Ensuring a generalizable machine learning model for forecasting reservoir inflow in Kurdistan region of Iraq and Australia

Sarmad Dashti Latif[1,2] · Ali Najah Ahmed[3,4]

Abstract

Correct inflow prediction is a critical non-engineering measure for ensuring flood control and increasing water supply efficiency. In addition, accurate inflow prediction can offer reservoir planning and management guidance since inflow is the major input into reservoirs. This study aims at generalizing a machine learning model for forecasting reservoir inflow. Daily, weekly, and monthly inflow and rainfall time-series data have been collected as two hydrological parameters to forecast reservoir inflow using a machine learning method, namely, support vector regression (SVR). Four different SVR kernels have been applied in this study. The kernels are radial basis function (RBF), linear, normalized polynomial, and sigmoid. Two scenarios for input selection have been implemented. Dokan dam in Kurdistan region of Iraq and Warragamba Dam in Australia were selected as the case studies for this research. For the purpose of generalization, the proposed models have been applied to two countries with a different climate condition. The findings showed that daily time-scale outperformed weekly and monthly, while RBF outperformed the other SVR kernels with root-mean-square error (RMSE) = 145.7 and coefficient of determination (R^2) = 0.85 for forecasting daily inflow at Dokan dam. However, RBF kernel could not perform well for forecasting daily inflow in Warragamba dam. The results showed that the proposed machine learning model performed well at Kurdistan region of Iraq only, while the result for Australia was not accurate. Therefore, the proposed models could not be generalized.

✉ Sarmad Dashti Latif
Sarmad.latif@soran.edu.iq; Sarmad.latif@komar.edu.iq

1 Scientific Research Center, Soran University, Soran, Erbil, Kurdistan Region, Iraq

2 Civil Engineering Department, College of Engineering, Komar University of Science and Technology, Sulaimany, Kurdistan Region, Iraq

3 Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional (UNITEN), Selangor, Malaysia

4 Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Selangor, Malaysia

1 Introduction

Climate change and weather patterns, rising water demand, and poor water resource management practices are all factors have led to the present global water crisis (Herslund & Mguni, 2019; Kooy et al., 2020; Marlow et al., 2013; Sosa-Rodriguez et al., 2019). Water management is a critical component of urban development's long-term viability. The scientific community has expressed worry about urban water-related issues all around the world (Jia et al., 2015). Water concerns currently include increasing urban floods, over-exploitation of groundwater, urban water shortages, the waste of rainfall resources, and water contamination as a result of fast urbanization and extreme weather events (Bábek et al., 2020; Nguyen et al., 2019; Wang et al., 2018).

One of the most essential elements in the development, maintenance, and sustainability of riparian ecosystems is reservoir inflow. Inflow may be thought of as a "master variable" that regulates riverine species' abundance and distribution (Latif, Ahmed, et al. 2021). Weather (rainfall and temperature) interacts with geology, topography, soil, and vegetation to impact infiltration, evaporation, and run-off generation, all of which influence reservoir inflow. The number and timing of reservoir inflows are key components of river system environmental fluxes and ecological integrity. This "master variable" also shapes river ecosystems and affects fish eating, migratory, nesting, and spawning conditions (Dhungel et al., 2016; O'Keeffe et al., 2019; U.S. Environmental Protection Agency (U.S. EPA) and US EPA, 2008; Xu et al., 2020).

Correct inflow forecast is an essential non-engineering measure to confirm flood-control protection and to raise the efficiency of water supply use. In addition, since inflow is the main input into reservoirs, good inflow forecast may provide direction for reservoir development and management (Apaydin et al., 2020; More et al., 2019; Qi et al., 2019). Because of its importance, numerous reservoir inflow forecasting models and techniques have been created and tested in real-world scenarios (Apaydin et al., 2020).

Inflow prediction has been proposed using a variety of hydrologic models over the past decade, but there is no silver bullet: Various techniques will perform better for particular watersheds, lead times, and types of occurrences (Tikhamarine et al., 2020). Since inflow is the primary input into reservoirs, accurate inflow prediction is not only an important non-engineering method to assure flood-control safety and enhance water resource use efficiency, but it may also give direction for reservoir development and management. Therefore, the need to have a capable model for predicting reservoir inflow is crucial (Amnatsan et al., 2018; Yan et al., 2018).

According to recent research, Iraq will face greater issues in the future, with the water deficit situation worsening over time and the Tigris and Euphrates Rivers anticipated to be dry by 2040. The estimated discharge of the two rivers in 2025 will be drastically reduced (Zakaria et al., 2013). In Australia, overall urban water consumption is expected to rise by at least 39% between 2009 and 2026, following a population increase of more than 24% between 2007 and 2026. Climate change will probably certainly exacerbate the situation on a global and regional basis (Yan et al., 2018). This study focuses on implementing a generalizable model for both countries for forecasting reservoir inflow.

Nowadays, hydrologists focuses on machine learning algorithms for forecasting hydrological parameters (Lai et al., 2020; Latif & Ahmed, 2021; Latif et al., 2020, 2021a, 2021b; Najah et al., 2021). For example, Babaei et al., 2019, conducted a study in Zayandehroud dam reservoir in Iran to predict the dam reservoir inflow, and their input parameters were monthly inflow and rainfall. They have applied ANN and SVR as their proposed method.

Their findings showed that the proposed model has the lowest error for inflow prediction, with the SVR model's products outperforming those of the ANN model. Another study was conducted by Zhang et al., 2020, in the Huanren reservoir in China to produce an ensemble of 10-day inflow forecasts. The time scale was 10 days with different input combinations such as inflow, precipitation, relative humidity, minimum temperature, maximum temperature, and precipitation forecast. They have implemented ANN, SVR, and ANFIS for their methods. The decomposition outcome of their study showed that the input set is the dominant source of uncertainty. They found out the contribution of the data-driven model is limited and has a substantial seasonal variation which is more significant in winter and summer but more minor in spring and autumn. Furthermore, Y. Yu et al., 2017, proposed a study in Three Gorges Reservoir (TGR), China. They have developed a novel model, combining monthly inflow forecasting and multi-objective ecological reservoir operations. The objective of their research was to improve the efficiency of water resource allocation. For the monthly time scale, meteorological and hydrological data were used as inputs. A hybrid model based on SVR and singular spectrum analysis (SSA), namely, SSA-SVR, was applied for the method. The results of the simulations revealed that the proposed coupled model for the TGR will outperform actual TGR operations; moreover, multi-objective ecological operations based on inflow forecasts may help relieve water shortages. Meanwhile, Al-Suhili & Karim, 2015, developed five ANN models for predicting daily inflow at Dokan dam. According to their findings, their proposed model was capable of forecasting daily inflow with the highest correlation coefficient of 0.94. Moreover, Y. Wang et al., 2014, conducted a study in order to forecast monthly inflow at Three Gorges Reservoir. Three machine learning models, namely, SVR, genetic programming (GP), and seasonal autoregressive (SAR), have been implemented in their study. RBF has been adopted in their SVR prediction model as an effective kernel. Their findings showed that SVR and GP model performance significantly improves when coupled with the SSA for predicting the inflow series. On the other hand, Halik et al., 2015, utilized wavelet support vector machine (WSVM) with the adaptation of RBF for forecasting inflow at Sutami Reservoir, Indonesia. Their findings showed that WSVM performed better in forecasting inflow with utilizing RBF kernel.

The area of research is based on the primary data in Dokan dam, Iraq, and the secondary data in Warragamba dam, Sydney, Australia. In this study, reservoir inflow and rainfall as two different scenarios have been utilized as the input parameters for the proposed machine learning models. In Dokan dam, the four kernels of SVR are not applied for forecasting reservoir inflow to check the most accurate kernel. Therefore, this study aims to fill this gap in the literature by contributing a new idea of applying four different kernels of SVR in order to ensure the most accurate kernel for forecasting reservoir inflow.

2 Materials and methods

2.1 Dokan dam

Dokan dam is located on the Lesser Zab tributary, approximately 295-km north of Baghdad and 65-km southeast of Sulaymaniyah (Fig. 1) (Sulaiman et al., 2021). At a typical functioning level of 511 m above sea level, the dam height is approximately 116 m, with a total storage capacity of 6.87 109 m3 (6.14 109 m3 living storage and 0.73 109 m3 dead storage) (Ezz-Aldeen et al., 2018). The historical daily time-series inflow and rainfall data
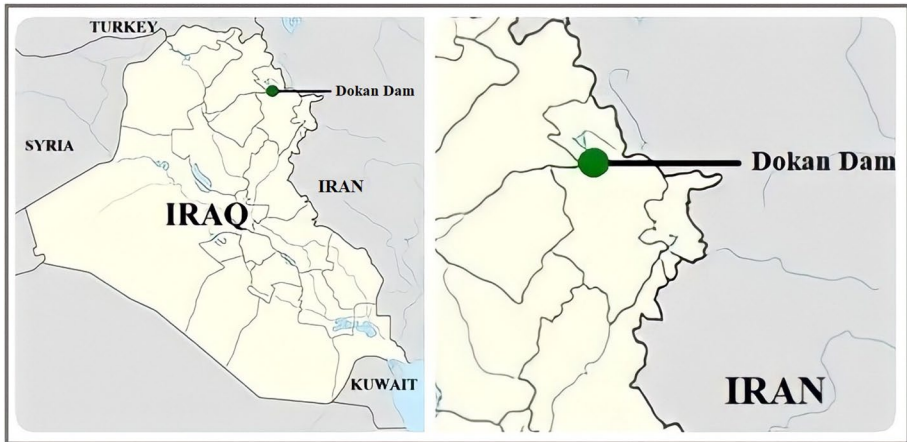
Fig. 1 First study area location

are collected from the Ministry of Agriculture and Water Resources, Kurdistan regional government, Iraq, for the duration of January 1, 1988, to December 31, 2015 (Fig. 2). The basic statistical characteristics of the utilized inflow data of Dokan dam are shown in Table 1.

2.2 Warragamba dam

Warragamba dam is a heritage-listed dam in the Wollondilly Shire of New South Wales in Australia, near the outer southwestern Sydney suburb of Warragamba (Fig. 3). One of the wild rivers that will be flooded after the Warragamba dam wall is lifted is the Kowmung River (Division, 2005). The historical daily time-series inflow data have been collected from January 1, 1988, to December 31, 2015 (Fig. 4). The basic statistical characteristics of the utilized data of Warragamba dam are shown in Table 2.

2.3 Statistical analysis for datasets

The annual data for both locations were split into two subsamples to test for homogeneity of the overall series by performing a t-test for significant change in the mean values and an F-test for the variances. According to the t-test and F-test values, there are no significant changes in these parameters; therefore, it is justified to forecast using the entire set of data. In the current study, 80% of the data was used for training, and the remained 20% was used for testing. Furthermore, the data for both locations have been collected as a daily time-series data; then, it was converted to weekly and monthly data. Table 3 shows the values for t-test and F-test for Dokan and Warragamba dams.

