# Performance improvement of machine learning models via wavelet theory in estimating monthly river streamflow

Kegang Wang, Shahab S. Band, Rasoul Ameri, Meghdad Biyari, Tao Hai, Chung-Chian Hsu, Myriam Hadjouni, Hela Elmannai, Kwok-Wing Chau & Amir Mosavi

Taylor & Francis
Taylor & Francis Group

∂ OPEN ACCESS   Check for updates

# Performance improvement of machine learning models via wavelet theory in estimating monthly river streamflow

Kegang Wang[a], Shahab S. Band[b], Rasoul Ameri[b], Meghdad Biyari[b], Tao Hai[c,d,e], Chung-Chian Hsu[f], Myriam Hadjouni[g], Hela Elmannai[h], Kwok-Wing Chau [i] and Amir Mosavi [j,k,l]

[a]School of Electronics and Information Engineering, Ankang University, Ankang, People's Republic of China; [b]Future Technology Research Center, National Yunlin University of Science and Technology, Douliou, Taiwan; [c]School of Electronics and Information Engineering, Ankang University, Ankang, People's Republic of China; [d]School of Computer Sciences, Baoji University of Arts and Sciences, Baoji, People's Republic of China; [e]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Shah Alam, Malaysia; [f]Department of Information Management, International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu, Taiwan; [g]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; [h]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; [i]Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China; [j]John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary; [k]Institute of the Information Society, University of Public Service, Budapest, Hungary; [l]Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia

## ABSTRACT

River streamflow is an essential hydrological parameters for optimal water resource management. This study investigates models used to estimate monthly time-series river streamflow data at two hydrological stations in the USA (Heise and Irwin on Snake River, Idaho). Five diverse types of machine learning (ML) model were tested, support vector machine-radial basis function (SVM-RBF), SVM-Polynomial (SVM-Poly), decision tree (DT), gradient boosting (GB), random forest (RF), and long short-term memory (LSTM). These were trained and tested alongside a conventional multiple linear regression (MLR). To improve the estimation and model performance, hybrid models were designed by coupling the models with wavelet theory (W). The models performance was assessed using root mean square error (RMSE), mean absolute error (MAE), coefficient of determination ($R^2$), Nash-Sutcliffe efficiency (NSE), and Willmott's index (WI). A side-by-side performance assessment of the stand-alone and hybrid models revealed that the coupled models exhibit better estimates of monthly river streamflow relative to the stand-alone ones. The statistical parameter values for the best model (W-LSTM4) during the test phase was RMSE = 36.533 m$^3$/s, MAE = 26.912 m$^3$/s, R$^2$ = 0.947, NSE = 0.946, WI = 0.986 (Heise station), and RMSE = 33.378 m$^3$/s, MAE = 24.562 m$^3$/s, R$^2$ = 0.952, NSE = 0.951, WI = 0.987 (Irwin station).

## 1. Introduction

Precise estimation of streamflow time series data for rivers is one of the most important issues for optimal management of surface water resources; in particular, making appropriate decisions when dealing with floods and droughts. The river streamflow phenomenon appears complex, non-stationary, and non-linear (Adnan et al., 2019; Bayazit, 2015; Meira Neto et al., 2018); because the river streamflow time series data can be influenced by a variety of parameters such as temperature, rainfall, and evaporation, these render it nearly impossible to estimate accurately. However, in the age of multi-threaded parallel computing, it is now possible

to deploy powerful mathematical and machine learning approaches to the issue.

In general, two different are used, these are conceptual (physical) models, and machine learning (data-driven). These are both proposed, and used by hydrologists, for estimating river streamflow (He et al., 2014; Reis et al., 2021; Zhang et al., 2016). Conceptual paradigms are usually complex models that require many hydrological and climatological parameters as inputs, many of which may be unavailable for certain locations. In addition, the inherent complexity of streamflow processes make it challenging to use physical models; accordingly, in recent years, researchers have shown increasing interest

**CONTACT**   Shahab S. Band   ✉ shamshirbands@yuntech.edu.tw; Chung-Chian Hsu ✉ hsucc@yuntech.edu.tw; Amir Mosavi ✉ amir.mosavi@nik.uni-obuda.hu

in machine learning models to estimate these time series solutions (Kalra et al., 2013; Mehdizadeh & Sales, 2018; Qu et al., 2021; Rasouli et al., 2012; Sahour et al., 2021; Wang et al., 2019; Xiang & Demir, 2020; Zhu et al., 2020). The most important reason for using machine learning, is that these models can estimate target parameters using relatively limited data series, without the need to for additional data regarding the (potentially) complex relationships between inputs and outputs (Deng et al., 2021; Deng et al., 2022; Singh et al., 2022).

More recently, hybrid models have received attention from researchers for river streamflow estimation. For example, Kilinc and Haznedar (Kilinc & Haznedar, 2022) proposed a hybrid approach that integrated long short-term memory (LSTM) and a genetic algorithm (GA) for streamflow estimation of the Euphrates River, Turkey. Particle swarm optimization (PSO) was coupled with LSTM by Kilinc (Kilinc, 2022), to develop a hybrid PSO-LSTM for forecasting the streamflow of the Orontes River basin, Turkey. Feng et al. (2021) optimized the parameters of an LSTM with a PSO algorithm for runoff prediction in the Jiulong River Basin, China. Huang et al. (2014) successfully predicted monthly streamflow at three hydrometric stations in China, using a hybrid empirical and decomposition-support vector machine (EMD-SVM). Wang et al. (2015) increased the forecasting accuracy of a time series model using an autoregressive integrated moving average (ARIMA) variant that also leveraged ensemble empirical mode decomposition (EEMD). This was able to predict the runoff values of three reservoirs, China. Wang et al. (2013) simulated the rainfall-runoff process of a catchment in the Yellow River, China, by coupling both a PSO and EEMD on the stand-alone SVM. A modified form of EMD was used by Meng et al. (2019), where they combined EMD and SVM to predict streamflow in the Wei River Basin, China. The performance of his model was then compared with the stand-alone SVM and ANN. Zhao and Chen (2015) developed two hybrid models via coupling the EMD and EEMD on an auto-regressive (AR) model, for forecasting runoff of from the Fenhe River basin, China. Rezaie-Balf et al. (2019) integrated EEMD into multivariate adaptive regression splines (MARS) and M5Tree models, to forecast the daily river streamflow's of two river basins in South Korea. Fu et al. (2020) tested the performance of an LSTM for simulating the daily streamflow of the Kelantan River, Malaysia.

As alternative approach uses bio-inspired algorithms. For example, Yaseen et al. (2020) used PSO, GA, and differential evolution (DE) to form a hybrid adaptive neuro-fuzzy inference system-based models to forecast the streamflow of the Pahang River, Malaysia. Al-Sudani et al. (2019) integrated the DE algorithm into a MARS

system to estimate the streamflow of the Tigris River, Iraq. Liu et al. (2020) also proposed a hybrid model, by coupling the EMD and Encoder Decoder LSTM to complete a streamflow prediction of the Yangtze River, China. A new model was proposed by Hadi et al. (2019) using a combination of extreme gradient boosting (XGB) and extreme learning machine (ELM); then compared its capabilities with stand-alone models for streamflow predictions of the Goksu-Himmeti basin, Turkey. Aside from river streamflow estimation, other types of hybrid model have been also used by scholars in various fields, such as sediment yield estimation (Meshram et al., 2019), shield movement prediction (Lin et al., 2022), and rainfall simulation (Chen et al., 2022).

It is already well documented, that coupled models can demonstrate improved performance when compared with stand-alone ones. This remains true for many types of time series estimation, not just hydrological parameters such as this study. One method of data pre-processing is to use wavelet transforms. This is an approach that has become widely used to generate wavelet-based hybrid models. One of the main advantages of this method is that wavelets offer simultaneous localization in both the time and frequency domains. The other main advantage is that it applies a fast wavelet transform, which is computationally quick. Wavelets are also able to separate fine details within a signal in a manner similar to an enhanced Fourier transform (Sifuzzaman et al., 2009). Wavelet transforms are an efficient mathematical transformation for signal processing and data pre-processing because they can transform a signal into a set of basic signal functions. The mother wavelet functions similarly and can be used for signal analysis, techniques to achieve this include Daubechies, Symlet, Haar, etc. Utilization of the wavelet theory depends on a number of basic principles, for example observational time series must first be decomposed into several sub-series, and then the generated sub-series need to be considered as new inputs for the machine learning model being used.

With reference to river streamflow data, for the first part of this study, the monthly river streamflow time series for the Heise and Irwin hydrometric stations located on the Snake River, USA, were estimated using five machine learning-based models. These were support vector machine-radial basis function (SVM-RBF), SVM-Polynomial (SVM-Poly), decision tree (DT), gradient boosting (GB), random forest (RF), and long short-term memory (LSTM). A conventional multiple linear regression (MLR) is also commonly applied to most predictive models (Adnan et al., 2019; Kadam et al., 2019). These models are foundational, but that also makes them controversial. Various methods are still debated,

such as SVM being used for non-stationary streamflow prediction, (Adnan et al., 2021; Meng et al., 2019), RF was implemented in (Latif & Ahmed, 2021), GB was used in (Ni et al., 2020), and LSTM being used in (Dong et al., 2020) for streamflow prediction.

Wavelet theory is a valuable pre-processing technique, especially when hybridized with the aforementioned stand-alone models. Its primary role, is improving model performance. A survey of published literature on estimating river streamflow time series models, shows that there are still few studies reporting or evaluating the performance of these models, particularly those that are hybrids and using wavelet theory.

## 2. Materials and methods

### 2.1. Case study and data used

Monthly river streamflow data from two gauging hydrometric stations at Heise and Irwin on Snake River, USA, were used for this study. The data was obtained from the United States Geological Survey (USGS) and is available at https://waterdata.usgs.gov/nwis/. Snake River is one of the largest rivers in the Pacific Northwest region of the USA. The drainage basin of Snake River covers six states, however both selected stations for this study are located in Idaho. Heise station (USGS 13037500) is at latitude 43°36′45″ and longitude 111°39″36″, with a drainage area of 5,752 square miles; and Irwin station (USGS 13032500) is at lat. 43°21″03″, long. 111°13′08″, and is 5,225 square miles. The map in Figure 1 shows the relevant geographical locations.

The data used in this study utilized monthly river streamflow values from Heise and Irwin stations for the period Oct. 1960 to Sep. 2020 (i.e. 720 data). The data was divided into training and testing datasets. In the applications, data from Oct. 1960 to Sep. 2005 (i.e. 540 data) at both stations were utilized as a training set, while the remaining data between Oct. 2005 and Sep. 2020 (i.e. 180 data) constituted the model. Table 1 is an overview of some statistical information regarding the data for both locations, during both the training and testing phases. In general, similar statistics can be observed for the training and testing stages. Monthly river streamflow data for the sites are shown in Figure 2.

### 2.2. Methodology

Wavelet (W) theory was used as a noise removal system for the data. This was implemented as a pre-processing technique for the applied machine learning models. The models used were multiple linear regression (MLR), support vector machine (SVM), decision tree (DT), random forest (RF), gradient boosted decision trees or gradient boosting (GB), and long short-term memory (LSTM). Explanations of the applied methods follow.

