




The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River

Lamine Diop^{1,2} · Ansoumana Bodian³ · Koffi Djaman⁴ · Zaher Mundher Yaseen⁵  · Ravinesh C. Deo⁶ · Ahmed El-shafie⁷ · Larry C. Brown²

Received: 15 November 2017 / Accepted: 20 February 2018 / Published online: 27 February 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Ideal prediction and modeling of stream-flow and its hydrological applications are extremely significant for decision-making tasks and proper planning of water resource and hydraulic engineering. In the last two decades, the potential of soft computing approaches has increased dramatically in engineering and science problems. In this research, the utility of two soft computing approaches, namely support vector regression (SVR) model and generalized regression neural network (GRNN), is validated to predict 1 day ahead daily river flow data in the upper Senegal River basin at the Bafing Makana station in West Africa. The modeling is conducted by including the climatological information in the modeled stream-flow patterns. Correlation procedure is established and applied to obtain the modeling of the input variables with statistically significant lagged datasets at $t - 1$, $t - 2$, and $t - 3$ used as three input combination for each case study scenario. Different statistical indicators are used to evaluate the accuracy of the prediction models. The results show that the accuracy of the models varied by the scenario and the input datasets, where the SVR model yielded the best results for both modeling scenarios. It is also evident that combining the historical stream-flow data with the rainfall and evapotranspiration can ameliorate substantially the accuracy of the two models for predicting 1-day ahead stream-flow. A comparison of the optimal SVR and GRNN models in this problem indicates that SVR exhibits superior performance to the GRNN model in estimating the daily stream-flow data, irrespective of the modeling scenario and the datasets that is applied. The findings offer an opportunity to apply SVR model for predicting daily stream-flow, with less data requirement for the investigated Senegal River basin.

Keywords Climatic inputs · Stream-flow pattern · Forecasting · SVR · GRNN · Senegal River basin

✉ Zaher Mundher Yaseen
yaseen@tdt.edu.vn

¹ UFR S2ATA « Sciences Agronomiques, de l'Aquaculture et des Technologies Agroalimentaires », Université Gaston Berger (UGB), BP 234, Saint Louis, Senegal

² Department of Food, Agricultural and Biological Engineering, The Ohio State University, 590 Woody Hayes Dr., Columbus, OH 43210, USA

³ Laboratoire Leïdi « Dynamique des territoires et développement », Université Gaston Berger (UGB), BP 234, Saint Louis, Senegal

⁴ Department of Plant and Environmental Sciences, Agricultural Science Center at Farmington, New Mexico State University, P.O. Box 1018, Farmington, NM 87499, USA

⁵ Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁶ School of Agricultural, Computational and Environmental Sciences, International Centre for Applied Climate Sciences (ICACS), Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, Toowoomba, Australia

⁷ Department of Civil Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

Introduction

Stream-flow modeling is extremely important for the management of water resources (Yaseen et al. 2015b). Precise river flow forecasts are essential for the design of hydraulic structures (e.g., dams and irrigation schemes) that are inherently associated to stream-flow and flood gates and drought impacts study. This can provide decision-makers very significant information for rural and urban development projects, environmental impact assessments, irrigation schemes design and management, and reservoir operations (Badrzadeh et al. 2013). Development of accurate and reliable predictive models that can help extracting quantitative information from the antecedent patterns that are embedded in data related to stream-flow can facilitate authorities administer water reservoirs in an optimal manner for water management, hydropower generation, agriculture and domestic and industrial water planning (Kisi and Cimen 2011).

There are two general categories of river flow forecasting models: (1) process-driven (physical or hydrological based) models and (2) data-driven (soft computing-based models) (Wang et al. 2006). The first category of models attempt applied for stream-flow predictions have been developed to describe and represent the physical processes in terms of relatively complex, deterministic-type mathematical equations, whereas the second category of models are purely of black box nature and do not need the understanding of the underlying physical process which governs the phenomena. Instead, data-driven models, to forecast the stream-flow values, utilize machine learning algorithms to extract pertinent data patterns and attributes. The only requirement in data-driven models is a set of hydrological variables related to stream-flow (e.g., rainfall, evaporation, etc.) that can provide the features for predictive modeling. With a sufficient amount of data, empirical equations are developed from the calibrated dataset, thus, providing distinct advantage in terms of their simplicity, low computational cost, and competitive performance relative to process-driven models.

Recently, the use of data-driven techniques to forecast stream-flow has received particular consideration from water resource specialists and researchers (Prairie et al. 2006; Toth and Brath 2007; Kashid et al. 2010; Kuo et al. 2010; Ni et al. 2010; Guimarães Santos and Silva 2014; Makwana and Tiwari 2014; Taormina and Chau 2015). These models, which were applied in a number of geographically diverse hydro-climatic zones, have shown good capability to generate reasonably accurate modeling results of stream-flow (Kumar et al. 2016). Among the primary predictive methods that have been considered lately, the application of support vector machine (SVM) and artificial neural network (ANNs) models has become prominent due to their broad application to diverse scientific domains (Yaseen et al. 2015a).

In general, ANNs models applied in hydrological modeling include (1) the generalized regression neural network (GRNN), (2) the radial basis function (RBF), and (3) the feed-forward back propagation (FFBP) models (Nourani et al. 2014). Over the last two decades, the application of FFBP, RBF, and GRNN models has increased especially in hydrological sciences (Fahimi et al. 2016). In comparison with many statistical modeling techniques, various forms of ANN models have showed significantly superior accuracy, particularly applications in stream-flow discharge forecasting, prediction of surface runoff and flood, stream-flow and water level predictions (Kagoda et al. 2010; Shiri and Kisi 2010). In a recent study, Tayyab et al. (2016) investigated stream-flow discharge forecasting by applying variations of the artificial neural network model including the training algorithms based on FFBPNN, RBFNN, and GRNN for the case of Jinsha River Basin in China (Tayyab et al. 2016). The results showed that the ANN model performance exceeded the performance of the statistical autoregressive (AR) model. By employing a number of test cases, the performance of the FFBPNN model was superior to all other models; however, for most cases, the GRNN model performed better than the RBFNN. In this paper, the GRNN model has been adopted for stream-flow forecasting.

Similar to an ANN model, a SVR model that utilizes a different modeling framework in terms of an application of kernel function for feature extraction has good potential to analyze unknown relationships between a set of input variables and the objective variable (Raghavendra and Deka 2014). Basically, SVR model can yield solutions in forecasting and predicting problems by means of pattern recognition techniques based on the structural risk minimization, and therefore, avoids the issue of over fitting the dataset (Liu and Lu 2014). Consequently, SVR model has been applied in hydrology and environmental applications in recent decades (Ch et al. 2013; Deo et al. 2016). Most recently, Wen et al. (2015) used limited climatic datasets to forecast the daily reference evapotranspiration by developing a SVR model (Wen et al. 2015). Their results showed that the SVR model was superior to the well-known empirical models that are commonly (e.g., the Priestley–Taylor, Hargreaves, and Ritchie model). In temperate and arid climate zones, Kim et al. (2012) tested the multilayer perception neural networks (MLP), GRNN, and SVR to simulate evapotranspiration (Kim et al. 2012), while (Gong et al. 2016) estimated monthly evapotranspiration using an SVR model compared with GRNN, multivariate adaptive regression spline (MARS) fuzzy genetic (FG), MLR, and adaptive neuro-fuzzy inference systems with grid partition (ANFIS-GP). In spite of numerous applications of these models in hydrology, to our knowledge, no studies have applied an ANN or SVR model for forecasting stream-flow in upper Senegal River basin, which falls in an important ecological zone in West Africa.