2.4 Model combinations and input selection

In machine learning application for forecasting purposes, one of the important steps is selecting the appropriate input parameters to the models. In this study, autocorrelation function has been utilized in order to select the most correlated input parameters for the
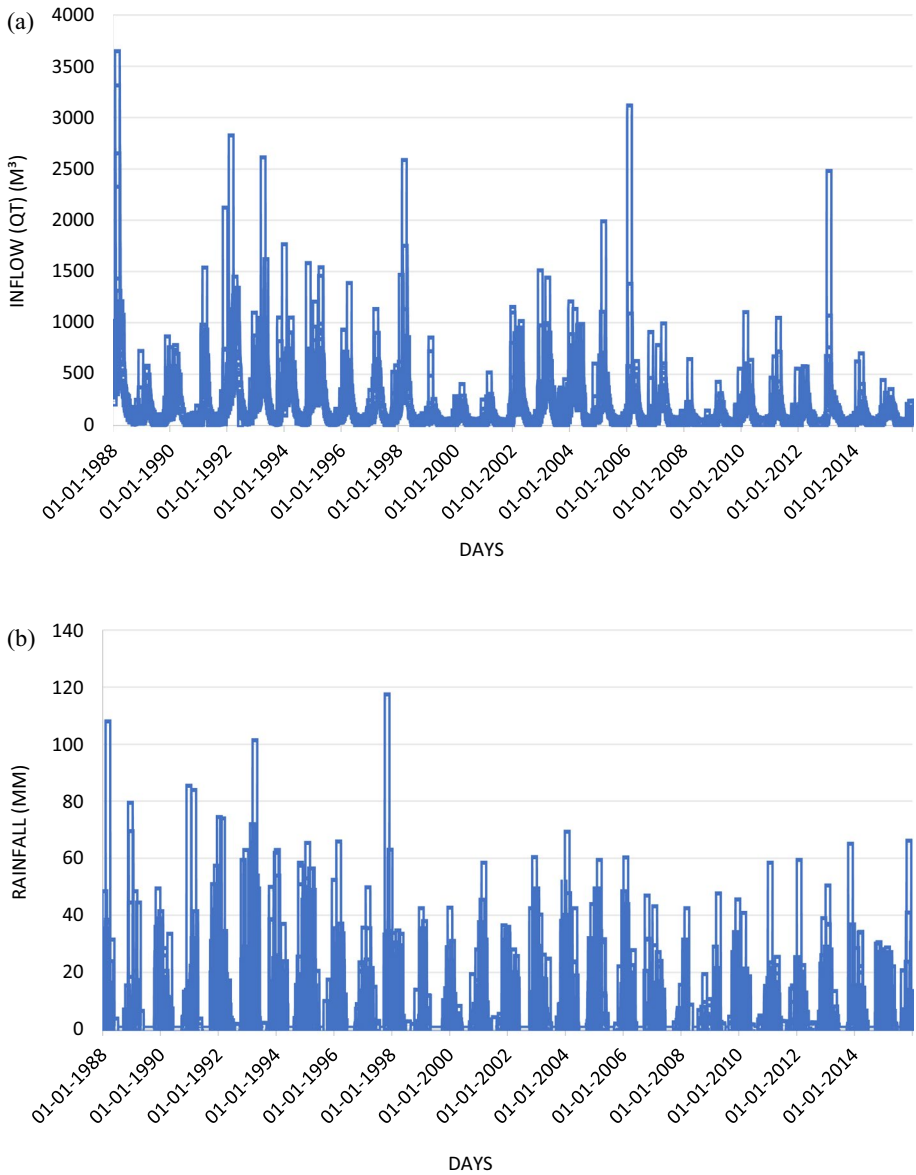
(a)

(b)

Fig. 2 **a** The daily inflow at Dokan dam and **b** the daily rainfall at Dokan dam

Table 1 Statistical characteristics of Dokan dam inflow data

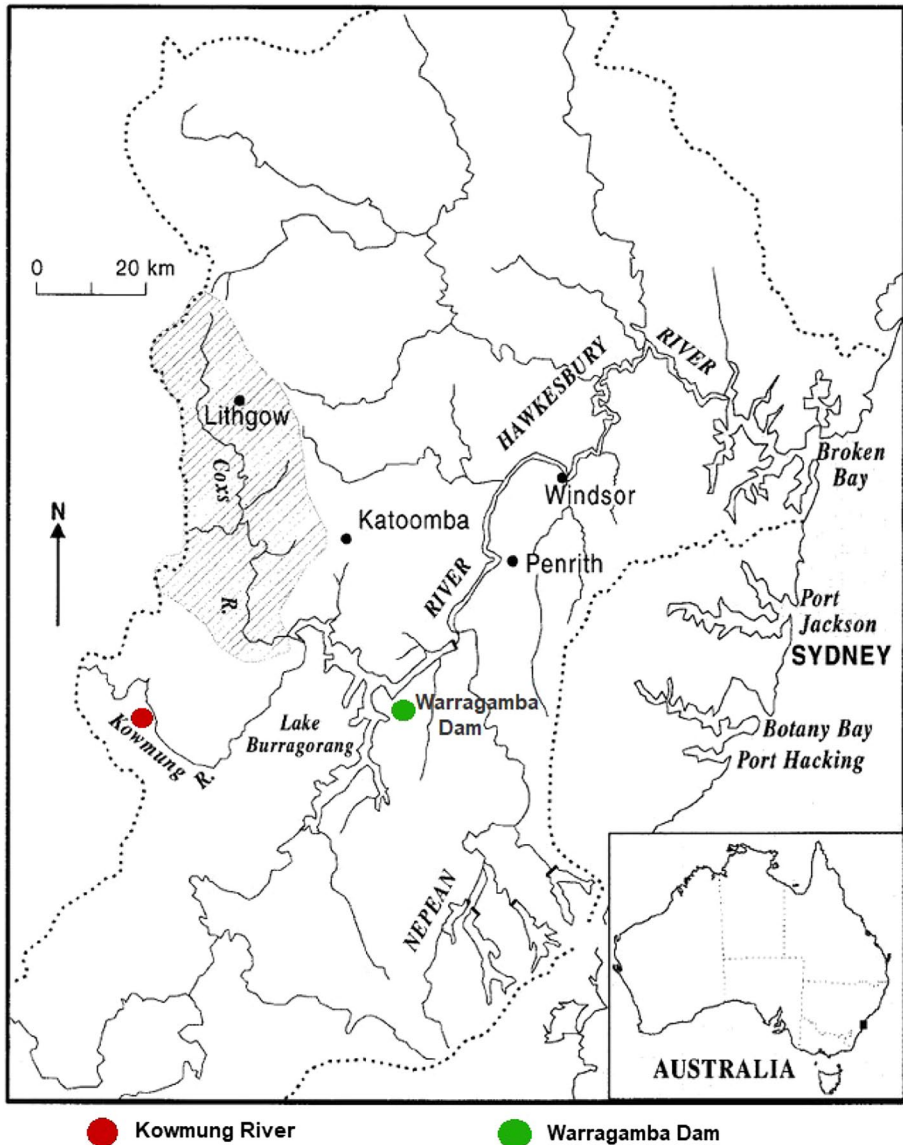| Time indices	Mean	Standard deviation	Minimum	Maximum
Daily	165.83	227.47	3	3608
Weekly	165.83	206.60	7.57	2298
Monthly	166.62	190.25	8.74	1569.16

Fig. 3 Second study area location (Latif & Ahmed, 2021)

proposed models. Autocorrelation function (ACF) is an important method that shows the correlation between two values in a time-series matter. ACF is widely used for hydrological prediction modeling since it will select the most appropriate input combination. According to ACF, five models with five time-lags have been selected (Table 4).

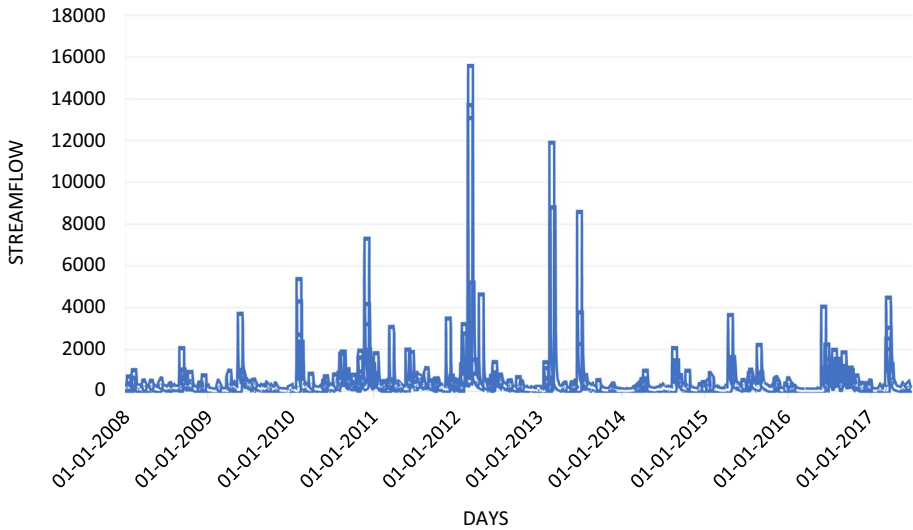Where Qt represents reservoir inflow, while Qt-1 represents reservoir inflow for previous 1-day time-lag.

Fig. 4 Daily streamflow of the Warragamba dam from January 1, 2008, to July 1, 2017

Table 2 Statistical characteristics of Warragamba dam inflow data

Time indices	Mean	Standard deviation	Minimum	Maximum
Daily	306.29	712.32	0.128	15,372.14

Table 3 T-test and F-test values for Dokan and Warragamba dams

Dams	T-test value	F-test value
Dokan dam	1.05254	5.8477
Warragamba dam	0.00025	1.4104

Table 4 Model combinations for daily inflow time-series data

Model	Target variable	Input combination
Model-1	Q_t	Q_{t-1}
Model-2	Q_t	Q_{t-1}, Q_{t-2}
Model-3	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}
Model-4	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}
Model-5	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}

Regarding the input selection, two scenarios were proposed. In the first scenario, inflow rate has been utilized as input selection. In the second scenario, inflow and rainfall have been combined in order to check which scenario achieve better performance for the proposed models.

2.5 Support vector regression (SVR)

SVR is one of the machine learning algorithms widely used in prediction (Lai et al., 2019). In general, it is a suitable method for the prediction of regression and time series in hydrological studies. This model typically defines the learning function for inputs and outputs. Support vectors are the training points nearest to the separating hyperplane, and the general definition of SVR is demonstrated in Fig. 5. For example, there are accountable decision functions, hyperplanes capable of delineating positive and negative data that defined the maximum margins. It displays the variance from the nearest positive to a hyperplane sample and maximizes the variance between the nearest negative sample and the hyperplane.

$$y(x) = w^T \phi(x) + b \tag{1}$$

where $\phi(x)$ represents the spaces of the high-dimensional function, which is mapped non-linearly from the x input space. By minimizing the regularized function R(C), the coefficients w and b are estimated:

$$R(C) = C\frac{1}{N} \sum_{i=1}^{N} L\varepsilon, (d_i, y_i) + \frac{1}{2} \parallel w \parallel^2 \tag{2}$$

where

$$L\varepsilon (d_i, y_i) = \{ |d_i - y_i| - \varepsilon \quad \text{if } |d_i - y_i| \geq \varepsilon \tag{3}$$

where b is the bias, ε is insensitive loss function, and w is the weight vector.

Four common types of SVR kernels as mentioned below are introduced in this study to train the SVR models for the first assessment to investigate the ability of the SVR model to mimic and learn on the pre-processed data. The four types of SVR kernels are RBF, linear, NP, and sigmoid, which are introduced as follows:

$$\text{RBF Kernel: } K(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|^2\right), \gamma > 0 \tag{4}$$

$$\text{Linear Kernel: } K(x_i, x_j) = x_i^T \times x_j \tag{5}$$

$$\text{NP Kernel: } K(x_i, x_j) = \left(\gamma x_i^T \times x_j + r\right)^d, \gamma > 0 \tag{6}$$

Fig. 5 The basic concept of SVR (Latif, 2021)
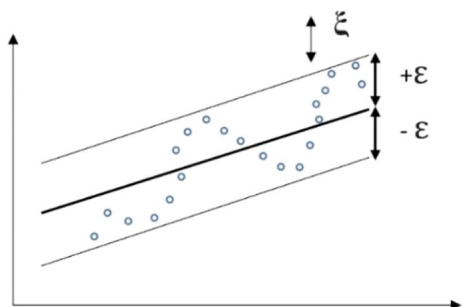
$$\text{Sigmoid Kernel: } K\left(x_i, x_j\right) = \tanh\left(x_i^T \times x_j + r\right) \tag{7}$$

Here γ, r, and d are kernel parameters.