### 2.2.1. Pre-processing-based wavelet theory (W)

Wavelet (W) transformation is a powerful technique used in signal processing (Starck & Murtagh, 2001), which is often used to de-noise, compress or decompress data (Daubechies, 2009). Some features of wavelet transformation and Fourier transformation are alike, and thus in time and frequency domain, wavelet analysis can be considered as a type of multi-resolution analysis. In order to decompose signals into different resolutions, both shifting and scaling of a wavelet basis function is required and this creates the mother wavelet (Ebrahimi & Rajaee, 2017).

Streamflow time series are accompanied by noise, which comes in the form of signals with high-frequency. Wavelet transformation is used to remove these and extract high-frequency signals from raw signals, and the process occurs in three main steps. First, by using the selected mother wavelet and level of decomposition, wavelet transformation of the input signals is performed. Second, a threshold is determine and applied to find the amount of high-frequency wavelet transformation. Third, the denoised time series signals are obtained by using low-frequency and high-frequency wavelet coefficients. In this study, the Daubechies wavelet (db4) model was used to determining the optimum level for the mother wavelet, and trial-and-error was used to find the level of decomposition, respectively.

One of the main advantages of wavelets is that they offer simultaneous localization in the time and frequency domains. The second main advantage of wavelets is that, using a fast wavelet transform, it is possible to make calculations very quickly. Wavelets have the great advantage of being able to separate the fine details in a signal. Wavelet transform is a highly efficient mathematical transformation function in signal processing (data pre-processing) and decomposes a signal into its basic signal functions (Singh et al., 2020).

### 2.2.2. Multiple linear regression (MLR)

MLR is implemented to identify the possible existence of relationships between independent and dependent variables. This method is often used as a tool to prove a correlation between the inputs and outputs of a given system (Clarke et al., 1959).

Linear regressions are a form of bivariate model used to predict an independent variable ($y$) from a dependent one ($x$). By extending the model to include more than one explanatory variable ($x_1, x_2, \ldots, x_p$) such as in MLR, a
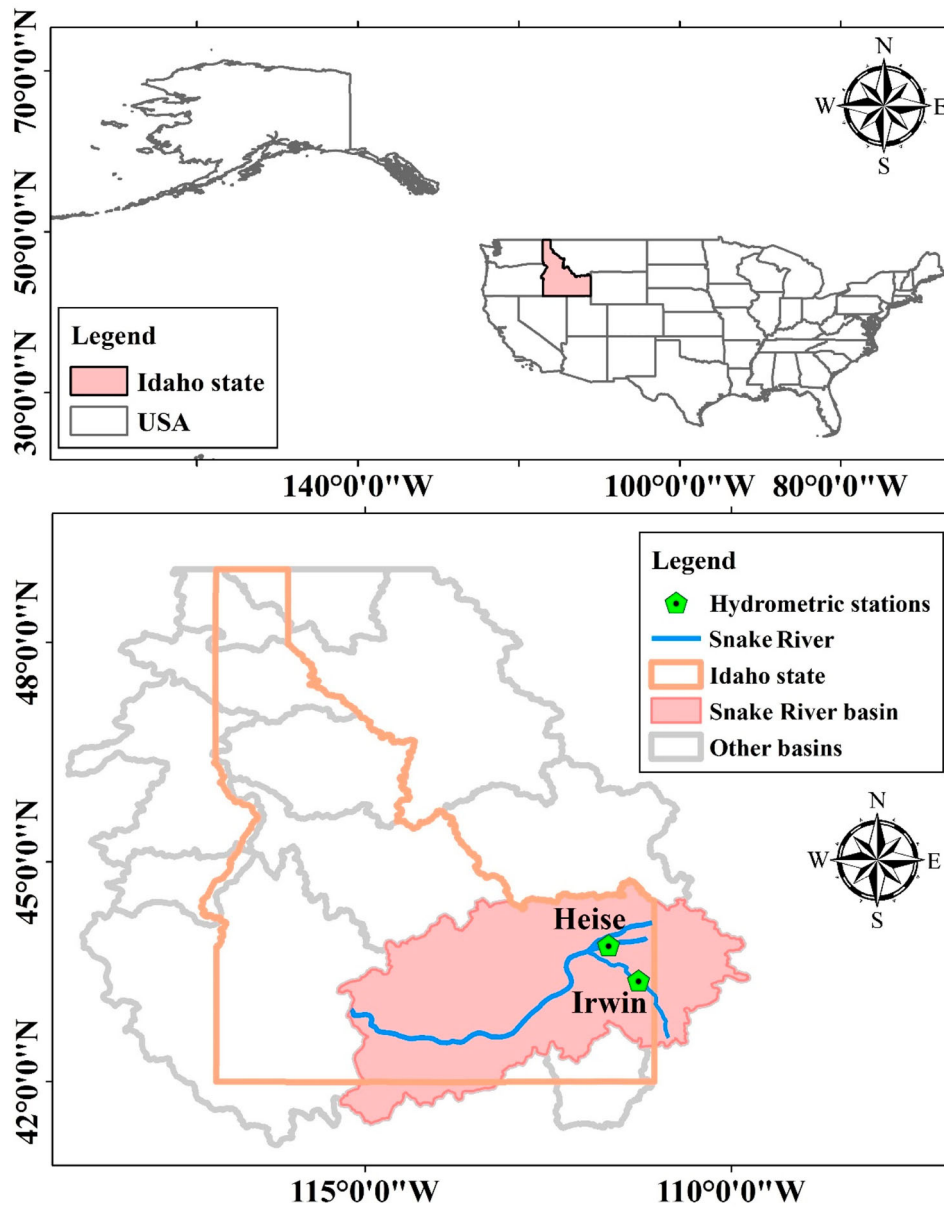
**Figure 1.** Geographical positions of the Heise and Irwin stations on Snake River, USA.

**Table 1.** Monthly time series streamflow statistical data at the Heise and Irwin stations.

| Stations | Stage | Minimum (m$^3$/s) | Maximum (m$^3$/s) | Mean (m$^3$/s) | Standard deviation (m$^3$/s) | Coefficient of variation |
|---|---|---|---|---|---|---|
| Heise | Training | 27.830 | 897.366 | 199.434 | 151.097 | 0.758 |
| | Testing | 31.573 | 616.178 | 203.759 | 157.593 | 0.773 |
| Irwin | Training | 17.180 | 836.767 | 182.457 | 144.450 | 0.792 |
| | Testing | 22.651 | 590.693 | 185.724 | 151.315 | 0.815 |

multivariate model is produced. MLR can be used to discern a linear relationship between two or more independent variables and a dependent variable. Since hydrological variables such as river streamflow can depend heavily on lagged data, river streamflow is generally considered as a dependent variable, $y$, and lagged streamflow data is considered the independent variable, $x_1, x_2, \ldots, x_p$. Thus,

the regression equation is expressed as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p \qquad (1)$$

where $b_0$ is a constant, and $b_1, b_2, \ldots, b_p$ are partial regression coefficients that can be fitted using a least squares approach (Uyanık & Güler, 2013).

**Figure 2.** Monthly time series river streamflow data at Heise and Irwin stations for the training and test periods.

### 2.2.3. Support vector machine (SVM)

Machine learning algorithms such as classifiers, regressions, and outlier detectors utilize SVMs as a set of supervised learning methods (Niu & Feng, 2021). Vapnik et al. introduced this in 1995 to do classification and regression on a set of data points (Cortes & Vapnik, 1995). Since the cost function of a model is no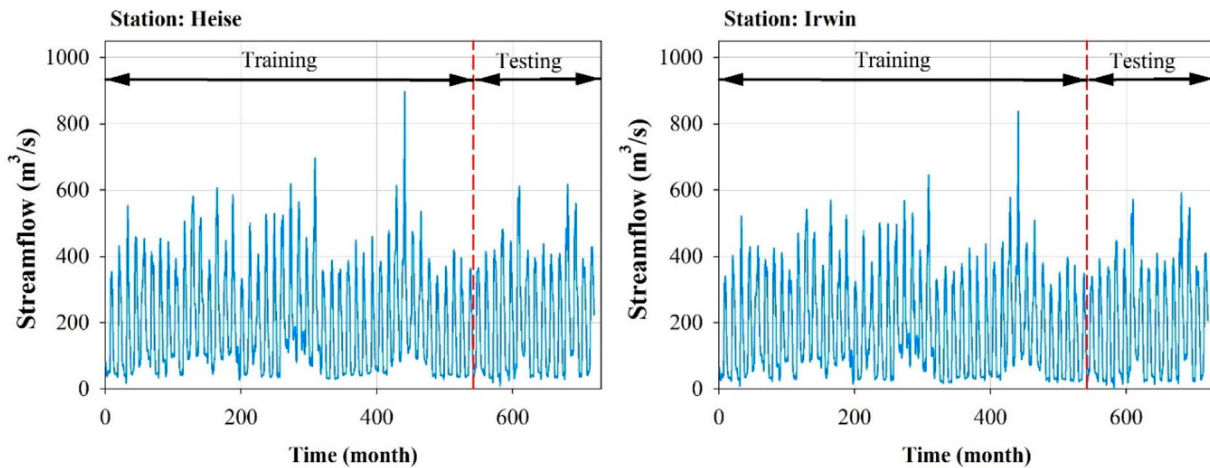t sensitive to training data point positions, models produced using SVM rely on a subset of training data beyond the margin. The intuition behind SVM, is to first obtain an optimized hyper-plan, which is as far as possible from both classes' actual samples. In other words, the learning method maximizes the class margin, based on the choice of the type of margin, which may be either soft or hard. The former SVM misclassifications would probably happen whereas it is not accepted with hard margin SVM (Hamasuna et al., 2008). SVM uses a kernel function to build expert knowledge, which can be considered one of the main advantages of this method.

The most commonly used SVM kernel functions are linear, polynomial, sigmoid, and Radial Basis Function (RBF). This study adopts the RBF and polynomial kernels to estimate the river streamflow, following a previous study by (Adnan et al., 2020; Baesens et al., 2000; Leong et al., 2021), where the efficiency and output quality of the RBF and polynomial kernels were compared with different data sets and yielded good results.

### 2.2.4. Decision tree (DT)

Decision trees are another important, and frequently used, method for supervised learning. It can be used to solve both regression and classification tasks, but the latter is more common. There are three types of nodes, (i) a root node, which represents the entire sample and is the initial node and may split into other nodes. The features of the dataset contain branches that represent the various decision rules, are these are determined using (ii) interior nodes. The final output is delivered at the (iii) leaf nodes.

This algorithmic approach can also be implemented to solve decision-related problems. To check an algorithm with a particular data point, all nodes in the tree need pass a conditional test (true/false) until the leaf node is reached. The final prediction is taken as the average value of the dependent variable in a specific leaf node. In order to predict precise values for the given data points, the algorithm runs for multiple iterations on the tree (Safavian & Landgrebe, 1991).

### 2.2.5. Random forest (RF)

Random forest regressions are another supervised learning algorithm, however this one implements an ensemble learning method to conduct the regression. Ensemble learning methods combine the predictions from multiple machine learning algorithms to make a precise prediction from a single model (Bernard et al., 2009). A meta-estimator, which fits a series of classifying decision trees on different subsets of a given dataset, plays a key role in this algorithm. This type of estimator benefits from averaging techniques to increase its predictive accuracy, and gain control of the over-fitting problem. The trees in a random forests model run in parallel, there is no interaction while they are being built. During training, multiple trees are constructed and then the classification to determine the related class of the data points is performed. After this the whole model can be regressed or particular trees can be selected (Mosavi, et al., 2022). The followings are steps required to implement a random forest algorithm:

(1) Choose $k$ data points randomly from the training set.
(2) Build the corresponding decision tree for the given data.

