Advances in Water Resources 141 (2020) 103595

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms

A.T.M. Sakiur Rahman[a,*], Takahiro Hosono[b,c], John M. Quilty[d], Jayanta Das[e], Amiya Basak[e]

[a] Department of Earth and Environmental Science, Kumamoto University, Kumamoto, Japan
[b] Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
[c] International Research Organization for Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
[d] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada
[e] Department of Geography and Applied Geography, University of North Bengal, Darjeeling 734013, India

ARTICLE INFO

Keywords:
Groundwater level
Forecasting
Maximal overlap discrete wavelet transform
Extreme gradient boosting machine
Random forests
Support vector machine

ABSTRACT

Groundwater level (GWL) forecasting is crucial for irrigation scheduling, water supply and land development. Machine learning (ML) (*e.g.*, artificial neural networks) has been increasingly adopted to forecast GWL due to its ability to model nonlinearities between GWL and its drivers (*e.g.*, rainfall). Although ML approaches have been successful at forecasting GWL, they are often inaccurate when GWL exhibits multiscale changes (*e.g.*, due to urbanization). To address this shortcoming, wavelet transforms (WT) are routinely coupled with ML methods. Unfortunately, researchers frequently neglect key issues associated with WT that render such forecasts useless for real-world scenarios. This study demonstrates how new ML methods, such as eXtreme Gradient Boosting and Random Forests, can be properly coupled with WT to generate accurate GWL forecasts (1-–3 months ahead) for 7 wells in Kumamoto City in Southern Japan that can be used to help address current pressing issues such as groundwater quality and land subsidence.

1. Introduction

Groundwater level (GWL) forecasting is a crucial task that allows water resources managers to create effective irrigation schedules, manage water supply, and plan land development. Simulation of GWL is often performed by processes-oriented models with MODFLOW (McDonald and Harbaugh, 1988) being the most popular (Mohanty et al., 2013). Advanced, fully coupled and distributed models, *e.g.*, HydroGeosphere (Therrien et al., 2010) and GETFLOWS (Mori et al., 2015, Hosono et al., 2019), are becoming more popular for surface water (*e.g.,* Mori et al., 2015; Ala-aho et al., 2015) and groundwater simulation (Maxwell et al., 2015; Ala-aho et al., 2015, Hosono et al., 2019). These models require a range of spatio-temporal data for approximating the complex hydrological processes, heterogeneous subsurface systems, and anthropogenic activities (Therrien et al., 2010; Barthel and Banzhaf, 2016; Kollet et al., 2017) and consequently involves high cost and long computational time (Maxwell et al., 2015). In addition, achieving accurate simulations through process-based models is very challenging (Sun et al., 2016; Sahoo et al., 2017) due to the over-simplification of complex hydrological processes (Therrien et al., 2010); discretization of model domains (White et al., 2019) and availability of data. Researchers are, therefore, exploring alternative approaches such as, data-driven (*i.e.,* statistical and machine learning (ML)) approaches over the last two decades (see review by Rajaee et al., 2019). ML approaches avoid having to specify the underlying complex physical hydrological processes and instead rely on only the statistical relationships between explanatory (*e.g.*, rainfall) and response variables (*e.g.*, GWL). Despite this lack of physical knowledge, ML has been successfully applied to GWL forecasting (Rajaee et al., 2019).

Some studies (*e.g.*, Adamowski and Chan, 2011; Barzegar et al., 2016; Belayneh et al., 2016; Quilty et al., 2018) have also shown that hybrid approaches that combine different ML approaches in various stages of the model development can be more accurate than standalone ML since particular patterns in the data (*e.g.,* trends, periodicities, level shifts) can be better captured by different approaches (Ghaemi et al., 2019). For instance, the wavelet transform (WT), a time-frequency localization method that is able to extract time-varying behavior (trends, periodicities, etc.), or multiscale change, from time series has often been coupled with ML for GWL forecasting, with many studies showing increased performance from this approach over standalone ML (Nourani et al., 2014; Rajaee et al., 2019). However, recent studies (Zhang et al., 2015; Du et al., 2017, Quilty and Adamowski, 2018) have shown that these hybrid WT-based ML models have been incorrectly developed and cannot be used properly for real-world applications. This

* Corresponding author.
E-mail address: shakigeo@gmail.com (A.T.M.S. Rahman).

https://doi.org/10.1016/j.advwatres.2020.103595
Received 28 December 2019; Received in revised form 25 March 2020; Accepted 18 April 2020
Available online 5 May 2020

study aims to demonstrate how a recent WT-based forecasting approach (the Wavelet Data-Driven Forecasting Framework, WDDFF (Quilty and Adamowski, 2018)) can be coupled with emerging ML methods for real-world GWL forecasting.

To date, there has been a plethora of ML approaches applied to GWL forecasting (see list in Rajaee et al., 2019) such as: Multiple Linear Regression (MLR) (*e.g.,* Sahoo and Jha, 2013; Sun et al, 2016; Ebrahimi and Rajaee, 2017), Artificial Neural Networks (ANN) (*e.g.*, Daliakopoulose et al. 2005; Zhang et al., 2018; Guzman et al. 2017; Wunsch et al. 2018; Natarajan and Sudheer, 2019; Mirarabi et al., 2019), Extreme Learning Machine (Alizamir et al., 2018; Malekzadeh et al., 2019; Natarajan and Sudheer, 2019), Fuzzy Logic (FL) (*e.g.,* Alvisi et al., 2006; Shiri and Kisi, 2011; Nadiri et al., 2019), Adaptive Neuro-Fuzzy Inference System (ANFIS) (Kholghi and Hosseini, 2009; Moosavi et al., 2013; Gong et al., 2018), Support Vector Machine (SVM) (*e.g.,* Huang et al. 2017; Natarajan and Sudheer, 2019; Tang et al., 2019; Mirarabi et al. 2019), genetic programming (GP) (*e.g.,* Shiri et al., 2013; Fallah-Mehdipour et al., 2013; Natarajan and Sudheer, 2019), Gradient Boosting Machine (GBM) (Kenda et al., 2018) along with statistical approaches (such as ARIMA model) (Adamowski and Chan, 2011). However, two very recent and promising approaches for GWL forecasting include Random Forests (RF, Wang et al., 2018; Tang et al., 2019; Kenda et al., 2018; Koch et al., 2019) and eXtreme Gradient Boosting (XGB). RF and XGB are both tree-based ML algorithms: the former is based on bagged regression trees while the latter is based on boosted regression trees. A comparison between these two methods has yet to be carried out for GWL forecasting nor have these methods been explored within the context of the WDDFF.

More recently, to address multiscale change in hydrology and water resources (Koutsoyiannis, 2013), researchers have coupled WT with ML to generate hybrid models for improving forecast accuracy of GWL (Adamowski and Chan, 2011), streamflow (Adamowski and Sun, 2010; Tiwari and Chatterjee, 2010), water quality (Barzegar et al., 2016), urban water demand (Quilty et al., 2019), and water balance components (Rahman et al., 2018) . WT is the most popular approach for developing the accurate hybrid models for hydrological and water resources forecasting (Nourani et al., 2014; Afan et al., 2016; Fahimi et al., 2017; Quilty and Adamowski, 2018). However, a recent study by Quilty and Adamowski (2018) revealed that many studies incorrectly developed wavelet-based hybrid models which are not applicable for real-world forecasting problems. The authors proposed a set of best practices to address this issue, resulting in the Wavelet Data-Driven Forecasting Framework (WDDFF), which can be used to correctly forecast real-world processes, such as GWL. To date, WDDFF has only been used with simple least-squares regression methods (*e.g.*, multiple linear regression, Volterra series models, Extreme Learning Machines) and has only been explored for urban water demand forecasting, irrigation flow forecasting (Mouatadid et al., 2019), and monthly pan evaporation prediction (Ghaemi et al., 2019) while it has high potential for GWL forecasting.

Hence, there is a need to develop correct wavelet-based ML models based on WDDFF that can be properly applied for a variety of critical real-world problems, such as GWL forecasting. Moreover, new ML methods, such as XGB and RF are only beginning to be explored in the hydrology and water resources domains. For example, the tree-based version of XGB (XGBT) has only been explored for evapotranspiration (Wu et al., 2019; Wu and Fan, 2019), tropical cyclone (Jin et al., 2019) and streamflow (Hadi et al., 2019; Tyralis et al. 2019; Li et al., 2019; Zhang et al., 2019) modelling while the linear XGB (XGBL) has yet to be investigated for hydrological or water resources modelling. Although RF is one of most the successful ML methods across a variety of domains, it has not received enough attention for hydrological and water resources modelling yet (Tyralis et al., 2019). Recent studies (Wang et al., 2018; Tang et al., 2019; Kenda et al., 2018; Koch et al., 2019) have explored the potential of standalone RF model for GWL forecasting and found that RF can generate accurate GWL forecasts. However, XGBT, XGBL and wavelet-based RF have yet to be explored in the context of GWL forecasting. In our current study, we explored, for the first time in the literature, the potential of XGBT, XGBL, and RF as well as XGBT, XGBL, and RF coupled with the WDDFF (from Quilty and Adamowski, 2018) for GWL forecasting. We compared these new approaches against benchmark methods: support vector regression (SVR), and their WDDFF counterparts. Additionally, we employ a Bayesian optimization routine based on Gaussian Processes (Snoek et al., 2012) for 'automatic' hyper-parameter selection in XGBT, XGBL, RF, and SVR (including their WT counterparts), which as far as we are aware, has yet to be explored in earlier studies employing XGB and RF. This Bayesian optimization routine is an effective hyper-parameter selection approach that has been used in many fields of science and engineering with considerable success (Falkner et al., 2018).

Therefore, the main goal of this study is to demonstrate the potential usefulness of XGBT, XGBL, WT-XGBT, WT-XGBL, and WT-RF for GWL forecasting through a comparison with methods developed earlier in the literature (i.e., RF, SVR and WT-SVR). For our case study, we forecast average monthly GWL at 1, 2, and 3-months ahead for seven wells in Kumamoto City, Japan, using meteorological variables as input data along with previous GWL measurements.

Our main contributions to the literature include, the first exploration of:

1. XGBT, XGBL, WT-XGBT, WT-XGBL, and WT-RF for GWL forecasting.
2. XGBT, XGBL, RF, and SVR coupled with WDDFF (i.e., from Quilty and Adamowski, 2018).
3. Bayesian optimization for the 'automatic' selection of hyper-parameters in XGB and RF-based models.

The rest of this study is organized as follows: Section 2 provides information on the study area and data; Section 3 outlines our methodology; Section 4 presents results and discussion and section 5 provides conclusions of our findings.

2. Study area and data

We conducted our experimental study using seven wells where GWL observations were available in Kumamoto City, which is in the center of Kyushu Island in the southern part of Japan (Fig. 1). Kumamoto city is one of the regions with the highest groundwater use in Japan. About one million city dwellers depend completely on groundwater for their domestic purposes (Oshima, 2010; Shimada, 2012), and approximately 8×10^7 m^3 groundwater per year is withdrawn from 58 pump stations to meet their water demand (Hosono et al., 2013). There are two types of aquifers in this region, unconfined and semi-confined to confined aquifers. The former is composed of alluvium sedimentary deposits and unwelded pyroclastic deposits overlying an impermeable aquiclude of the lacustrine sedimentary layer. The latter is semi-confined to confined (Taniguchi et al., 2003; Hosono et al., 2013) in nature and it consists of pyroclastic flow deposits and volcanic lavas. This aquifer is locally referred to as a second aquifer and its depth below the ground surface generally varies between 60 m and 200 m (Hossain et al., 2016). Majority groundwater is withdrawn from the second aquifer for water supply in the city, hence, this is the most important aquifer in the area. A detailed hydrogeological description of the study area can be found in Hosono et al. (2013) and Kagabu et al. (2017). The majority of groundwater in the second aquifer is recharged through infiltration of precipitation in the recharge areas (such as Kikuchi, Ueki and Takayubaru highlands) as well as by river and artificial ponding waters in paddy fields in and around the midsection of the Shira River (Hosono et al., 2013; Taniguchi et al., 2019) (Fig. 1). Groundwaters in the recharge areas are then transported through highly porous pyroclastic flow deposits (Aso-2 and Aso-3) and discharged mostly to Lake Ezu.