The tuning parameters of RBF, linear, NP and sigmoid kernels are summarized in Table 5 in order to obtain the most optimal parameters to train and test in three different inputs designed for SVR.

Where γ: gamma; C: cost; d: degree; and r: coefficient. γ is a hypermeter that is established before the training model and is used to give the decision boundary curvature weight. Low values indicate "far" while big values indicate "near." The gamma parameter controls how far a single training example's impact reaches. The inverse of the radius of effect of samples picked by the model as support vectors are the gamma parameters. C is also a hypermeter that is used to regulate errors and is set before the training model. d is a parameter used when kernel is set to polynomial kernel. It is basically the degree of the polynomial used to find the hyperplane to split the data. r is the coefficient of the kernel.

There are two forms of SVR regression; both have the general Eqs. (3 and 4). Form 1 or epsilon is regarded as the first phase of SVR regression. The formulation shown in Eq. (4) gives this type of error function. Form 2 of regression is known as Nu (Aljanabi et al., 2018; Ehteram et al., 2019; Yahya et al., 2019).

Generally, if the model was created using the SVR approach, and V = 5, the data would be divided into five subsets of similar size by the model, and the training process would be repeated five times. The model runs for each training phase, using subsets for training and leaving one for assessing the model error.

In this research, two types of SVR were used, namely, SVR regression type 1 (also known as epsilon-SVR regression) and SVR regression type 2 (also known as nu-SVR regression). The nu-SVR is used to calculate the percentage of support vectors to keep in the solution compared to the total number of samples in the dataset. While epsilon-SVR is brought into the design of the optimization problem and is automatically approximated.

Two types of SVR models may be distinguished based on the specification of this error function:

a. Epsilon-SVR regression.

For this type of *SVR,* the error function is as follows:

$$\frac{1}{2}wTw + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^{\cdot} \tag{8}$$

which we minimize subject to:

Table 5 Tuning components for four different kernels in SVR

Type of kernel functions	Tuning or affecting parameters
RBF	γ and C
Linear	C
Normalized polynomial kernel (NP)	d, γ, r, and C
Sigmoid	γ, r, and C

$$w^T \emptyset(X_i) + b - y_i \leq \varepsilon + \xi \cdot_i$$
$$y_i - w^T \phi(X_i) + b_i \leq \varepsilon + \xi_i \quad (9)$$
$$\xi_i, \xi \cdot_i \geq 0, i = 1, \dots, N$$

b. nu-SVR regression.

For this *SVR* model, the error function is given by:

$$\frac{1}{2} w^T w - C \left(v\varepsilon + \frac{1}{N} \sum_{i=1}^{N} (\xi_i + \xi \cdot_i) \right) \quad (10)$$

which we minimize subject to:

$$w^T \emptyset(X_i) + b) - y_i \leq \varepsilon + \xi_i$$
$$y_i - \left(w^T \phi(X_i) + b_i \right) \leq \varepsilon + \xi_i \quad (11)$$
$$\xi_i, \xi \cdot_i \geq 0, i = 1, \dots, N, \varepsilon$$

2.6 Statistical measurements

2.6.1 Root-mean-square error (RMSE)

The residuals standard deviation is abbreviated as RMSE (predictive errors). To validate experimental data, the RMSE is commonly employed in climate analysis, prediction, and regression testing.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{Qio} - \text{Qip})^2} \quad (12)$$

The mean value of the observed inflow is where *Qip* and *Qio* are observed and predicted inflow values. The closer the RMSE value to zero, the better accuracy is going to show.

2.6.2 Coefficient of determination (R.^{2})

A major performance of the regression analysis is the coefficient of determination (denoted by R^2). This is the fraction of variation in the dependent variable that is predicted from the independent variable.

$$R^2 = \{(1/N)* \sum [(xi - X)*(yi - Y)/(\sigma x * \sigma y)\}^2 \quad (13)$$

2.6.3 Nash–Sutcliffe model efficiency coefficient (NSE)

NSE is a normalized statistic that assesses the amount of the residual variance in relation to the variance of the measured data. In the case of zero error of the proposed model, NSE

should be equal to 1. NSE is utilized in order to assess the predictive skill of the proposed models. It is defined as follows:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{T}\left(Q_m^t - Q_o^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q}_o\right)^2} \tag{14}$$

where $\overline{Q}_o$ is the mean of observed inflow, and Q_m^t is modeled inflow. Q_o^t is observed inflow at time t.

2.7 Sensitivity analysis (SA)

SA is valuable to investigate the uncertainty, especially during model development (Borgonovo, 2017). The meteorological parameter is one of the factors that contribute to reservoir inflow. Therefore, it is important to comprehensively study the meteorological input parameter contribute on reservoir inflow at study locations. The performance evaluation of the various possible combinations of the parameters is utilizing RMSE, R^2, and NSE approaches to determine the most effective parameters on the output. Utilizing these evaluations, the model can be observed if the parameter under consideration is missing or included in the analysis. As a result, the most important parameters will give a higher R^2 and NSE with the lower RMSE values. Then, it is indicated that the parameters are the most effective tool for the performance of the models. Sensitivity analysis is applied in order to have different models for comparison purposes in terms of accuracy. Figure 6 represents the sensitivity analysis of the proposed study.

2.8 Strength and limitation of the proposed techniques

In prediction modeling, SVR has various advantages. For instance, it has a superior generalization performance compared to other machine learning algorithms. Moreover, SVR can easily deal with nonlinear process through utilizing kernel functions since it is able
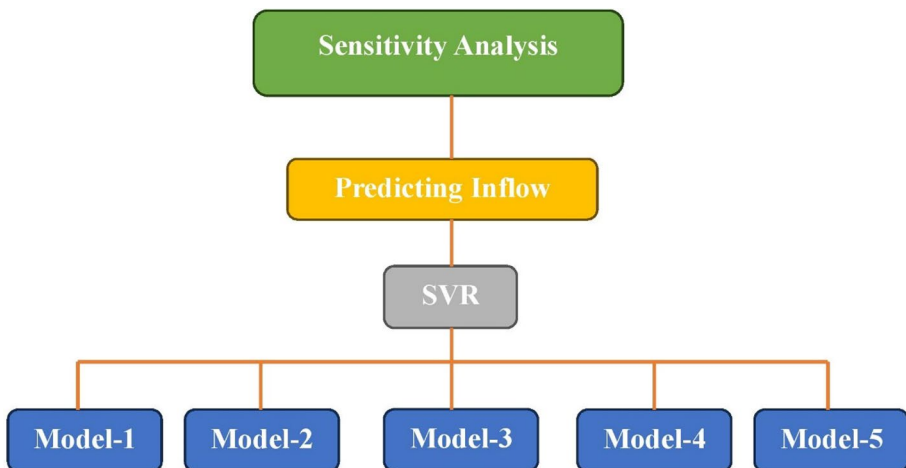

Fig. 6 Sensitivity analysis of the current study

to consider nonlinear relationships between observed and target values. Furthermore, predicted values can exceed observed values in the training data in SVR. However, there are various limitations of using SVR for prediction modeling. For instance, most of the time SVR is not suitable for a vast amount of data. Also, the training period is longer than the other machine learning algorithms. In addition, SVR is not very accurate for predicting extreme events. In summary, SVR should be well improved through utilizing different kernels and improving the pre-processing of observed data. In this study, epistemic uncertainty has been realized since the rainfall was not a suitable parameter as input for predicting inflow.

3 Results and discussion

SVR models are developed and compared in terms of RMSE, R^2, and NSE with different kernel functions and input parameters designed. The model that yields lower errors will reflect higher performance in this prediction of reservoir inflow. The first assessment is to scrutinize the optimization of RBF, linear, NP, and sigmoid kernels. SVR plays an important role in regression. Different kernel parameters were used as tuning parameters to improve the model accuracy. Several tuning or affecting parameters were used in SVR kernels.

The execution of optimizations is shown in the following section with the model performance of RMSE, R^2, and NSE. In this study, two scenarios were proposed for selecting input parameters of the proposed models. The first scenario is selecting reservoir inflow as a single input parameter, while the second scenario is combining reservoir inflow and rainfall as input parameters.

In this study, the SVR technique has been implemented on three-time horizons (daily, weekly, and monthly) with its four different kernels, namely, RBF, linear, NP, and sigmoid kernels. These four kernels have been applied in order to find out the most accurate kernel. Five models with different input combinations have been applied to the three different time horizons in two different scenarios (inflow and inflow + rainfall). Also, two different types of SVR (SVR regression type 1 and SVR regression type 2) were applied. According to the results, 36 models have been run. The best model, time horizon, scenario, SVR type, and kernel have been selected among all the 36 models. Model-2 outperformed all the other four models, while daily outperformed weekly and monthly time horizons. On the other hand, the first scenario (selecting only inflow as model input) outperformed the second scenario. In addition, SVR regression type 2 outperformed SVR regression type 1. The RBF kernel outperformed the other three kernels. The second-best kernel was the linear kernel, while the NP kernel was the third-best kernel. The kernel with the least performance was the sigmoid kernel, among the others. The results of SVR model are used in selecting the best scenario since it has four different kernels with two different regression types.

3.1 Forecasting reservoir inflow utilizing RBF kernel

RBF is a well-known kernel function that may be found in a variety of kernelized learning methods. It is widely used in classification and regression using SVR. The RBF kernel is a function whose value is proportional to the distance between the origin and a given location. The RBF kernel with different input designs is performed a pre-processing process in order to obtain the most optimum RBF tuning components. First, SVR

regression type 1 (also known as epsilon-SVR regression) was selected in the beginning. Secondly, SVR regression type 2 (also known as nu-SVR regression) has been implemented. The accuracy between SVR regression type 1 and SVR regression type 2 has been compared in order to select the most accurate type. Tables 6 shows the implementation of summary results of the overall model performance RMSE, R^2, and NSE with different input designs for SVR regression type 1. The tuning or affecting parameters of RBF kernel components are gamma and C. Figure 7 shows the performance of the overall set of Model-2 SVR regression type 1 RBF kernel for forecasting daily reservoir inflow.

Based on the outcomes achieved, Model-2 had significant value compared to the other models. Therefore, Model-2 is considered as the best-selected model to be applied for the next techniques. Therefore, SVR regression type 2 will be applied on Model-2.

In terms of RMSE and NSE, SVR type 2 regression outperformed type 1 regression. Therefore, from now on, SVR regression type 2 will be selected for the remaining SVR analysis. Table 7 shows the comparison results of the overall, training, and testing set for the SVR regression type 1 and type 2 for RBF kernel on the Model-2. Figure 8 shows the overall set of Model-2 performance for SVR regression type 2 RBF kernel.

Based on the previous results, from now on SVR regression type 2 will be selected for the remaining SVR analysis since it could successfully outperform the SVR type 1 regression.

3.2 Forecasting reservoir inflow utilizing linear kernel

In this stage, the pre-processing of predicted inflow at the Dokan dam is executed with the linear kernel. The tuning or affecting parameters of the linear kernel component is only *C*. The splitting ratio of train-to-test is set up at 80:20. Table 8 shows the execution of summary results of the model performance RMSE, R^2, and NSE for the Model-2 SVR linear kernel. Figure 9 shows the overall set of Model-2 performance for SVR regression type 2 linear kernel.