(3) Repeat steps 1 and 2 to build $N$ trees.

(4) After building $N$ trees, for any new data point, the related prediction of the value of $y$ as it corresponds to the data point can be made. It is then possible to assign the new data point an average value across all the obtained $y$ values.

### 2.2.6. Gradient boosting (GB)

In the field of machine learning, the gradient boosting algorithm can be considered one of the most powerful algorithms. The most common types of allocated errors in machine learning algorithms are bias error and variance error (Islam & Amin, 2020). To decrease the bias error of a model, the gradient boosting algorithm can be used. Gradient boosting algorithm can also target a continuous variable as a regressor, as well as a categorical target variable (as a classifier) (Friedman, 2001). In the former case, the mean square error (MSE) is the cost function, while in the latter it appears as log loss. For both regression and classification problems, GB can show notable outcomes. Steps to implement GB are as follows:

(1) Determine the average value of the target label.
   1. Compute the residuals using 'residual = actual value−predicted value'.
   2. Construct a decision tree.
   3. Use all presented trees within the ensemble to predict the target label.
   4. Compute the new residuals, such that the number of iterations matches the number of estimators by repeating steps 3 through 5.
   5. Assign the value of the target variable using all trees from the training step.

### 2.2.7. Long short-term memory (LSTM)

A suitable performance for streamflow prediction has also been demonstrated by deep learning models that use an appropriate architecture (Ghimire et al., 2021; Lin et al., 2021). Using dynamic system modeling in diverse application areas such as speech recognition, image processing, manufacturing, communication or energy consumption and autonomous systems, recurrent neural network (RNN)-based deep learning models, especially LSTM, play a vital role (Lipton et al., 2017). The setup of prediction models according to time series data can be used to predict non-linear, time variant system outputs for many data types (Lindemann et al., 2021).

LSTMs, which rely on memory blocks in their hidden layers, and perform the same role as neurons in hidden layers, are a class of RNN. There are input, output, and forget gates in the memory blocks. These gates are used for controlling and updating the information through the memory blocks (Hochreiter & Schmidhuber, 1997; Sainath et al., 2015). The equations for determining the gates in an LSTM are:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \qquad (2)$$

where $i_t$, $f_t$, and $o_t$ are the input, forget, and output gates, respectively. $\sigma$ denotes a sigmoid function (gates in LSTM are the sigmoid activation functions). $w_x$ and $b_x$ are weights and biases for the correspondence gate ($x$) neurons. $x_t$ denotes the input at the current timestamp, and $h_{t-1}$ is used for the output of the previous LSTM block at timestamp $t$-1.

The equations for the cell state, candidate cell state and final output are as follows:

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c)$$
$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t$$
$$h_t = o_t \times \tanh(c^t) \qquad (3)$$

where $c_t$ represents the cell state (memory) at timestamp $t$, and $\tilde{c}_t$ represents the candidate for all states at timestamp $t$ (Salem, 2018; Wang et al., 2020).

Overall, LSTMs are an efficient, gradient-based method to handle complex, artificial long-time-lag tasks. In addition, LSTMs are the RNN variant that is capable of learning long-term dependencies because their cells have the ability to retain previous time step information.

### 2.3. Performance evaluation of models

Five evaluation metrics are often used to determine model performance: root mean square error (RMSE), mean absolute error (MAE), coefficient of determination ($R^2$), Nash-Sutcliffe efficiency (NSE), and Willmott's index (WI). Accordingly, these were employed in this study to determine the relative performances of the stand-alone, and coupled models. These determinants are described in the equations below, and follow the form described by (Kim et al., 2019; Willmott et al., 2012; Yaseen et al., 2020):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_{o,i} - Q_{e,i})^2}{N}} \qquad (4)$$

$$MAE = \frac{\sum_{i=1}^{N}|Q_{o,i} - Q_{e,i}|}{N} \qquad (5)$$

$$R^2 = \left[ \frac{\sum_{i=1}^{N}(Q_{o,i} - \overline{Q_o}) \cdot (Q_{e,i} - \overline{Q_e})}{\sqrt{\sum_{i=1}^{N}(Q_{o,i} - \overline{Q_o})^2 \cdot \sum_{i=1}^{N}(Q_{e,i} - \overline{Q_e})^2}} \right]^2$$

(6)

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_{o,i} - Q_{e,i})^2}{\sum_{i=1}^{N}(Q_{o,i} - \overline{Q_o})^2}$$

(7)

$$WI = 1 - \frac{\sum_{i=1}^{N}(Q_{o,i} - Q_{e,i})^2}{\sum_{i=1}^{N}(|Q_{o,i} - \overline{Q_o}|) + (|Q_{o,i} - \overline{Q_o}|)^2}$$

(8)

where $Q_{o,i}$ and $Q_{e,i}$ describe the observed and estimated daily river streamflow for the $i^{th}$ month, respectively. $\overline{Q_o}$ and $\overline{Q_e}$ indicate the average values of the observed and estimated daily river streamflow, respectively. $N$ shows the total number of observations. A better model should present lower RMSE and MAE values, and a higher $R^2$, NSE, and WI.

## 3. Results and discussion

This study utilizes historical monthly streamflow data from 1 to 4 months as input predictors for the models, as shown in Table 2. Setting parameters for each model accordingly will improve the performance of the models. Table 3 summarizes the parameter settings for all the models used in this paper.

Determination of the optimum parameter values changes the efficiency of the model design. To accomplish this, trial-and-error was used to select the optimum value each parameter. For all experiments, the model parameters were independently drawn from a 10-fold cross-validation run on the training set. The parameter ranges are shown in Table 3.

First, a standard MLR and stand-alone machine learning models using SVM-RBF, SVM-Poly, DT, RF, GB, and LSTM were established using the input configurations shown in Table 2. RMSE, MAE, $R^2$, NSE, and WI values for the Heise and Irwin hydrometric stations are presented in Tables 4 and 5, respectively. These stand-alone models were able to estimate daily streamflow using streamflow data of previous daily periods. The tables clearly show that the performance of the models generally improved as the number of inputs increased (from M1 to M4). As a result, the stand-alone models exhibited the potential for estimating river streamflow in the current month, using antecedent monthly streamflow data.

**Table 2.** Input patterns defined in the present study.

| Models abbreviations | Models | Inputs | Output |
|---|---|---|---|
| M1 | MLR1, SVM-RBF1, SVM-Poly1, DT1, RF1, GB1, LSTM1, W-MLR1, W-SVM-RBF1, W-SVM-Poly1, W-DT1, W-RF1, W-GB1, W-LSTM1 | $Q_{t-1}$ | $Q_t$ |
| M2 | MLR2, SVM-RBF2, SVM-Poly2, DT2, RF2, GB2, LSTM2, W-MLR2, W-SVM-RBF2, W-SVM-Poly2, W-DT2, W-RF2, W-GB2, W-LSTM2 | $Q_{t-1}, Q_{t-2}$ | $Q_t$ |
| M3 | MLR3, SVM-RBF3, SVM-Poly3, DT3, RF3, GB3, LSTM3, W-MLR3, W-SVM-RBF3, W-SVM-Poly3, W-DT3, W-RF3, W-GB3, W-LSTM3 | $Q_{t-1}, Q_{t-2}, Q_{t-3}$ | $Q_t$ |
| M4 | MLR4, SVM-RBF4, SVM-Poly4, DT4, RF4, GB4, LSTM4, W-MLR4, W-SVM-RBF4, W-SVM-Poly4, W-DT4, W-RF4, W-GB4, W-LSTM4 | $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$ | $Q_t$ |

**Table 3.** Parameter settings of the models used in this study.

| Models | Parameters | Range |
|---|---|---|
| Support vector machine (SVM) kernel = RBF | C (Regularization parameter) 40 epsilon = 0.1 | $[10^{-2}, 10^2]$ $[10^{-2}, 10^1]$ |
| Support vector machine (SVM) kernel = Poly | degree = 5 coef0 (independent term in kernel function) 0.6 C (Regularization parameter) 5 | [1,2,3,4,5] [0.1,0.9] $[10^{-2}, 10^2]$ |
| Decision Tree (DT) | min_samples_split = 2 criterion = 'mse' splitter = 'best' | [2,10] [mse, mae] |
| Random Forest (RF) | min_samples_split = 4 n_estimators = 100 criterion = 'mse' min_samples_leaf = 1, | [2,10] [10:10:100] [mse, mae] [1,10] |
| Gradient Boosted Decision Trees (GBDT) | loss = 'squared_error' learning_rate = 0.1 max_iter = 100 max_leaf_nodes = 31 min_samples_leaf = 20 validation_fraction = 0.1 | [squared_error, absolute_error] [0.01, 1] [10, 100] [1, 50] [1, 100] [0.1, 0.2] |

**LSTM Structures**

| Layer (type) | Output Shape | Param # | |
|---|---|---|---|
| lstm (LSTM) | (None, 1, 6) | 264 | LSTM inputs: [4, 50] |
| lstm_1 (LSTM) | (None, 4) | 176 | |
| dense (Dense) | (None, 4) | 20 | |
| dense_1 (Dense) | (None, 1) | 5 | |
| Total params: 465 | | | |
| Trainable params: 465 | | | |
| Non-trainable params: 0 | | | |
| epochs = 500 | | | [50, 1000] |
| batch size = 4 | | | [4,8,16,32] |
| validation split = 0.3 | | | |
| optimizer = adam | | | |
| loss = mean square error | | | |