Recurrent droughts of the seventies in the Sahel regions have deeply affected crop production. This situation has brought the global and local communities to look for sustainable solutions in order to mitigate drought effects in these regions (Ndiaye 2010). For instance, in 1972, the Senegal River basin riparian states (Senegal, Mauritania, and Mali) have joined their efforts to create a river basin management organization called “Organisation pour la Mise en Valeur du fleuve Senegal (OMVS).” The main goal of this organization is to better manage the water resources, to develop irrigated agriculture, and to generate substantiate hydropower. OMVS led the construction of two dams: Manantali in the upstream of the river and Diama in the delta of the same river, and recently a new one (Felou) in Mali. Diama and Manantali dams have allowed the availability of 375,000 ha irrigable land in the Senegal River basin and thus facilitated the development of irrigated agriculture in this zone (Varis and Lahtela 2002).

It is particularly notable that the Manantali dam is built on the Bafing tributary and its primary purpose is to generate electric power, and to store water in the wet season for augmenting dry-season outflows for the benefit of agriculture and related irrigation practices. Its power production in terms of hydropower energy is estimated to be approximately 800 Gwh/year, guaranteed 9 out of 10 years for three of the country members of the OMVS (Mali, Mauritania, and Senegal). Therefore, any negative impacts of water supply with respect to the Manantali inflows is likely to affect the energy production as well as the agriculture sectors of those countries, which are the two keys elements of the African economy. Thus, it is crucial that we monitor, evaluate, and predict with good level of accuracy the availability of water in the Senegal River basin, particularly in its upper region that has influence of water management practices downstream of the river system.

Considering the importance of stream-flow knowledge in irrigations, water management and hydraulic infrastructural design in the Senegal River basin, this paper aims to investigate the capability of the SVR (Vapnik 1995) and GRNN (Specht 1991) models for forecasting and predicting daily stream-flow in the upper Senegal River basin at Bafing Makana station. To optimize the forecasting models, predictor variables based on antecedent stream-flow, rainfall, and evapotranspiration (1961–2014) within the Senegal River basin are applied.

Case study and methodological background

Study area

This research is conducted at the upper Senegal River basin (bounded by latitude 10°30′ and 12°30′N and longitude

12°30′ and 9°30′W). The total area of the study zone (the upper Senegal River basin at Bafing Makana station) is 21,290 km² covering a part of Guinea Conakry and Mali (Fig. 1). It has a dense hydrographic network (Kane and Diallo 2005; Bodian et al. 2016); but considering the groundwater resources, the nature of the soil and the geological formations are not favorable to the existence of large aquifers. The area is characterized by the movement of the Inter Tropical Convergence Zone from south to north which directs the penetration of the West African monsoon driven by the thermal contrast between the Atlantic Ocean and the continent (Dione 1996). The climate of the basin is Guinean–Sudanese with a majority of the rainfall falling from April to October. The average rainfall of the basin is 1490 mm/year (Bodian et al. 2016). The elevation varies from 215 to 1389 m, and the slope indices decrease from upstream to downstream which points out the significance of the mountainous region of Fouta Djallon.

Data

Daily rainfall data from 12 rain gauges and daily temperature data from 5 meteorological stations are used to develop the present stream-flow forecasting models. The data are collected from Mali and Guinea National Meteorological Agencies. Daily stream flows from 1961 to 2014 at the station of Bafing Makana are obtained from the Senegal River Basin Organization (OMVS). The meteorological information including in this study are the rainfall and evapotranspiration datasets over the period (1961–2004). It can be noted, the most recent data information did not included in the modeling. This was mainly due a lack of the most recent period data. Therefore, a total of 43 years of the predictor data series that encompassed both the wet and the dry periods in this region are considered in this research paper.

Support vector regression (SVR)

Support vector regression (SVR) has been introduced by Vapnik as a novel statistical learning tool applied in complex prediction problems (Vapnik 1995). The basic idea of this technique is to map the data X into a high-dimensional feature space via a nonlinear mapping function to perform linear regression in this space (Wang et al. 2009). SVR is composed of a computer algorithm that learns the predictor data by examples to deduce the best function for the classifier/hyperplane in order to divide and analyze the linearly and nonlinearly separable data in the input space (Ghorbani et al. 2016). Figure 2a gives an example of the linearly separated data by means of support vectors within a hyperplane region. The use of kernel functions to map input data to a higher dimensional space makes the strength of the SVR

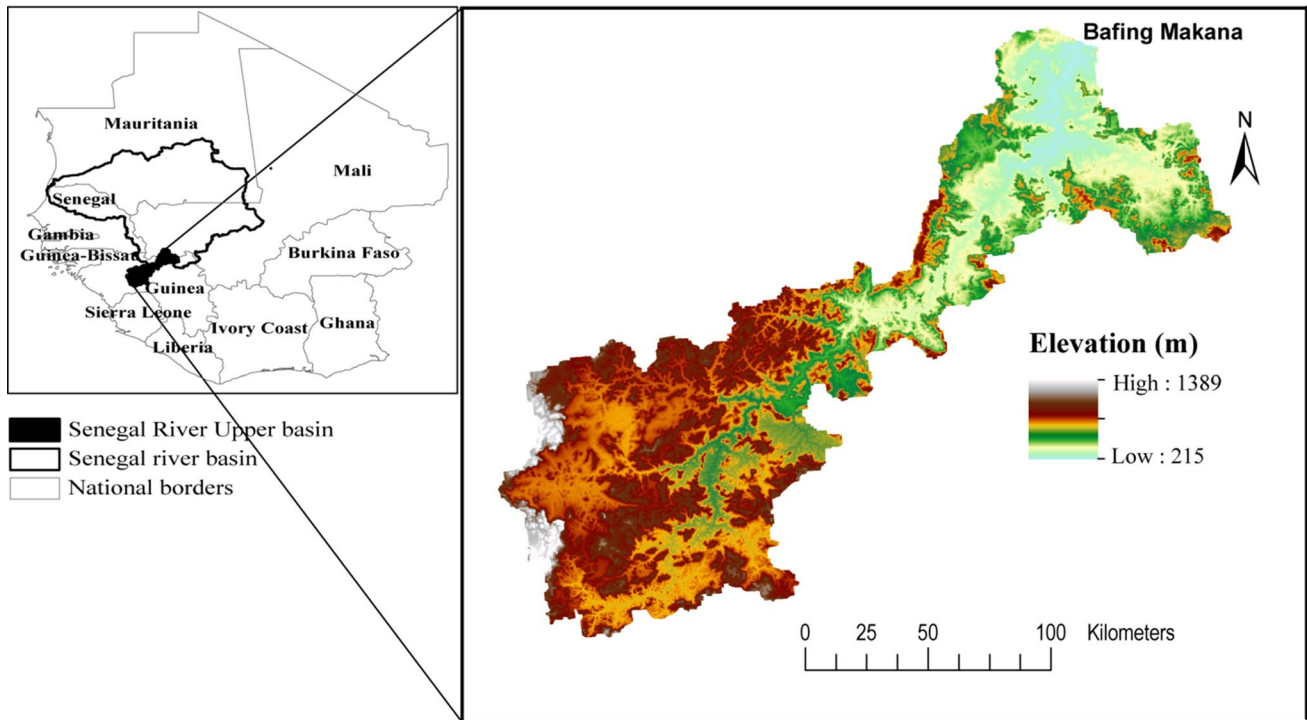


Fig. 1 Location map of the upper Senegal River basin

model (Jain 2012), both for data classification and regression purposes.