3.3 Forecasting reservoir inflow utilizing NP kernel

First, the pre-processing of predicted inflow at Dokan dam is executed with the NP kernel. The tuning or affecting parameters of NP kernels component are degree, gamma, and coefficient. The splitting ratio of train-to-test is set up at 80:20. Table 9 shows the execution of summary results of the model performance RMSE, R^2, and NSE for

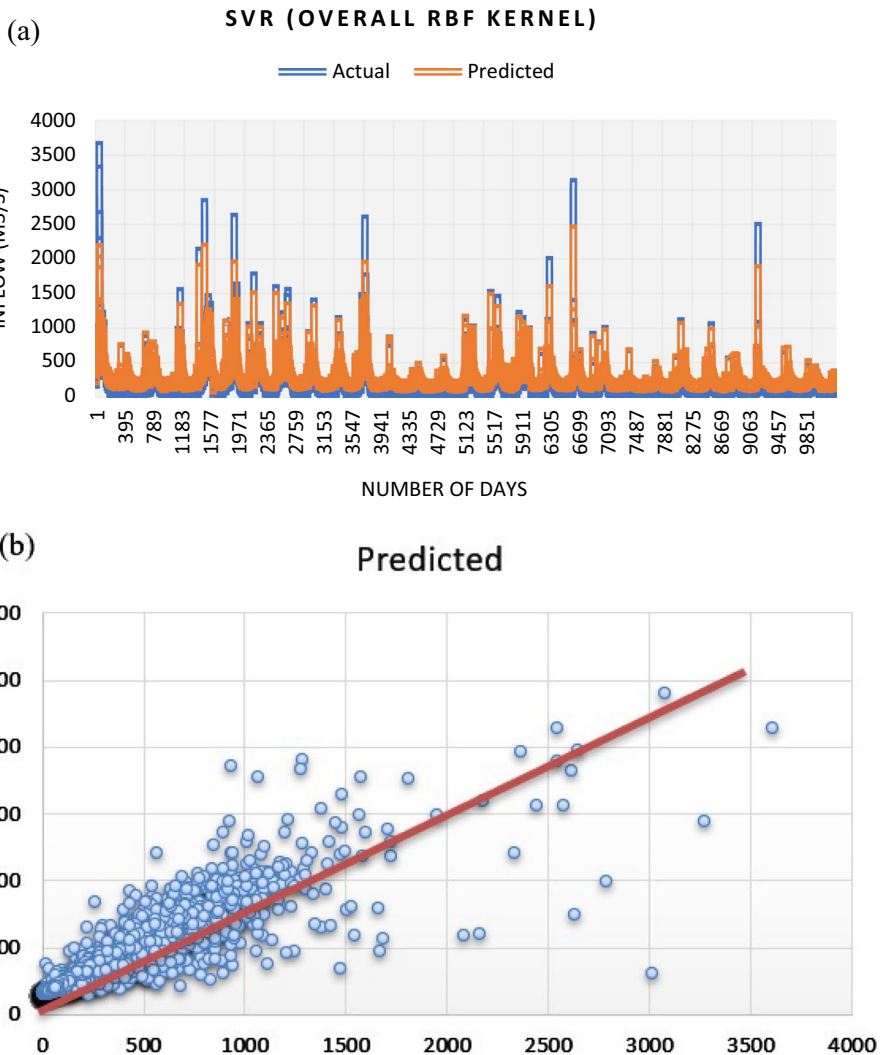| SVR (RBF kernel)	RMSE	R^2	NSE
Model-1	148.2	0.84	0.57
Model-2	148.0	0.85	0.58
Model-3	149.2	0.85	0.57
Model-4	156.1	0.85	0.53
Model-5	161.6	0.85	0.49

Table 6 Summary results of daily overall RBF kernel of SVR regression type 1

(a)

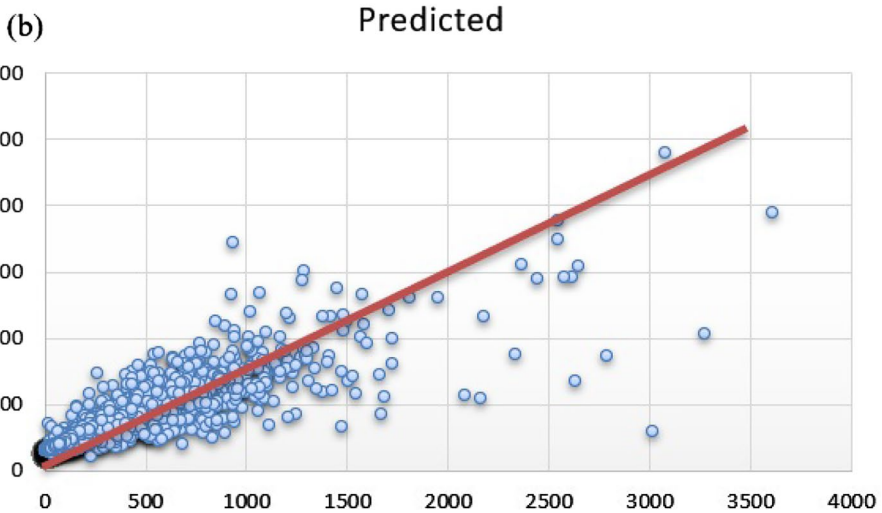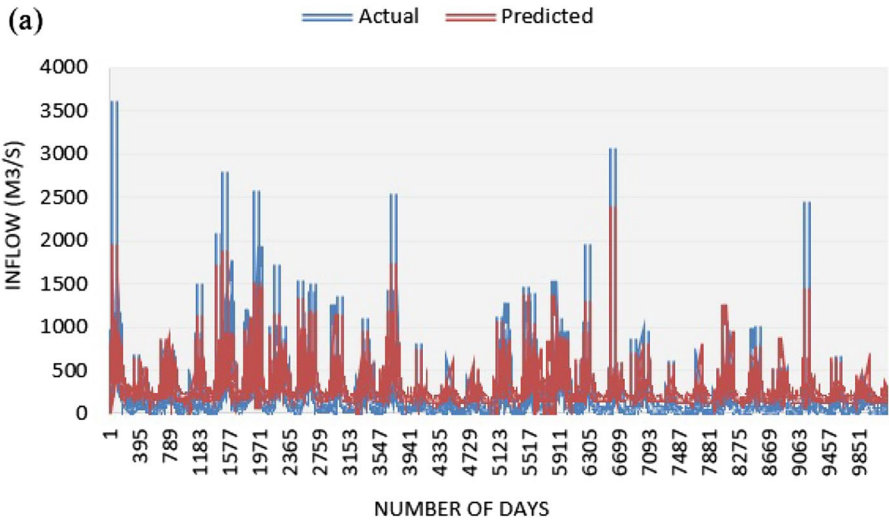**SVR (OVERALL RBF KERNEL)**

(b)

Fig. 7 The performance of overall set of Model-2 for SVR regression type 1 RBF kernel **a** actual vs predicted daily reservoir inflow and **b** scatter plot

Table 7 Summary results of daily SVR regression type 1 and type 2 for RBF kernel

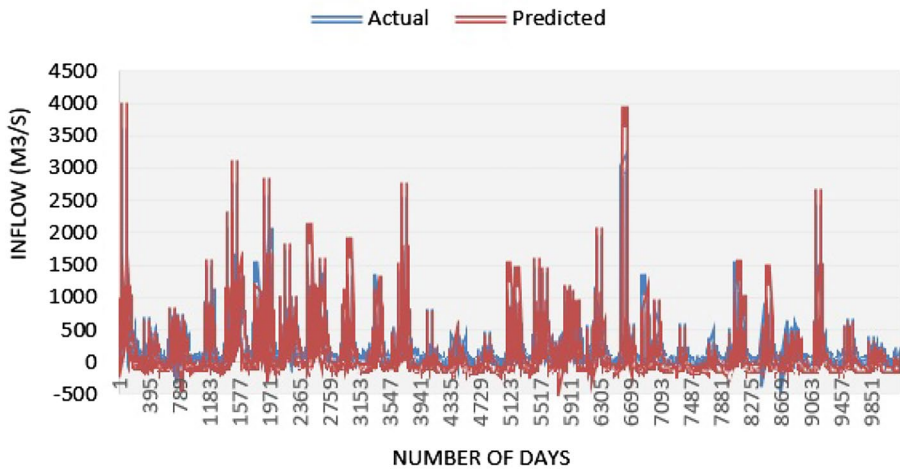| Types of SVR regression	SVR regression type 2 (Model-2)	RMSE	R^2	NSE
Type 1	Overall	148	0.84	0.57
	Training	148	0.83	0.56
	Testing	148.1	0.87	0.61
Type 2	Overall	145.7	0.85	0.59
	Training	145.1	0.84	0.58
	Testing	145.5	0.88	0.62

Fig. 8 The performance of overall Model-2 of SVR regression type 2 RBF kernel **a** actual vs predicted daily reservoir inflow and **b** scatter plot

SVR regression type 2 (Model-2) linear kernel	RMSE	R^2	NSE
Overall	157.5	0.84	0.52
Training	159.1	0.82	0.50
Testing	152.6	0.87	0.59

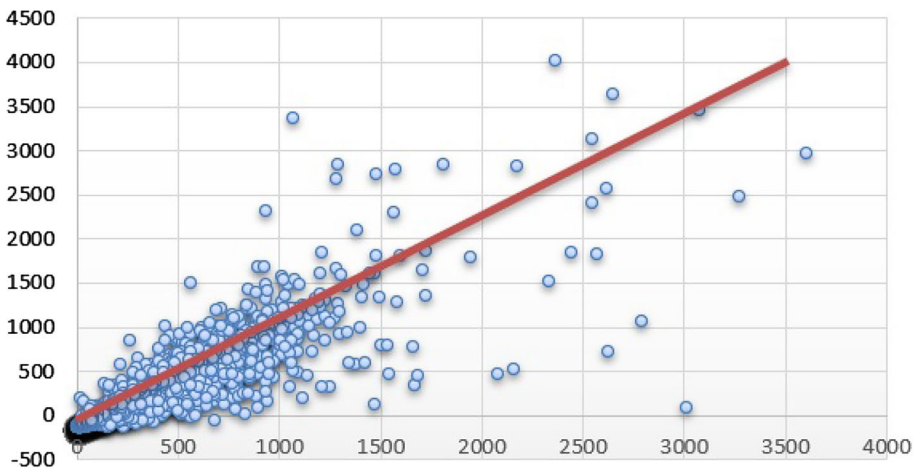**Table 8** Summary results of daily SVR regression type 2 linear kernel

Fig. 9 The performance of overall set of Model-2 for regression type 2 SVR linear kernel **a** actual vs predicted daily reservoir inflow and **b** scatter plot

Table 9 Summary results of daily SVR regression type 2 NP kernel

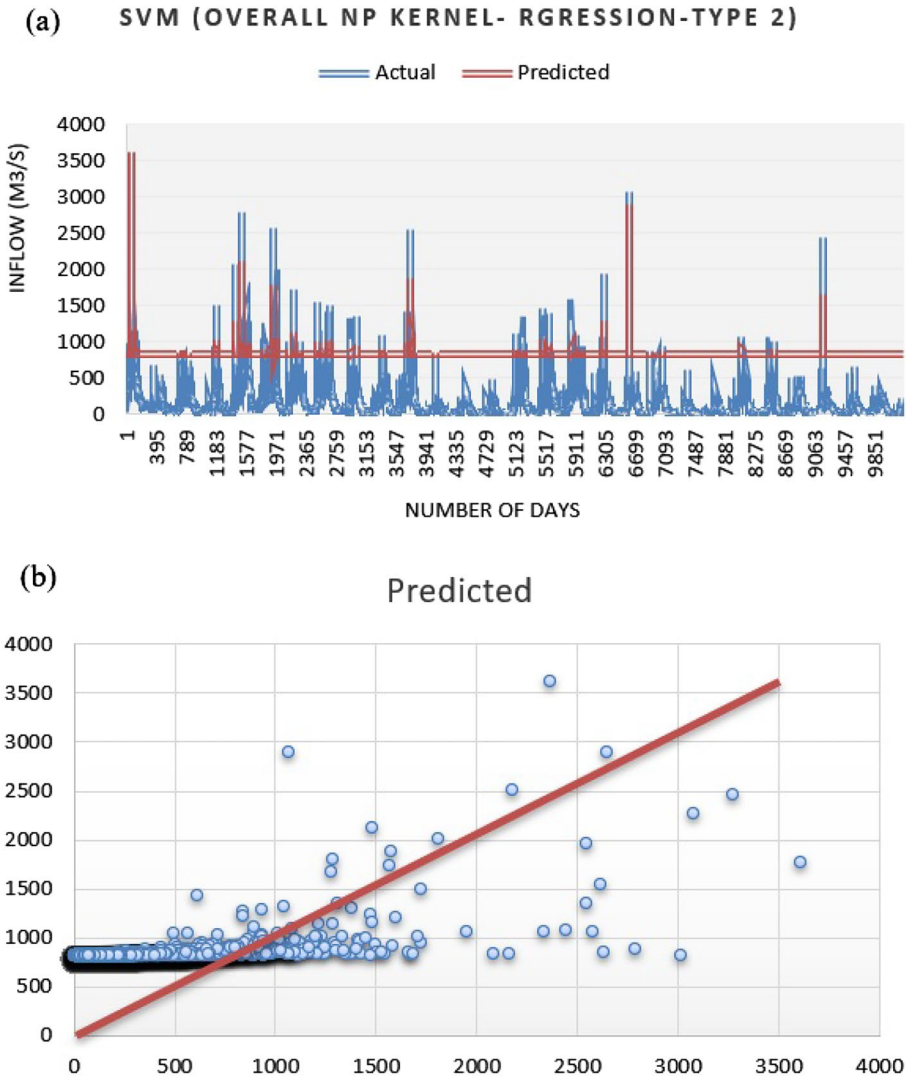| SVR regression type 2 (Model-2) NP kernel	RMSE	R^2	NSE
Overall	695.2	0.24	− 8.34
Training	694.2	0.23	− 8.61
Testing	698.0	0.26	− 7.62

Fig. 10 The performance of the overall set of Model-2 for regression type 2 SVR NP kernel **a** actual vs predicted daily reservoir inflow and **b** scatter plot

Model-2 SVR NP kernel. Figure 10 shows the overall set of Model-2 performance for SVR regression type 2 NP kernel.