**Table 4.** Values of evaluation metrics for the Heise station model.

| Models abbreviations | Models | Training | | | | | Testing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI |
| M1 | MLR1 | 94.224 | 71.026 | 0.610 | 0.610 | 0.868 | 94.895 | 71.001 | 0.636 | 0.635 | 0.879 |
| | SVM-RBF1 | 97.534 | 62.256 | 0.621 | 0.583 | 0.865 | 96.823 | 61.111 | 0.650 | 0.620 | 0.883 |
| | SVM-Poly1 | 98.680 | 64.385 | 0.613 | 0.573 | 0.866 | 97.859 | 62.502 | 0.641 | 0.612 | 0.882 |
| | DT1 | 76.218 | 54.133 | 0.745 | 0.745 | 0.923 | 98.105 | 66.908 | 0.632 | 0.610 | 0.890 |
| | RF1 | 54.125 | 39.121 | 0.876 | 0.871 | 0.962 | 97.585 | 67.443 | 0.628 | 0.614 | 0.889 |
| | GB1 | 81.614 | 59.970 | 0.708 | 0.708 | 0.906 | 90.625 | 63.664 | 0.669 | 0.667 | 0.898 |
| | LSTM1 | 92.231 | 67.765 | 0.627 | 0.627 | 0.874 | 93.158 | 66.873 | 0.650 | 0.649 | 0.887 |
| | W-MLR1 | 85.963 | 66.626 | 0.676 | 0.676 | 0.896 | 88.079 | 66.630 | 0.686 | 0.686 | 0.900 |
| | W-SVM-RBF1 | 91.661 | 61.011 | 0.668 | 0.631 | 0.883 | 93.690 | 61.446 | 0.674 | 0.645 | 0.889 |
| | W-SVM-Poly1 | 91.699 | 61.422 | 0.669 | 0.631 | 0.886 | 94.134 | 61.405 | 0.671 | 0.641 | 0.890 |
| | W-DT1 | 67.639 | 46.091 | 0.799 | 0.799 | 0.942 | 96.210 | 69.202 | 0.640 | 0.625 | 0.892 |
| | W-RF1 | 49.099 | 35.442 | 0.897 | 0.894 | 0.970 | 97.377 | 70.746 | 0.630 | 0.616 | 0.888 |
| | W-GB1 | 76.054 | 56.296 | 0.747 | 0.746 | 0.921 | 92.409 | 68.576 | 0.656 | 0.654 | 0.892 |
| | W-LSTM1 | 85.306 | 66.274 | 0.681 | 0.681 | 0.896 | 88.065 | 66.385 | 0.686 | 0.686 | 0.900 |
| M2 | MLR2 | 80.064 | 58.394 | 0.719 | 0.719 | 0.913 | 78.456 | 55.756 | 0.752 | 0.751 | 0.923 |
| | SVM-RBF2 | 71.928 | 44.481 | 0.790 | 0.773 | 0.931 | 77.892 | 45.048 | 0.778 | 0.754 | 0.927 |
| | SVM-Poly2 | 84.628 | 52.512 | 0.712 | 0.686 | 0.904 | 88.758 | 53.179 | 0.707 | 0.681 | 0.903 |
| | DT2 | 52.251 | 34.082 | 0.880 | 0.880 | 0.967 | 74.458 | 47.755 | 0.780 | 0.776 | 0.939 |
| | RF2 | 32.012 | 20.751 | 0.956 | 0.955 | 0.988 | 68.101 | 43.069 | 0.814 | 0.812 | 0.948 |
| | GB2 | 44.941 | 29.875 | 0.912 | 0.911 | 0.976 | 65.396 | 40.975 | 0.828 | 0.827 | 0.952 |
| | LSTM2 | 61.152 | 42.186 | 0.836 | 0.836 | 0.953 | 63.428 | 41.731 | 0.838 | 0.837 | 0.953 |
| | W-MLR2 | 56.532 | 39.845 | 0.860 | 0.860 | 0.961 | 52.105 | 36.812 | 0.892 | 0.890 | 0.969 |
| | W-SVM-RBF2 | 49.367 | 31.872 | 0.902 | 0.893 | 0.969 | 46.166 | 30.617 | 0.923 | 0.914 | 0.976 |
| | W-SVM-Poly2 | 62.797 | 39.158 | 0.836 | 0.827 | 0.951 | 61.979 | 39.705 | 0.855 | 0.844 | 0.956 |
| | W-DT2 | 38.435 | 24.605 | 0.935 | 0.935 | 0.983 | 53.028 | 36.682 | 0.886 | 0.886 | 0.969 |
| | W-RF2 | 20.946 | 13.844 | 0.981 | 0.981 | 0.995 | 43.708 | 29.733 | 0.923 | 0.923 | 0.979 |
| | W-GB2 | 31.483 | 20.011 | 0.957 | 0.957 | 0.989 | 44.451 | 31.367 | 0.920 | 0.920 | 0.979 |
| | W-LSTM2 | 39.872 | 27.256 | 0.930 | 0.930 | 0.982 | 38.668 | 28.630 | 0.940 | 0.939 | 0.984 |

| Models abbreviations | Models | Training | | | | | Testing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI |
| M3 | MLR3 | 78.565 | 57.330 | 0.729 | 0.729 | 0.916 | 75.184 | 53.031 | 0.773 | 0.771 | 0.930 |
| | SVM-RBF3 | 72.096 | 43.634 | 0.792 | 0.772 | 0.929 | 77.832 | 44.347 | 0.790 | 0.755 | 0.925 |
| | SVM-Poly3 | 82.722 | 51.633 | 0.721 | 0.700 | 0.907 | 85.660 | 51.086 | 0.730 | 0.703 | 0.908 |
| | DT3 | 49.484 | 31.697 | 0.893 | 0.893 | 0.971 | 68.696 | 44.180 | 0.811 | 0.809 | 0.947 |
| | RF3 | 29.089 | 18.819 | 0.964 | 0.963 | 0.990 | 62.654 | 39.320 | 0.842 | 0.841 | 0.956 |
| | GB3 | 38.016 | 24.890 | 0.937 | 0.937 | 0.983 | 63.728 | 41.559 | 0.838 | 0.836 | 0.953 |
| | LSTM3 | 67.684 | 47.859 | 0.799 | 0.799 | 0.941 | 68.000 | 45.386 | 0.815 | 0.813 | 0.945 |
| | W-MLR3 | 55.913 | 39.163 | 0.863 | 0.863 | 0.962 | 51.112 | 37.280 | 0.897 | 0.894 | 0.970 |
| | W-SVM-RBF3 | 50.164 | 32.305 | 0.899 | 0.890 | 0.968 | 46.850 | 30.790 | 0.924 | 0.911 | 0.974 |
| | W-SVM-Poly3 | 56.261 | 35.872 | 0.865 | 0.861 | 0.962 | 60.144 | 39.104 | 0.861 | 0.854 | 0.958 |
| | W-DT3 | 37.582 | 24.171 | 0.938 | 0.938 | 0.984 | 53.092 | 36.719 | 0.886 | 0.886 | 0.969 |
| | W-RF3 | 19.831 | 12.655 | 0.983 | 0.983 | 0.996 | 41.234 | 27.981 | 0.932 | 0.931 | 0.981 |
| | W-GB3 | 26.882 | 16.464 | 0.968 | 0.968 | 0.992 | 43.289 | 30.161 | 0.924 | 0.924 | 0.980 |
| | W-LSTM3 | 40.191 | 28.152 | 0.930 | 0.929 | 0.981 | 38.682 | 28.297 | 0.940 | 0.939 | 0.984 |
| M4 | MLR4 | 77.151 | 57.226 | 0.739 | 0.739 | 0.920 | 71.737 | 50.569 | 0.793 | 0.792 | 0.937 |
| | SVM-RBF4 | 69.957 | 42.731 | 0.803 | 0.785 | 0.933 | 74.559 | 42.905 | 0.808 | 0.775 | 0.930 |
| | SVM-Poly4 | 81.422 | 50.617 | 0.729 | 0.709 | 0.910 | 84.384 | 50.863 | 0.745 | 0.712 | 0.909 |
| | DT4 | 43.156 | 28.406 | 0.918 | 0.918 | 0.978 | 69.364 | 42.013 | 0.810 | 0.805 | 0.946 |
| | RF4 | 28.742 | 18.109 | 0.966 | 0.964 | 0.990 | 57.935 | 34.238 | 0.866 | 0.864 | 0.963 |
| | GB4 | 33.640 | 21.253 | 0.951 | 0.950 | 0.987 | 62.208 | 38.401 | 0.849 | 0.843 | 0.955 |
| | LSTM4 | 67.532 | 47.235 | 0.800 | 0.800 | 0.942 | 62.934 | 42.862 | 0.842 | 0.840 | 0.954 |
| | W-MLR4 | 51.079 | 33.680 | 0.886 | 0.886 | 0.969 | 42.856 | 30.439 | 0.927 | 0.926 | 0.980 |
| | W-SVM-RBF4 | 50.932 | 32.556 | 0.900 | 0.886 | 0.966 | 49.867 | 33.118 | 0.922 | 0.899 | 0.970 |
| | W-SVM-Poly4 | 56.087 | 34.670 | 0.867 | 0.862 | 0.962 | 59.020 | 36.011 | 0.870 | 0.859 | 0.960 |
| | W-DT4 | 36.541 | 22.910 | 0.941 | 0.941 | 0.985 | 53.851 | 36.600 | 0.883 | 0.883 | 0.968 |
| | W-RF4 | 19.203 | 12.038 | 0.985 | 0.984 | 0.996 | 41.294 | 28.036 | 0.932 | 0.931 | 0.981 |
| | W-GB4 | 24.522 | 14.680 | 0.974 | 0.974 | 0.993 | 43.894 | 31.224 | 0.923 | 0.922 | 0.979 |
| | W-LSTM4 | 37.995 | 26.387 | 0.937 | 0.937 | 0.983 | **36.533** | **26.912** | **0.947** | **0.946** | **0.986** |

N.B. Values in boldface indicate the statistical parameters of the best model during testing.

In addition to the stand-alone models, the models can be fine-tuned to increase accuracy. To achieve this purpose, wavelet theory (W) was included. The hybrid W-MLR, W-SVM-RBF, W-SVM-Poly, W-DT, W-RF, W-GB, and W-LSTM details are established in this section.

Daubechies (db4) is used in this study as the mother wavelet, and values of RMSE, MAE, $R^2$, NSE and WI are used to measure the product of the model. The results are shown in Tables 4 and 5, for Heise and Irwin stations, respectively. Comparing the

