space \(\chi\) to fetch the inner products into feature space \(\mathfrak{F}\). Any function satisfying Mercer’s theorem (Vapnik, 1999) can be used as kernels.

Finally, the decision function of non-linear SVR with the allowance of the kernel trick is expressed as follows:

\[
g(u) = \sum_{j=1}^{n} (u_i - \overline{u_i}) K(u_j \cdot u) + b
\]  

(9)

The parameters that impact over the effectiveness of the non-linear SVR are the cost constant \(C\), the radius of the insensitive tube \(\varepsilon\), and the kernel parameters. These parameters are mutually dependent over one another and hence, altering the value of one parameter affects the other linked parameters also (Raghavendra & Deka, 2014b).

Platt (1998) proposed an algorithm called sequential minimal optimization (SMO) for solving the problem of regression with SVMs. SMO breaks the large quadratic programming (QP) problem into a series of smallest possible QP problems. For the evaluation of the decision function, SMO devotes significant time, rather than performing QP. The SMO algorithm has a much naïve formulation and is trouble-free to implement. Shevade, Keerthi, Bhattacharyya, and Murthy (2000) proposed an improvement that enhances the algorithm to perform significantly faster by avoiding the use of threshold “b” in checking KKT conditions.
3. Theory of Wavelet packet transform (WPT)

Wavelets are mathematical functions that slice up data into distinct frequency domains, and then examine each domain with a resolution matched to its scale. They have benefits over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes (Shinde & Hou, 2004).

The WPT is a generalization of the discrete wavelet transform (DWT), and it is convenient to introduce the WPT from the DWT. The DWT is a basis transformation, i.e. it estimates the coordinates of a data vector (spectrum or signal) in the so-called wavelet basis (Coifman & Wickerhauser, 1992). A wavelet is a function that appears like a small wave, a ripple of the baseline, thus its name. The wavelet basis is generated by stretching out the wavelet to fit different scales of the signal and by moving it to cover all parts of the signal. The DWT is said to give a time-scale, or time-frequency, analysis of signals. In wavelet transform, signals split into a detail and an approximation. The approximation obtained from the first level is split into new detail and approximation, and this process is repeated. Because of the fact that DWT decomposes only the approximations of the signal, it may pose problems while applying DWT in certain applications where the valuable information is localized in higher frequency components. The major dissimilarity between DWT and WPT is that WPT splits not only approximations but also details. The peak level of the WPT is the time representation of the signal, whereas, the base level has a better frequency resolution. Thus, with the aid of WPT, a better frequency resolution can be achieved from the decomposed signal. In addition, the use of WPT mines much more features about the signal. Consequently, the Wavelet packet analysis provides better check of frequency resolution and more features about signal than DWT (Lei, Meyer, & Ryan, 1994).

Each component in the Wavelet packet tree can be regarded as a filtered component with a bandwidth of a filter, decreasing with increasing level of decomposition and the whole tree can be considered as a filter bank. At the crown of the tree, the time resolution of the WP components is reasonable, but at the cost of poor frequency resolution, whereas at the tail end with the use of Wavelet packet analysis, the frequency resolution of the decomposed component with high-frequency content can be increased. As a result, the Wavelet packet analysis provides better check of frequency resolution for the decomposition of the signal (Learned & Willsky, 1995). A Wavelet packet is represented as a function, \( \psi_{i,j,k}^n(t) \), where “\( n \)” is the modulation parameter, “\( j \)” is the dilation parameter, and “\( k \)” is the translation parameter.

\[
\psi_{i,j,k}^n(t) = 2^{j/2} \psi^i \left( 2^j \cdot t - k \right)
\]

where \( i = 1, 2, \ldots, j \), and “\( n \)” is the level of decomposition in the Wavelet packet tree. The wavelet is obtained by the following recursive relationships:

\[
\psi_i^2(t) = \frac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} h(k) \cdot \psi_i \left( \frac{t}{2} - k \right)
\]

\[
\psi_i^{2i+1}(t) = \frac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} g(k) \cdot \psi_i \left( \frac{t}{2} - k \right)
\]

where \( \psi_i(t) \) is called a mother wavelet and the discrete filters \( h(k) \) and \( g(k) \) are quadrature mirror filters associated with the scaling function and the mother wavelet function. These two filters, \( h(k) \) and \( g(k) \), are also called group conjugated orthogonal filters.