As the dataset used for stream-flow forecasting is numeric and exhibits statistical relationships between predictors, this paper has applied a regression form of the SVR model: support vector regression (SVR) algorithm. Further theoretical detail of SVR model can be found in the recent work of (Raghavendra and Deka 2014).

Figure 2a, b illustrates the basic form of an SVR model. Considering a couple of series of data $(x_i, y_i) \in (X \times Y)$ where i varying from 1 to m (the total number of data patterns), $x_i \in X = R^n$ is the predictor vector and $y_i \in Y = R^1$ is the matching output (here, stream-flow), and the SVR model is described as follows (Raghavendra and Deka 2014):

$$f(X) = W_i \cdot \emptyset(X) + b \tag{1}$$

where W_i , $\emptyset(X)$, and b are the weight vector, the nonlinear transfer function that maps the input vectors into a high-dimensional feature space, and the bias, respectively.

In order to forecast the objective variable (i.e., stream-flow), the magnitudes of weight vector and bias are derived by minimizing the performance error function (Vapnik 1995):

$$\frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \tag{2}$$

this is subject to:

$$\begin{aligned} W^T \cdot \Phi(X_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - W^T \cdot \Phi(X_i) - b &\leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, \dots, N \end{aligned} \tag{3}$$

The degree of the penalized loss when a training error is detected is determined by the parameter C (positive constant). Φ is the kernel function, N is the sample size, and ξ_i and ξ_i^* are slack variables specifying the upper and lower training error subject to an error tolerance ε . In the regression problem, most data samples are expected to be within the ε -tube. If a data sample is not within the tube, then, an error ξ_i and ξ_i^* will exist. Subsequently, the coefficients ω and b are determined by minimizing $r(C)$: the regularized risk function (Raghavendra and Deka 2014):

$$r(C) = C \frac{1}{N} \sum_{i=1}^N L_\varepsilon(f(x_i), y_i) + \frac{1}{2} \|\omega\|^2 \tag{4}$$

The ε -insensitive loss function: $L_\varepsilon(f(x_i), y_i)$ is defined (Raghavendra and Deka 2014):

$$L_\varepsilon(f(x_i), y_i) = \begin{cases} |f(x_i) - y| - \varepsilon & \text{if } |f(x_i) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

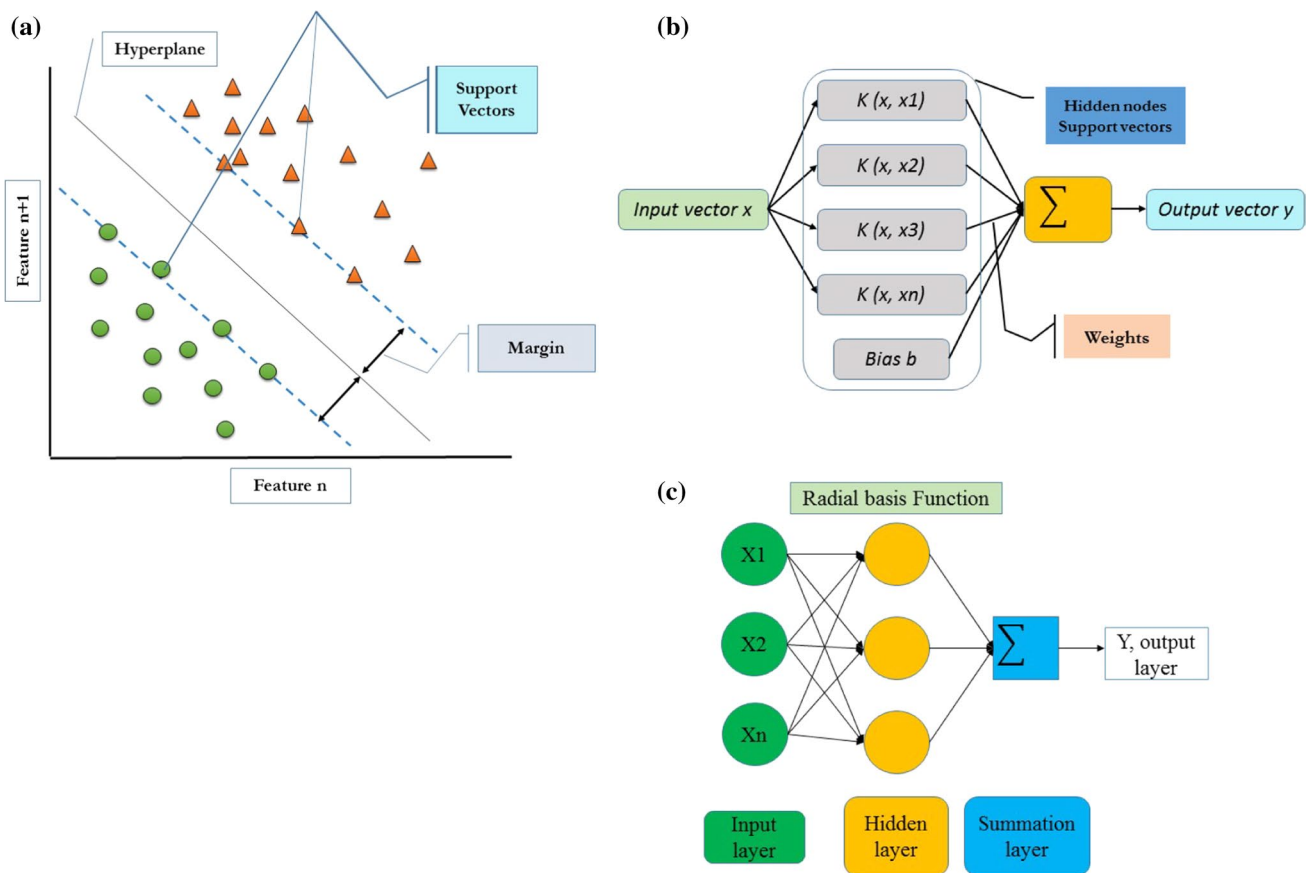


Fig. 2 **a** The schematic concept of the SVM model (Ghorban et al. 2016), **b** network architecture of the SVR model (Sujay and Deka 2015), **c** schematic diagram of generalized regression neuron network (GRNN) model

$C > 0$ is the penalty parameter and $\frac{1}{2} \|\omega\|^2$ represents the regularization term. The loss function $L_\epsilon(f(x_i), y_i) = 0$ if the difference between the predicted $f(x_i)$ and the measured value $y_i < \epsilon$. A nonlinear regression function (Eq. 6) is given by a function that minimizes Eq. (4):

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{6}$$

α_i and α_i^* are the introduced Lagrange multipliers, and $k(x_i, x)$ refers to kernel function. The kernel function given by Eq. (7) describes the inner product in the D-dimension feature space.

$$k(x_i, x) = \sum_{j=1}^D \varphi_j(x_i) \varphi_j(x) \tag{7}$$

Based on the hydrological literature, radial basis function (RBF) has been used broadly in optimizing the kernel function of SVR model (Rubio et al. 2011). For the purpose of achieving good results, RBF is used in this research. The

three parameters (penalty parameter C, error exceeding ϵ , and kernel function's parameter γ) of the RBF equation were determined by a suitable grid search algorithm using the Matlab software and with harmony to the latest research conducted by (Yaseen et al. 2016).