**Table 5.** Values of evaluation metrics for the Irwin station model.

| Models abbreviations | Models | Training | | | | | Testing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI |
| M1 | MLR1 | 90.849 | 68.819 | 0.604 | 0.604 | 0.865 | 91.877 | 69.071 | 0.630 | 0.629 | 0.876 |
| | SVM-RBF1 | 93.433 | 60.298 | 0.615 | 0.581 | 0.867 | 93.058 | 59.425 | 0.646 | 0.620 | 0.884 |
| | SVM-Poly1 | 95.143 | 62.579 | 0.608 | 0.565 | 0.865 | 94.647 | 61.002 | 0.638 | 0.607 | 0.880 |
| | DT1 | 71.780 | 49.896 | 0.753 | 0.753 | 0.925 | 92.976 | 63.932 | 0.629 | 0.620 | 0.886 |
| | RF1 | 50.802 | 37.240 | 0.881 | 0.876 | 0.964 | 97.686 | 66.099 | 0.594 | 0.581 | 0.875 |
| | GB1 | 78.396 | 57.535 | 0.706 | 0.705 | 0.904 | 92.664 | 63.985 | 0.625 | 0.623 | 0.881 |
| | LSTM1 | 88.721 | 65.727 | 0.622 | 0.622 | 0.873 | 89.854 | 65.125 | 0.646 | 0.645 | 0.886 |
| | W-MLR1 | 82.264 | 64.079 | 0.675 | 0.675 | 0.896 | 84.461 | 64.514 | 0.687 | 0.687 | 0.900 |
| | W-SVM-RBF1 | 88.264 | 59.389 | 0.670 | 0.626 | 0.878 | 90.447 | 60.378 | 0.678 | 0.641 | 0.884 |
| | W-SVM-Poly1 | 87.558 | 59.535 | 0.668 | 0.632 | 0.887 | 90.178 | 59.965 | 0.671 | 0.643 | 0.891 |
| | W-DT1 | 64.836 | 46.004 | 0.798 | 0.798 | 0.941 | 103.050 | 74.156 | 0.564 | 0.534 | 0.865 |
| | W-RF1 | 45.844 | 33.886 | 0.902 | 0.899 | 0.971 | 102.724 | 74.405 | 0.562 | 0.537 | 0.863 |
| | W-GB1 | 71.893 | 54.206 | 0.752 | 0.752 | 0.923 | 91.140 | 69.175 | 0.638 | 0.635 | 0.886 |
| | W-LSTM1 | 81.647 | 64.493 | 0.680 | 0.680 | 0.896 | 84.527 | 64.985 | 0.686 | 0.686 | 0.900 |
| M2 | MLR2 | 78.244 | 57.159 | 0.706 | 0.706 | 0.908 | 77.725 | 55.233 | 0.736 | 0.735 | 0.917 |
| | SVM-RBF2 | 70.165 | 43.332 | 0.780 | 0.764 | 0.929 | 76.242 | 43.725 | 0.768 | 0.745 | 0.924 |
| | SVM-Poly2 | 81.250 | 50.953 | 0.707 | 0.683 | 0.903 | 85.920 | 51.833 | 0.699 | 0.676 | 0.900 |
| | DT2 | 52.417 | 34.414 | 0.868 | 0.868 | 0.964 | 74.547 | 48.374 | 0.758 | 0.756 | 0.930 |
| | RF2 | 31.025 | 20.761 | 0.956 | 0.954 | 0.988 | 65.864 | 41.919 | 0.810 | 0.809 | 0.947 |
| | GB2 | 44.150 | 29.818 | 0.907 | 0.906 | 0.974 | 67.710 | 43.624 | 0.799 | 0.799 | 0.942 |
| | LSTM2 | 63.281 | 44.568 | 0.808 | 0.808 | 0.945 | 67.501 | 45.882 | 0.801 | 0.800 | 0.941 |
| | W-MLR2 | 52.810 | 37.089 | 0.866 | 0.866 | 0.963 | 49.732 | 35.216 | 0.893 | 0.891 | 0.970 |
| | W-SVM-RBF2 | 45.919 | 30.353 | 0.905 | 0.899 | 0.971 | 44.512 | 29.676 | 0.920 | 0.913 | 0.976 |
| | W-SVM-Poly2 | 58.020 | 36.559 | 0.846 | 0.838 | 0.955 | 57.802 | 37.650 | 0.861 | 0.853 | 0.959 |
| | W-DT2 | 35.864 | 23.452 | 0.938 | 0.938 | 0.984 | 49.795 | 34.474 | 0.892 | 0.891 | 0.971 |
| | W-RF2 | 19.906 | 13.165 | 0.981 | 0.981 | 0.995 | 39.403 | 28.186 | 0.933 | 0.932 | 0.982 |
| | W-GB2 | 29.021 | 18.566 | 0.960 | 0.960 | 0.989 | 43.657 | 31.012 | 0.916 | 0.916 | 0.978 |
| | W-LSTM2 | 46.431 | 31.519 | 0.897 | 0.896 | 0.972 | 43.839 | 31.243 | 0.916 | 0.916 | 0.978 |

| Models abbreviations | Models | Training | | | | | Testing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI | RMSE (m³/s) | MAE (m³/s) | $R^2$ | NSE | WI |
| M3 | MLR3 | 76.313 | 55.851 | 0.720 | 0.720 | 0.913 | 73.830 | 52.272 | 0.762 | 0.761 | 0.926 |
| | SVM-RBF3 | 70.184 | 42.370 | 0.781 | 0.763 | 0.928 | 75.386 | 42.451 | 0.781 | 0.750 | 0.924 |
| | SVM-Poly3 | 79.894 | 49.519 | 0.717 | 0.694 | 0.904 | 84.115 | 49.783 | 0.714 | 0.689 | 0.903 |
| | DT3 | 47.743 | 29.328 | 0.891 | 0.891 | 0.970 | 73.174 | 44.720 | 0.770 | 0.765 | 0.935 |
| | RF3 | 28.738 | 18.794 | 0.962 | 0.960 | 0.989 | 61.999 | 38.701 | 0.832 | 0.831 | 0.953 |
| | GB3 | 38.462 | 25.525 | 0.930 | 0.929 | 0.981 | 61.320 | 39.838 | 0.836 | 0.835 | 0.953 |
| | LSTM3 | 66.287 | 46.693 | 0.789 | 0.789 | 0.938 | 67.877 | 45.252 | 0.800 | 0.798 | 0.940 |
| | W-MLR3 | 52.315 | 36.495 | 0.869 | 0.869 | 0.964 | 48.989 | 35.484 | 0.896 | 0.895 | 0.970 |
| | W-SVM-RBF3 | 46.971 | 30.457 | 0.902 | 0.894 | 0.970 | 45.237 | 29.666 | 0.922 | 0.910 | 0.974 |
| | W-SVM-Poly3 | 53.227 | 33.857 | 0.868 | 0.864 | 0.963 | 56.395 | 36.165 | 0.866 | 0.860 | 0.961 |
| | W-DT3 | 37.627 | 24.818 | 0.932 | 0.932 | 0.982 | 48.072 | 33.848 | 0.899 | 0.899 | 0.972 |
| | W-RF3 | 19.295 | 12.603 | 0.983 | 0.982 | 0.995 | 39.500 | 27.564 | 0.932 | 0.931 | 0.982 |
| | W-GB3 | 24.957 | 15.678 | 0.970 | 0.970 | 0.992 | 41.096 | 29.054 | 0.926 | 0.926 | 0.980 |
| | W-LSTM3 | 36.499 | 26.348 | 0.937 | 0.936 | 0.983 | 39.345 | 29.489 | 0.934 | 0.932 | 0.982 |
| M4 | MLR4 | 74.803 | 55.522 | 0.731 | 0.731 | 0.917 | 70.203 | 49.800 | 0.785 | 0.784 | 0.934 |
| | SVM-RBF4 | 68.418 | 41.232 | 0.792 | 0.775 | 0.931 | 72.660 | 41.235 | 0.800 | 0.768 | 0.929 |
| | SVM-Poly4 | 79.401 | 49.445 | 0.721 | 0.697 | 0.904 | 83.035 | 50.099 | 0.728 | 0.697 | 0.904 |
| | DT4 | 50.252 | 30.013 | 0.879 | 0.879 | 0.967 | 65.130 | 40.460 | 0.817 | 0.814 | 0.948 |
| | RF4 | 27.368 | 18.138 | 0.966 | 0.964 | 0.990 | 59.657 | 35.047 | 0.846 | 0.844 | 0.956 |
| | GB4 | 33.306 | 21.685 | 0.948 | 0.947 | 0.986 | 56.944 | 34.709 | 0.863 | 0.858 | 0.960 |
| | LSTM4 | 65.103 | 45.719 | 0.797 | 0.797 | 0.940 | 61.534 | 42.051 | 0.836 | 0.834 | 0.951 |
| | W-MLR4 | 48.168 | 31.876 | 0.889 | 0.889 | 0.970 | 41.830 | 29.692 | 0.924 | 0.923 | 0.979 |
| | W-SVM-RBF4 | 47.602 | 30.645 | 0.901 | 0.891 | 0.968 | 47.180 | 31.255 | 0.919 | 0.902 | 0.971 |
| | W-SVM-Poly4 | 53.683 | 33.151 | 0.868 | 0.862 | 0.962 | 56.458 | 34.107 | 0.872 | 0.860 | 0.961 |
| | W-DT4 | 36.743 | 23.623 | 0.935 | 0.935 | 0.983 | 48.889 | 33.729 | 0.895 | 0.895 | 0.972 |
| | W-RF4 | 18.938 | 12.072 | 0.984 | 0.983 | 0.996 | 38.451 | 26.742 | 0.937 | 0.935 | 0.982 |
| | W-GB4 | 22.561 | 13.712 | 0.976 | 0.976 | 0.994 | 40.805 | 29.284 | 0.928 | 0.927 | 0.980 |
| | W-LSTM4 | 35.604 | 24.953 | 0.940 | 0.939 | 0.984 | **33.378** | **24.562** | **0.952** | **0.951** | **0.987** |

N.B. Values in boldface indicate the statistical parameters of the best model during testing.

estimation accuracy of stand-alone and hybrid models in Tables 4(a) and Table 5(a) (i.e. using only the streamflow data of the preceding month), it can be observed that wavelet theory shows relatively little ability to improve the performance of the single models, and in very few cases, it even slightly reduces estimation accuracy. However, wavelets can enhance the accuracy of stand-alone models using monthly data
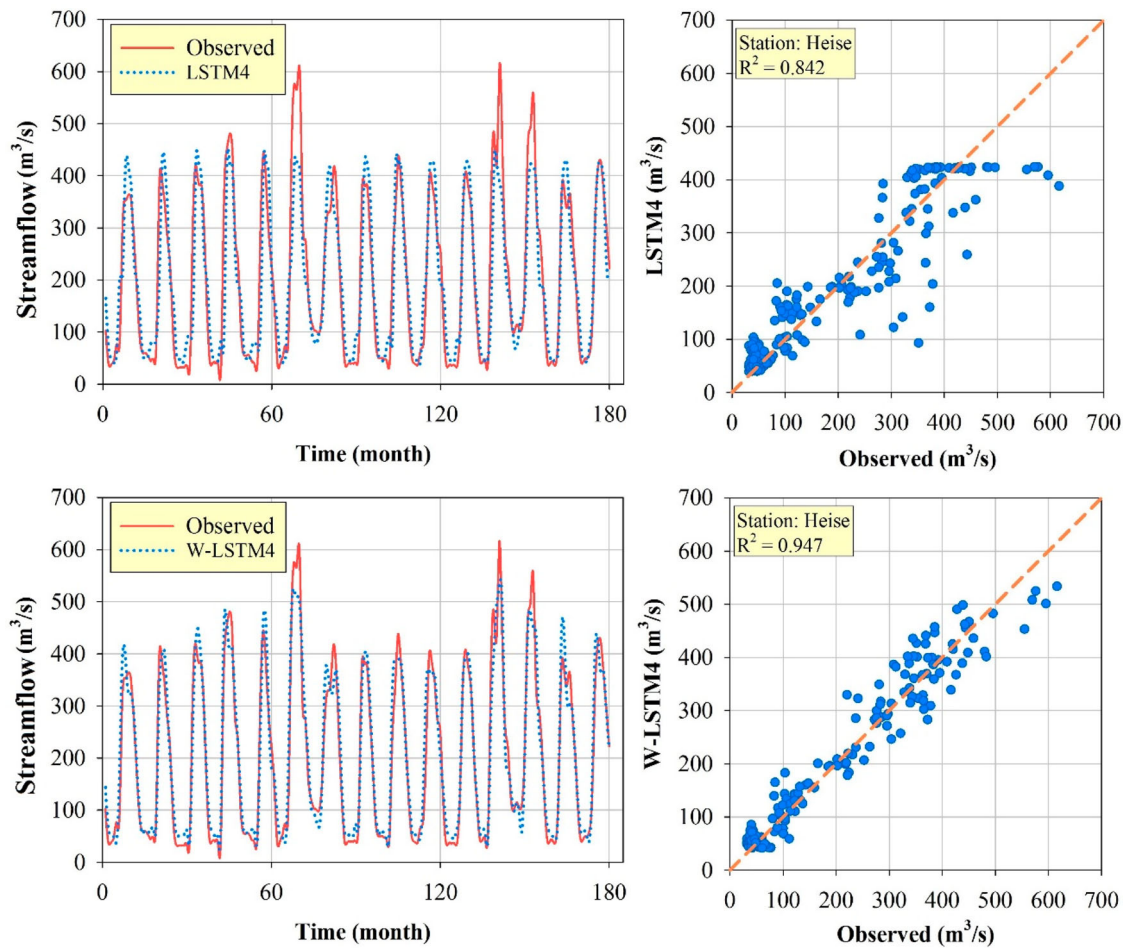
**Figure 3.** Scatter and time series plots of the observed and estimated river streamflow data for Heise station, using the best-performing hybrid model during testing (W-LSTM4) and the corresponding stand-alone LSTM4.

inputs with two to four lags. Examining Table 4, it can be seen that RMSE = 61.152 m$^3$/s, MAE = 42.186 m$^3$/s, R$^2$ = 0.836, NSE = 0.836, WI = 0.953 (training phase) and RMSE = 63.428 m$^3$/s, MAE = 41.731 m$^3$/s, R$^2$ = 0.838, NSE = 0.837, WI = 0.953 (testing phase), for the stand-alone LSTM2 model at Heise station, this improves to RMSE = 39.872 m$^3$/s, MAE = 27.256 m$^3$/s, R$^2$ = 0.930, NSE = 0.930, WI = 0.982 (training phase) and RMSE = 38.668 m$^3$/s, MAE = 28.630 m$^3$/s, R$^2$ = 0.940, NSE = 0.939, WI = 0.984 (testing phase) in the hybrid W-LSTM2 model.