The Wavelet packet coefficients \( C_{j,k}^i \) corresponding to the signal \( f(t) \) can be obtained by

\[
C_{j,k}^i = \int_{-\infty}^{\infty} f(t) \cdot \psi_{j,k}^i(t) \cdot dt
\]
Provided the wavelet coefficients satisfy the orthogonality condition (Fan & Zuo, 2006). The Wavelet packet component of the signal at a particular node can be obtained as:

\[ f_j(t) = \sum_{k=-\infty}^{\infty} C_{j,k} \cdot \psi_{j,k}(t) \, dt \]  

(14)

After performing Wavelet packet decomposition up to \( j \)th level, the original signal can be represented as a summation of all Wavelet packet components at \( j \)th level as shown in equation:

\[ f(t) = \sum_{j=1}^{2^j} f_j(t) \]  

(15)

In Wavelet packet, standard structure composed of low- and high-pass filters is used for perfect reconstruction (Wickerhauser, 1994). Wavelet packets are particular linear combinations of wavelets which form bases that hold back many of the orthogonality, smoothness, and localization properties of their parent wavelets. The coefficients in the linear combinations are computed by a recursive algorithm, making each newly computed Wavelet packet coefficient sequence, the root of its own analysis tree (Graps, 1995).

4. Study area and data analysis

Dakshina Kannada is a maritime district located in the southwestern part of Karnataka adjoining the Arabian Sea. The geographical area is 4770 km\(^2\) extending from 12°30'00″ to 13°11'00″ north latitudes and 74°35’00″ and 75°33’30″ east longitudes. The observation wells selected for the current research are located in two adjacent micro-watersheds that come under the Pavanje and Gurpura river catchments. The observation well located at Surathkal lies at 12°59’00″ north latitude and 74°48’10″ east longitude, the well location at Mangalore lies at 12°53’23″ north latitude and 74°51’36″ east longitude, and lastly, the well near Ganjimatta is located at 12°59’02″ north latitude and 74°57’15″ east longitude as shown in Figure 1. The study area is dominated by the southwest monsoon (June–September) and non-monsoon period (October–May). The average annual rainfall over the watershed is around 3,500 mm.

In the study area, Lateritic soil is the predominant with highly porous and permeable nature. Due to this lateritic soil property, the infiltration rate is high and the shallow wells give very quick response to rainfall with water table rising immediately, but subsequent drastic draw down is also observed within a short period of time.
Main input to the groundwater in the micro-watershed is from the monsoon rains. The point rainfall observed at the rain gauge stations established at National Institute of Technology Karnataka (NITK) campus, Surathkal; Bajpe and Mangalore (Kavoor) are being utilized in this study. The rainfall data of these stations for the years 1996–2006 have been used in this study. The monthly total rainfall, average temperature, and mean tidal-level data are used to correlate the measurement of groundwater levels in the observation wells on a monthly basis.

The water-level data of the observation wells located at Surathkal, Mangalore, and Ganjimatta for the years 1996–2006 used in the study were obtained from Department of Mines and Geology, Dakshina Kannada Dist. The tidal-level data of the study period are obtained from “Tide Tables” published by the Geological Survey of India.