The monthly GWL data spans 1980–2017 and was collected from the Government office of Kumamoto City, Kumamoto City Waterworks and Sewerage Bureau. We analyzed GWL data of 7 wells that are located along the regional groundwater flows from recharge to discharge areas (Fig. 1). All wells are installed into the second (deep) aquifer with

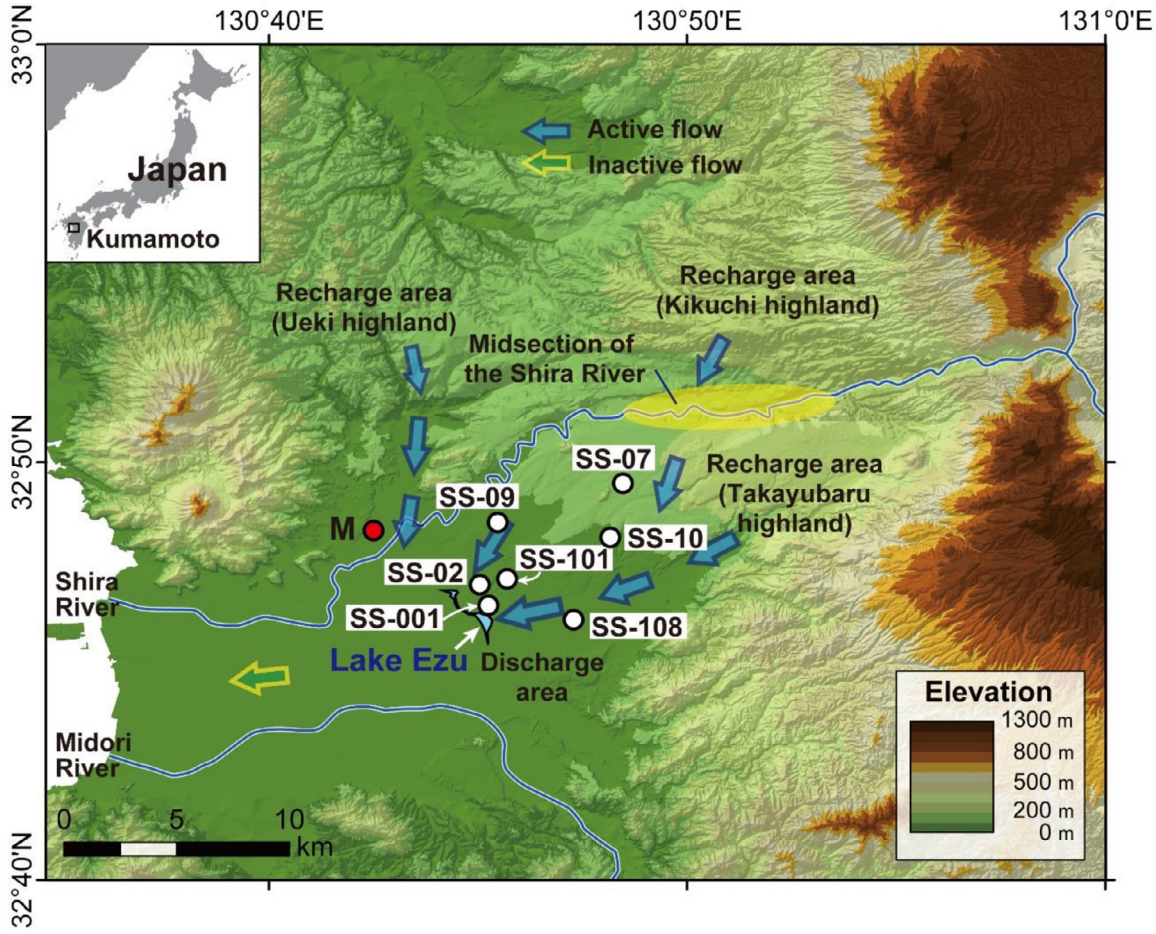

Fig. 1. Map showing study area and the location of 7 wells analyzed for groundwater levels. The location of Kumamoto meteorological station (indicated by M) is also shown in the map.

their well depths ranging from 31 m (Well no.: SS-09) to 124 m (Well no.: SS-108) withan average of 68.28 m. The screen depths range from 21 to 31 m (Well no.: SS-108) and 81.5 to 114.50 m (Well no.: SS-10) from the surface, respectively. The observed GWL data contains few missing (<2%) records. Missing data for a particular month were filled by the average of the previous and following month (i.e., neighbouring months). Furthermore, meteorological data (monthly mean temperature and monthly total rainfall of the Kumamoto meteorological station) were gathered from the Japan Meteorological Agency (JMA) website (https://www.jma.go.jp/jma/indexe.html) as potential inputs for the ML models developed in this study.

3. Methodology

Methodology section arranges for brief descriptions of the ML approaches such as SVM, RF, XGB and WT as well as descriptions of the steps followed for the model development.

3.1. Support vector machine

Support Vector Machine (SVM) is a statistical learning theory. The basic algorithm of SVM is discussed in Cortes and Vapnik (1995). SVM is also known as support vector regression (SVR) when applied to regression problems. The Library of SVM (LIBSVM) (Chang and Lin, 2011) is commonly used for developing SVR models. For regression problems, LIBSVM implements the sequential minimal optimization algorithm, and it supports epsilon (ε-SVR) and nu (γ-SVR) regressions. In this study, SVR models were developed through the R-programing language environment (R Core Development Team, 2019) using the LIBSVM adopted R-package "e1071" (Meyer et al., 2019). The ε-SVR approach was adopted since it has been shown to be an accurate approach for predicting hydrological and water resources variables (Tabari et al., 2012; Raghavendra and Deka, 2014; Mehdizadeh et al., 2017; Huang et al., 2017; Ferreira et al., 2019).

In SVR, the selection of the hyper-parameters is of key importance in developing accurate models. Hyper-parameters include: kernel type (e.g., linear, polynomial, radial basis function, which controls the similarity between a current input and each input in the training set), *epsilon* (ε - the width of insensitive tube, which determines the 'support vectors' that are considered in the model), gamma (γ- controls the shape of the hyperplane separation) and cost (c - a regularization parameter that controls the sparsity of the model). Aside from the selection of the kernel, hyper-parameter selection is discussed in Section 3.5.4. Several studies have demonstrated that radial basis function (RBF) kernel provides strong performance for predicting hydrological and water resources variables (Mehdizadeh et al., 2017; Ferreira et al., 2019) and it has been used in recent studies for GWL forecasting (Yoon et al., 2016; Huang et al., 2017; Tang et al., 2019; Guzman et al., 2019). Further details on theoretical background regarding the ε-SVR and RBF can be found in Chang and Lin (2011).

3.2. Random forests

Random Forest (Breiman, 2001) is a ML algorithm derived from the "classification and regression trees" (CART) paradigm (Breiman et al., 1984) with some additional randomization, such as 'bagging' input data samples and variables (Tyralis et al., 2019). RF is a popular ML approach used for prediction and classification in numerous domains due to its

high precision and capability to handle a large number of input variables (Pal, 2016; Tyralis and Papacharalampous, 2017; Tyralis et al., 2019). There exist several varieties of RF based on the CART-criterion (Biau and Scornet, 2016; Tyralis et al., 2019). All variations of RF are not applicable for solving regression/prediction problems (Geurts et al., 2006; Wright and Ziegler, 2017), and many of them are tested mainly for classification (Geurts et al., 2006; Biau and Scornet, 2016). The regression variant of RF is a nonlinear model in which samples are partitioned at each node of a tree based on the value of a selected input variable (Breiman, 1984). In this study, extremely randomized trees (details in Geurts et al., 2006), applicable for regression (Geurts et al., 2006; Biau and Scornet, 2016; Wright and Ziegler, 2017), was applied for tree splitting. This approach is the most computationally efficient variation of RF (Geurts et al., 2006; Wright and Ziegler, 2017), and has been shown to achieve similar performance with other similar RF-based algorithms (Biau and Scornet, 2016). The performance of RF also depends on several hyper-parameters (Biau and Scornet, 2016; Wright and Ziegler, 2017; Tyralis and Papacharalampous, 2017; Probst et al., 2019; Tyralis et al., 2019) such as: *number of trees* (n) (indicating the size of the forest), *mtry*(representing the number of input variables to split in each node), *minimal node size*(the minimum number of observations in a terminal node) and *sample size* (the number of observations randomly selected for growing each tree).

3.3. eXtreme gradient boosting

The eXtreme Gradient Boosting is a scalable tree boosting ML method (Chen and Guestrin, 2016) that can capture the nonlinear relationships between predictor and target variables (Chen et al., 2019; Hadi et al., 2019). XGB is also based on CART and resembles Gradient Boosting Machines with some variations in model structure (Chen and Guestrin, 2016). GBM and XGB differ from RF in how trees are grown. For example, GBM grows tree on the residuals of the former tree, whereas RF trains each tree individually (Just et al., 2018). XGB implements second-order derivatives, whereas ordinary GBM uses first-order derivatives (Chen and Guestrin, 2016). XGB is more efficient compared to GBM and is less likely to result in model over-fitting since it incorporates improved regularization features (Chen and Guestrin, 2016; Chen et al., 2019). The parallel processing of the computation makes it not only the fastest among the gradient boosting algorithms but also affords improved performance over ordinary GBM (Chen et al., 2019).

In the generic version of CART, a tree ensemble model with k additive functions is used to predict the target variable ($y_i \in \mathbb{R}$), such as GWL, using input variable(s) ($(\mathrm{x}_i \in \mathbb{R}^d)$) (where $i = 1, 2, \ldots, N$ and N are the number of input-target variable pairs) via (Chen and Guestrin, 2016):

$$\hat{y}_i = \phi(\mathrm{x}_i) = \sum_{k=1}^{K} f_k(\mathrm{x}_i),\ f_k \in \mathcal{F} \tag{1}$$

Here, $\phi(\mathrm{x}_i)$ maps the model inputs to the model prediction ($\hat{y}_i \in \mathbb{R}$) where $\mathcal{F} = \{f(x) = w_{q(x)}\}(q : \mathbb{R}^m \to T, w \in \mathbb{R}^T)$ denotes the space for the regression tree; q represents the structure of each tree; f_k corresponds to the individual tree structure; T is the number of leaves in the tree f_k, and w is the leaf weights. In XGB, the objective function incorporates a loss function and a regularization term, given as (Chen and Guestrin, 2016):

$$\mathcal{L}(\phi) = \sum_i l\ (\hat{y}_i, y_i) + \sum_k \Omega\ (f_k) \tag{2}$$

where:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{3}$$

Here, l is the function for the differentiable convex loss used for measuring the difference between the target (y_i) and prediction ($\hat{y}_i$) during training. The second term (Ω) denotes the regularization term, which is used for model complexity penalization.