Generalized regression neural network (GRNN)

The long-standing ANN model, of which the GRNN model is a special case, is a nonlinear modeling technique which is suitable for modeling over a range of variables (Babu and Reddy 2014). ANNs are able to identify a complex nonlinear relationship between the predictor (inputs) and output datasets. Their basic functioning units are composed of neurons. Each neuron receives and processes the input data before transforming it into output forms. There are two possibilities of input data: (1) pure collected data or (2) input results from other neurons, while the output data forms may be either the results of the final process or the input data of other neurons (Kim and Kim 2008).

In this research, we applied the GRNN developed initially by Specht (1991). GRNN is a variation of the radial basis

neural network based on kernel regression network, requiring no iterative training procedure as with the case of back propagation ANN (Hannan et al. 2010). Instead, the GRNN is capable of approximating any relation between the input and output vectors and estimates the function directly from training dataset (Kisi 2006). Figure 2c shows a schematic view of GRNN. It is comprised of four nodal layers (input, pattern, summation, and output). Each of them is connected to adjacent ones by a set of weights between nodes. Without the need to iteratively tune the model as with the case of traditional ANN models, the GRNN model architecture is characterized by its fast learning and convergence to the optimal regression surface (Kisi 2006). It is also imperative to mention that the local minima problem is not a concern in GRNN-based ANN model, as with the case of other neural network models and they do not generate ambiguous predictions (Tayyab et al. 2016). Therefore, the proposed GRNN model provides an alternative framework for fast and accurate stream-flow forecasting in this study.

Forecasting model development

In this research, different predictive modeling scenarios based on the input attributes combinations have been considered. Scenario A denotes the univariate forecasting model development that considers the antecedent values of the stream-flow (Q) only, whereas Scenario B represents the multivariate prediction model that includes the stream-flow (Q), rainfall (R), and evapotranspiration (E) datasets as the predictor variables, carrying the climatological information required to model daily stream-flow. Table 1 shows the details of the model structures.

In order to determine the number of antecedent observations that are able to provide effective inputs to the prescribed GRNN and SVR models, partial autocorrelation functions of the daily stream-flow series at Bafing Makana station are computed (Fig. 3). It is evident that at the confidence level of 95%, the lag 1, lag 2, and lag 3 are highly

Table 1 Investigated modeling scenarios and their data set combinations

Sets	Scenario A	Scenario B
Set 1	$Q_{(t-1)}$	$Q_{(t-1)}, R_{(t-1)},$ and $E_{(t-1)}$
Set 2	$Q_{(t-1)}$ and $Q_{(t-2)}$	$Q_{(t-1)}, R_{(t-1)}, E_{(t-1)}, Q_{(t-2)}, R_{(t-2)},$ and $E_{(t-2)}$
Set 3	$Q_{(t-1)}, Q_{(t-2)},$ and $Q_{(t-3)}$	$Q_{(t-1)}, R_{(t-1)}, E_{(t-1)}, Q_{(t-2)}, R_{(t-2)}, E_{(t-2)}, Q_{(t-3)}, R_{(t-3)},$ and $E_{(t-3)}$

$Q_{(t-1)}, Q_{(t-2)},$ and $Q_{(t-3)}$ denote the different lag time series of river flow at time $t - 1, t - 2,$ and $t - 3,$ respectively; $R_{(t-1)}, R_{(t-2)},$ and $R_{(t-3)}$ represent different time lagged series of rainfall at time $t - 1, t - 2,$ and $t - 3,$ respectively, and $E_{(t-1)}, E_{(t-2)},$ and $E_{(t-3)}$ represent different lag time series of evapotranspiration for $t - 1, t - 2,$ and $t - 3,$ respectively

significant in terms of their association with the stream-flow variations. Therefore, for this paper, the three antecedent days are considered for stream-flow forecasting (Table 1).

Model evaluation criteria

In theory, there exist several model evaluation criteria used to assess the forecasting accuracy; however, there is no consensus standard on the choice of one metric over the other as each metric is expected to reflect one or more characteristics of the forecasting method and the datasets used (Cheng et al. 2015). In this study, the prescribed GRNN and SVR are evaluated and compared with each other by using four common metrics: root mean square error (RMSE), mean absolute percentage error (MAPE), Willmott’s Index of agreement (WI), and the coefficient of determination (R^2) expressed as:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Q_{simi} - Q_{obsi})^2}{N}} \tag{8}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Q_{obsi} - Q_{simi}|}{|Q_{obsi}|} \tag{9}$$

$$WI = 1 - \frac{\sum_{i=1}^n (Q_{simi} - Q_{obsi})^2}{\sum_{i=1}^n (|Q_{simi} - \overline{Q_{obsi}}| + |Q_{obsi} - \overline{Q_{obsi}}|)^2} \tag{10}$$

$$R^2 = \frac{\sum_{i=1}^n [(Q_{obsi} - \overline{Q_{obsi}})(Q_{simi} - \overline{Q_{simi}})]}{\sqrt{\sum_{i=1}^n (Q_{obsi} - \overline{Q_{obsi}})^2 \sum_{i=1}^n (Q_{simi} - \overline{Q_{simi}})^2}} \tag{11}$$

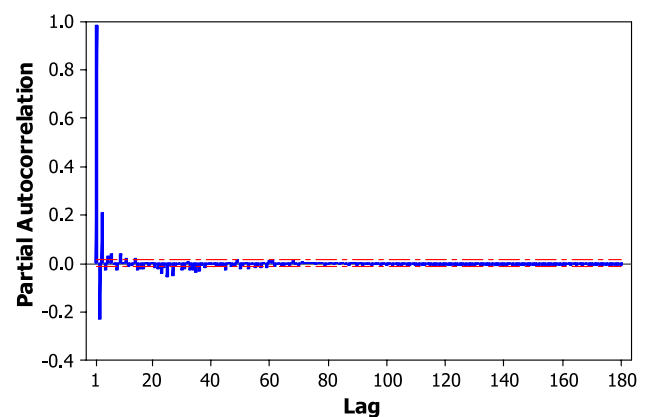


Fig. 3 Partial autocorrelation function for daily stream-flow data at Bafing Makana station (with 5% significance limits indicated in red)

In Eqs. (8–11), Q_{obsi} and Q_{simi} are the observed and the forecasted stream-flow values and n is the number of observations or the time period over which the errors are forecasted in the testing data set. The RMSE and MAPE are deduced to measure the residual error, and R^2 is applied to examine the level of statistical agreement between forecasted and observed stream-flow data in terms of the variance in the test set. In general, smaller values of the RMSE and MAPE and larger values of R^2 indicate better performances of the forecasting models. In this paper, we also employed the Willmott’s Index, WI, which provides an alternative way to assess the performances of the forecasting model, especially in terms of its ability to address the limitations faced by RMSE (Willmott 1981).