From the data analysis, the maximum, minimum, mean, standard deviation $S_d$, variance, and correlation values of all the variables influencing groundwater table fluctuation is being tabulated in Tables 1–3. The mean and variance of all the variables do not stay steady over time and this proves that time series data employed to be non-stationary. The predictive relationship between two dependent variables is represented in terms of correlation. The tidal levels of the sea have also exhibited a fair correlation with the monthly groundwater level time series, which needles its use during modeling. In this context, it is seen that rainfall is negatively correlated with groundwater level time series.

| Table 1. Statistical parameters of groundwater level, rainfall, temperature, and tidal level data with respect to well location at Surathkal |
|-------------------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Data-set | Variable | $X_{\text{mean}}$ | $X_{\text{max}}$ | $X_{\text{min}}$ | $S_d$ | Variance | CC with GWT |
| Training | GWT | 8.33 | 11.04 | 3.14 | 1.65 | 2.71 | 1 |
| | Rainfall | 314.70 | 1,523.00 | 0.00 | 427.94 | 183,130.20 | −0.87 |
| | Temperature | 27.68 | 35.12 | 25.15 | 1.54 | 2.36 | 0.45 |
| | Tidal Level | 7.08 | 7.25 | 6.78 | 0.10 | 0.01 | 0.49 |
| Testing | GWT | 8.52 | 10.98 | 5.52 | 1.18 | 1.39 | 1 |
| | Rainfall | 261.50 | 885.20 | 0.00 | 322.61 | 104,074.09 | −0.81 |
| | Temperature | 27.85 | 31.38 | 26.19 | 1.27 | 1.60 | 0.52 |
| | Tidal Level | 7.08 | 7.22 | 6.89 | 0.09 | 0.01 | 0.47 |

| Table 2. Statistical parameters of groundwater level, rainfall, temperature, and tidal level data with respect to well location at Mangalore |
|-------------------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Data-set | Variable | $X_{\text{mean}}$ | $X_{\text{max}}$ | $X_{\text{min}}$ | $S_d$ | Variance | CC with GWT |
| Training | GWT | 10.43 | 12.13 | 6.26 | 1.11 | 1.23 | 1 |
| | Rainfall | 310.28 | 1489.20 | 0.00 | 423.87 | 179,662.76 | −0.59 |
| | Temperature | 27.53 | 31.21 | 23.81 | 1.51 | 2.29 | 0.43 |
| | Tidal Level | 7.08 | 7.25 | 6.78 | 0.10 | 0.01 | 0.41 |
| Testing | GWT | 10.24 | 11.91 | 7.94 | 1.03 | 1.07 | 1 |
| | Rainfall | 296.58 | 1,129.80 | 0.00 | 369.43 | 136,479.98 | −0.68 |
| | Temperature | 27.41 | 30.10 | 25.90 | 1.12 | 1.25 | 0.39 |
| | Tidal Level | 7.08 | 7.22 | 6.89 | 0.09 | 0.01 | 0.35 |
5. Methodology

5.1. Development of basic SVR model

In this study, the SMO-based SVR model is adopted for the prediction of monthly groundwater table using various input combinations. Monthly total rainfall, average temperature, mean tidal levels, and previous groundwater levels are used as input variables. The model is trained utilizing seven years (1996–2003) of data and tested with three years (2003–2006) of data. The SVR models were developed using the WEKA software (Hall et al., 2009). There is no way in advance to know which kernel function will be best for an application; henceforth, the RBF kernel is being adopted for prediction of groundwater levels considering its effective generalization capability. The accuracy of an RBF kernel-based SVR model is principally dependent on the selection of the model parameters such as $C$—regularization parameter, Gamma ($\gamma$)—kernel parameter, and epsilon parameter ($\varepsilon$). WEKA provides two methods for finding optimal parameter values, firstly, a grid search and the other one is cross-validation parameter selection. Grid search tries to find values of each parameter across the specified search range using geometric steps (Cherkassky & Ma, 2004). Usually, grid searches are computationally expensive as the model is evaluated at various points within the grid for each parameter. If cross-validation parameter selection is opted, a V-fold cross-validation is used by the search to calculate the optimal parameters using the error computed from the training data. The input–output combination of the SMO-SVR model is given below:

$\Rightarrow$ GW Level + Rainfall = 1 month lead GW Level forecast
$\Rightarrow$ GW Level + Rainfall + Temperature = 1 month lead GW Level forecast
$\Rightarrow$ GW Level + Rainfall + Temperature + Tidal levels = 1 month lead GW Level forecast

Initially, the time series data of input variables are normalized in SVR model and the data are fitted using RBF kernel function. The best RBF-based SVR model is obtained from a grid search of hyperparameters—($C$, $\gamma$, and $\varepsilon$). The developed SVR model is trained using the optimal values of hyperparameters of RBF kernel. The model is then simulated again with testing data-set and evaluated. The predicted values obtained from the SMO-SVR model is then compared with the known observed values.

5.2. Development of hybrid WP–SVR model

The time series data of all input variables, namely, monthly total rainfall, average temperature, mean tidal levels, and previous groundwater levels are considered for Wavelet packet denoising. These time series signals are highly non-stationary and non-linear. They possess high peaks and sudden downfalls. This affects the predicting capability of an SVR model. Henceforth, the application of denoising of input signals serves as a suitable data pre-processing technique to address the non-linear time series data. All the input signals are denoised using WPT and the resulting Wavelet...
packet coefficients from the denoised signals are used as inputs for the prediction of the target variable. Note, the target variable (i.e. one month lead GW level) is unchanged i.e. original observed time series is retained without any transformation. To perform wavelet packet analysis, haar, db2, db3, and db4 wavelets have been selected as mother Wavelets and various decomposition levels have been evaluated. Figure 2 depicts the flowchart of Wavelet packet analysis. Each input variable is denoised separately using Wavelet packet analysis in order to obtain Wavelet packet coefficients from the denoised signal. MATLAB-Wavelet Design and Analysis Tool (Misiti & Misiti, 1996) was used for Wavelet packet analysis. After obtaining output in the form of coefficients from the discrete Wavelet packet transformation, the next step is to train the SVR model as explained in the above Section 5.1 using WP coefficients of all input variables of the training phase. Once the model is trained, the WP coefficients of test phase are used to simulate the model and thereby tested. Figure 3 represents the methodology of the hybrid WP–SVR model. Wavelet packets serve as a powerful tool for the task of signal denoising. The ability to decompose a signal into different scales is very important for denoising, and it improves the analysis of the signal significantly.

6. Performance evaluation

The statistical indices assess the degree of confidence one can have in the predictions of the model and whether the model displays any bias which could lead to “fail-dangerous” predictions. Models usually involve prediction or measurement of a number of variables. The usual practice in texts is to develop the concept of error propagation in connection with precision (random) errors. To evaluate the performance of SVR and hybrid WP–SVR models, the following statistical indices were adopted.

(1) Normalized Root Mean Square Error (NRMSE)

\[
NRMSE = \frac{RMSE}{(X_{\text{max}} - X_{\text{min}})} \times 100 \quad \text{where, } RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}}
\]
Figure 3. Schematic portrait depicting WP–SVR methodology.

Data Preprocessing via Wavelet Packet

Wavelet Packet Coefficients of Rain(t-1), Temp(t-1), Tide(t-1) and GWL(t-1)

TRAINING

Support Vector Regression with RBF Kernel function

TESTING

WP-SVR Model

Forecasted monthly GWL(t)

Input Parameters

Total Monthly Rainfall [Rain(t-1)]

Mean Monthly Temperature [Temp(t-1)]

Mean Monthly Tide levels [Tide(t-1)]

Monthly Groundwater Levels [GWL(t-1)]

(2) Normalized Mean Bias

\[ \text{NMB} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{Y_i - X_i}{X_i} \right) = \left( \frac{\bar{Y}_i}{\bar{X}_i} - 1 \right) \]  \hspace{1cm} (17)

(3) Absolute Relative Error (ARE) and Average Absolute Relative Error (AARE)

\[ \text{ARE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - X_i}{X_i} \right| \times 100; \quad \text{AARE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - X_i}{X_i} \right| \times 100 \] \hspace{1cm} (18)