Since model optimization cannot be achieved using traditional methods for the tree ensemble approach (XGBT) using Eq. (2), the additive approach is applied in model training for XGBT (Chen and Guestrin, 2016). Moreover, the function (f_t) is added for improving the model (Eq. (2)) for predicting the $\hat{y}_i^{(t)}$ after the t-th iteration for the i-th instance as (Chen and Guestrin, 2016):

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, y_i^{(t-1)} + f_t\ (\mathrm{x}_i)\right) + \Omega(f_t) \tag{4}$$

For quick optimization of the objective function, Taylor second-order expansion is applied by following Friedman et al. (2000):

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{N} \left[l\ \left(y_i, y_i^{(t-1)}\right) + g_i f_t\ (\mathrm{x}_i) + \frac{1}{2}\ h_i\ f_t^2\ (\mathrm{x}_i)\right] + \Omega(f_t) \tag{5}$$

Here, the gradient statistics of the loss function for the first (g_i) and second (h_i) order are as follows:

$$g_i\ = \partial_{\hat{y}^{(t-1)}}\quad l\left(y_i, \hat{y}_i^{(t-1)}\right) \tag{6}$$

and

$$h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(i, \hat{y}_i^{(t-1)}\right) \tag{7}$$

The objective is simplified by removing the constant term for step t as:

$$\widetilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{N} \left[g_i f_t\ (\mathrm{x}_i) + \frac{1}{2}\ h_i\ f_t^2\ (\mathrm{x}_i)\right] + \Omega(f_t) \tag{8}$$

The Eq. (8) can be rewritten by defining the $I_j = \{i|q(\mathrm{x}_i) = j\}$ and expanding Ω as:

$$\begin{aligned}\widetilde{\mathcal{L}}^{(t)} &= \sum_{i=1}^{n} \left[g_i f_t\ (x_i) + \frac{1}{2}\ h_i\ f_t^2\ (x_i)\right] + \gamma T + \frac{1}{2}\ \lambda \sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T} \left[\left(\sum_{i\in \mathrm{I}_j} g_i\right) w_j\ + \frac{1}{2}\left(\sum_{i\in \mathrm{I}_j} h_i + \ \lambda\right) w_j^2\ \right] + \gamma T\end{aligned} \tag{9}$$

Now, the weight (w_j^*) and the subsequent optimal value $\{(\widetilde{\mathcal{L}}^{(t)}(q))\}$ of a leaf (j) for a fixed structure $q(\mathrm{x})$ can be estimated as:

$$w_j^* = -\frac{\sum_{i\in \mathrm{I}_j} g_i}{\sum_{i\in \mathrm{I}j} h_i + \lambda} \tag{10}$$

and

$$\widetilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{\left(\sum_{i\in \mathrm{I}_j} g_i\right)^2}{\left(\sum_{i\in \mathrm{I}_j} h_i + \lambda\right)} + \gamma T \tag{11}$$

A tree structure(q) quality can be measured using Eq. (11). However, computing all the trees and simultaneously finding the best one from all possible options is computationally intractable, and the problem is solved by using an algorithm that adds branches to the tree at each iteration from a single leaf. If the tree splitting for right and left nodes are I_R and I_L respectively, the loss reduction by adding them as $I = I_L \cup I_R$ can be estimated as (Chen and Guestrin, 2016):

$$\mathcal{L}_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i\in \mathrm{I}_L} g_i\right)^2}{\sum_{i\in \mathrm{I}_L} h_i + \lambda} + \frac{\left(\sum_{i\in \mathrm{I}_R} g_i\right)^2}{\sum_{i\in \mathrm{I}_R} h_i + \lambda} - \frac{\left(\sum_{i\in \mathrm{I}} g_i\right)^2}{\sum_{i\in \mathrm{I}} h_i + \lambda}\right] - \gamma \tag{12}$$

In this study, we applied two variants of XGB such as linear boosting of XGB (*i.e.,* XGBL) and tree boosting (*i.e.*, XGBT). The main difference between these two variations of XGB is the base learner. The base leaner is the linear learner for the XGBL, whereas the base learner is the tree learner for the XGBT (Chen, 2014). The interested reader can find further details in Chen (2014).

A number of hyper-parameters need to be set for developing XGB models. The hyper-parameters considered for XGB model are: *Eta* (representing the learning rate); *gamma* (representing the minimum split loss);

max depth (indicating the maximum depth of a tree); *min child weight* (the minimum sum of the weights); *subsample* (ratio of the samples used for model training); *column sample by tree* (parameter for subsampling of columns, i.e., variables); *nround* (number of rounds); and two regularization parameters, *lambda* (which controls L2- regularization) and *alpha* (which controls for L1-regularization).

3.4. Wavelet transformation

Wavelet transforms are a mathematical tool that is particularly useful for identifying important time-frequency localized features in a time series (*e.g.,* periodicities, transients, level shifts). WT can decompose a time series into different sub-time series, which provide coarse and fine-grained details on the multiscale nature of the time series. Incorporating these sub-time series into a ML model has often improved the accuracy of hydrological and water resources forecasts when compared to standalone ML models (Afan et al., 2016; Fahimi et al., 2017; Mouatadid et al., 2019). There are two variations of WT: Continuous (CWT) and Discrete (DWT). Several studies have already discussed the advantages and disadvantages of the two variations of WT (Fugal, 2009; Nourani et al., 2014) and recommended that DWT is better for decomposing hydrological and water resources time series data (Nourani et al., 2014; Rezaie-balf et al., 2017; Samadianfard et al., 2018; Mouatadid et al., 2019). Recent studies have noted that DWT with multiresolution analysis (DWT-MRA) is the most common WT that is adopted in hydrological and water resources modelling (Nourani et al., 2014; Quilty and Adamowski, 2018; Rajaee et al., 2019). However, the DWT-MRA has been demonstrated to have several drawbacks for when used in real-world applications (Maheswaran and Khosa, 2012; Du et al., 2017; Quilty and Adamowski, 2018). For example, DWT is sensitive to the length of time series and adding new data points to the time series (*i.e.,* in an online system as new data is collected) and performing wavelet decomposition causes the wavelet and scaling coefficients computed at a given time to change (Quilty and Adamowski, 2018). To overcome these drawbacks, the maximal overlap DWT (MODWT) proposed by Percival and Walden (2000), was implemented in this study as recommended by Quilty and Adamowski (2018).

The MODWT applies low and high pass filters to a given time series, resulting in wavelet and scaling coefficients. If $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ (here, l denotes the length of the filter, $j = 1,\ 2, \ldots, J$ represents the j-th scale, and J is the level of decomposition) are the wavelet and scaling filters, the MODWT can be given as (Percival and Walden, 2000):

$$\tilde{h}_{j,l} = \frac{h_{j,l}}{2^{\frac{j}{2}}} \tag{13}$$

$$\tilde{g}_{j,l} = \frac{g_{j,l}}{2^{\frac{j}{2}}} \tag{14}$$

The MODWT wavelet ($\tilde{W}_{j,i}$) and scaling ($\tilde{V}_{j,i}$) coefficients for a given time series ($X = X_i,\ i = 0, 1, \ldots\ldots N-1$)can be determined via (Percival and Walden, 2000):

$$\tilde{W}_{j,i} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{i-l \bmod N} \tag{15}$$

$$\tilde{V}_{j,i} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{i-l \bmod N} \tag{16}$$

Here, $L_j = (2^j - 1)(L - 1) + 1$. Further details on the MODWT can be obtained from earlier studies (e.g., Percival and Walden, 2000; Quilty and Adamowski, 2018; Mouatadid et al., 2019).

3.5. Model development

This section gives details on the ML-based models (XGBT, XGBL, RF, SVR, WT-XGBT, WT-XGBL, WT-RF, and WT-SVR) developed in this study for GWL forecasting.

3.5.1. Dataset pre-processing and boundary corrections

For our case study, the target (response) variable was average monthly GWL at lead times of 1, 2, and 3 months ahead. For the standalone ML models (XGBT, XGBL, RF and SVR), candidate input (explanatory) variables included time-lagged (from lag-1 up to lag-12; depending on the target variable forecasting step, more details in Supplementary Material (SM) Table S1) data such as average monthly GWL, average monthly air temperature (Tavg), average monthly total rainfall (R) and cumulative monthly rainfall (CR) (see also Section 3.4). These are the common explanatory variables considered for GWL forecasting in several earlier studies (*e.g.,* Adamowski and Chan, 2011; Wunsch et al. 2018; Nadiri et al., 2019). Furthermore, water fluxes into and out of the reservoir are considered by incorporating variables such as rainfall and air temperature (related to evapotranspiration) in the ML models.

For the WT-based models (WT-XGBT, WT-XGBL, WT-RF, and WT-SVR), original and decomposed data were used as the candidate input variables. MODWT was applied for decomposing the explanatory variables (see Section 3.4). Since the wavelet-decomposed data is affected by "boundary conditions" (Quilty and Adamowski, 2018; Percival and Walden, 2000; Basta, 2014), it was necessary to remove the first L_J wavelet and scaling coefficients (see Section 3.4) affected by boundary conditions due to the selected wavelet filter length (L) and decomposition level (J) at the beginning of the decomposed data. However, since this study compares non-wavelet-based and WT models that vary based on decomposition level and wavelet filter, it was necessary to remove the same number of boundary conditions affected coefficients from the beginning of both the target and all input variables for both non-wavelet and WT models to ensure the same training and validation indices were used in the development and evaluation of the different models. This is based on the best practices established in Quilty and Adamwoski (2018). The number of boundary conditions effected coefficients that were removed from the beginning of the target and input variables in all models (L_f) was determined according to (Quilty and Adamowski, 2018):

$$L_f = \left(2^{J_{max}} - 1\right)(L - 1) + 1 \tag{17}$$

The careful selection of J_{max} and L is necessary to ensure the number of data available for model training and validation is sufficient. In this study, J_{max} and L were set to four. Hence, the total number of the boundary conditions effected coefficient was 46 which was removed from the beginning of the target and input variables for all models (*e.g.,* Fig. 2). A decomposition level of four was selected to ensure that the prominent annual cycle present in all GWL time series was captured in the level four wavelet coefficients.

After adjusting the original dataset (456 records) for the lagged target and input variables, the historical record available ranged from March 1981 to December 2017 (442 records). The number of available records (396) for training and validating the different ML models after adjusting for time lags and boundary conditions were divided into the training (85%) and validation (15%) sets. While there is no set rule for dataset partitioning during model training and validation (Deo et al., 2017), it is typically accepted that the validation partition should be within 10–40% of the total dataset length (Barzegar et al., 2019).

3.5.2. Input variable selection and importance of input variables

Input variable selection is one of the most important steps for developing ML models (Tiwari and Chatterjee, 2010, Galelli et al., 2014, Quilty et al., 2016). Both linear (*e.g.*, Sudheer et al., 2002; Yaseen et al., 2015) and nonlinear approaches (Quilty et al., 2016) have been applied for input variable selection in hydrological and water resources modelling (Galelli et al., 2014). However, linear approaches based on partial auto-correlation function (PACF) and auto-correlation function (ACF) are typically not the most suitable approaches for hydrological and water resources modelling since such problems are generally nonlinear (Hadi et al., 2019). Nonlinear approaches based on mutual information (MI) (Shannon, 1948) typically perform better than linear

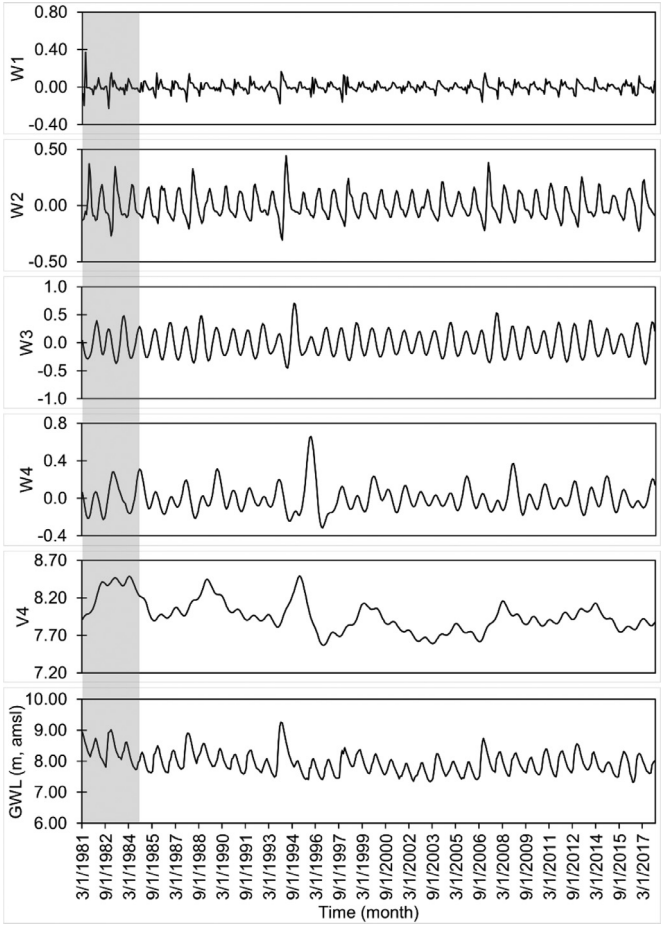

Fig. 2. Mean monthly groundwater level (GWL) time series (m, above mean sea level) (Well: SS-001) decomposed by the MODWT using the best localized (BL4) wavelet filter and a decomposition level of four. Here, 'Wj' and 'Vj' represent the wavelet and scaling coefficients at level j, respectively. The shaded area shows the boundary affected maximum number (46) of coefficients.

approaches for hydrological and water resources modelling (see details in Quilty et al., 2016; Gong et al., 2018; Taormina et al., 2016) as well as other science and engineering applications (Peng et al., 2005; Galelli et al., 2014).