This shows a dependable potential for W theory to capture the monthly streamflow time series. This result is also observed for the other hybrid models. The most important reason for the improved performance is that the wavelet removes unwanted high-frequency signals from the raw signal, improving output overall. The same results are observed for stand-alone and hybrid models. The coupled methods perform the best for both locations, when utilizing longer streamflow data, this is

particularly evident in the M4 models which draw on 4-months of streamflow data.

It is possible to visualize the performance of these models using scatter, time series and Taylor diagrams. These comparison diagrams are shown in Figures 3 and 4, respectively. In this regard, the superior coupled models during the test phase for both locations (i.e. W-LSTM4) and the stand-alone models (i.e. LSTM4) were considered. Dashed lines in the scatter graphs denote the perfect line. It can be seen in the scatter plots that there is a lower dispersion in the hybrid W-LSTM4 models compared to the stand-alone LSTM4. This indicates that the hybrid models offer reliability and compatibility with the data for estimating monthly river streamflow time series. The time series plots also illustrate that peak-points of streamflow data can be estimated accurately using hybrid W-LSTM4 models, this is true for both stations and is an improvement over the LSTM4 models. Figure 5 also shows the Taylor diagrams for the observed and estimated streamflow data using the best hybrid models and
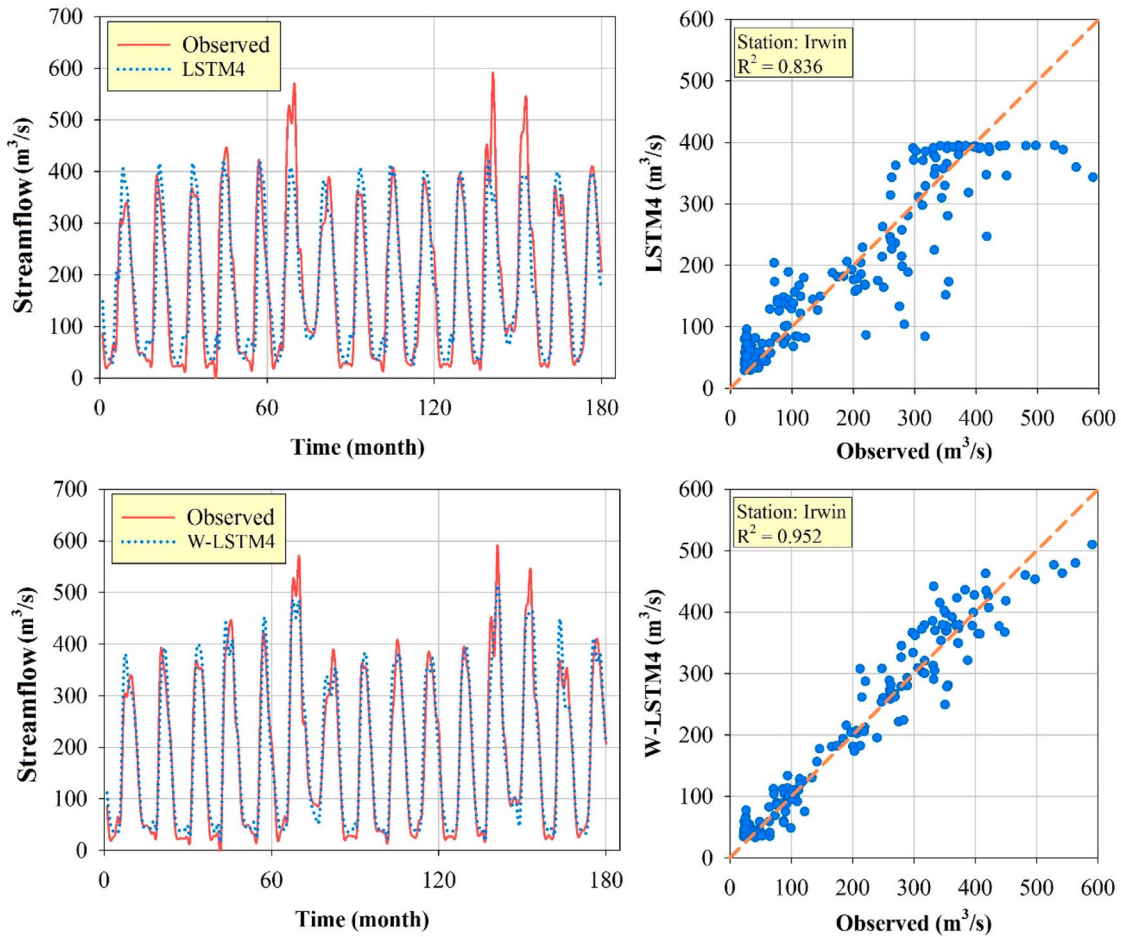
**Figure 4.** Scatter and time series plots of the observed and estimated river streamflow data for Irwin station using the best-performing hybrid model during testing (W-LSTM4) and the corresponding stand-alone LSTM4.
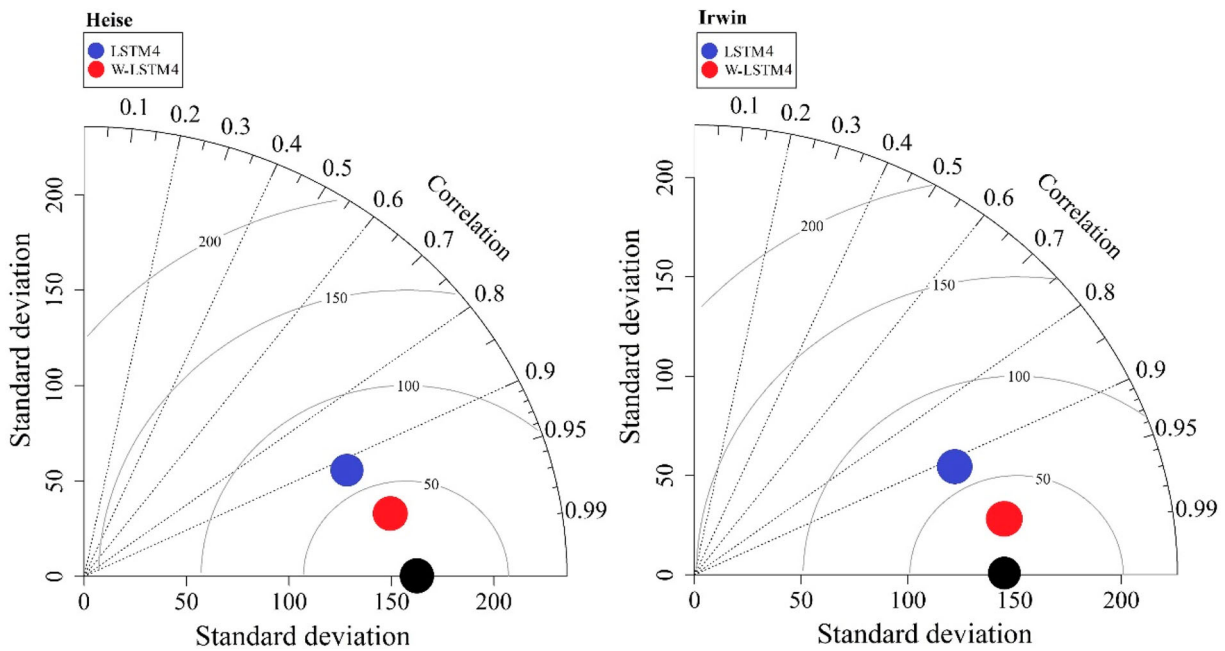


**Figure 5.** Taylor diagrams of the observed and estimated river streamflow data using the best-performing hybrid model during testing (W-LSTM4) and the corresponding stand-alone LSTM4, at Heise and Irwin stations.

**Table 6.** Accuracy rankings of the applied models using Heise station data.

| Models | M1 Training | M1 Testing | M2 Training | M2 Testing | M3 Training | M3 Testing | M4 Training | M4 Testing | Total Training | Total Testing |
|---|---|---|---|---|---|---|---|---|---|---|
| MLR | 12 | 8 | 13 | 13 | 13 | 12 | 13 | 12 | 51 | 45 |
| SVM-RBF | 13 | 10 | 12 | 12 | 12 | 13 | 12 | 13 | 49 | 48 |
| SVM-Poly | 14 | 13 | 14 | 14 | 14 | 14 | 14 | 14 | 56 | 55 |
| DT | 5 | 14 | 8 | 11 | 7 | 11 | 7 | 11 | 27 | 47 |
| RF | 2 | 12 | 3 | 10 | 3 | 8 | 3 | 7 | 11 | 37 |
| GB | 6 | 3 | 6 | 9 | 5 | 9 | 4 | 9 | 21 | 30 |
| LSTM | 11 | 5 | 10 | 8 | 11 | 10 | 11 | 10 | 43 | 33 |
| W-MLR | 8 | 2 | 9 | 5 | 9 | 5 | 9 | 3 | 35 | 15 |
| W-SVM-RBF | 9 | 6 | 7 | 4 | 8 | 4 | 8 | 5 | 32 | 19 |
| W-SVM-Poly | 10 | 7 | 11 | 7 | 10 | 7 | 10 | 8 | 41 | 29 |
| W-DT | 3 | 9 | 4 | 6 | 4 | 6 | 5 | 6 | 16 | 27 |
| W-RF | 1 | 11 | 1 | 2 | 1 | 2 | 1 | 2 | 4 | 17 |
| W-GB | 4 | 4 | 2 | 3 | 2 | 3 | 2 | 4 | 10 | 14 |
| W-LSTM | 7 | 1 | 5 | 1 | 6 | 1 | 6 | 1 | 24 | 4 |