(4) Nash–Sutcliffe coefficient (NSE)

\[ \text{NSE} = 1 - \frac{\sum_{i=1}^{N} (Y_i - X_i)^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \] \hspace{1cm} (19)
(5) Threshold Statistics (TS)

\[ TS_\chi = \frac{n_x}{N} \times 100 \]  

(6) Correlation Coefficient (CC)

\[ CC = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}} \]

where \( x \) is the Observed/Actual Values; \( y \) is the Computed Values; \( \bar{x} \) is the Mean of Actual Data; \( n_x \) is the Number of Data points whose ARE value is less than \( x \% \); \( N \) is the Total Number of Data Points.

7. Results and discussions

The aim of using the SMO-SVR is to test its ability to predict groundwater level fluctuation in observation wells located at Surathkal, Mangalore, and Ganjimatta. The analysis is being performed for one month lead time prediction of groundwater levels for the above-mentioned input combinations (Section 5.1) at all the three well locations. The optimal parameters obtained after tuning the SVR model is as tabulated in Table 4. The number of support vectors (nsv) provides the information regarding the efficacy of the model developed i.e. whether the model is underfitting or overfitting. The SVR models developed above have “nsv” in 30–40% range and this indicates that the models are neither underfitting nor overfitting. Table 5 presents the statistical results of the SVR models developed for all the three input cases. The magnitude of the NMB and AARE computations infers that the SVR model trained with four input variables closely predicts the observed groundwater level time series than the SVR models trained with two and three input variables. The negative and positive NMB values can be interpreted as the model underestimates and overestimates, respectively. The performance in terms of Threshold Statistics of ARE computed for 5 and 10%, also indicates that the SVR model with four input variables is performing better. During the test phase, the TS < 5% and TS < 10% for an SVR model with four input variables are 66 and 86.11%, respectively with respect to well at Surathkal. The NRMSE and NSE statistics of the SVR models with different input cases is presented in Figures 4 and 5, respectively. The degree of residual variance portrayed using NRMSE statistic shows the global goodness of fit between the computed and observed groundwater levels for input combination with four variables. Since the observation wells considered in the study are near to the coast, the effect of tidal levels can have a considerable effect on the shoreline aquifers. The results obtained during this modeling demonstrate the relevance of inclusion of mean tidal level data as an input variable.

### Table 4. Optimal SVR parameters for different input combinations

<table>
<thead>
<tr>
<th>Well location</th>
<th>Input combinations</th>
<th>Optimal values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( C )</td>
</tr>
<tr>
<td>Surathkal</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN + TEMP</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN</td>
<td>15</td>
</tr>
<tr>
<td>Mangalore</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN + TEMP</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN</td>
<td>7.5</td>
</tr>
<tr>
<td>Ganjimatta</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN + TEMP</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>GWL + RAIN</td>
<td>6.5</td>
</tr>
</tbody>
</table>
During Wavelet packet analysis, Haar, Daubechies 2 (db2), db3, db4 wavelets were selected as mother wavelet. Due to the satisfactory performance of SVR model with four input variables, namely, monthly time series of groundwater level, total rainfall, average temperature and mean tidal levels, the same input combination is retained and considered for Wavelet packet analysis. The optimal SVR parameters obtained while modeling using Wavelet packet coefficients as input, for selected mother wavelets with respect to well locations at Surathkal, Mangalore, and Ganjimatta are presented in Tables 6–8, respectively. The individual performance of different mother wavelets is evaluated using correlation coefficient (CC) statistic. The correlation (CC) statistic infers that the db4 mother wavelet performed more efficiently than other mother wavelets at level 4. Hence, the db4 Wavelet packet...
Figure 5. NSE of SVR models with different input combinations.