As will be discussed below, since SVR does not internally measure the importance of input variables, it is important to perform input variable selection prior to developing an SVR model. In this study, we applied one of the most widely used MI approaches for input variable selection "minimal-redundancy-maximal-relevance" (mRMR) proposed by Peng et al. (2005). This approach selects input variables from a set of candidates by using MI to identify candidates that are relevant but not redundant. This approach has been shown to select more appropriate input variables than other similar approaches (Berrendero et al., 2016). Peng et al. (2005) defined an operator $\Phi(D, R)$ for simultaneously optimizing the maximum-relevance (D) and minimal-redundancy (R) to select an input subset (S) from the d input variables in x as:

$$\max \Phi(D, R), \Phi = D - R \tag{18}$$

The complete mathematical derivations for mRMR can be obtained from Peng et al. (2005).

However, the input variable selection is not mandatory for approaches like XGB and RF. Both approaches internally perform input variable selection as well as quantify the variables' importance (VI) (Chen and Guestrin, 2016; Tyralis et al., 2019; Hadi et al., 2019, Jin et al., 2019). However, SVM does not produce any internal measures of VI (Boehmke and Greenwell, 2019). Hadi et al. (2019) pointed out that XGB is a very robust modelling approach even with a high number of explanatory variables. The performance of RF is also not overly influenced by additional superfluous input variables i.e., granted the model is provided with some input variables that are useful in predicting the target variable (Biau and Scornet, 2016; Pal, 2016; Probst et al., 2019; Tyralis et al., 2019). The VI in RF was estimated by the advanced permutation method (Strobl et al., 2009) and in XGBT, VI was measured via the impurity approach (Chen and Guestrin, 2016; Boehmke and Greenwell, 2019), while XGBL estimate VI by measuring the weight (Chen, 2014).

3.5.3. Data scaling

Approaches like XGBL and SVR are affected by the scale of the different input variables. In other words, if input variables are on widely different scales, this can lead to the model favoring those variables that have a large range while ignoring those with a smaller range, regardless of their importance in predicting the target variable. Hence, input variable scaling was performed prior to the development of the XGBL and SVR models. Since the SVR models were developed using the "e1071" R-package, it internally performs input data scaling by scaling all variables to have unit variance and zero mean (Meyer et al., 2019). For the XGBL models, the input variables were scaled such that they ranged from 0 to 1 using the min-max normalization approach, which has been successfully applied in many earlier hydrological and water resources ML models (*e.g.*, Yaseen et al., 2016; Jeong and Park, 2019). Min-max normalization for a given input variable is determined using:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{19}$$

Here, X is the original input variable and X' is the normalized input variable. X_{min} and X_{max} are the minimum and maximum values of the input variable, respectively.

3.5.4. Hyper-parameter optimization

Most ML algorithms require the careful selection of a set of hyper-parameters (Probst et al., 2019). Examples of hyper-parameters include: the kernel function in SVR, the number of trees in RF, the learning rate in XGB, etc. Hyper-parameters for each ML model used in this study are mentioned in the relevant sections (3.1 for SVR, 3.2 for RF, and 3.3 for XGB). The performance of ML models is highly influenced by hyper-parameters settings (Eggensperger et al., 2013; Zhang et al., 2016) and their improper selection can lead to poorly performing models. Many studies select hyper-parameters through trial-and-error (Suryanarayana et al., 2014; Zhang et al., 2016; Jeong and Park, 2019), grid search and/or random search (Bergstra and Bengio, 2012), and some use heuristic optimization algorithms such as particle swarm optimization and genetic algorithms (Snoek et al., 2012; Bergstra and Bengio, 2012; Feurer and Hutter, 2019). Ideally, an accurate and efficient automated hyper-parameter optimization method is desirable (Zhang et al., 2016) and useful for permitting a fair comparison between ML alternative (Sculley et al., 2018). Furthermore, when exploring different ML models, they can be equally compared only if they all receive the same degree of optimization (or level of attention) for the problem at hand (Feurer and Hutter, 2019).

Bayesian Hyper-parameter Optimization (BHO) (Snoek et al., 2012) is an emerging state-of-the-art approach for global and local optimization of hyper-parameters (Falkner et al., 2018; Feurer and Hutter, 2019). BHO has been shown to outperform other approaches such as grid search and random search (Bergstra and Bengio, 2012) on several challenging optimization benchmarks (Jones, 2001). BHO has broad applicability to various problem settings (Falkner et al., 2018; Feurer and Hutter, 2019), can be used with integer and real-valued hyper-parameters, and in many cases, can outperform domain experts for finding optimal hyper-parameters (Snoek et al., 2012; Eggensperger et al., 2013; Feurer and Hutter, 2019).

In this study, the BHO algorithm (Snoek et al., 2012 and software implementation in Wilson, 2019) was implemented for hyper-parameter selection. BHO was used for hyper-parameters optimization by using K-fold cross-validation on the training data. K-fold cross-validation has been successfully applied for hyper-parameter selection in several hydrological and water resources modelling studies (*e.g.*, Barzegar et al., 2019; Wu and Fan 2019). In this study we set K = 5.

For performing BHO, it is necessary to choose a prior and an acquisition function. The former adopts a Gaussian process prior and the latter uses Expected Improvement (EI) (Snoek et al., 2012). The usefulness of Gaussian processes is solely dependent on the covariance function, which makes use of the Matern52 kernel (Snoek et al., 2012).

3.5.5. Coupling of ML and MODWT

To develop the WT models, the ML approaches (XGBT, XGBL, RF, and SVR) were integrated within the WDDFF by following the direct approach and using the MODWT for wavelet decomposition of the input variables (Approach-4; Quilty and Adamowski, 2018). Approach 4 includes both un-decomposed and wavelet decomposed boundary condition corrected time series data as input variables. Furthermore, there is no single wavelet filter and decomposition level (J) combination that has been demonstrated to provide the best performance across all hydrological and water resources modelling applications (Maheswaran and Khosa, 2012; Mouatadid, et al., 2019). In this study, three different wavelet families, such as *haar, daubechies* (d2) and *best localized* (BL-4) with J from 1 up to 4, were tested to identify the best WT models. Therefore, the total number of hybrid models for the seven GWL monitoring wells and the three forecasting lead times (1, 2, and 3- months ahead) was 1008 along with 84 standalone models (see Table S2).

3.5.6. Model accuracy

The accuracy of applied ML models was evaluated by a number goodness of fit statistics such as: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), RMSE-observations standard deviation ratio (RSR) (Moriasi et al., 2007), coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009). RSR can be used to interpret a given model's forecasts as: very good ($0.00 < RSR \leq 0.50$), good ($0.50 < RSR \leq 0.60$), satisfactory ($0.60 < RSR \leq 0.70$) and unsatisfactory (RSR > 070) (Moriasi et al., 2007) and NSE as very good ($0.75 < NSE \leq 1$), good ($0.65 < NSE \leq 0.75$), satisfactory ($0.50 < NSE \leq 0.65$) and unsatisfactory ($NSE \leq 0.50$) (Moriasi et al., 2007).

4. Results and discussion

The results of the ML approaches for 1, 2 and 3 month(s) ahead lead time forecasting of GWL was evaluated using several goodness of fit statistics (MSE, MAE, RMSE, RSR, NSE, KGE and R^2), and graphical tools (hydrographs, scatterplots, and boxplots). The experiment results showed a good trade-off between training and validation performance, confirming the stable generalization capacity of the ML approaches. For the standalone and wavelet-based models, the results obtained for the validation period are presented due to the large number of models and to focus on the generalization abilities of the various models.

4.1. Standalone models

The results (RMSE, NSE and R^2) of the standalone ML approaches for forecasting GWL at 1, 2 and 3 month(s) ahead can be found in Table 1 and results from additional statistics including MAE, MSE, RSR and KGE are also provided in Tables S3-S6. The average accuracy across the seven wells for the standalone ML approaches showed very good performance in terms of NSE (NSE > 0.75) and RSR (RSR < 0.50) (Moriasi et al., 2007) with higher values of KGE and lower values of MSE and RMSE for 1 and 2 (except RF for 2-step) month(s) ahead forecasts. The NSE values ranged between 0.84 and 0.96 with an average of 0.93 for SVR, 0.85 and 0.97 with an average of 0.92 for XGBL, 0.75 and 0.95 with an average of 0.86 for XGBT, and 0.55 and 0.93 with an average of 0.81 for RF, respectively, for 1 month ahead GWL forecasting (Table 1). The NSE values for 2 months ahead GWL forecasts were slightly lower than 1 month ahead forecasts with average NSE values across all seven wells of 0.86 for SVR, 0.81 for XGBL, 0.75 for XGBT and 0.70 for RF, respectively, (Table 1). Although the accuracy further decreased with increasing lead time, SVR (avg. NSE = 0.78) and XGBL (avg. NSE = 0.75) showed very good performance for 3 months ahead GWL forecasting, while XGBT (avg. NSE = 0.67) and RF (avg. NSE = 0.66) showed satisfactory performance. An example of comparison between observed and forecasted GWL (for 1, 2, and 3 months ahead) is illustrated in Fig. 3 using hydrographs and scatterplots for well SS-001. It is noticeable that all the methods showed decaying accuracy as well as higher deviations from the 1:1 line with increasing time steps (Fig. 3 and Table 1). Other GWL forecasting case studies also found similar results though considering different ML approaches (Yoon et al., 2011; Rezaie-balf et al., 2017; Barzegar et al., 2017).

The best performing model was the SVR for all lead time forecasting based on the average goodness of fit statistics. A number of studies (Yoon et al., 2011; Suryanarayana et al. 2014; Huang et al., 2017; Mukherjee and Ramachandran, 2018; Guzman et al. 2019; Mirarabi et al. 2019) have reported that SVR is the best performing model for GWL forecasting, while comparisons were made against several ML approaches such as MLR, ANN and RF (Rajaee et al., 2019). However, most of the relevant studies only tested the capability of the different ML approaches for limited number (~4) of observation wells.

However, it can be argued from the obtained results of this study that different ML approaches showed the best performance for different observation wells (Table 1). For example, SVR was the best model for a well (SS-07), whereas XGBL was the best model for another well (SS-10) (Table 1). Furthermore, SVR and XGBL showed similar performance for some wells (SS-001 and SS-09), while RF exhibited the lowest accuracy for all lead times for almost all observation wells (Table 1) with an unsatisfactory performance for a well (SS-07). The RF models show, for certain cases, lower accuracy than highly parameterized tree-based models (such as XGB) due to the automation of the model (Efron and Hastie, 2016; Tyralis et al. 2019). It is worth mentioning again that it is necessary to select input variables for SVR models using another algorithm, while XGB and RF models do not require such extra efforts for modeling. Therefore, it can be inferred that the standalone XGB model can effectively manage a high number of input variables without application of other algorithms which is the main advantage of XGB variants, and they can perform equally with other high performing models like SVR.

4.2. WT coupled models

The results of the best performing wavelet-based hybrid model for each ML (such as WT-SVR, WT-RF, WT-XGBL and WT-XGBT) approach can be found in Table 2 and Tables S3-S6. Interestingly, nearly half of the hybrid models (556) showed less accuracy than standalone models, and many wavelet-based models (198) showed similar results. Furthermore, the same results were also found for the best models with different WT decomposition levels and wavelet filters (Table 2). The reason behind this is that wavelet-based models are not guaranteed to outperform standalone models. In this study, only a small number of wavelet filters were explored due to the boundary condition at the beginning of the time series (see Eq. (17)), which needs to be carefully considered in light of the length of the dataset. Unfortunately, many studies overlook this boundary condition and consider wavelet filters that are too wide and decomposition levels that are too high causing errors to be introduced in the wavelet and scaling coefficients (sometimes leading to unrealistically high performance than what is achievable in real-world scenarios) (Quilty and Adamowski, 2018), which this study avoided by

Table 1
Accuracy of standalone ML approaches for 1, 2 and 3 month(s) ahead forecasts of monthly average GWL.