**Table 7.** Accuracy rankings of the applied models using Irwin station data.

| Models | M1 Training | M1 Testing | M2 Training | M2 Testing | M3 Training | M3 Testing | M4 Training | M4 Testing | Total Training | Total Testing |
|---|---|---|---|---|---|---|---|---|---|---|
| MLR | 12 | 7 | 13 | 13 | 13 | 12 | 13 | 12 | 51 | 44 |
| SVM-RBF | 13 | 10 | 12 | 12 | 12 | 13 | 12 | 13 | 49 | 48 |
| SVM-Poly | 14 | 11 | 14 | 14 | 14 | 14 | 14 | 14 | 56 | 53 |
| DT | 4 | 9 | 8 | 11 | 8 | 11 | 9 | 11 | 29 | 42 |
| RF | 2 | 12 | 3 | 8 | 3 | 9 | 3 | 9 | 11 | 38 |
| GB | 6 | 8 | 5 | 10 | 6 | 8 | 4 | 8 | 21 | 34 |
| LSTM | 11 | 3 | 11 | 9 | 11 | 10 | 11 | 10 | 44 | 32 |
| W-MLR | 8 | 1 | 9 | 5 | 9 | 6 | 8 | 4 | 34 | 16 |
| W-SVM-RBF | 10 | 5 | 6 | 4 | 7 | 4 | 7 | 5 | 30 | 18 |
| W-SVM-Poly | 9 | 4 | 10 | 7 | 10 | 7 | 10 | 7 | 39 | 25 |
| W-DT | 3 | 14 | 4 | 6 | 5 | 5 | 6 | 6 | 18 | 31 |
| W-RF | 1 | 13 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 18 |
| W-GB | 5 | 6 | 2 | 2 | 2 | 3 | 2 | 3 | 11 | 14 |
| W-LSTM | 7 | 2 | 7 | 3 | 4 | 1 | 5 | 1 | 23 | 7 |

most relevant stand-alone methods. A Taylor diagram consists of three error measures, standard deviation, correlation, and centred RMSE. The short distance between the point of the hybrid W-LSTM4 models at both stations (marked in red) and the observational data (black) indicates the superior performance of the hybrids.

To better understand the estimation accuracy of the various models, all were ranked using their RMSE values. Tables 6 and 7 show the rankings for all models used to analyse data from Heise and Irwin stations, respectively. In this case, lower ranked models denote a superior model performance for estimating streamflow. Here, the total ranking for all input combinations is obtained (i.e. M1-M4) for each region. Examining the total ranking for models concerning Heise station, in Table 6, it can be seen that W-RF, followed by W-GB, outperformed the other standalone and coupled models during the training period. Meanwhile, W-LSTM models showed superior performance during the testing phase of all M1-M4 input patterns; this had the lowest total ranking of 4. After that, W-GB, W-MLR, and W-RF were lower ranked, indicating their better performance.

The same result was observed at Irwin station, and the rankings are shown in Table 7. In this context, the W-RF and W-GB methods (training), as well as W-LSTM, and W-GB, and W-MLR (testing) showed lower rankings, showing their superior performance.

It can be seen that the coupled W-GB model was one of the superior models during both training and testing of the hydrometric station data. Therefore, this is the recommended process for precisely estimating river streamflow data. Conversely, the stand-alone SVM-Poly illustrates the highest total ranking, showing that it offered the poorest performance. Examining Tables 6 and 7 in more detail, the stand-alone SVM-RBF and hybrid W-SVM-RBF models were superior to the SVM-Poly and W-SVM-Poly ones for both locations. Furthermore, of the tree-based models, the stand-alone RF, GB and their hybridized forms with the wavelets (W-RF and W-GB) were superior to the DT and W-DT models in training and testing. It is possible to conclude that the hybrid models exhibit better river streamflow estimations than stand-alone methods. It was also shown several papers that coupled techniques

frequently outperformed single models. For example, pre-processing with EMD or EEMD will improve the performance of machine learning models for estimation of river streamflow (Huang et al., 2014; Liu et al., 2020; Meng et al., 2019; Rezaie-Balf et al., 2019; Wang et al., 2013). Different optimization algorithms have been also been proposed to develop new hybrids (Feng et al., 2021; Kilinc & Haznedar, 2022; Yaseen et al., 2019).

The potential of wavelet theory to improve the performance of stand-alone machine learning techniques has been widely reported for river streamflow estimation (Hadi & Tombul, 2018; Ravansalar et al., 2017; Sun et al., 2019). The superiority of wavelet and hybrid models is not limited to riverflow, as other authors have shown it can estimate other time series hydro-climatological parameters such as precipitation (Kumar et al., 2021; Paul et al., 2020), groundwater levels (Band et al., 2021; Yosefvand & Shabanlou, 2020) evapotranspiration (Kisi & Alizamir, 2018), and soil temperature (Mehdizadeh et al., 2020; Samadianfard et al., 2018).

## 4. Conclusion

Monthly river streamflow time series for two hydrometric stations on Snake River, USA were estimated. Stand-alone machine learning (ML) models, SVM-RBF, SVM-Poly, DT, RF, GB, LSTM, and a traditional MLR were used. It was found that the river streamflow for each month could be estimated using lagged monthly data as inputs, using lags of 1 to 4 months. Wavelet theory was incorporated into the stand-alone models to establish a wavelet-based hybrid model. The study results indicate that the coupled models generally performed better than their corresponding stand-alone counterparts. [when measured by …] In general, input patterns of the models benefitted from a larger number of inputs, especially the M4 models. These exhibited better estimates than other input combinations. The final performance rankings of models showed that W-RF and then W-GB were the best-performing methods in the training phase. Conversely, the W-LSTM and W-GB models exhibited the lowest rankings during testing, indicating that they were highly dependable. The db4 was used in this study to produce the mother wavelet, when hybridizing the stand-alone models. Future studies could test the efficiency of other mother wavelets when estimating hydrological parameters. Seeing the successful application of wavelets here, many possibilities for complex estimations open up. Wavelet theory can be combined effectively with other ML-based models and determining which places they can be used would be beneficial for researchers. Also, different ML techniques, in combination with other

pre-processing methods, such as empirical mode decomposition, ensemble empirical mode decomposition, may also be used to produce bio-inspired optimization algorithms.

## Disclosure statement

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

## ORCID

*Kwok-Wing Chau* 🄳 http://orcid.org/0000-0001-6457-161X
*Amir Mosavi* 🄳 http://orcid.org/0000-0003-4842-0613

## References

Adnan, R. M., Kisi, O., Mostafa, R., Ahmed, A. N., & El-Shafie, A. (2021). The potential of a novel support vector machine trained with modified mayfly optimization algorithm for streamflow prediction. *Hydrological Sciences Journal*(just-accepted).

Adnan, R. M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., & Li, B. (2020). Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydrometeorological data as inputs. *Journal of Hydrology*, *586*, 124371. https://doi.org/10.1016/j.jhydrol.2019.124371

Adnan, R. M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., & Kisi, O. (2019). Daily streamflow prediction using optimally pruned extreme learning machine. *Journal of Hydrology*, *577*, 123981. https://doi.org/10.1016/j.jhydrol.2019.123981

Al-Sudani, Z. A., Salih, S. Q., & Yaseen, Z. M. (2019). Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *Journal of Hydrology*, *573*, 1–12. https://doi.org/10.1016/j.jhydrol.2019.03.004

Baesens, B., Viaene, S., Van Gestel, T., Suykens, J. A., Dedene, G., De Moor, B., & Vanthienen, J. (2000). An empirical assessment of kernel type performance for least squares support vector machine classifiers. (Ed.), KES'2000. *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*. Proceedings (Cat. No. 00TH8516).

Band, S. S., Heggy, E., Bateni, S. M., Karami, H., Rabiee, M., Samadianfard, S., Chau, K.-W., & Mosavi, A. (2021). Groundwater level prediction in arid areas using wavelet

analysis and Gaussian process regression. *Engineering Applications of Computational Fluid Mechanics*, *15*(1), 1147–1158. https://doi.org/10.1080/19942060.2021.1944913

Bayazit, M. (2015). Nonstationarity of hydrological records and recent trends in trend analysis: A state-of-the-art review. *Environmental Processes*, *2*(3), 527–542. https://doi.org/10.1007/s40710-015-0081-7

Bernard, S., Heutte, L., & Adam, S. (2009). On the selection of decision trees in random forests. (Ed.). *2009 International Joint Conference on Neural Networks*.

Chen, C., Zhang, Q., Kashani, M. H., Jun, C., Bateni, S. M., Band, S. S., Dash, S. S., & Chau, K.-W. (2022). Forecast of rainfall distribution based on fixed sliding window long short-term memory. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 248–261. https://doi.org/10.1080/19942060.2021.2009374

Clarke, A., Clarke, A. M., & Brown, R. (1959). Regression to the mean-A confused concept. (Ed.), BULLETIN OF THE BRITISH PSYCHOLOGICAL SOCIETY.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Daubechies, I. (2009). *The wavelet transform, time-frequency localization and signal analysis*. Princeton University Press.

Deng, T., Chau, K. W., & Duan, H. F. (2021). Machine learning based marine water quality prediction for coastal hydro-environment management. *Journal of Environmental Management*, *284*, 112051. https://doi.org/10.1016/j.jenvman.2021.112051

Deng, T., Duan, H. F., & Keramat, A. (2022). Spatiotemporal characterization and forecasting of coastal water quality in the semi-enclosed Tolo Harbour based on machine learning and EKC analysis. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 694–712. https://doi.org/10.1080/19942060.2022.2035257

Dong, L., Fang, D., Wang, X., Wei, W., Damaševičius, R., Scherer, R., & Woźniak, M. (2020). Prediction of streamflow based on dynamic sliding window LSTM. *Water*, *12*(11), 3032. https://doi.org/10.3390/w12113032

Ebrahimi, H., & Rajaee, T. (2017). Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. *Global and Planetary Change*, *148*, 181–191. https://doi.org/10.1016/j.gloplacha.2016.11.014

Feng, R., Fan, G., Lin, J., Yao, B., & Guo, Q. (2021). Enhanced long short-term memory model for runoff prediction. *Journal of Hydrologic Engineering*, *26*(2), 04020063. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002035

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *11*(8), 1189–1232.

Fu, M., Fan, T., Ding, Z. A., Salih, S. Q., Al-Ansari, N., & Yaseen, Z. M. (2020). Deep learning data-intelligence model based on adjusted forecasting window scale: Application in daily streamflow simulation. *IEEE Access*, *8*, 32632–32651. https://doi.org/10.1109/ACCESS.2020.2974406

Ghimire, S., Yaseen, Z. M., Farooque, A. A., Deo, R. C., Zhang, J., & Tao, X. (2021). Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, *11*(1), 1–26. https://doi.org/10.1038/s41598-020-79139-8

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. https://doi.org/10.1109/ACCESS.2019.2943515

Hadi, S. J., & Tombul, M. (2018). Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. *Journal of Hydrology*, *561*, 674–687. https://doi.org/10.1016/j.jhydrol.2018.04.036

Hamasuna, Y., Endo, Y., & Miyamoto, S. (2008). Support Vector Machine for data with tolerance based on Hard-margin and Soft-Margin. (Ed.),. *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*.