Table 6. Optimal SVR parameters for forecasting using Wavelet packet coefficients with respect to well location at Surathkal

<table>
<thead>
<tr>
<th>Well at Surathkal</th>
<th>Haar</th>
<th>db2</th>
<th>db3</th>
<th>db4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>14</td>
<td>18</td>
<td>7.25</td>
<td>9.25</td>
</tr>
<tr>
<td>Gamma (γ)</td>
<td>0.25</td>
<td>1.633</td>
<td>2.25</td>
<td>3.5</td>
</tr>
<tr>
<td>Epsilon (ε)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Train CC</td>
<td>0.75</td>
<td>0.77</td>
<td>0.82</td>
<td>0.97</td>
</tr>
<tr>
<td>Test CC</td>
<td>0.63</td>
<td>0.68</td>
<td>0.75</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 7. Optimal SVR parameters for forecasting using Wavelet packet coefficients with respect to well location at Mangalore

<table>
<thead>
<tr>
<th>Well at Mangalore</th>
<th>Haar</th>
<th>db2</th>
<th>db3</th>
<th>db4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>9</td>
<td>12</td>
<td>21</td>
<td>7.75</td>
</tr>
<tr>
<td>Gamma (γ)</td>
<td>2.25</td>
<td>3.8</td>
<td>4.25</td>
<td>5.25</td>
</tr>
<tr>
<td>Epsilon (ε)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Train CC</td>
<td>0.55</td>
<td>0.67</td>
<td>0.84</td>
<td>0.96</td>
</tr>
<tr>
<td>Test CC</td>
<td>0.51</td>
<td>0.59</td>
<td>0.79</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 8. Optimal SVR parameters for forecasting using Wavelet packet coefficients with respect to well location at Ganjimatta

<table>
<thead>
<tr>
<th>Well at Ganjimatta</th>
<th>Haar</th>
<th>db2</th>
<th>db3</th>
<th>db4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>12</td>
<td>14</td>
<td>20</td>
<td>12.25</td>
</tr>
<tr>
<td>Gamma (γ)</td>
<td>2.276</td>
<td>2.88</td>
<td>3.75</td>
<td>1.75</td>
</tr>
<tr>
<td>Epsilon (ε)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Train CC</td>
<td>0.69</td>
<td>0.74</td>
<td>0.8</td>
<td>0.97</td>
</tr>
<tr>
<td>Test CC</td>
<td>0.61</td>
<td>0.62</td>
<td>0.73</td>
<td>0.97</td>
</tr>
</tbody>
</table>
coefficients of input variables are considered for the development of hybrid WP–SVR model in order
to generate one month ahead groundwater level forecasts. Table 9 presents the optimal SVR
parameters of hybrid WP–SVR model.

The comparative evaluation of results obtained from the WP–SVR and SVR models for one month
lead GW-level prediction is presented in the form of various performance indices in Table 10 and in
the form of graphs (Figures 6–8). The advantage and robustness of WP–SVR model are being checked
and the results were found to be quite satisfactory over the basic SVR model.

Study area 1 – Well location at Surathkal: From the time series graph and scatter plot as presented
in Figures 9 and 12 for 1 month lead GW-level prediction, it is observed that both SVR and WP–SVR
model results are closely following the observed time series. The proposed WP–SVR model is almost

Table 9. Optimal SVR parameters for forecasting using db4–Wavelet packet coefficients

<table>
<thead>
<tr>
<th>Well location</th>
<th>Input combinations</th>
<th>Efficient mother wavelet-db4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Surathkal</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>9.25</td>
</tr>
<tr>
<td>Mangalore</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>7.75</td>
</tr>
<tr>
<td>Ganjimatta</td>
<td>GWL + RAIN + TEMP + TIDE</td>
<td>12.25</td>
</tr>
</tbody>
</table>