Well Number	Method	1-month			2-months			3-months		
		RMSE (m)	NSE	R^2	RMSE (m)	NSE	R^2	RMSE (m)	NSE	R^2
SS-001	SVR	**0.07**	**0.93**	**0.94**	**0.10**	**0.85**	**0.88**	**0.10**	**0.84**	**0.87**
	RF	0.09	0.88	0.90	0.10	0.84	0.87	0.11	0.83	0.86
	XGBL	**0.07**	**0.93**	**0.94**	0.12	0.78	0.79	0.12	0.77	0.84
	XGBT	0.08	0.90	0.92	**0.10**	**0.85**	0.87	0.13	0.76	0.82
SS-02	SVR	**0.06**	**0.94**	**0.94**	0.10	0.85	0.88	0.11	0.84	0.87
	RF	0.09	0.89	0.90	0.10	0.85	0.88	**0.10**	**0.86**	**0.88**
	XGBL	0.07	0.93	0.94	**0.10**	**0.86**	**0.90**	0.11	0.83	0.87
	XGBT	0.08	0.91	0.91	0.11	0.83	0.86	0.12	0.80	0.85
SS-07	SVR	**0.33**	**0.93**	**0.94**	**0.58**	**0.79**	**0.80**	**0.64**	**0.74**	**0.77**
	RF	0.85	0.55	0.72	1.09	0.27	0.51	1.19	0.12	0.47
	XGBL	0.45	0.88	0.88	0.65	0.74	0.76	0.83	0.57	0.62
	XGBT	0.64	0.75	0.84	0.90	0.50	0.67	1.00	0.38	0.59
SS-09	SVR	**0.16**	**0.92**	**0.93**	0.28	0.79	0.80	0.58	0.73	0.78
	RF	0.28	0.74	0.81	0.34	0.62	0.73	0.36	0.57	0.71
	XGBL	**0.16**	**0.92**	**0.93**	**0.24**	**0.80**	**0.85**	**0.28**	**0.73**	**0.80**
	XGBT	0.22	0.84	0.86	0.30	0.69	0.77	0.39	0.49	0.65
SS-10	SVR	**0.15**	**0.97**	**0.97**	**0.18**	**0.96**	**0.96**	**0.29**	**0.90**	**0.90**
	RF	0.27	0.92	0.92	0.33	0.87	0.87	0.33	0.87	0.87
	XGBL	0.16	**0.97**	**0.97**	0.26	0.92	0.93	**0.29**	**0.90**	**0.90**
	XGBT	0.22	0.94	0.94	0.33	0.87	0.87	0.34	0.86	0.86
SS-101	SVR	**0.07**	**0.96**	**0.96**	**0.08**	**0.96**	**0.96**	0.13	0.89	0.90
	RF	0.10	0.93	0.94	0.12	0.91	0.91	**0.12**	**0.90**	**0.91**
	XGBL	0.08	0.95	0.96	0.12	0.89	0.92	0.13	0.88	0.91
	XGBT	0.09	0.95	0.95	0.13	0.88	0.89	0.13	0.89	0.90
SS-108	SVR	0.07	0.84	0.85	**0.06**	**0.79**	**0.81**	0.13	0.52	0.53
	RF	0.09	0.74	0.80	0.12	0.56	0.60	0.14	0.46	0.48
	XGBL	**0.07**	**0.85**	**0.87**	0.10	0.70	0.75	**0.12**	**0.60**	**0.65**
	XGBT	0.09	0.75	0.75	0.11	0.66	0.67	0.13	0.49	0.50

Table 2
Accuracy of wavelet-based ML approaches for 1, 2 and 3 month(s) ahead forecasts of monthly average GWL.

Well Number	Method	1-month				2-months				3-months			
		WT	RMSE	NSE	R^2	WT	RMSE	NSE	R^2	WT	RMSE	NSE	R^2
SS-001	WT-SVR	D2,J4	0.06	0.94	0.94	D,J3	0.08	0.90	0.91	H,J4	0.10	0.86	0.88
	WT-RF	D,J1	0.06	0.89	0.90	H,J1	0.07	0.85	0.87	D,J1*	0.08	0.83	0.86
	WT-XGBL	D,J2	0.07	0.93	0.94	H,J2	0.10	0.85	0.89	D,J2	0.10	0.84	0.88
	WT-XGBT	H,J2*	0.08	0.91	0.92	B,J1	0.11	0.83	0.87	B,J4	0.11	0.81	0.88
SS-02	WT-SVR	H,J1*	0.07	0.94	0.94	H,J4	0.08	0.89	0.92	B,J3	0.09	0.88	0.89
	WT-RF	D,J1	0.06	0.90	0.91	H,J1	0.07	0.86	0.89	H,J3	0.08	0.86	0.88
	WT-XGBL	B,J2	0.07	0.93	0.94	D,J2	0.10	0.86	0.90	H,J2*	0.10	0.85	0.88
	WT-XGBT	H,J2*	0.07	0.92	0.93	B,J3	0.10	0.86	0.89	B,J4	0.11	0.82	0.90
SS-07	WT-SVR	B,J1	0.34	0.93	0.93	H,J1	0.55	0.81	0.81	H,J1	0.62	0.76	0.76
	WT-RF	D,J4	0.08	0.84	0.89	D,J1	1.03	0.22	0.50	D,J1*	1.10	0.12	0.46
	WT-XGBL	H,J3*	0.38	0.91	0.91	H,J4	0.10	0.85	0.89	D,J2	0.11	0.83	0.87
	WT-XGBT	D,J1	0.55	0.81	0.86	H,J2	0.87	0.53	0.68	D,J2	1.00	0.38	0.64
SS-09	WT-SVR	H, J1	0.15	0.92	0.93	D,J4	0.22	0.84	0.87	H,J4	0.27	0.76	0.81
	WT-RF	D,J1*	0.19	0.74	0.81	D,J1	0.25	0.61	0.73	D,J3	0.27	0.59	0.70
	WT-XGBL	H,J1	0.16	0.92	0.93	D,J2	0.24	0.80	0.85	B,J4	0.27	0.75	0.81
	WT-XGBT	H,J2*	0.20	0.87	0.90	H,J2	0.31	0.68	0.76	H,J4	0.34	0.62	0.77
SS-10	WT-SVR	H,J4	0.15	0.97	0.97	H,J4	0.23	0.94	0.94	B,J3	0.26	0.92	0.92
	WT-RF	D,J4	0.19	0.93	0.93	D,J4	0.23	0.90	0.90	H,J4	0.25	0.88	0.88
	WT-XGBL	B,J1*	0.17	0.97	0.97	D,J1	0.25	0.92	0.93	H,J2	0.30	0.89	0.90
	WT-XGBT	H,J3	0.19	0.96	0.96	D,J3	0.29	0.90	0.90	H,J1	0.30	0.89	0.89
SS-101	WT-SVR	H,J2	0.08	0.96	0.96	D,J3	0.10	0.93	0.93	B,J3	0.11	0.91	0.92
	WT-RF	D, J1*	0.07	0.94	0.94	D,J1	0.09	0.92	0.92	B,J3	0.09	0.91	0.92
	WT-XGBL	D,J3	0.09	0.95	0.96	D,J2	0.13	0.89	0.92	H,J2*	0.13	0.88	0.90
	WT-XGBT	H,J3	0.08	0.96	0.96	H,J3	0.11	0.92	0.93	H,J3	0.12	0.90	0.92
SS-108	WT-SVR	H,J2*	0.07	0.85	0.85	B, J4	0.10	0.73	0.74	B,J4	0.10	0.68	0.69
	WT-RF	D,J1*	0.07	0.75	0.81	D,J1	0.10	0.55	0.59	B,L4	0.11	0.46	0.48
	WT-XGBL	D,J3	0.07	0.86	0.87	H,J4	0.10	0.71	0.72	H,J2	0.11	0.61	0.65
	WT-XGBT	D,J1	0.08	0.83	0.83	H,J1	0.11	0.63	0.65	B,J1	0.14	0.46	0.46

Note: B - Best localized, D- Daubechies 2, H-Haar, J-decomposition level,

* indicates similar results obtained by different wavelet filters and decomposition level or same wavelet filters but different decomposition level.

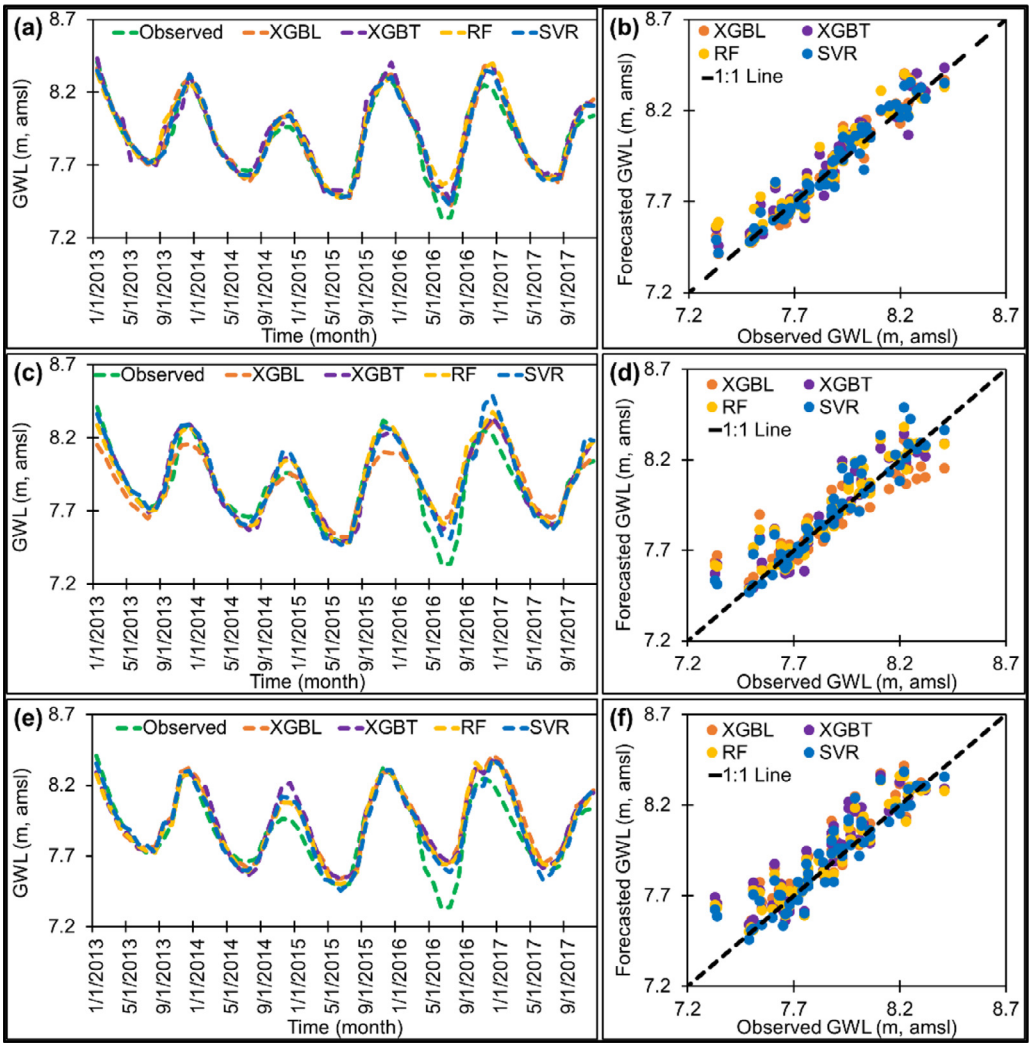

Fig. 3. Comparison between observed and forecasted (by standalone ML approaches) mean monthly GWL illustrated by hydrograph and 1:1 line for well SS-001 for (a-b) 1 month ahead lead time, (c-d) 2 months ahead lead time, and (e-f) 3 months ahead lead time, respectively.

following the best practices in Quilty and Adamowski (2018). In this study it was confirmed that no particular wavelet filter and decomposition level universally led to improved performance of the hybrid models over the standalone models for GWL forecasting. Several studies also found similar results for other hydrological variables (*e.g.*, Witten et al., 2011; Quilty and Adamowski, 2018).