3.4 Forecasting reservoir inflow utilizing sigmoid kernel

The sigmoid kernel is derived from the area of neural networks, where the bipolar sigmoid function is frequently employed as an artificial neuron activation function. It is worth noting that a sigmoid kernel function SVR model is equal to a two-layer perceptron neural network. Because of its origins in neural network theory, this kernel was extremely popular

Table 10 Summary results of daily SVR regression type 2 sigmoid kernel	SVR regression type 2 (Model-2) sigmoid kernel	RMSE	R^2	NSE
	Overall	716.4	0.84	− 8.92
	Training	715.7	0.83	− 9.22
	Testing	718.3	0.87	− 8.12

for SVR. First, the pre-processing of predicted inflow at Dokan dam is executed with the sigmoid kernel. The tuning or affecting parameters of the sigmoid kernel's component are gamma, coefficient, and *C*. The splitting ratio of train-to-test is set up at 80:20. Table 10
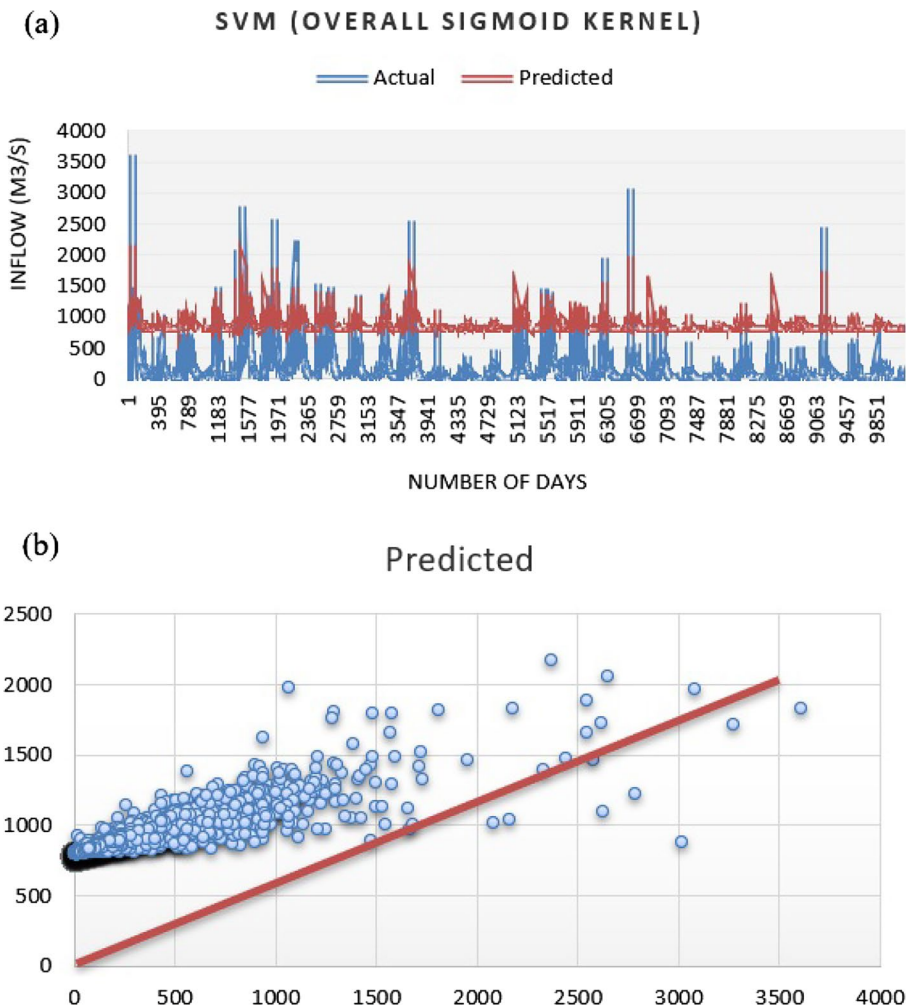

Fig. 11 The performance of overall set of Model-2 for regression type 2 SVR sigmoid kernel **a** actual vs predicted daily reservoir inflow and **b** scatter plot

Table 11 Comparison performance of SVR kernels (daily time horizon)

Data partition	SVR kernels	RMSE	R^2	NSE
Training	Linear	159.1	0.82	0.50
	NP	694.2	0.23	− 8.61
	RBF	145.1	0.84	0.58
	Sigmoid	715.7	0.83	− 9.22
Testing	Linear	152.6	0.87	0.59
	NP	698.0	0.26	− 7.62
	RBF	145.5	0.88	0.62
	Sigmoid	718.3	0.87	− 8.12
Overall	Linear	157.5	0.84	0.52
	NP	695.2	0.24	− 8.34
	RBF	145.7	0.85	0.59
	Sigmoid	716.4	0.84	− 8.92

Table 12 The accuracy of daily (inflow only) reservoir inflow prediction.

Techniques	Regression Type	Kernels	Models
			Model-1
			Model-2
	R-1	RBF	Model-3
			Model-4
SVR			Model-5
	R-2	RBF	Model-2
	R-2	Liner	Model-2
	R-2	NP	Model-2
	R-2	Sigmond	Model-2

shows the execution of summary results of the model performance RMSE, R^2, and NSE for Model-2. Figure 11 shows the accuracy of overall set of Model-2 performance for SVR regression type 2 sigmoid kernel for forecasting daily reservoir inflow.

According to the results of the overall datasets using sigmoid kernel, it is clear that sigmoid kernel is not capable of predicting the current data accurately. Sigmoid kernel could not forecast extreme events. The results of the overall dataset of sigmoid kernel are similar to the results of NP kernel. However, the RBF and linear kernels showed better performance.

3.5 Comparison performance of RBF, linear, NP, and sigmoid kernels

In order to understand the performance of the SVR kernels, the results of SVR kernels have been summarized using three statistical indices, namely, RMSE, R^2, and NSE. Four SVR kernels have been implemented to find the best-performed kernel for forecasting reservoir inflow. Table 11 shows the results of the selected variables used in the final models and SVR hyperparameter values in terms of comparison of the four types of kernels of the proposed SVR method for the daily reservoir inflow prediction.

Based on the results achieved, RBF outperformed linear, NP, and sigmoid kernels for the daily time-lag. Thus, RBF will be selected to apply in the other time-lags. Moreover, SVR regression type 2 outperformed SVR regression type 1. Furthermore, Model-2 will be

applied for other techniques such as RF, BRT, and LSTM because Model-2 outperformed the other models, namely, Model-1, Model-3, Model-4, and Model-5. Table 12 shows the first stage of the SVR technique results for forecasting the daily reservoir inflow.

3.6 Predicting weekly reservoir inflow utilizing RBF kernel

Forecasting reservoir inflow through a weekly time-lag has been implemented through the SVR regression type 2 technique and the selected Model-2 with the RBF kernel. The outcomes showed that the weekly time-lag was not as accurate as the daily time-lag. Table 13 represents the results of the weekly reservoir inflow prediction for the overall, training, and testing dataset, respectively. Figure 12 shows the accuracy of overall set of Model-2 performance for SVR regression type 2 RBF kernel for forecasting weekly reservoir inflow.

3.7 Predicting monthly reservoir inflow utilizing RBF kernel

Predicting reservoir inflow through a monthly time-lag has been applied over the SVR regression type 2 technique and the selected Model-2 with the RBF kernel. The results showed that the monthly time-lag was not as accurate as the daily and weekly time-lag. As it is shown in Table 14, the results of the RMSE in the monthly time-lag are a bit more accurate compared to the weekly time-lag. In contrast, the results of the R^2 and NSE in the weekly time-lag are significantly more accurate compared to the monthly time-lag. Therefore, the weekly time-lag is considered more accurate compared to the monthly time-lag, and the daily time-lag is considered as the most accurate compared to weekly and monthly time-lags. Figure 13 shows the overall set of Model-2 performance for SVR regression type 2 RBF kernel for forecasting monthly reservoir inflow.

Depending on the above results, the daily time-lag showed the best performance compared to weekly and monthly time-lags. However, the weekly time-lag outperformed the monthly time-lags. Based on this result, daily time-lag will be selected for the next applications. The reason behind this achievement is that Dokan dam has semi-arid weather. Therefore, daily data provide more time-series values and support the models to accurately train the dataset.

3.8 The second scenario for forecasting daily reservoir inflow

The second scenario of input combinations is to select inflow and rainfall parameters for forecasting inflow using different machine learning and deep learning methods, namely, SVR, RF, BRT, and LSTM. As previously explained, daily, weekly, and monthly reservoir

Table 13 Summary results of weekly SVR regression type 2 RBF kernel	SVR regression type 2 (Model-2) RBF kernel (weekly)	RMSE	R^2	NSE
	Overall	131.9	0.71	0.59
	Training	136.2	0.70	0.59
	Testing	118.4	0.74	0.58

(a)

WEEKLY SVM (OVERALL RBF KERNEL)