Table 10. Statistical results of WP–SVR and SVR models

<table>
<thead>
<tr>
<th>Well location</th>
<th>Model</th>
<th>NRMSE</th>
<th>NMB</th>
<th>AARE</th>
<th>NSE</th>
<th>TS &lt; 5%</th>
<th>TS &lt; 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surathkal</td>
<td>WP–SVR</td>
<td>TRAIN</td>
<td>-0.22</td>
<td>3.82</td>
<td>0.94</td>
<td>77.38</td>
<td>95.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-0.70</td>
<td>3.58</td>
<td>0.91</td>
<td>75</td>
<td>97.22</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>TRAIN</td>
<td>-0.41</td>
<td>4.21</td>
<td>0.89</td>
<td>75</td>
<td>92.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-0.75</td>
<td>5.29</td>
<td>0.8</td>
<td>66</td>
<td>86.11</td>
</tr>
<tr>
<td>Mangalore</td>
<td>WP–SVR</td>
<td>TRAIN</td>
<td>-0.49</td>
<td>2.63</td>
<td>0.92</td>
<td>79.76</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-0.58</td>
<td>2.45</td>
<td>0.91</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>TRAIN</td>
<td>-0.54</td>
<td>3.66</td>
<td>0.82</td>
<td>72.61</td>
<td>96.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-1.01</td>
<td>4.18</td>
<td>0.79</td>
<td>72.22</td>
<td>94.44</td>
</tr>
<tr>
<td>Ganjimatta</td>
<td>WP–SVR</td>
<td>TRAIN</td>
<td>0.13</td>
<td>3.65</td>
<td>0.94</td>
<td>82.14</td>
<td>96.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-0.29</td>
<td>4.68</td>
<td>0.93</td>
<td>77.77</td>
<td>94.44</td>
</tr>
<tr>
<td></td>
<td>SVR</td>
<td>TRAIN</td>
<td>0.19</td>
<td>4.61</td>
<td>0.89</td>
<td>70.23</td>
<td>90.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST</td>
<td>-0.38</td>
<td>5.65</td>
<td>0.83</td>
<td>66.86</td>
<td>88.88</td>
</tr>
</tbody>
</table>

Figure 6. Normalized mean bias of WP–SVR and SVR forecast.
following the trend of observed time series, while the SVR foresee slightly deviates at peaks and base, not properly catching the higher and lower values. The WP–SVR model performance appears to have accepted accuracy during training and testing phase as observed from different statistical indices presented in Table 10 and from the scatter plot of test phase as shown in Figure 12. Here, the TS < 5% and TS < 10% of WP–SVR model being 75 and 97.22%, respectively, during test phase verifies the close agreement of WP–SVR model with the observed groundwater levels. The other statistical evaluations of NMB and NRMSE as presented in Figures 6 and 7, respectively, demonstrate the efficacy of WP–SVR prediction to that of the SVR model.

Study area 2—Well location at Mangalore: Figure 10 displays the time series plot of WP–SVR and SVR model outputs in test stage, wherein it can be visualized that the WP–SVR prediction closely
follows the observed groundwater level time series. Here, NRMSE of 4.94 and 7.55%, NSE of 0.92 and 0.91 observed during training and test phase, respectively, from WP–SVR model depicts the significance of the hybrid model developed. Figure 13 shows the closely spaced scatters of the computed and observed groundwater levels of WP–SVR and SVR models during the test phase. It can be observed that the SVR predictions have more number of outliers than that of WP–SVR predictions during the test phase. Referring to Figure 8, it can be seen that the WP–SVR yields 75% forecast lower than 5% absolute relative error (ARE) and 100% forecast less than 10% ARE during test phase, thereby clearly ascertaining the consistency and robust performance of the hybrid model that is developed.
Study area 3—Well location at Ganjimatta: From the time series plot of WP–SVR and SVR model outputs of the test phase as represented in Figure 11, it can be inferred that the WP–SVR prediction is capable of catching the lower peak values efficiently and it also diligently shadows the observed groundwater level time series. During training phase, positive NMB values are seen in both WP–SVR and SVR forecasts, thereby indicating the overestimation of observed data values (Figure 6). However, in the test phase, both the models exhibit underestimation of observed values. The SVR predictions show slightly weak correlation with observed groundwater levels when compared with WP–SVR predictions during the test phase as depicted from the scatter plot (Figure 14).