He, Z., Wen, X., Liu, H., & Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, *509*, 379–386. https://doi.org/10.1016/j.jhydrol.2013.11.054

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, S., Chang, J., Huang, Q., & Chen, Y. (2014). Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, *511*, 764–775. https://doi.org/10.1016/j.jhydrol.2014.01.062

Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques. *Journal of Big Data*, *7*(1), 1–22. https://doi.org/10.1186/s40537-020-00345-2

Kadam, A., Wagh, V., Muley, A., Umrikar, B., & Sankhua, R. (2019). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*, *5*(3), 951–962. https://doi.org/10.1007/s40808-019-00581-3

Kalra, A., Ahmad, S., & Nayak, A. (2013). Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns. *Advances in Water Resources*, *53*, 150–162. https://doi.org/10.1016/j.advwatres.2012.11.003

Kilinc, H. C. (2022). Daily streamflow forecasting based on the hybrid particle swarm optimization and long short-term memory model in the Orontes Basin. *Water*, *14*(3), 490. https://doi.org/10.3390/w14030490

Kilinc, H. C., & Haznedar, B. (2022). A hybrid model for streamflow forecasting in the Basin of Euphrates. *Water*, *14*(1), 80. https://doi.org/10.3390/w14010080

Kim, S., Seo, Y., Rezaie-Balf, M., Kisi, O., Ghorbani, M. A., & Singh, V. P. (2019). Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: Peninsula vs continent. *Theoretical and Applied Climatology*, *137*(1), 693–712. https://doi.org/10.1007/s00704-018-2627-x

Kisi, O., & Alizamir, M. (2018). Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks. *Agricultural and Forest Meteorology*, *263*, 41–48. https://doi.org/10.1016/j.agrformet.2018.08.007

Kumar, Y. P., Maheswaran, R., Agarwal, A., & Sivakumar, B. (2021). Intercomparison of downscaling methods for daily precipitation with emphasis on wavelet-based hybrid

models. *Journal of Hydrology*, *599*, 126373. https://doi.org/10.1016/j.jhydrol.2021.126373

Latif, S. D., & Ahmed, A. N. (2021). Application of deep learning method for daily streamflow time-series prediction: A case study of the Kowmung River at Cedar Ford, Australia. *International Journal of Sustainable Development and Planning*, *16*(3), 497–501. https://doi.org/10.18280/ijsdp.160310

Leong, W. C., Bahadori, A., Zhang, J., & Ahmad, Z. (2021). Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *International Journal of River Basin Management*, *19*(2), 149–156. https://doi.org/10.1080/15715124.2019.1628030

Lin, S.-S., Zhang, N., Zhou, A., & Shen, S.-L. (2022). Time-series prediction of shield movement performance during tunneling based on hybrid model. *Tunnelling and Underground Space Technology*, *119*, 104245. https://doi.org/10.1016/j.tust.2021.104245

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., & Dai, Y. (2021). A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology*, *601*, 126636. https://doi.org/10.1016/j.jhydrol.2021.126636

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, *99*, 650–655. https://doi.org/10.1016/j.procir.2021.03.088

Lipton, Z., Kale, D., Elkan, C., & Wetzel, R. (2017). Learning to Diagnose with LSTM Recurrent Neural Networks [arXiv: 1511.03677]. *arXiv preprint arXiv:1511.03677*.

Liu, D., Jiang, W., Mu, L., & Wang, S. (2020). Streamflow prediction using deep learning neural network: Case study of Yangtze River. *IEEE Access*, *8*, 90069–90086. https://doi.org/10.1109/ACCESS.2020.2993874

Mehdizadeh, S., Ahmadi, F., & Kozekalani Sales, A. (2020). Modelling daily soil temperature at different depths via the classical and hybrid models. *Meteorological Applications*, *27*(4), e1941. https://doi.org/10.1002/met.1941

Mehdizadeh, S., & Sales, A. K. (2018). A comparative study of autoregressive, autoregressive moving average, gene expression programming and Bayesian networks for estimating monthly streamflow. *Water Resources Management*, *32*(9), 3001–3022. https://doi.org/10.1007/s11269-018-1970-0

Meira Neto, A. A., Oliveira, P. T. S., Rodrigues, D. B., & Wendland, E. (2018). Improving streamflow prediction using uncertainty analysis and Bayesian model averaging. *Journal of Hydrologic Engineering*, *23*(5), 05018004. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001639

Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L., & Wang, L. (2019). A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. *Journal of Hydrology*, *568*, 462–478. https://doi.org/10.1016/j.jhydrol.2018.11.015

Meshram, S. G., Ghorbani, M., Deo, R. C., Kashani, M. H., Meshram, C., & Karimi, V. (2019). New approach for sediment yield forecasting with a two-phase feedforward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. *Water Resources Management*, *33*(7), 2335–2356. https://doi.org/10.1007/s11269-019-02265-0

Mosavi, A., Mohammadzadeh, A, Rathinasamy, S, Zhang, C, Reuter, U, Levente, K, & Adeli, H (2022). Deep learning fuzzy immersion and invariance control for type-I diabetes. *Computers in Biology and Medicine*, *149*. 105975. https://doi.org/10.1016/j.compbiomed.2022.105975

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, *586*, 124901. https://doi.org/10.1016/j.jhydrol.2020.124901

Niu, W.-j., & Feng, Z.-k. (2021). Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society*, *64*, 102562. https://doi.org/10.1016/j.scs.2020.102562

Paul, R. K., Paul, A., & Bhar, L. (2020). Wavelet-based combination approach for modeling sub-divisional rainfall in India. *Theoretical and Applied Climatology*, *139*(3), 949–963. https://doi.org/10.1007/s00704-019-03026-0

Qu, J., Ren, K., & Shi, X. (2021). Binary Grey wolf optimization-regularized extreme learning machine wrapper coupled with the boruta algorithm for monthly streamflow forecasting. *Water Resources Management*, *35*(3), 1029–1045. https://doi.org/10.1007/s11269-021-02770-1

Rasouli, K., Hsieh, W. W., & Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, *414*, 284–293. https://doi.org/10.1016/j.jhydrol.2011.10.039

Ravansalar, M., Rajaee, T., & Kisi, O. (2017). Wavelet-linear genetic programming: A new approach for modeling monthly streamflow. *Journal of Hydrology*, *549*, 461–475. https://doi.org/10.1016/j.jhydrol.2017.04.018

Reis, G. B., da Silva, D. D., Fernandes Filho, E. I., Moreira, M. C., Veloso, G. V., de Souza Fraga, M., & Pinheiro, S. A. R. (2021). Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. *Journal of Environmental Management*, *290*, 112625. https://doi.org/10.1016/j.jenvman.2021.112625

Rezaie-Balf, M., Kim, S., Fallah, H., & Alaghmand, S. (2019). Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in Iran and South Korea. *Journal of Hydrology*, *572*, 470–485. https://doi.org/10.1016/j.jhydrol.2019.03.046

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(3), 660–674. https://doi.org/10.1109/21.97458

Sahour, H., Gholami, V., Torkaman, J., Vazifedan, M., & Saeedi, S. (2021). Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environmental Earth Sciences*, *80*(22), 1–14. https://doi.org/10.1007/s12665-021-10054-5

Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). Convolutional, long short-term memory, fully connected deep neural networks. (Ed.), *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*.

Salem, F. M. (2018). Slim lstms. *arXiv preprint arXiv:1812.11391*.

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 and Tillage Research*, *175*, 37–50. https://doi.org/10.1016/j.still.2017.08.012

Sifuzzaman, M., Islam, M. R., & Ali, M. (2009). Application of wavelet transform and its advantages compared to Fourier transform.

Singh, S., Parmar, K. S., Makkhan, S. J. S., Kaur, J., Peshoria, S., & Kumar, J. (2020). Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. *Chaos, Solitons & Fractals*, *139*(4), 110086–110099. https://doi.org/10.1016/j.chaos.2020.11008

Singh, V. K., Panda, K. C., Sagar, A., Al-Ansari, N., Duan, H. F., Paramaguru, P. K., Vishwakarma, D. K., Kumar, A., Kumar, D., Kashyap, P. S., Singh, R. M., & Elbeltagi, A. (2022). Novel genetic algorithm (GA) based hybrid machine learning-pedotransfer function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 1082–1099. https://doi.org/10.1080/19942060.2022.2071994

Starck, J.-L., & Murtagh, F. (2001). Astronomical image and signal processing: Looking at noise, information and scale. *IEEE Signal Processing Magazine*, *18*(2), 30–40. https://doi.org/10.1109/79.916319

Sun, Y., Niu, J., & Sivakumar, B. (2019). A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stochastic Environmental Research and Risk Assessment*, *33*(10), 1875–1891. https://doi.org/10.1007/s00477-019-01734-7

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Wang, C., Du, W., Zhu, Z., & Yue, Z. (2020). The real-time big data processing method based on LSTM or GRU for the smart job shop production process. *Journal of Algorithms & Computational Technology*, *14*(5), 174–188. https://doi.org/10.1177/1748302620962390

Wang, L., Li, X., Ma, C., & Bai, Y. (2019). Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. *Journal of Hydrology*, *573*, 733–745. https://doi.org/10.1016/j.jhydrol.2019.03.101

Wang, W.-c., Chau, K.-w., Xu, D.-m., & Chen, X.-Y. (2015). Improving forecasting accuracy of annual runoff time series

using ARIMA based on EEMD decomposition. *Water Resources Management*, *29*(8), 2655–2675. https://doi.org/10.1007/s11269-015-0962-6

Wang, W.-c., Xu, D.-m., Chau, K.-w., & Chen, S. (2013). Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics*, *15*(4), 1377–1390. https://doi.org/10.2166/hydro.2013.134

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419

Xiang, Z., & Demir, I. (2020). Distributed long-term hourly streamflow predictions using deep learning–A case study for State of Iowa. *Environmental Modelling & Software*, *131*, 104761. https://doi.org/10.1016/j.envsoft.2020.104761

Yaseen, Z. M., Mohtar, W. H. M. W., Ameen, A. M. S., Ebtehaj, I., Razali, S. F. M., Bonakdari, H., Salih, S. Q., Al-Ansari, N., & Shahid, S. (2019). Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region. *IEEE Access*, *7*, 74471–74481. https://doi.org/10.1109/ACCESS.2019.2920916

Yaseen, Z. M., Naganna, S. R., Sa'adi, Z., Samui, P., Ghorbani, M. A., Salih, S. Q., & Shahid, S. (2020). Hourly river flow forecasting: Application of emotional neural network versus multiple machine learning paradigms. *Water Resources Management*, *34*(3), 1075–1091. https://doi.org/10.1007/s11269-020-02484-w

Yosefvand, F., & Shabanlou, S. (2020). Forecasting of groundwater level using ensemble hybrid wavelet–self-adaptive extreme learning machine-based models. *Natural Resources Research*, *29*(5), 3215–3232. https://doi.org/10.1007/s11053-020-09642-2

Zhang, H., Singh, V. P., Wang, B., & Yu, Y. (2016). CEREF: A hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. *Journal of Hydrology*, *540*, 246–256. https://doi.org/10.1016/j.jhydrol.2016.06.029

Zhao, X.-h., & Chen, X. (2015). Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. *Water Resources Management*, *29*(8), 2913–2926. https://doi.org/10.1007/s11269-015-0977-z

Zhu, S., Luo, X., Yuan, X., & Xu, Z. (2020). An improved long short-term memory network for streamflow forecasting in the upper Yangtze River. *Stochastic Environmental Research and Risk Assessment*, *34*(9), 1313–1329. https://doi.org/10.1007/s00477-020-01766-4