The prediction modeling scenarios results

In this section, the accuracy of the GRNN and SVR models applied for 1-day-ahead stream-flow forecasting for the case of the upper Senegal River basin at Bafing Makana station is presented. Two modeling scenarios (A and B) are considered with three sets of input data for each scenario to judge the versatility of the prescribed models using statistical performance metrics and distribution of errors. Results from both scenarios are compared in light of the model’s performance and datasets used for forecasting stream-flow values.

Scenario A

For this scenario, three different combination inputs data are used: Set 1, Set 2, and Set 3. Table 2 lists the results from the GRNN and SVR model simulation for the three input combinations. The results show that the RMSE is about 88.4, 154.25, and 142.44 m³/s (81.90, 77.87, and 79.14 m³/s) forecasted by the GRNN (SVR) for Set 1, Set 2, and Set 3. This indicates that the Set 1 and Set 2 input combinations yield the lowest value of the RMSE (88.4 and 77.87 m³/s) for the case of the GRNN and SVR, respectively. Looking at the MAPE metric, one can note that for Set 3, the error values of about MAPE = 26.67 m³/s and for Set 2, a value of MAPE = 24.98 are recorded for the best performing GRNN and SVR models. When the optimal GRNN model (Set 1) and the optimal SVR model (Set 2) are compared, it is evident that SVR performs better than GRNN model in terms of

lower RMSE and larger value of WI and R^2 . From Table 2, it is clearly shown that the SVR and the GRNN model provide different levels of performance for the different data sets. Therefore, the model accuracy depends on the input dataset but differences among the two models also exist for any common dataset that are used.

For a direct comparison of the forecasted and the observed stream-flow, Fig. 4 shows the hydrographs as well as the scatter plots of the optimal GRNN and SVR models used for 1-day-ahead flow forecasting. The SVR model is able to forecast the mean and peak flow data more accurately than the GRNN model. Notwithstanding this performance difference, both models show generally good ability to predict daily stream-flow data as the time series of the forecasted and observed stream-flow are in reasonably good agreement.

As additional measure of agreement between observed and simulated stream-flow data, the scatter plots of the optimum GRNN and SVR models are presented in Fig. 4. It should be noted that large deviations from the 45° line will indicate lesser prediction accuracy of the models. According to Fig. 4, the scatter plot of the SVR model falls close to 45° line, whereas that of the GRNN shows pronounced deviation from the 45° line. This result indicates that the SVR model accurately forecasts the stream-flow data better than the GRNN model. Also, it is noticeable that, according to the coefficient of determination and the best fit line, the SVR model is able to forecast better than the GRNN model. Considering all sets in this study, the SVR is able to generate better forecasting results than the GRNN model.

Further evaluation of the models is undertaken by investigating the observed and forecasted stream-flow from the optimal GRNN and SVR models using the five descriptive statistical metrics. Table 3 shows the values of the minimum stream-flow, median stream-flow, first and third quartile flows (25th percentile, Q1; 75th percentile, Q3, respectively), and maximum stream-flow for both models.

GRNN and SVR models are seen to overestimate the minimum stream-flow by about 12.73 and 5.87 times, respectively (Table 3), which indicates that neither of the two data-driven models are able to simulate the lowest values of stream-flow with very good degree of accuracy. This is not an unexpected result since the extremely low input features from the historical stream-flow and the climate inputs are usually more intermittent (i.e., generally rare) compared

Table 2 Performance criteria values of Scenario A

Input sets	GRNN model				SVR model			
	RMSE (m ³ /s)	MAPE (m ³ /s)	WI	R ²	RMSE (m ³ /s)	MAPE (m ³ /s)	WI	R ²
Set 1	88.44	35.05	0.97	0.949	81.90	25.73	0.98	0.951
Set 2	154.25	28.02	0.94	0.861	77.87	24.98	0.98	0.956
Set 3	142.44	26.67	0.95	0.87	79.14	26.33	0.98	0.956

Fig. 4 Scatter plots and hydrographs of 1-day ahead stream-flow forecasting for the optimal GRNN and SVM models for Scenario A, cms cubic meter second

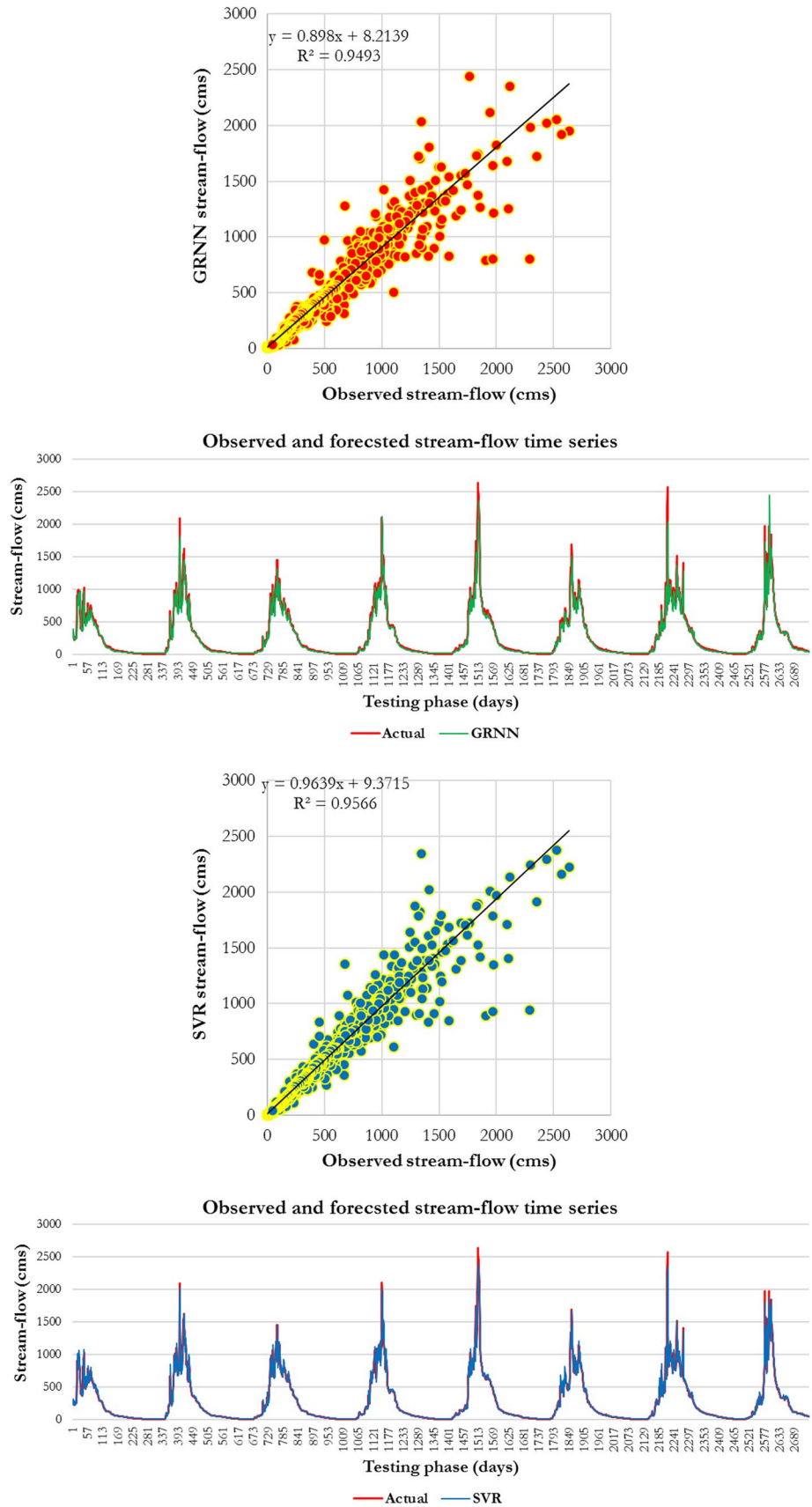


Table 3 Significant descriptive statistics for Scenario A for the optimal GRNN and SVR model compared with observed data