It can be observed that the performance is satisfactory with RBF Kernel-based SVR models and the WP–SVR model forecasted the water levels with superior accuracy in terms of all the statistical indices during calibration and validation period at all well locations. It is obvious from the fit line equations and $R^2$ values that the WP–SVR model performs much better than the basic SVR model (Figures 12–14). In overall, the Wavelet packet and SVR conjunction model framed by integrating two methods, Discrete WPTs and SVR, seems to be more adequate than the single SVR model for forecasting monthly groundwater levels. The denoised periodic components obtained from the Wavelet packet technique is found to be most effective in yielding more accurate forecast when used as inputs in the SVR models.
8. Summary and conclusions
In this study, a novel hybrid model of WP–SVR has been developed to forecast monthly groundwater levels of three different wells located in Dakshina Kannada district, southwest coast of India. From the advantages of discrete Wavelet packet transformation, the meteorological data obtained for these three well locations were decomposed and the denoised coefficients were used with SVR for groundwater level prediction. The results obtained from the proposed WP–SVR model was compared with the basic SMO-SVR model for each well location. From Tables 7–9, it is clear that among the Wavelet packet coefficients the daubechis 4 wavelet with decomposition level 4 is giving better results. The SVR model with 4 input variables is performing better. The Wavelet packet denoising of input variables is found to improve the forecasting accuracy of SVR model. The developed WP–SVR model is relatively more efficient than the basic SVR prediction in groundwater level forecasting. The Wavelet packet decomposition better addresses the meteorological variables with non-stationary and non-linear features.

This study analysis and results reveal the following conclusions.

⇒ The proposed WP–SVR model outperforms the basic SMO-SVR model for a month lead time forecast.
⇒ The SVR modeling is mainly dependent on tuning of hyperparameters, selection of kernel function, and formulation of suitable optimization algorithm.
⇒ The improvement of results in WP–SVR model is due to dividing the data-set into multi-frequency bands and denoising the unnecessary peaks using discrete WPT.
⇒ The efficiency of any regression analysis mainly depends on the quality of the data that are being fed into the model and henceforth, proper pre-processing of the data using techniques like wavelets, Wavelet packets, or discretization using any other methods helps in improving the model performance of regression analysis. The work done in this study is a best example for the aforementioned conclusion.

The work can be extended for higher lead time forecasting and also models can be developed for monitoring seasonal groundwater level fluctuations by dividing the data according to the monsoon and non-monsoon periods.

9. Limitations of the study

• The wells selected for this study are monitored by Department of Mines and Geology, Government of Karnataka essentially to measure groundwater levels. No pumping is done. Pumping from wells will have a significant impact on the groundwater level fluctuations. Since this factor is unaccounted in the study, the model suitability is limited and doesn’t hold good to real-time situations where there is pumping from wells.
• Other influential factors that are supposed to be included in modeling such as evaporation and evapotranspiration data could have been beneficial if available during modeling.

Acknowledgments
The authors would like to thank the Department of Mines and Geology, Government of Karnataka; Indian Meteorological Department, Pune, for providing the necessary data required for research and the Department of Applied Mechanics & Hydraulics, National Institute of Technology Karnataka for the necessary infrastructural support. The authors would like to thank two anonymous reviewers for their valuable suggestions and comments.

Funding
The authors received no direct funding for this research.

Author details
Sujay Raghavendra. N1 E-mail: sujayraghavendra@ymail.com
ORCID ID: http://orcid.org/0000-0002-0482-1936
Paresh Chandra Deka1 E-mail: pareshdeka@yahoo.com
ORCID ID: http://orcid.org/0000-0003-1771-2496
1 Department of Applied Mechanics and Hydraulics, National Institute of Technology Karnataka, Surathkal, Mangalore 575025, India.

Citation information

Cover image
Outline of hybrid wavelet packet–SVR model.
Source: Author.
References