Actual   Predicted

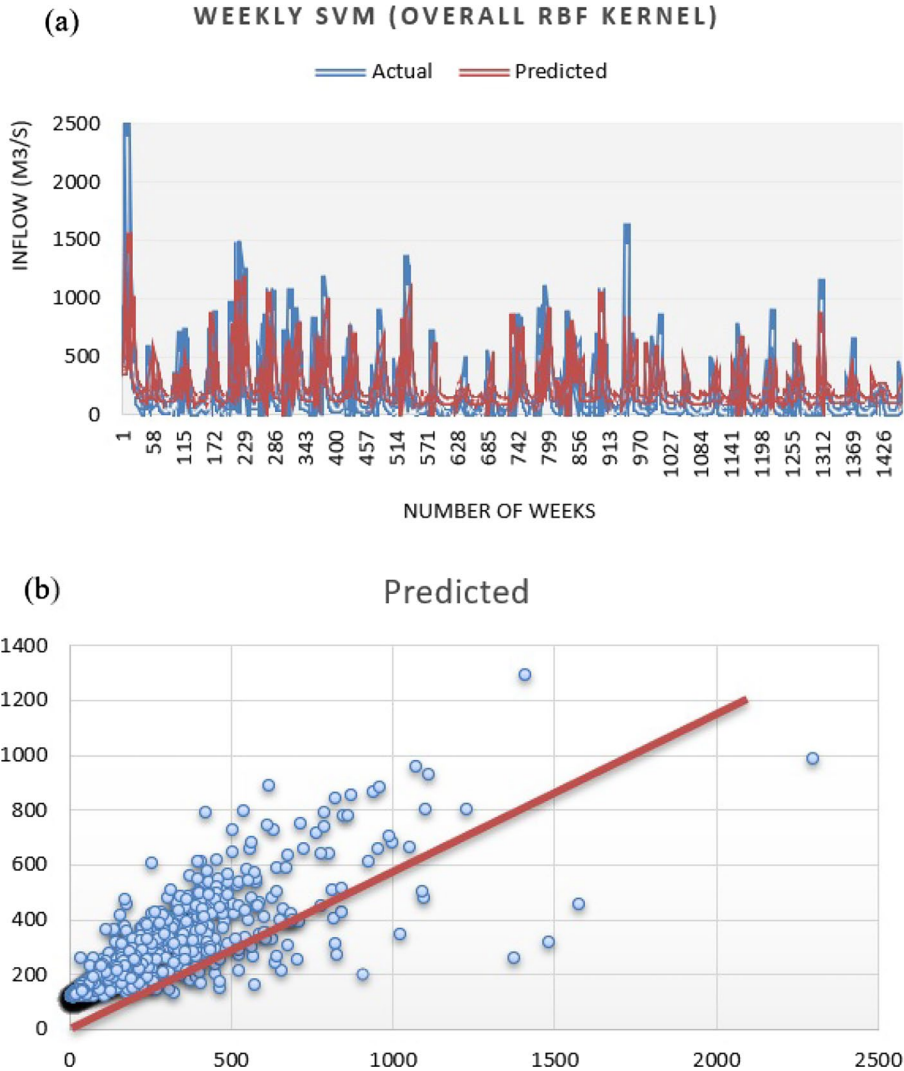

NUMBER OF WEEKS

(b)

Predicted

Fig. 12 The performance of overall set of Model-2 for regression type 2 SVR RBF kernel **a** actual vs predicted weekly reservoir inflow and **b** scatter plot

inflow time-series data have been selected to check the accuracy of the proposed models. Twenty-eight years daily, weekly, and monthly datasets for reservoir inflow and rainfall parameters have been collected and selected as the input parameters. ACF is used for selecting the most suitable input combinations. According to the previous results for the SVR method, the RBF kernel outperformed the other three kernels (linear, NP, and sigmoid). Therefore, for this scenario, the RBF kernel will be selected for the SVR method to forecast the daily, weekly, and monthly reservoir inflow at Dokan dam, Iraq.

The tuning or affecting parameters of RBF kernel components are gamma and C. The splitting ratio of train-to-test is set up at 80:20. Table 15 shows the execution of summary

Table 14 Summary results of monthly SVR regression type 2 RBF kernel

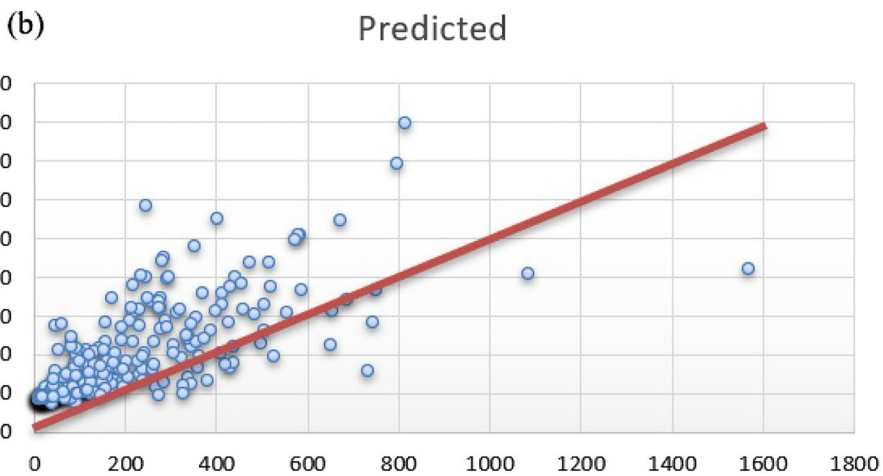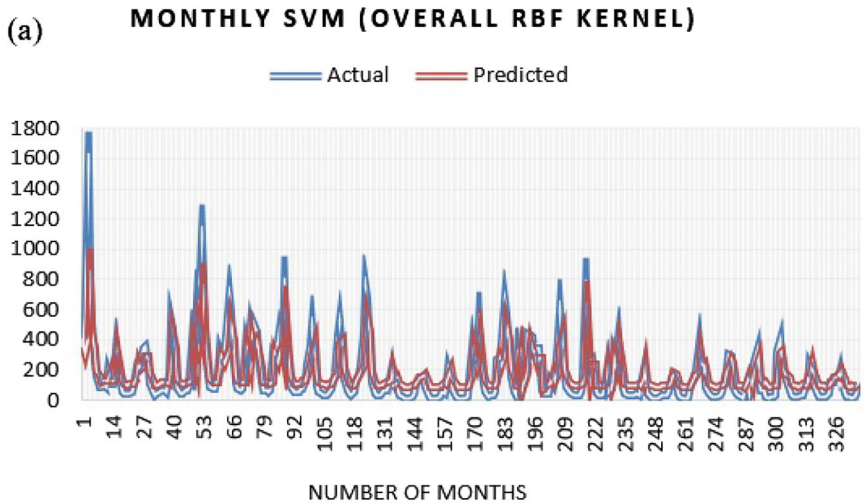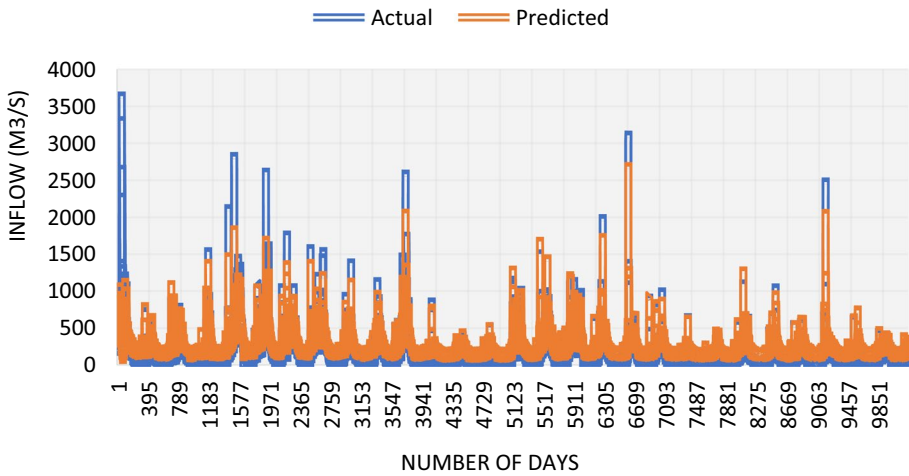| SVR regression type 2 (Model-2) RBF kernel (monthly)	RMSE	R^2	NSE
Overall	131.3	0.55	0.52
Training	135.5	0.54	0.50
Testing	117.6	0.61	0.60

Fig. 13 The performance of overall set of Model-2 for SVR regression type 2 RBF kernel **a** actual vs predicted monthly reservoir inflow and **b** scatter plot

results of the model performance of RMSE, R^2, and NSE for the proposed models. Based on the results, Model-2 has the highest accuracy, the RMSE = 150.328, R^2 = 75, and NSE = 0.56, respectively, for the overall dataset. Figure 14 shows the overall set of Model-2

Table 15 Results of overall set of SVR RBF kernel regression type 2 (second scenario) for forecasting daily reservoir inflow

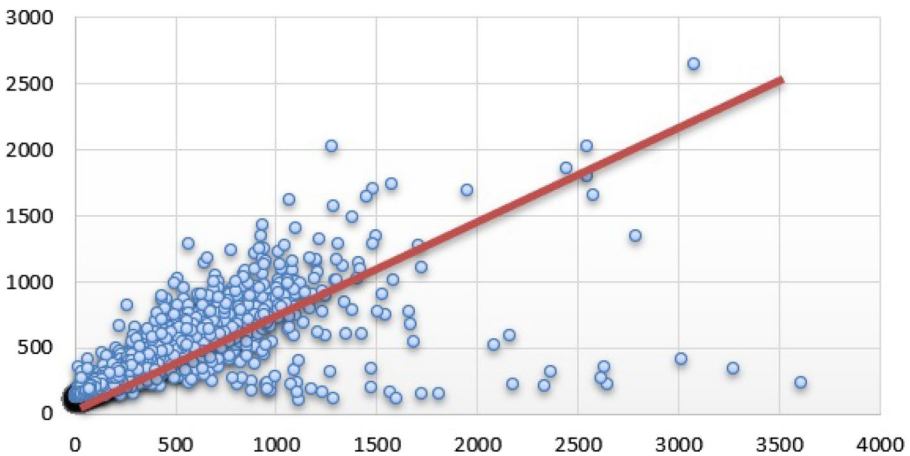| SVR (RBF kernel)	RMSE	R^2	NSE
Model-1	151.643	0.73	0.55
Model-2	150.328	0.75	0.56
Model-3	151.02	0.74	0.56
Model-4	154.90	0.71	0.54
Model-5	155.92	0.70	0.53

Fig. 14 The performance of overall set of Model-2 for SVR regression type 2 RBF kernel (second scenario) **a** actual vs predicted daily reservoir inflow and **b** scatter plot

performance for SVR regression type 2 (second scenario) RBF kernel for forecasting daily reservoir inflow.

So based on the results, the first scenario performed better accuracy than the second scenario; therefore, the first scenario will be used for the rest of the models. Depending on the outcomes and the achieved results, SVR regression type 2 was better than SVR regression type 1. Besides, among all the SVR kernels, RBF attained the best result; thus, it is considered the best kernel. The second-best kernel was the linear kernel, and each sigmoid and NP came after them. Relying on the different models, Model-2 was the best among the rest. While Model-3, Model-1, Model-4, and Model-5 performed from the best to the least performance, respectively. Depending on the different time-lags, the data analyzed showed that the daily reservoir inflow outperformed the weekly and monthly time-lags. In contrast, the weekly time-lag showed better performance than the monthly time-lag. Based on the two different scenarios, the first scenario that has reservoir inflow as the input parameter showed a significant outcome and the best result compared to the second scenario, which was the combination of reservoir inflow and rainfall as input parameters. Thus, the first scenario will be selected to be applied in other techniques.

3.9 Analysis of SVR results

The results of the SVR method are similar to the previous studies as it has some limitations. According to Zhang et al., 2020, who have conducted a study at Huanren reservoir in China, the contribution of the data-driven model through SVR is limited and has substantial seasonal variation. It is more significant in winter and summer but more minor in spring and autumn. According to the different time scales, results from Wushan and Weijiabao hydrologic stations, China, that presented by Hu et al., 2020, revealed that the reliability of the forecasting decreased as the foresight period increased. This indicates that the SVR prediction model could constantly achieve virtuous performance in the testing stage and had relative stability. On the other hand, Mohsenzadeh Karimi et al., 2021, revealed that implementing SVR could predict river-flow time series with decent accuracy in Alaviam Dam, Soofi-Chai River in Iran. This is also relevant to the current research because the outcome of implementing SVR indicates that this machine learning method can be applicable and reliable but not with significant accuracy. Moreover, Yu et al., 2020, showed the result of their experiment on Three Gorges Dam (TGD) in China and demonstrated that a hybrid model that consists of Fourier transform and SVR is able to drive near-perfect 10-day streamflow forecasting. Meanwhile, Al-Suhili & Karim, 2015, revealed that their developed ANN model was successfully capable of predicting daily inflow at Dokan dam with a correlation coefficient of 0.94. The result of the current study is consistent with the results of the previous studies in the literature. For example, Babaei et al., 2019, applied ANN and SVR for predicting monthly inflow and SVR outperformed ANN according to the findings. On the other hand, Wang et al., 2014, could achieve a well-performed results for predicting monthly inflow utilizing SVR with the adaptation of RBF. Moreover, Halik et al. 2015, showed that WSVM with selecting RBF could accurately predict inflow. It can be mentioned that the only suitable kernel for SVR was RBF since the other kernels could not perform well with the three statistical indices (RMSE, R^2, and NSE). Therefore, it is recommended to develop SVR with adaptation of RBF for predicting other hydrological parameters such as rainfall, temperature, wind, evaporation, and humidity.