Variable	Mean (m ³ /s)	Minimum (m ³ /s)	Q1 (m ³ /s)	Median (m ³ /s)	Q3 (m ³ /s)	Maximum (m ³ /s)
Observed	267.89	1.15	19.95	85.30	406.10	2634.00
GRNN	248.79	14.64	21.40	62.71	385.90	2443.36
SVR	267.59	6.75	22.05	80.40	410.95	2382.52

to the features representing the mean stream-flow data. Of course, a lack of sufficient features can lead to a low accuracy for the forecasted minimum stream-flow values. The implications of high inaccuracy of the minimum simulated stream-flow must be considered in the design of hydrological structures and water management decisions. For example, if data-driven model results are adopted in a future period where current stream-flow and rainfall values are dramatically low, such results should be subjected to “a conservative application and interpretation for core decision-making” and appropriate precautions should be taken to ensure that the decisions do not undermine their practical usage. Further benchmarking of the predicted results should be performed with additional information before such forecasts are implemented.

It is also noticeable that the SVR model exhibits a mean stream-flow very close to observed values (267.59 against 267.89) compared to the GRNN model which underestimates the mean value (248.79 against 267.89). Moreover, the evaluations of statistical parameters indicate that GRNN model underestimated the flows under the 50th and 75th percentiles more pronounced than SVR for the median. SVR slightly underestimates the third quartile. However, GRNN model predicts the maximum stream-flow closer to the observed values compared to the SVR. These results also show that both models overestimate low stream-flows and underestimate the high stream-flows. Generally, distribution of the low stream-flow values indicates that SVR has better predictions in comparison with the GRNN model.

Scenario B

For this scenario, stream-flow, rainfall, and evapotranspiration data were taken into account as predictor (input) variables of the GRNN and SVR model. Three lag times ($t - 1$, $t - 2$, and $t - 3$) are considered for all inputs, and each of lag time represented a data set input (i.e., lag ($t - 1$), ($t - 2$), and ($t - 3$)) and corresponded to Set 1, Set 2, and

Set 3 respectively (Table 1). In this scenario, data that could impact the hydrological cycle are added and an improvement in the forecasting performance is expected. Table 4 presents the models, input sets (Set 1, Set 2, and Set 2), and the prediction skills for each set.

In the Set 1, the input layer consists of three values of stream-flow, rainfall, and evapotranspiration both of them at time $t - 1$. For this Set, SVR performed better than GRNN if the RMSE (55.26 against 58.57), MAPE (20.11 against 29.20), and R^2 (0.974 against 0.971) are considered. It is evident that this superiority of SVR is confirmed to be better for both Set 2 and Set 3. For the GRNN models, on average Set 1 shows the best performance with RMSE, MAPE, WI, and R^2 of 58.57, 29.20, 0.99, and 0.971, respectively, and Set 3 shows the poorest results according to the same performance indicators (Table 4).

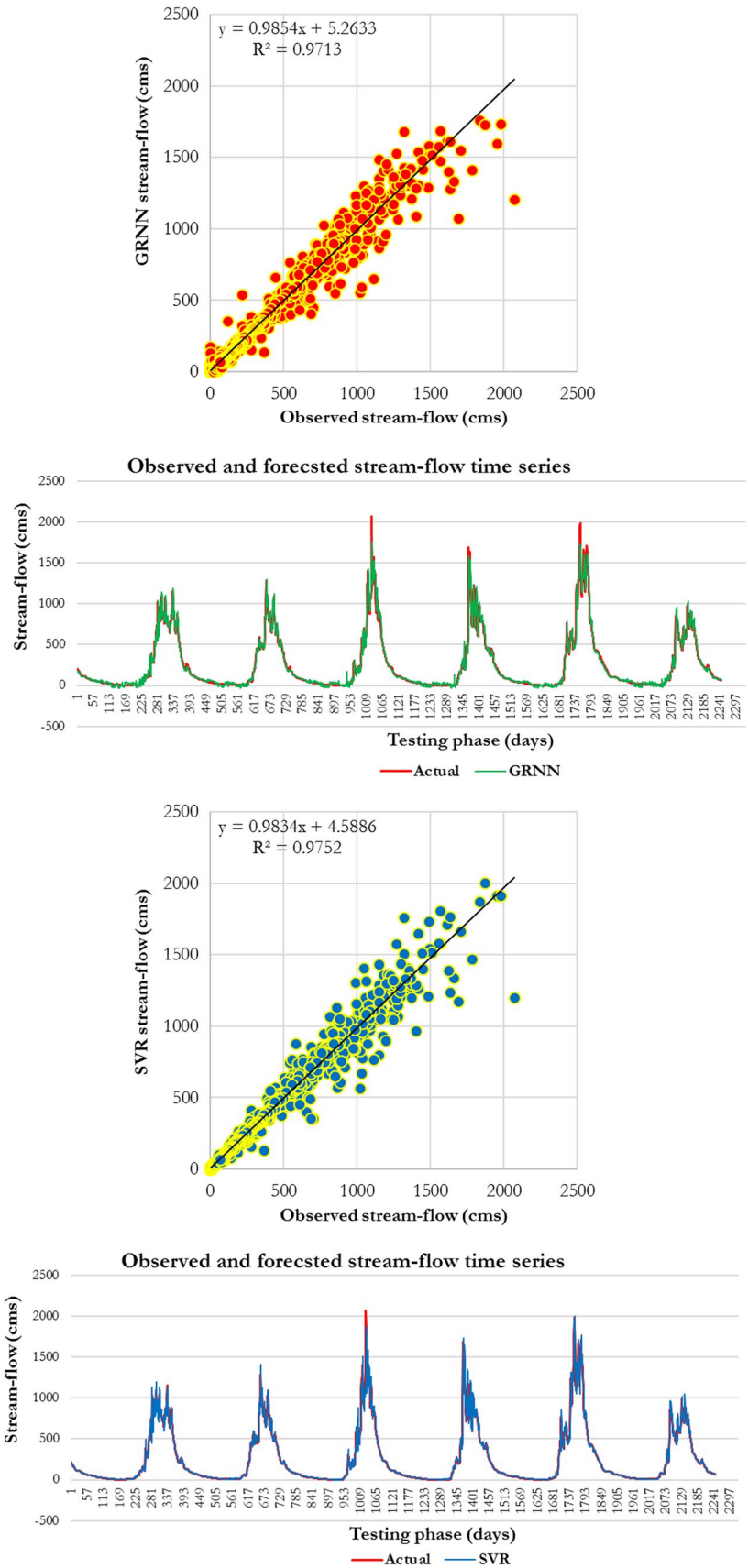
Concerning the SVR models, Set 3 performed better among the three sets with values of RMSE = 54.14, MAPE = 19.2, WI = .99, and $R^2 = 0.975$. Results from Table 4 show that the GRNN model with Set 1 can provide the best result among the GRNN with respect to RMSE, MAPE, WI, and R^2 criteria and Set 3 for SVR gave the best results with respect to the same criteria.