The most significant finding in Table 2 is that the WT-XGBL showed the best performance for many wells for 1 month ahead forecasting. Furthermore, the performance of WT-XGBL was also better than standalone XGBL for most cases (Tables S3-S6). The average NSE and R^2 values for the seven wells at the 2 month ahead were 0.86 and 0.87 for WT-SVR, 0.84 and 0.87 for WT-XGBL, 0.76 and 0.80 for WT-XGBT, and 0.70 and 0.77 for WT-RF, respectively, while for 3 months ahead GWL forecasts average NSE and R^2 values were 0.78 and 0.82 for WT-SVR, 0.81 and 0.84 for WT-XGBL, 0.70 and 0.78 for WT-XGBT, and 0.66 and 0.74 for WT-RF, respectively. When the comparisons between standalone models and wavelet-based models were made (Fig. 4, Fig. S1, and Tables S3-S6), the improvement was more evident for 2 and 3 months ahead forecasts except few cases than that of 1 month ahead forecasting. The improvement in accuracy measures based on average NSE were 4% (3 months ahead) for WT-SVR; 5% (3 months ahead) for WT-XGBL; 3% (3 months ahead) for WT-XGBT; and almost no improvement for RF 3 months ahead forecasts. It is worthwhile to note that these percentages were calculated by considering the maximum NSE value (equal to 1). This finding is very crucial to improving water resources management since the mid-term forecasting (such as 2 and 3 months ahead) is useful for irrigation planning, land development and other environmental considerations.

4.3. Variables of Importance

RF and XGB variants naturally provide the importance of explanatory variables in predicting the target variable (see section 3.5.2). It was found that some variables have significant relative contributions to the models and most of them showed very low contributions or no contribution as shown in Fig. 5 (explanatory variables with at least 1% contribution for wavelet-based model are only shown). Similar findings have also been reported by Hadi et al. (2019), while XGBT was applied for streamflow modelling. It was generally observed that lagged GWL (Lag-1), as expected, was the most important variables for 1 month ahead forecasting. However, cumulative rainfall (CR) was generally the most important variable for 2 and 3 months ahead forecasting (Figs. S2-S3). It is noticeable that the number of variables for XGBT was much lower than that of XGBL and RF. The variable importance plots (Fig. 5b,d,f) clearly demonstrate the importance of WT. For example, Fig. 5b indicates that lagged GWL (Lag-1) was the most important (about 36% contribution) variable, followed by wavelet scaling coefficients (V) of CR, Tavg and R for WT-XGBT model and WT-RF also showed similar tendency (Fig. 5f).

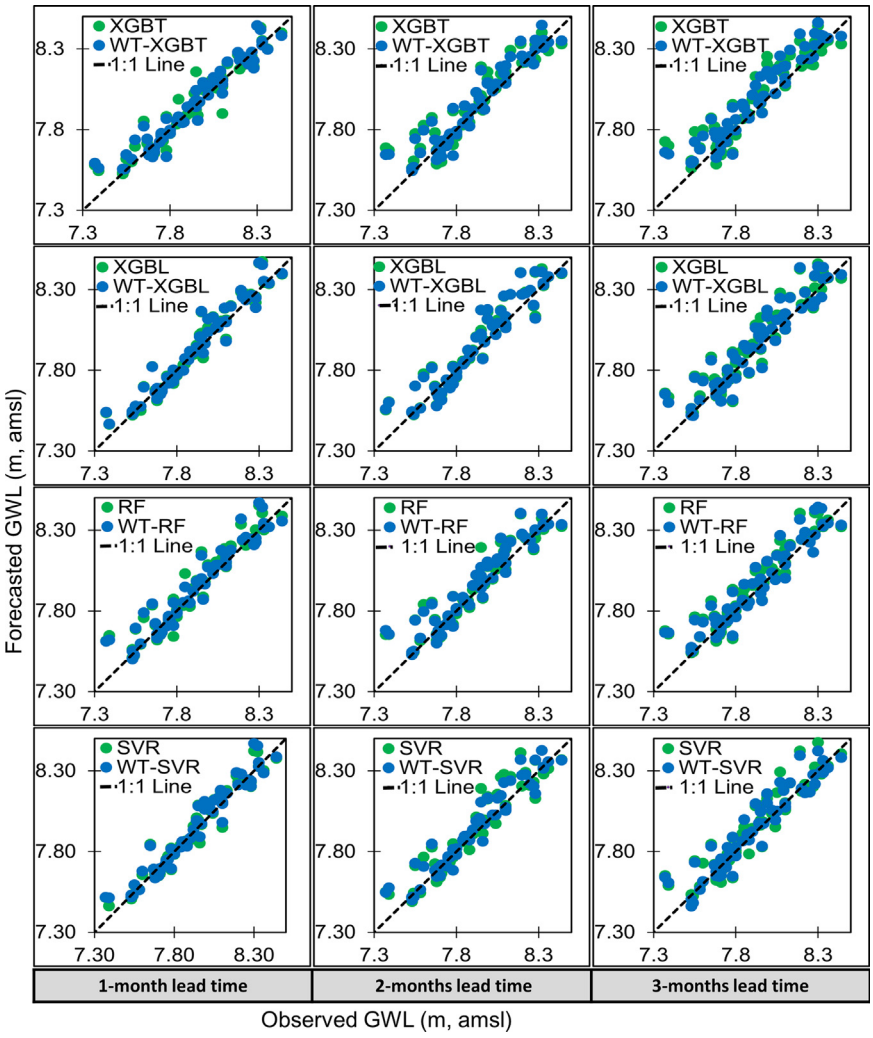

Fig. 4. Comparison of results of standalone ML approaches and best wavelet-based ML approaches for multi-steps ahead monthly lead time forecasting for well SS-02: left panels represent the 1 month ahead lead time, middle panels show 2 months ahead lead time, and right panels illustrate 3 months ahead lead time, respectively. Observed and forecasted GWL are shown in x-axis and y-axis, respectively.

However, the main contributing variables for XGBL were the wavelet coefficients (W) along with lagged GWL (Lag-1) (Fig. 5d) for 1 month ahead GWL forecasting. Another interesting finding of this study is that RF and XGB variants were less dependent on lagged GWL for longer lead time forecasts (Figs. S2-S3). This is interesting since these approaches could be explored for long-term GWL forecasting (six months to several years, Mouatadid et al. 2019) granted that adequate uncertainty assessment to be undertaken. Other studies have successfully applied ML approaches for long-term forecasting of hydrological variables such as GWL by SVR (Salem et al., 2018) and stream flow by SVR and ELM (Zhu et al., 2019). Long-term forecasting is important for sustainable water resources management, though it is challenging due to high uncertainties (Mouatadid et al. 2019).

5. Limitations and future trends

Despite significant progress in GWL modeling using ML approaches over the last two decades (Rajaee et al., 2019), there are still some important issues that should be addressed to promote consistency between ML models and their conceptual and physical-based counterparts (Koch et al. 2019). So far, several studies have drawn attention to the development of hybrid approaches for capturing different patterns (*e.g.,* trends, periodicities, transients, or level shifts) in time series data (*e.g.,* Adamowski and Chan, 2011, Suryanarayana et al. 2014, Rezaiebalf et al. 2017, Rahman et al., 2018, Malekzadeh et al. 2019) or to develop multi-model ensemble approaches (*e.g.,* Barzegar et al., 2017, Nadiri et al. 2019) for improving model performance, mainly for short-term GWL forecasting. For example, this study developed extensive ML models for short-term forecasting of GWL by coupling pre-processing techniques such as WT. We also coupled BHO technique for optimizing model parameters for GWL forecasting. The developed ML models demonstrated high performance for almost all cases and error generally increased with increasing forecast lead time. However, model performance for mid-term and long-term forecasting could likely be further enhanced by incorporating time series data of groundwater recharge, groundwater pumping and evapotranspiration. Incorporating these relevant variables may improve the model performance for mid and long-term forecasting. For our study, we only considered lagged GWL, rainfall and average air temperature as model input variables which are generally considered as input variables for GWL forecasting using ML approaches (Rajaee et al., 2019). In addition, relevant studies generally ignore incorporating certain important variables such as well depths, geological characteristics such as formation porosity and permeability, and proximity to sink or source wells. Further exploration is necessary for systematically incorporating these important hydrological and geological variables in ML models in order to improve their performance for mid and long-term forecasting as well as to ensure these models are consistent with conceptual and physical-based models. Such considerations could provide further physical insights and would likely lead to a greater acceptance of ML models in the hydrological and water resources communities. Furthermore, an alternative approach for generating GWL forecasts could be tested by considering an average GWL history based on all the available GWL historical records at all wells. This average GWL can then be forecasted by ML approaches and the GWL forecasts

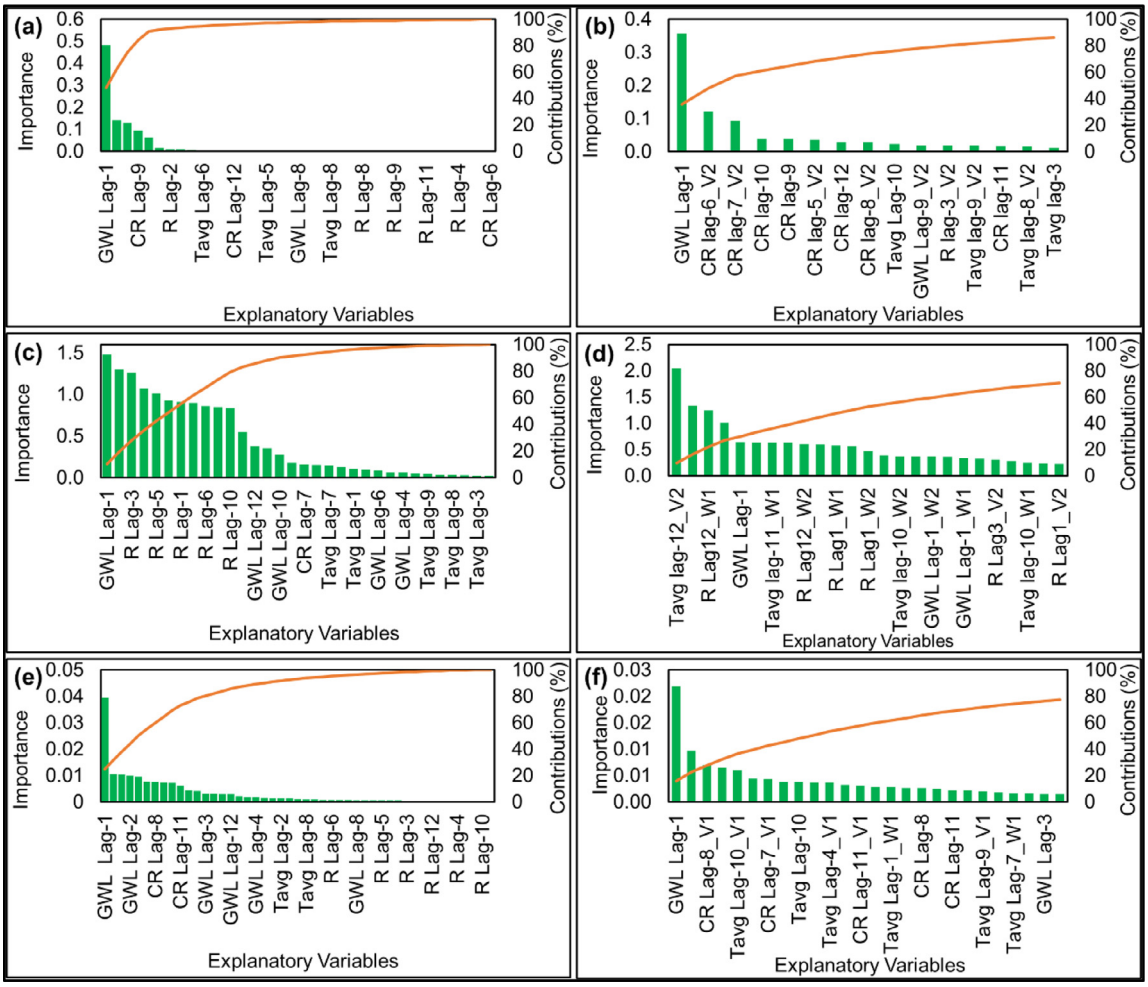

Fig. 5. Importance of input (explanatory) variables for 1 month ahead lead time forecasting for well SS-001 using (a) XGBT, (b) WT-XGBT, (c) XGBL, (d) WT-XGBL, (e) RF and (f) WT-RF, respectively. Here, 'Wj' and 'Vj' represent the wavelet and scaling coefficients at level j, respectively. For improved visualization variables with less than 1% contribution were discarded from the wavelet-based model plots. The cumulative contributions in plots shown in panels b, d and f are less than 100% as illustrated by the orange line; green bars represent the importance of each variable.

at each well can be obtained from historical deviations of the wells' GWL from the averaged value. This approach only requires the development of a single ML model, can be deployed at a regional scale, and could provide competitive performance with the models developed in this study.