Table 16 Model-2 daily first scenario

Model-2	SVR		
	RMSE	R^2	MSE
Overall	145.7	0.85	2.07
Training	145.1	0.84	2.74
Testing	145.5	0.88	8.51

Table 17 T-test and F-test values for Dokan dam

Dams	T-test value	F-test value
Dokan dam	0.1	8.225

Table 18 Extreme values for observed and forecasted inflow of Dokan dam

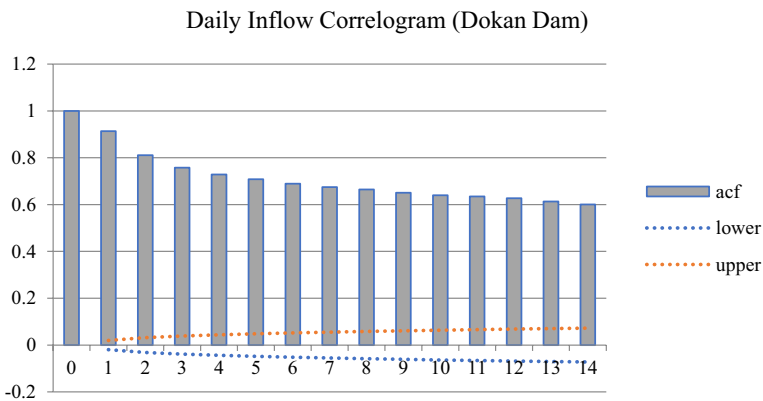| Time indices	Minimum value		Maximum value	
	Actual	Predicted	Actual	Predicted
Daily	3	21.2	3608	2650
Weekly	7.57	27.8	2298	692.97
Monthly	8.74	74.34	1569.16	795.38

Fig. 15 Autocorrelation function value for the daily time-series inflow data at Dokan dam

Based on the previous literature and the current study, it is shown that the SVR model is one of the good models for forecasting reservoir inflow but not the best one. Table 16 shows the best model (Model-2) for the first scenario of Dokan dam.

In order to check if the model can resemble the overall mean and variance, the daily means, and variances were performed using the t-test and F-test (Table 17).

When the null hypothesis is correct, there is a good chance of getting a t-value between -2 and + 2. The model performs better with the larger *F* value. Therefore, the

Table 19 Summarization of all SVR results for different scenarios based on hyperparameters, time horizons, and proposed models

Timescale	SVM kernels	Models	RMSE	R^2	NSE
Daily	Linear	Model-1	159.1	0.83	0.50
		Model-2	157.5	0.84	0.52
		Model-3	160.7	0.83	0.50
		Model-4	160.4	0.83	0.50
		Model-5	160.2	0.83	0.50
	NP	Model-1	701.3	0.22	−8.67
		Model-2	695.2	0.24	−8.34
		Model-3	698.9	0.22	−8.65
		Model-4	699.3	0.21	−8.66
		Model-5	700.2	0.21	−8.66
	RBF	Model-1	148.2	0.84	0.57
		Model-2	145.7	0.85	0.59
		Model-3	149.2	0.85	0.57
		Model-4	156.1	0.85	0.53
		Model-5	161.6	0.85	0.49
	Sigmoid	Model-1	719.1	0.83	−8.98
		Model-2	716.4	0.84	−8.92
		Model-3	720.7	0.83	−8.99
		Model-4	718.9	0.82	−8.95
		Model-5	721.3	0.83	−9.01
Weekly	Linear	Model-1	148.9	0.78	0.46
		Model-2	142.6	0.80	0.47
		Model-3	149.8	0.77	0.46
		Model-4	148.3	0.78	0.46
		Model-5	148.1	0.78	0.46
	NP	Model-1	631.8	0.20	−8.25
		Model-2	629.4	0.21	−8.23
		Model-3	630.9	0.20	−8.24
		Model-4	631.4	0.20	−8.25
		Model-5	630.5	0.20	−8.24
	RBF	Model-1	133.2	0.70	0.58
		Model-2	131.9	0.71	0.59
		Model-3	133.5	0.70	0.58
		Model-4	134.7	0.70	0.58
		Model-5	133.9	0.70	0.58
	Sigmoid	Model-1	644.8	0.80	−8.22
		Model-2	641.3	0.81	−8.16
		Model-3	643.2	0.80	−8.21
		Model-4	644.9	0.80	−8.22
		Model-5	645.1	0.79	−8.23

Table 19 (continued)

Timescale	SVM kernels	Models	RMSE	R^2	NSE
Monthly	Linear	Model-1	143.7	0.77	0.44
		Model-2	141.8	0.78	0.45
		Model-3	143.2	0.77	0.44
		Model-4	144.1	0.77	0.44
		Model-5	143.7	0.77	0.44
	NP	Model-1	620.4	0.18	−8.19
		Model-2	618.7	0.19	−8.17
		Model-3	620.3	0.18	−8.20
		Model-4	621.1	0.18	−8.20
		Model-5	620.8	0.18	−8.19
	RBF	Model-1	133.2	0.54	0.50
		Model-2	131.3	0.55	0.52
		Model-3	133.8	0.54	0.50
		Model-4	132.9	0.54	0.50
		Model-5	133.4	0.54	0.50
	Sigmoid	Model-1	640.4	0.76	−8.38
		Model-2	638.1	0.78	−8.35
		Model-3	641.1	0.77	−8.33
		Model-4	640.8	0.77	−8.32
		Model-5	640.5	0.77	−8.30

results of t-test and F-test for observed and forecasted inflow was acceptable. In order to check if the proposed model could accurately predict the extreme values, the minimum and maximum values for observed and forecasted inflow were compared (Table 18). According to the results, the model was not capable of forecasting extreme values; however, it was capable of forecasting the overall values except extreme values. Moreover, the ACF resemblance by comparing the correlogram for the original and the forecasted time-series up to 14 lags was performed (Fig. 15).

In order to show the performance of the proposed models clearly, Table 19 represents the summarization of all SVR results for different scenarios based on hyperparameters, time horizons, and proposed models.

3.9.1 Warragamba dam results

SVR techniques were not performed acceptable results. Figure 16 shows the performance accuracy of the overall set of Model-2 for SVR, for forecasting daily reservoir inflow at Warragamba dam in Australia.
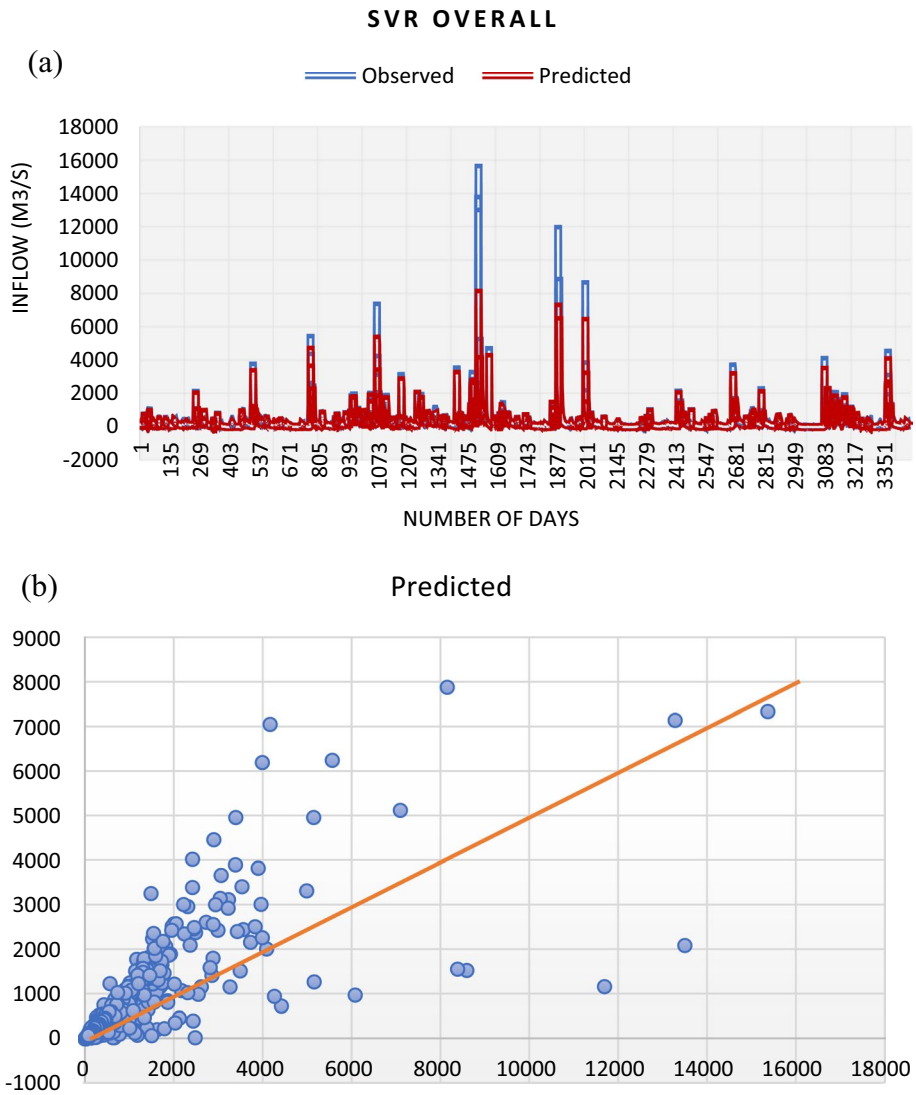
SVR OVERALL

(a)

(b)

Predicted

Fig. 16 The performance of overall set of Model-2 for SVR regression type 2 RBF kernel (first scenario) **a** actual vs predicted daily reservoir inflow and **b** scatter plot

Model-2	SVR		
	RMSE	R^2	MSE
Overall training + testing	429.32	0.66	53.0

Table 20 Daily time-series reservoir inflow prediction at Warragamba dam (first scenario)

In contrast, SVR had good accuracy. Depending on the results, the developed SVR model in this study provided a suitable accuracy in Kurdistan region of Iraq. Furthermore, the current SVR models could not be utilized to forecast inflow for ensuring flood protection and increasing water supply efficiently in Warragamba dam. Since inflow is a primary input into reservoirs, its accurate forecasting can aid reservoir development and management. Based on the previous studies, the majority of the results showed the highest performance when using SVR model. Table 20 represents the overall result of the first scenario at Warragamba dam.

3.9.2 The most appropriate selections for the best-performed model

Selecting inflow leads to superior performance than selecting inflow and rainfall together. The main reason is that rainfall and inflow rates at Dokan dam are different from one another. Sometime, there is only precipitation since Dokan dam will not get inflow from Tigris river as the sharing source from Iran side. Sometime, there will be precipitation, as well as receiving inflow. Therefore, selecting inflow as the only parameter will lead to better accuracy prediction in the proposed models. Meanwhile, daily time-series data will get better performance in terms of accuracy since the more data added to the proposed models, the model will train better. SVR regression type 2 outperformed SVR regression type 1 with a very little rate. Regarding the input selection, Model-2 has a significant result since it is the most correlated value to the actual values.