Comparison of the optimal GRNN (Set 1) and SVR (Set 3) indicates that SVR performs better than GRNN model. Figure 5 shows hydrographs as well as scatter plots of the best GRNN and SVR models in daily flow prediction. The SVR is capable to predict the mean and peak stream-flow data more precisely than the GRNN model. The scatter plot of the SVR model falls close to 45° line, whereas that of the GRNN model presents a pronounced deviation. Also, the coefficient of determination (R^2) and fit line equations are better for the SVR model, suggesting that the SVR model forecasts are better than the GRNN model. For both the GRNN and SVR model, the simulated hydrographs reproduce quite well the observed hydrographs. If all sets, results indicate that the SVR model performed better forecasts of stream-flow than GRNN model.

Table 4 Performance criteria values of Scenario B

Input sets	GRNN model				SVR model			
	RMSE (m ³ /s)	MAPE (m ³ /s)	WI	R^2	RMSE (m ³ /s)	MAPE (m ³ /s)	WI	R^2
Set 1	58.57	29.20	0.99	0.971	55.26	20.11	0.99	0.974
Set 2	71.08	20.01	0.98	0.958	54.14	20.23	0.99	0.975
Set 3	78.08	44.70	0.98	0.953	54.14	19.21	0.99	0.975

Fig. 5 Scatter plots and hydrographs of 1-day ahead stream-flow forecasting for the optimal GRNN and SVM models for Scenario B, *cms* cubic meter second



Similar to Scenario A, the observed flows and optimal GRNN and SVR flows are evaluated by determining the minimum stream-flow, maximum stream-flow, median stream-flow, first quartile flow, and third quartile stream-flow values (Table 5).

The GRNN and SVR models have difficulty in simulating the low stream-flow data and gave negative values of -31.43 and -2.35 , respectively (Table 5). The GRNN model is the poorest in terms of estimating the low values. SVR and GRNN models exhibit mean flow values that are very close to observed values, but both models slightly overestimate the mean values by less than 1%. Moreover, the evaluations of statistics parameters indicate that both the GRNN and SVR models underestimate the stream-flows under the 50th, 75th percentiles, and maximum flow with values of SVR closer to observed values than GRNN model (Table 5). Generally, based on the evaluation criteria and the five significant statistics values, the SVR model has better predictions in comparison with the GRNN model.

Scenarios comparisons

The performances of the forecasting results of daily river flow are compared using the GRNN and SVR models. Tables 2 and 4 gave the values of performance measures for each of the scenario and data set input layers. The two optimal models for each scenario can be compared based on the performance indicators determined in “Scenario A” and “Scenario B” sections. For both scenarios, this study shows that the SVR model appears to be the best model and the trials with Set 2 and Set 3 are the best for the Scenarios A and B, respectively. The main difference between Scenarios A and B is the number of input layers. Scenario A considered only the antecedent (lagged) stream-flow data as an input, whereas Scenario B took into account the related rainfall and evapotranspiration data together with stream-flow. Results showed that the fact that model integrated the other input variables that are likely to affect the hydrological cycle (i.e., rainfall and evapotranspiration) and the accuracy of the models could be improved substantially. From these results, it can be seen that the best input combination is achieved when the SVR model incorporates antecedent stream-flow, rainfall, and evapotranspiration as the predictor variables.

Results analysis summary

Based on the obtained results, in general, GRNN and SVR models are valuable predictive tools for forecasting short-term stream-flow in the context of data scarce regions. Comparisons among the two models showed that the SVR model was the best model for all scenarios and the data set input used in this study. The optimal SVR model with stream-flow, rainfall, and evapotranspiration integrated information from lags $t - 3$, $t - 2$, and $t - 1$ (Scenario B and Set 3). Differences in the accuracy of forecasts were also associated with the different scenarios that were tested. The results indicated that the accuracy of the forecasts increased with an increase in the inputs combinations. For all scenarios and the data input set, the SVR model outperformed the GRNN model. However, the incorporation of rainfall and evapotranspiration data as predictor variables led to improve the accuracy which is explained by the tightly linked hydrological cycle to rainfall and evapotranspiration changes in the study region.

In addition, in the light of the results achieved all the way through the proposed two modeling methods and also the input pattern scenarios, two major interpretations could be distinguished. Firstly, the necessity of external climatological input variables are needed for carrying out a reliable and accurate forecasting model for stream-flow. Apparently, the more hydrological input variables influencing on the stream-flow are counted in the model, the higher forecasting accuracy could be achieved. Even for those hydrological variables which have insignificant influence on the stream-flow might have indirect impact on the model accuracy. Furthermore, with respect to feature of the study area (the location of the study area with respect to the global environmental and climatological zone), particular hydrological parameters could have significant impact on the stream-flow forecasting, while these variables might have trivial impact in other study area. Therefore, a special attention has been given for the selection of input variables while developing stream-flow forecasting model reflecting not only the hydrological features and characteristics but also the climatological variables.

The second interpretation is the selection of the best modeling method. In fact, the existing evaluation metrics for the forecasting model are generally indecorous to provide enough information for evaluating the modeling method and hence introduce a solid judgment on the model performance.

Table 5 Significant statistics values for Scenario B for the optimal GRNN and SVR models compared with observed data

Variable	Mean (m ³ /s)	Minimum (m ³ /s)	Q1 (m ³ /s)	Median (m ³ /s)	Q3 (m ³ /s)	Maximum (m ³ /s)
Observed	247.53	0.45	16.39	81.15	380.78	2069.00
GRNN	249.19	- 31.43	28.66	84.08	370.12	1761.24
SVM	248.09	- 2.35	18.30	81.57	375.59	2001.34

On the other hand, the evaluation of the model should be based on the purpose of the model based on the level the user of the model (water resources planner and decision-maker). Actually, the decision-makers usually give a great attention for the extreme stream-flow pattern in order to avoid having flood and/or drought period, and hence, it is preferable to achieve high forecasting accuracy for extreme events without precaution for the medium-high/medium stream-flow/medium-low categories. Also, the water resources planners usually keen for having homogenous distribution errors and then low RMSE for all categories of the stream-flow which give them the flexibility to introduce a proper plan in order to avoid the water deficit at any stream-flow categories.

In context of the present results, it is important to highlight practical significance of implementing data intelligence predictive models in the present study region using the case of upper Senegal River. One important consideration is that the electric power generated by Manantali dam that is built on the Bafing tributary is shared by the Senegal River basin riparian states on a daily basis. Hence, a prediction of the energy production at a horizon of one day ahead timescale is likely to generate valuable information for the river basin management organization (OMVS). Such information can be used by OMVS in their daily river basin management decisions since stream-flow a primary input required to estimate the power production. Therefore, developing accurate and reliable models for predicting stream-flow at short forecast horizons (e.g., daily) can help decision-makers to better manage the water resources. In addition, the established modeling can provide crucial information for energy production and management to avoid potential conflicts between the different riparian states.