6. Conclusion

Multiscale GWL forecasting is a vital task for sustainable water resources management. However, generating accurate GWL forecasts is often challenging due to nonlinear relationships between GWL and explanatory variables as well as their multiscale behavior that changes with time. One of the pre-requisites for developing accurate ML models is to select only the most useful input variables from a set of candidate input variables along with optimizing model parameter. For addressing these issues, this study tested the ability of newer ML approaches (such as Random Forests and variants of eXtreme Gradient Boosting), which can perform input variable selection internally. These models, which are able to capture nonlinear dependencies between model inputs, were fed with multiscale information extracted through wavelet transforms to enhance the forecasting models' ability to address multiscale change. Furthermore, these approaches were coupled, for the first time, with Bayesian hyper-parameter optimization for automatic model parameters estimation. The developed models were explored for 1, 2 and 3 months ahead GWL forecasting.

The performance of the wavelet-based models (WT-XGBL, WT-XGBT, and WT-RF) were compared against standalone (non-wavelet-based) models (XGBL, XGBT, and RF) including benchmark methods, such as SVR models and WT-SVR models, which often show the best accuracy for GWL forecasting in different parts of the world. However, the SVR model depends on additional external algorithms for selecting input variables. The comparative study of standalone ML approaches revealed that new approaches, especially the XGBL showed similar accuracy when compared to SVR. XGBT also showed acceptable accuracy although performance was slightly lower than that of SVR. RF generally showed the lowest performance for almost all cases.

The coupling of WT further improved the performance for all ML approaches, and the improvement was more prominent for 3 months ahead (3–5%) GWL forecasts than 1 month ahead lead forecasts, which is promising for forecasting multiscale processes commonly found in hydrology and water resources. Furthermore, XGB variants and RF provide an internal measure of variables' importance, which make the models more interpretative over other black box approaches such as SVR. The coupling of the ML approaches (especially XGB variants) with BHO and WT, therefore, is a new promising framework for GWL forecasting and may deserve further study in the field of hydrology and water resources both for short-term and long-term forecasting.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

A.T.M. Sakiur Rahman: Conceptualization, Methodology, Formal analysis, Software, Validation, Visualization, Writing - original draft. **Takahiro Hosono:** Conceptualization, Supervision, Writing - review & editing. **John M. Quilty:** Conceptualization, Software, Validation, Supervision, Writing - review & editing. **Jayanta Das:** Formal analysis. **Amiya Basak:** Formal analysis.

Acknowledgements

We sincerely thank the editor Prof. G.C. Sander, guest editors Profs. Pejman Tahmasebi and Muhammad Sahimi, and anonymous reviewers for their helpful comments that improved the quality of the final paper. A.T.M.S.R is grateful to the ministry of culture, sports, science and technology, Japan for their financial support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2020.103595.

References

Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 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. J. Hydrol. 390, 85–91.

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. 541, 902–913. https://doi.org/10.1016/j.jhydrol.2016.07.048.

Ala-aho, P., Rossi, P.M., Isokangas, E., Kløve, B., 2015. Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging. J. Hydrol. 522, 391–406.

Alizamir, M., Kisi, O., Zounemat-Kermani, M., 2018. Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. Hydrol. Sci. J. 63 (1), 63–73. https://doi.org/10.1080/02626667.2017.1410891.

Alvisi, S., Mascellani, G., Franchini, M., Bardossy, A., 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. Hydrol. Earth Syst. Sci. Discuss. 10 (1), 1–17.

Barthel, R., Banzhaf, S., 2016. Groundwater and surface water interaction at the regional-scale–a review with focus on regional integrated models. Water Resour. Manag. 30 (1), 1–32.

Barzegar, R., Fijani, E., Moghaddam, A.A., Tziritis, E., 2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. Sci. Total Environ. 599, 20–31. https://doi.org/10.1016/j.scitotenv.2017.04.189.

Barzegar, R., Ghasri, M., Qi, Z., Quilty, J., Adamowski, J., 2019. Using bootstrap ELM and LSSVM models to estimate river ice thickness in the mackenzie river basin in the northwest territories, Canada. J. Hydrol. 577, 123903. https://doi.org/10.1016/j.jhydrol.2019.06.075.

Barzegar, R., Moghaddam, A.A., Baghban, H., 2016. A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. Stoch. Environ. Res. Risk Assess. 30 (3), 883–899.

Bašta, M., 2014. Additive decomposition and boundary conditions in wavelet-based forecasting approaches. Acta Oeconomica Pragensia 2, 48–70.

Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos. Res. 172, 37–47.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Berrendero, J.R., Cuevas, A., Torrecilla, J.L., 2016. The mRMR variable selection method: a comparative study for functional data. J. Statist. Comput. Simulation 86 (5), 891–907. https://doi.org/10.1080/00949655.2015.1042378.

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227.

Boehmke, B., Greenwell, B.M., 2019. Hands-on machine learning with R. CRC Press.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees Monterey. Wadsworth and Brooks, CA.

Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 27. https://doi.org/10.1145/1961189.1961199.

Chen, T., 2014. Introduction to boosted trees. University of Washington Computer Science 22,115. https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, pp. 785–794.

Chen, Z.Y., Zhang, T.H., Zhang, R., Zhu, Z.M., Yang, J., Chen, P.Y., Ou, C.Q., Guo, Y., 2019. Extreme gradient boosting model to estimate PM2. 5 concentrations with missing-filled satellite data in China. Atmos. Environ. 202, 180–189.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. J. Hydrol. 309 (1–4), 229–240. https://doi.org/10.1016/j.jhydrol.2004.12.001.

Deo, R.C., Downs, N., Parisi, A.V., Adamowski, J.F., Quilty, J.M., 2017. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. Environ. Res. 155, 141–166.

Du, K., Zhao, Y., Lei, J., 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. J. Hydrol. 552, 44–51. https://doi.org/10.1016/j.jhydrol.2017.06.019.

Ebrahimi, H., Rajaee, T., 2017. Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. Global Planet. Change 148, 81–191.

Efron, B., Hastie, T., 2016. Computer Age Statistical Inference, Vol. 5. Cambridge University Press.

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., Leyton-Brown, K., 2013. Towards an empirical foundation for assessing bayesian optimization of hyperparameters. In: NIPS workshop on Bayesian Optimization in Theory and Practice, 10, pp. 1–5.

Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. Theor. Appl. Climatol. 128 (3–4), 875–903.

Falkner, S., Klein, A., Hutter, F., 2018. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden PMLR 80, 2018, arXiv preprint arXiv:1807.01774.

Fallah-Mehdipour, E., Haddad, O.B., Mariño, M.A., 2013. Prediction and simulation of monthly groundwater levels by genetic programming. J. Hydroenviron. Res. 7 (4), 253–260.

Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM–A new approach. J. Hydrol. 572, 556–570.

Feurer, M., Hutter, F., 2019. Hyperparameter Optimization. In: Hutter, F., Kotthoff, L., Vanschoren, J. (Eds.), Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham, pp. 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann. Stat. 28 (2), 337–407.

Fugal, D.L., 2009. Conceptual wavelets in digital signal processing: an in-depth. Practical Approach for the Non-mathematician. Space & Signals Technical Pub., University of California, USA.

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ. Model. Softw. 62, 33–51. https://doi.org/10.1016/j.envsoft.2014.08.015.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42.

Ghaemi, A., Rezaie-Balf, M., Adamowski, J., Kisi, O., Quilty, J., 2019. On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. Agric. For. Meteorol. 278, 107647. https://doi.org/10.1016/j.agrformet.2019.107647.

Gong, Y., Wang, Z., Xu, G., Zhang, Z., 2018. A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. Water 10 (6), 730. https://doi.org/10.3390/w10060730.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.

Guzman, S.M., Paz, J.O., Tagert, M.L.M., 2017. The use of NARX neural networks to forecast daily groundwater levels. Water Resour. Manag. 31 (5), 1591–1603. https://doi.org/10.1007/s11269-017-1598-5.

Guzman, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A.E., 2019. Evaluation of seasonally classified inputs for the prediction of daily groundwater levels: NARX networks vs support vector machines. Environ. Model. Assess. 24 (2), 223–234.

Hadi, S.J., Abba, S.I., Sammen, S.S., Salih, S.Q., Al-Ansari, N., Yaseen, Z.M., 2019. Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation. IEEE Access 7, 141533–141548.

Hosono, T., Tokunaga, T., Kagabu, M., Nakata, H., Orishikida, T., Lin, I.T., Shimada, J., 2013. The use of δ15N and δ18O tracers with an understanding of groundwater flow dynamics for evaluating the origins and attenuation mechanisms of nitrate pollution. Water Res. 47 (8), 2661–2675.

Hosono, T., Yamada, C., Shibata, T., Tawara, Y., Wang, C.-Y., Manga, M., Rahman, A.T.M.S., Shimada, J., 2019. Coseismic groundwater drawdown along crustal ruptures during the 2016 Mw 7.0 Kumamoto earthquake. Water Resour. Res. 55 (7), 5891–5903.