4 Conclusion

The objective of this study is to implement a machine learning method, namely, SVR for forecasting reservoir inflow. Two scenarios were proposed in this research. The first scenario includes reservoir inflow only as an input parameter. In contrast, the second scenario had a combination of inflow and rainfall as the input parameters. The first scenario outperformed the second scenario with a significant difference in the accuracy level. Daily, weekly, and monthly reservoir inflow data were the three selected time-lags in the current research. The findings showed that daily time-series reservoir inflow obtained the highest accuracy in comparison with the weekly and monthly time-lags. In contrast, weekly outperformed the monthly time-lags. Two locations (Kurdistan region of Iraq and Sydney, Australia) were selected in the current research in order to ensure if the proposed model could be generalized for different climates. The outcomes indicate that the proposed models could not be generalized. The proposed models performed well for Dokan dam only. This study has contributed to the field of water resources engineering in relation to forecasting models by directing the attention of researchers, instructors, and policymakers. Although many research studies have been conducted on a particular model or benchmarking models of reservoir inflow prediction, there has not been any study thus far, to the best knowledge of this researcher, to implement SVR for predicting reservoir inflow on both Kurdistan regions of Iraq and Australia in terms of accuracy. Therefore, the present study has served to fill this gap in the literature. It is recommended for future studies to apply other machine learning methods for different climate zones in order to ensure that it can be generalized.

Declarations

References

Aljanabi, Q. A., Chik, Z., Allawi, M. F., El-Shafie, A. H., Ahmed, A. N., & El-Shafie, A. (2018). Support vector regression-based model for prediction of behavior stone column parameters in soft clay under highway embankment. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-016-2807-5

Al-Suhili, R. H., & Karim, R. A. (2015). Daily inflow forecasting for Dukan reservoir in Iraq using artificial neural networks. *International Journal of Water, 9*(2), 194–208. https://doi.org/10.1504/IJW.2015.068961

Amnatsan, S., Yoshikawa, S., & Kanae, S. (2018). Improved forecasting of extreme monthly reservoir inflow using an analogue-based forecasting method: A case study of the Sirikit Dam in Thailand. *Water (switzerland)*. https://doi.org/10.3390/w10111614

Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., & Chau, K. W. (2020). Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water (switzerland)*. https://doi.org/10.3390/w12051500

Babaei, M., Moeini, R., & Ehsanzadeh, E. (2019). Artificial Neural Network and Support Vector Machine Models for Inflow Prediction of Dam Reservoir (Case Study: Zayandehroud Dam Reservoir). *Water Resources Management, 33*(6), 2203–2218. https://doi.org/10.1007/s11269-019-02252-5

Bábek, O., Kielar, O., Lenďáková, Z., Mandlíková, K., Sedláček, J., & Tolaszová, J. (2020). *Reservoir deltas and their role in pollutant distribution in valley-type dam reservoirs: Les Království Dam*. Elbe River, Czech Republic. https://doi.org/10.1016/j.catena.2019.104251

Borgonovo, E. (2017). Sensitivity analysis: An introduction for the management scientist. *Springer*. https://doi.org/10.1007/978-3-319-52259-3

Dhungel, S., Tarboton, D. G., Jin, J., & Hawkins, C. P. (2016). Potential effects of climate change on ecologically relevant streamflow regimes. *River Research and Applications*. https://doi.org/10.1002/rra.3029

Division, W. (2005). KOWMUNG RIVER KANANGRA-BOYD NATIONAL PARK Wild River Assessment, (June).

Ehteram, M., Singh, V. P., Ferdowsi, A., Mousavi, S. F., Farzin, S., Karami, H., et al. (2019). An improved model based on the support vector machine and cuckoo algorithm for simulating reference evapotranspiration. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0217499

Ezz-Aldeen, M., Hassan, R., Ali, A., Al-Ansari, N., & Knutsson, S. (2018). Watershed sediment and its effect on storage capacity: Case study of Dokan Dam Reservoir. *Water (switzerland), 10*(7), 1–16. https://doi.org/10.3390/w10070858

Halik, G., Anwar, N., Santosa, B., & Edijatno. (2015). Reservoir inflow prediction under GCM scenario downscaled by wavelet transform and support vector machine hybrid models. *Advances in Civil Engineering*. https://doi.org/10.1155/2015/515376

Herslund, L., & Mguni, P. (2019). Examining urban water management practices – challenges and possibilities for transitions to sustainable urban water management in Sub-Saharan cities. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2019.101573

Hu, H., Zhang, J., & Li, T. (2020). A comparative study of VMD-based hybrid forecasting model for nonstationary daily streamflow time series. *Complexity*. https://doi.org/10.1155/2020/4064851

Jia, H., Yao, H., Tang, Y., Yu, S. L., Field, R., & Tafuri, A. N. (2015). LID-BMPs planning for urban runoff control and the case study in China. *Journal of Environmental Management*. https://doi.org/10.1016/j.jenvman.2014.10.003

Kooy, M., Furlong, K., & Lamb, V. (2020). Nature based solutions for urban water management in Asian cities: Integrating vulnerability into sustainable design. *International Development Planning Review*. https://doi.org/10.3828/idpr.2019.17

Lai, V., Ahmed, A. N., Malek, M. A., Afan, H. A., Ibrahim, R. K., El-Shafie, A., & El-Shafie, A. (2019). Modeling the nonlinearity of Sea level oscillations in the Malaysian Coastal areas using machine learning algorithms. *Sustainability (switzerland)*. https://doi.org/10.3390/su11174643

Lai, V., Malek, M. A., Abdullah, S., Latif, S. D., & Ahmed, A. N. (2020). Time-series prediction of sea level change in the east coast of Peninsular Malaysia from the supervised learning approach. *International Journal of Design and Nature and Ecodynamics., 15*(3), 409–415. https://doi.org/10.18280/ijdne.150314

Latif, S. D. (2021). Concrete compressive strength prediction modeling utilizing deep learning long short-term memory algorithm for a sustainable environment. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-021-12877-y

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

Latif, S. D., Azmi, M. S. B. N., Ahmed, A. N., Fai, C. M., & El-Shafie, A. (2020). Application of artificial neural network for forecasting nitrate concentration as a water quality parameter a case study of feitsui reservoir Taiwan. *International Journal of Design and Nature and Ecodynamics, 15*, 647–652. https://doi.org/10.18280/ijdne.150505

Latif, S. D., Ahmed, A. N., Sathiamurthy, E., Huang, Y. F., & El-Shafie, A. (2021a). *Evaluation of deep learning algorithm for inflow forecasting : A case study of Durian Tunggal Reservoir*. Peninsular Malaysia: Natural Hazards. https://doi.org/10.1007/s11069-021-04839-x

Latif, S. D., Birima, A. H., Najah, A., Mohammed, D., Al-ansari, N., Ming, C., & El-shafie, A. (2021b). Development of prediction model for phosphate in reservoir water system based machine learning algorithms. *Ain Shams Engineering Journal*. https://doi.org/10.1016/j.asej.2021.06.009

Marlow, D. R., Moglia, M., Cook, S., & Beale, D. J. (2013). Towards sustainable urban water management: A critical reassessment. *Water Research*. https://doi.org/10.1016/j.watres.2013.07.046

Mohsenzadeh Karimi, S., Karimi, S., & Poorrajabali, M. (2021). Forecasting monthly streamflows using heuristic models. *ISH Journal of Hydraulic Engineering, 27*(1), 73–78. https://doi.org/10.1080/09715010.2018.1516575

More, D., Magar, R. B., & Jothiprakash, V. (2019). Intermittent reservoir daily inflow prediction using stochastic and model tree techniques. *Journal of the Institution of Engineers (india): Series A, 100*(3), 439–446. https://doi.org/10.1007/s40030-019-00368-w

Najah, A., Teo, F. Y., Chow, M. F., Huang, Y. F., Latif, S. D., Abdullah, S., et al. (2021). Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia. *International Journal of Environmental Science and Technology*. https://doi.org/10.1007/s13762-021-03139-y

Nguyen, T. T., Ngo, H. H., Guo, W., Wang, X. C., Ren, N., Li, G., et al. (2019). Implementation of a specific urban water management-Sponge City. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2018.10.168

O'Keeffe, J., Piniewski, M., Szcześniak, M., Oglęcki, P., Parasiewicz, P., & Okruszko, T. (2019). *Index-based analysis of climate change impact on streamflow conditions important for Northern Pike*. Chub and Atlantic salmon: Fisheries Management and Ecology. https://doi.org/10.1111/fme.12316

Qi, Y., Zhou, Z., Yang, L., Quan, Y., & Miao, Q. (2019). A Decomposition-ensemble learning model based on LSTM neural network for daily reservoir inflow forecasting. *Water Resources Management, 33*(12), 4123–4139. https://doi.org/10.1007/s11269-019-02345-1

Sosa-Rodriguez, F. S., Tapia Silva, F. O., & Alvarado Arriaga, V. Y. (2019). Urban Water and Sanitation. In: *Sustainable Cities and Communities* (pp. 945–953). https://doi.org/10.1007/978-3-319-95717-3_44

Sulaiman, S. O., Abdullah, H. H., Al-Ansari, N., Laue, J., & Yaseen, Z. M. (2021). Simulation model for optimal operation of Dokan Dam reservoir northern of Iraq. *International Journal of Design and Nature and Ecodynamics, 16*(3), 301–306. https://doi.org/10.18280/IJDNE.160308

Tikhamarine, Y., Souag-Gamane, D., Najah Ahmed, A., Kisi, O., & El-Shafie, A. (2020). Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *Journal of Hydrology, 582*, 124435. https://doi.org/10.1016/j.jhydrol.2019.124435

US Environmental Protection Agency (U.S. EPA), & US EPA (2008). Climate Change Effects on Stream and River Biological Indicators: a Preliminary Analysis EPA EPA 600-R-07-085 . *Global Change Research Program, National Center for Environmental Assessme. Washington, DC*.

Wang, Y., Guo, S., Chen, H., & Zhou, Y. (2014). Comparative study of monthly inflow prediction methods for the Three gorges reservoir. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-013-0772-4

Wang, H., Mei, C., Liu, J. H., & Shao, W. W. (2018). A new strategy for integrated urban water management in China: Sponge city. *Science China Technological Sciences*. https://doi.org/10.1007/s11431-017-9170-5

Xu, B., Huang, X., Zhong, P. A., & Wu, Y. (2020). Two-Phase Risk Hedging Rules for Informing Conservation of Flood Resources in Reservoir Operation Considering Inflow Forecast Uncertainty. *Water Resources Management*. https://doi.org/10.1007/s11269-020-02571-y

Yahya, A. S. A., Ahmed, A. N., Othman, F. B., Ibrahim, R. K., Afan, H. A., El-Shafie, A., et al. (2019). Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water (switzerland)*. https://doi.org/10.3390/w11061231

Yan, L., McManus, P., & Duncan, E. (2018). Understanding ethnic differences in perceptions, attitudes, and behaviours: A study of domestic water use in Sydney. *Geographical Research, 56*(1), 54–67. https://doi.org/10.1111/1745-5871.12244

Yu, Y., Wang, P., Wang, C., Qian, J., & Hou, J. (2017). Combined monthly inflow forecasting and multiobjective ecological reservoir operations model: Case study of the three gorges reservoir. *Journal of Water Resources Planning and Management, 143*(8), 05017004. https://doi.org/10.1061/(asce)wr.1943-5452.0000786

Yu, X., Wang, Y., Wu, L., Chen, G., Wang, L., & Qin, H. (2020). Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. *Journal of Hydrology, 582*, 124293. https://doi.org/10.1016/j.jhydrol.2019.124293

Zakaria, S., Mustafa, Y. T., Mohammed, D. A., Ali, S. S., Al-Ansari, N., & Knutsson, S. (2013). Estimation of annual harvested runoff at Sulaymaniyah Governorate. *Kurdistan Region of Iraq Natural Science, 05*(12), 1272–1283. https://doi.org/10.4236/ns.2013.512155

Zhang, X., Wang, H., Peng, A., Wang, W., Li, B., & Huang, X. (2020). Quantifying the uncertainties in data-driven models for reservoir inflow prediction. *Water Resources Management*. https://doi.org/10.1007/s11269-020-02514-7