Hossain, S., Hosono, T., Yang, H., Shimada, J., 2016. Geochemical processes controlling fluoride enrichment in groundwater at the western part of Kumamoto Area, Japan. Water Air Soil Pollut. 227 (10), 385.
Huang, F., Huang, J., Jiang, S.H., Zhou, C., 2017. Prediction of groundwater levels using evidence of chaos and support vector machine. J. Hydroinf. 19 (4), 586–606. https://doi.org/10.2166/hydro.2017.102.
Jeong, J., Park, E., 2019. Comparative applications of data-driven models representing water table fluctuations. J. Hydrol. 572, 261–273.
Jin, Q., Fan, X., Liu, J., Xue, Z., Jian, H., 2019. Using eXtreme gradient BOOSTing to predict changes in tropical cyclone intensity over the Western North Pacific. Atmosphere 10, 341.
Jones, D.R., 2001. A taxonomy of global optimization methods based on response surfaces. J. Global Optim. 21 (4), 345–383. https://doi.org/10.1023/A:1012771025575.
Just, A., De Carli, M., Shtein, A., Dorman, M., Lyapustin, A., Kloog, I., 2018. Correcting measurement error in satellite aerosol optical depth with machine learning for modeling PM2. 5 in the Northeastern USA. Remote Sens. 10 (5), 803. https://doi.org/10.3390/rs10050803.
Kagabu, M., Matsunaga, M., Ide, K., Momoshima, N., Shimada, J., 2017. Groundwater age determination using 85Kr and multiple age tracers (SF_6, CFCs, and 3H) to elucidate regional groundwater flow systems. J. Hydrol. Reg. Stud. 12, 165–180.
Kenda, K., Čerin, M., Bogataj, M., Senožetnik, M., Klemen, K., Pergar, P., Laspidou, C., Mladenić, D., 2018. Groundwater modeling with machine learning techniques: ljubljana polje aquifer. In: In Multidisciplinary Digital Publishing Institute Proceedings, 2, p. 697. https://doi.org/10.3390/proceedings2110697.
Kholghi, M., Hosseini, S.M., 2009. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. Environ. Model. Assess. 14 (6), 729.
Koch, J., Berger, H., Henriksen, H.J., Sonnenborg, T.O., 2019. Modelling of the shallow water table at high spatial resolution using random forests. Hydrol. Earth Syst. Sci. 23 (11), 4603–4619. https://doi.org/10.5194/hess-23-4603-2019.
Kollet, S., Sulis, M., Maxwell, R.M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E.T., Cordano, E., Endrizzi, S., Kikinzon, E., Mouche, E., 2017. The integrated hydrologic model intercomparison project, IH-MIP2: a second set of benchmark results to diagnose integrated hydrology and feedbacks. Water Resour. Res. 53 (1), 867–890. https://doi.org/10.1002/2016WR019191.
Koutsoyiannis, D., 2013. Hydrology and change. Hydrol. Sci. J. 58 (6), 1177–1197. https://doi.org/10.1080/02626667.2013.804626.
Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B., Wang, D., 2019. A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. J. Hydroinf.10.2166/hydro.2019.066.
Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295.
Malekzadeh, M., Kardar, S., Saeb, K., Shabanlou, S., Taghavi, L., 2019. A novel approach for prediction of monthly ground water level using a hybrid wavelet and non-tuned self-adaptive machine learning model. Water Resour. Manag. 33, 1609–1628. https://doi.org/10.1007/s11269-019-2193-8.
Maxwell, R.M., Condon, L.E., Kollet, S.J., 2015. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. Geosci. Model Dev. 8 (3), 923–937.
McDonald, M.G., Harbaugh, A.W., 1988. A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, Vol. 6. US Geological Survey, Reston, VA.
Mehdizadeh, S., Behmanesh, J., Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. Comput. Electron. Agric. 139, 103–114.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C. Lin, C.C. 2019. e1071: misc functions of the department of statistics, probability theory group, TU Wien. available at: https://cran.r-project.org/web/packages/e1071/index.html (last access: 1 April 2019).
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. Environ. Earth Sci. 78 (15), 489. https://doi.org/10.1007/s12665-019-8474-y.
Mohanty, S., Jha, M.K., Kumar, A., Panda, D.K., 2013. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi–Surua Inter-basin of Odisha, India. J. Hydrol. 495, 38–51.
Moosavi, V., Vafakhah, M., Shirmohammadi, B., Behnia, N., 2013. A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour. Manage. 27 (5), 1301–1321.
Mori, K., Tada, K., Tawara, Y., Ohno, K., Asami, M., Kosaka, K., Tosaka, H., 2015. Integrated watershed modeling for simulation of spatiotemporal redistribution of post–fallout radionuclides: application in radiocesium fate and transport processes derived from the Fukushima accidents. Environ. Model. Softw. 72, 126–146.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. Am. Soc. Agric. Biol. Eng. 50 (3), 885–900.
Mouatadid, S., Adamowski, J.F., Tiwari, M.K., Quilty, J.M., 2019. Wavelet-long short-term memory networks: an approach to irrigation flow forecasting. Agric. Water Manag.
Mukherjee, A., Ramachandran, P., 2018. Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: analysis of comparative performances of SVR, ANN and LRM. J. Hydrol. 558, 647–658.
Nadiri, A.A., Naderi, K., Khatibi, R., Gharekhani, M., 2019. Modelling groundwater level variations by learning from multiple models using fuzzy logic. Hydrol. Sci. J. 64 (2), 210–226.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models Part I a discussion of principles. J. Hydrol. 10 (3), 282–290.
Natarajan, N., Sudheer, C., 2019. Groundwater level forecasting using soft computing techniques. Neural Comput. Appl. 1–18. https://doi.org/10.1007/s00521-019-04234-5.
Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet – artificial Intelligence models in hydrology: a review. J. Hydrol. 514, 358–377.
Oshima, I., 2010. Administration for groundwater management in the Kumamoto area. J. Groundwater Hydrol. 52, 49–64 (in Japanese with English Abstract).
Pal, R., 2016. Predictive Modeling of Drug Sensitivity. Academic Press. https://doi.org/10.1016/C2015-0-04656-1.
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 8, 1226–1238.
Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis (Vol. 4). Cambridge University Press.
Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. rev.: Data Min. Knowl. Discov. 9 (3), e1301. https://doi.org/10.1002/widm.1301.
Quilty, J., Adamowski, J., Boucher, M.A., 2019. A stochastic data-driven ensemble forecasting framework for water resources: a case study using ensemble members derived from a database of deterministic wavelet-based models. Water Resour. Res. 55, 175–202. https://doi.org/10.1029/2018WR023205.
Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J. Hydrol. 563, 336–353.
Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI): a nonlinear input variable selection method for water resources modeling. Water Resour. Res. 52 (3), 2299–2326. https://doi.org/10.1002/2015WR016959.
Raghavendra, N.S., Deka, P.C., 2014. Forecasting monthly groundwater table fluctuations in coastal aquifers using support vector regression. In: International Multi Conference on innovations in engineering and technology (IMCIET-2014) (61-69). Elsevier Science and Technology, Bangalore.
Rahman, A.T.M.S., Ahmed, M.S., Adnan, H.M., Kamruzzaman, M., Khalek, M.A., Mazumder, Q.H., Jahan, C.S., 2018. Modeling the changes in water balance components of the highly irrigated western part of Bangladesh. Hydrol. Earth Syst. Sci. 22 (8), 4213–4228. https://doi.org/10.5194/hess-22-4213-2018.
Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2018.12.037.
Rezaie-balf, M., Naganna, S.R., Ghaemi, A., Deka, P.C., 2017. Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. J. Hydrol. 553, 356–373. https://doi.org/10.1016/j.jhydrol.2017.08.006.
Sahoo, S., Jha, M.K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol. J. 21 (8), 1865–1887.
Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour. Res. 53 (5), 3878–3895.
Salem, G.S.A., Kazama, S., Shahid, S., Dey, N.C., 2018. Impacts of climate change on groundwater level and irrigation cost in a groundwater dependent irrigated region. Agric. Water Manage. 208, 33–42.
Samadianfard, S., Asadi, E., Jarhan, S., Kazemi, H., Kheshtgar, S., Kisi, O., Sajjadi, S., Manaf, A.A., 2018. Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. Soil Tillage Res. 175, 37–50.
Sculley, D., Snoek, J., Wiltschko, A., Rahimi, A., 2018. Winner's curse? on pace, progress, and empirical rigor. In: International Conference on Learning Representations Workshop track (2018), published online: iclr.cc
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.
Shimada, J., 2012. Sustainable management of groundwater resources for 700,000-plus residents: a practical example of the transboundary management of groundwater resources in the Kumamoto area. In: Taniguchi, M., Shiraiwa, T. (Eds.), The Dilemma of Boundaries: Toward a New Concept of Catchment. Springer, JapanJapan, pp. 235–246.
Shiri, J., KişI, Ö, 2011. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Comput. Geosci. 37 (10), 1692–1701.
Shiri, J., Kisi, O., Yoon, H., Lee, K.K., Nazemi, A.H., 2013. Predicting groundwater level fluctuations with meteorological effect implications – a comparative study among soft computing techniques. Comput. Geosci. 56, 32–44.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In Advances in Neural Information Processing Systems, MITpress pp. 2951-2959.
Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods 14 (4), 323.
Sudheer, K.P., Gosain, A.K., Ramasastri, K.S., 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol. Processes 16 (6), 1325–1330.
Sun, Y., Wendi, D., Kim, D.E., Liong, S.Y., 2016. Application of artificial neural networks in groundwater table forecasting–a case study in a Singapore swamp forest. Hydrol. Earth Syst. Sci. 20 (4), 1405–1412. https://doi.org/10.5194/hess-20-1405-2016.

Suryanarayana, C., Sudheer, C., Mahammood, V., Panigrahi, B.K., 2014. An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing 145, 324–335.

Tabari, H., Kisi, O., Ezani, A., Talaee, P.H., 2012. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J. Hydrol. 444, 78–89. https://doi.org/10.1016/j.jhydrol.2012.04.007.

Tang, Y., Zang, C., Wei, Y., Jiang, M., 2019. Data-driven modeling of groundwater level with least-square support vector machine and spatial–temporal analysis. Geotech. Geol. Eng. 37 (3), 661–1670. https://doi.org/10.1007/s10706-018-0713-6.

Taniguchi, M., Burnett, K., Shimada, J., Hosono, T., Wada, C.A., Ide, K., 2019. Recovery of lost nexus synergy via payment for environmental services in Kumamoto, Japan. Front. Environ. Sci. 7, 28. https://doi.org/10.3389/fenvs.2019.00028.

Taniguchi, M., Shimada, J., Uemura, T., 2003. Transient effects of surface temperature and groundwater flow on subsurface temperature in Kumamoto Plain, Japan. Phys. Chem. Earth, Parts A/B/C 28 (9–11), 477–486.

Taormina, R., Galelli, S., Karakaya, G., Ahipasaoglu, S.D., 2016. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models. J. Hydrol. 542, 18–34. https://doi.org/10.1016/j.jhydrol.2016.07.045.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S.M., 2010. HydroGeoSphere: a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulations Group, University of Waterloo, Waterloo, ON.

Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J. Hydrol. 394, 458–470.

Tyralis, H., Papacharalampous, G., 2017. Variable selection in time series forecasting using random forests. Algorithms 10 (4), 114. https://doi.org/10.3390/a10040114.

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. Super learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms. arXiv preprint arXiv:1909.04131.

Wang, X., Liu, T., Zheng, X., Hui Peng, H., Xin, J., Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. Appl. Water Sci. (2018) 8, 125. https://doi.org/10.1007/s13201-018-0742-6.

White, J.T., Knowling, M.J., Moore, C.R., 2019. Consequences of groundwater-model vertical discretization in risk-based decision making. Groundwater. 10.1111/gwat.12957

Wilson, S., 2019. Parallel bayesian optimization of hyperparameters. A R-package, version 0.2.0. https://cran.r-project.org/web/packages/ParBayesianOptimization/ParBayesianOptimization.pdf.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufman, Burlington, MA.

Wright, M.N., Ziegler, A., 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 77 (1), 1–17. https://doi.org/10.18637/jss.v077.i01.

Wu, L., Fan, J., 2019. Comparison of neuron-based, kernel-based, tree-based and curve based machine learning models for predicting daily reference evapotranspiration. PLoS One 14 (5), e0217520.

Wu, L., Peng, Y., Fan, J., Wang, Y., 2019. Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. Hydrol. Res.10.2166/nh.2019.060.

Wunsch, A., Liesch, T., Broda, S., 2018. Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). J. Hydrol. 567, 743–758.

Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. J. Hydrol. 530, 829–844.

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-Shafie, A., 2016. Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J. Hydrol. 542, 603–614. https://doi.org/10.1016/j.jhydrol.2016.09.035.

Yoon, H., Hyun, Y., Ha, K., Lee, K.K., Kim, G.B., 2016. A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions. Comput. Geosci. 90, 144–155.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., Lee, K.K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol. 396 (1–2), 128–138.

Zhang, F., Deb, C., Lee, S.E., Yang, J., Shah, K.W., 2016. Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique. Energy Build. 126, 94–103.

Zhang, H., Yang, Q., Shao, J., Wang, G., 2019. Dynamic streamflow simulation via online gradient-boosted regression tree. J. Hydrol. Eng. 24 (10), 04019041. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001822.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018. Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929.

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137–152. https://doi.org/10.1016/j.jhydrol.2015.09.047.

Zhu, R., Yang, L., Liu, T., Wen, X., Zhang, L., Chang, Y., 2019. Hydrological responses to the future climate change in a data scarce region, Northwest China: application of machine learning models. Water 11 (8), 1588.