Conclusion

Accurate river flow forecasts are a vital component of power production, sustainable water planning and management. Particularly, the forecasting of stream-flow is crucial for the management of hydropower production, flood and reservoir management decisions. Several data-driven techniques are currently available for hydrological forecasting purposes. This study has investigated and compared the abilities of the GRNN and the SVR model in forecasting the daily stream-flow of Bafing Makana station on the upper Senegal River basin using the input combinations of rainfall, evapotranspiration and historical stream-flow at different lagged timescales ($t - 1$, $t - 2$, $t - 3$). Four standard statistical performance evaluation measures (i.e., RMSE, MAPE, d , and R^2) have been adopted to evaluate the performances of the data-driven models. These models with different input combinations were compared with each other in their ability

to estimate daily stream-flow data. The results showed that the data-driven models with a larger number of inputs combination generally led to a better accuracy and that the SVR model outperformed the GRNN model in predicting daily stream-flow for all inputs combinations and modeling scenario. The present study showed that SVR model can be a valuable predictive tool for stream-flow prediction and it can assist the Senegal River basin organization to better manage the Senegal River water resources, especially in the upper part of the basin.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Babu CN, Reddy BE (2014) A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Appl Soft Comput* 23:27–38. <https://doi.org/10.1016/j.asoc.2014.05.028>
- Badrzadeh H, Sarukkalige R, Jayawardena AW (2013) Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. *J Hydrol* 507:75–85. <https://doi.org/10.1016/j.jhydrol.2013.10.017>
- Bodian A, Dezetter A, Deme A, Diop L (2016) Hydrological evaluation of TRMM rainfall over the upper Senegal River basin. *Hydrology* 3:15. <https://doi.org/10.3390/hydrology3020015>
- Ch S, Anand N, Panigrahi BK, Mathur S (2013) Stream-flow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputing* 101:18–23. <https://doi.org/10.1016/j.neucom.2012.07.017>
- Cheng C, Niu W, Feng Z et al (2015) Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. *Water* 7:4232–4246. <https://doi.org/10.3390/w7084232>
- Deo RC, Wen X, Qi F (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 168:568–593. <https://doi.org/10.1016/j.apenergy.2016.01.130>
- Dione O (1996) Evolution Climatique Récente et Dynamique Fluviale dans les Hauts Bassins des Fleuves Sénégal et Gambie. *Science et changements planétaires/Sécheresse* 8:300–301
- Fahimi F, Yaseen ZM, El-shafie A (2016) Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-016-1735-8>
- Ghorbani MA, Zadeh HA, Isazadeh M, Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ Earth Sci* 75:1–14. <https://doi.org/10.1007/s12665-015-5096-x>
- Gong Y, Zhang Y, Lan S, Wang H (2016) A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida. *Water Resour Manag* 30:375–391

- Guimarães Santos CA, Da Silva GBL (2014) Daily stream-flow forecasting using a wavelet transform and artificial neural network hybrid models. *Hydrol Sci J* 59:312–324
- Hannan SA, Manza RR, Ramteke RJ (2010) Generalized regression neural network and radial basis function for heart disease diagnosis. *Int J Comput Appl* 7:975–8887. <https://doi.org/10.5120/1325-1799>
- Jain SK (2012) Modeling river stage–discharge–sediment rating relation using support vector regression. *Hydrol Res* 43:851–861. <https://doi.org/10.2166/nh.2011.101>
- Kagoda PA, Ndiritu JJ, Ntuli C, Mwaka C (2010) Application of radial basis function neural networks to short-term stream-flow forecasting. *Phys Chem Earth* 35:571–581. <https://doi.org/10.1016/j.pce.2010.07.021>
- Kane H, Diallo A (2005) Etude portant sur l'évaluation de l'état de l'environnement des ressources naturelles et des ressources en eau dans la partie guinéenne du bassin du fleuve Sénégal, en se servant du système d'indicateurs de l'Observatoire de l'environnement de l'OMVS. OMVS Report, Dakar, 2005 (In French)
- Kashid SS, Ghosh S, Maity R (2010) Stream-flow prediction using multi-site rainfall obtained from hydroclimatic teleconnection. *J Hydrol* 395:23–38. <https://doi.org/10.1016/j.jhydrol.2010.10.004>
- Kim S, Kim HS (2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *J Hydrol* 351:299–317. <https://doi.org/10.1016/j.jhydrol.2007.12.014>
- Kim S, Shiri J, Kisi O (2012) Pan evaporation modeling using neural computing approach for different climatic zones. *Water Resour Manag* 26:3231–3249. <https://doi.org/10.1007/s11269-012-0069-2>
- Kisi Ö (2006) Generalized regression neural networks for evapotranspiration modelling. *Hydrol Sci J* 51:1092–1105
- Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly stream-flow forecasting. *J Hydrol* 399:132–140. <https://doi.org/10.1016/j.jhydrol.2010.12.041>
- Kumar PS, Praveen TV, Prasad MA (2016) Artificial neural network model for rainfall-runoff: a case study. *Int J Hybrid Inf Technol* 9:263–272
- Kuo CC, Gan TY, Yu PS (2010) Seasonal stream-flow prediction by a combined climate-hydrologic system for river basins of Taiwan. *J Hydrol* 387:292–303. <https://doi.org/10.1016/j.jhydrol.2010.04.020>
- Liu M, Lu J (2014) Support vector machine-an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-014-3046-x>
- Makwana JJ, Tiwari MK (2014) Intermittent stream-flow forecasting and extreme event modelling using wavelet based artificial neural networks. *Water Resour Manag* 28:4857–4873. <https://doi.org/10.1007/s11269-014-0781-1>
- Ndiaye O (2010) The predictability of the sahelian climate: seasonal sahel rainfall and onset over Senegal. Columbia University, Columbia
- Ni Q, Wang L, Ye R et al (2010) Evolutionary modeling for stream-flow forecasting with minimal datasets: a case study in the west Malian river, China. *Environ Eng Sci* 27:377–385. <https://doi.org/10.1089/ees.2009.0082>
- Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377. <https://doi.org/10.1016/j.jhydrol.2014.03.057>
- Prairie JR, Rajagopalan B, Fulp TJ, Zagona EA (2006) Modified K-NN model for stochastic stream-flow simulation. *J Hydrol Eng* 11:371–378. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:4\(371\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:4(371))
- Raghavendra S, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput J* 19:372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rubio G, Pomares H, Rojas I, Herrera LJ (2011) A heuristic method for parameter selection in LS-SVM: application to time series prediction. *Int J Forecast* 27:725–739. <https://doi.org/10.1016/j.ijforecast.2010.02.007>
- Shiri J, Kisi O (2010) Short-term and long-term stream-flow forecasting using a wavelet and neuro-fuzzy conjunction model. *J Hydrol* 394:486–493. <https://doi.org/10.1016/j.jhydrol.2010.10.008>
- Specht DF (1991) A general regression neural network. *IEEE Trans Neural Netw* 2:568–576. <https://doi.org/10.1109/72.97934>
- Sujay RN, Deka PC (2015) Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. *Cogent Eng* 2(1). <https://doi.org/10.1080/23311916.2014.999414>
- Taormina R, Chau KW (2015) ANN-based interval forecasting of stream-flow discharges using the LUBE method and MOFIPS. *Eng Appl Artif Intell* 45:429–440. <https://doi.org/10.1016/j.engappai.2015.07.019>
- Tayyab M, Zhou J, Zeng X, Adnan R (2016) Discharge forecasting by applying artificial neural networks at the Jinsha river basin, China. *Eur Sci J* 12:1857–7881. <https://doi.org/10.19044/esj.2016.v12n9p108>
- Toth E, Brath A (2007) Multistep ahead stream-flow forecasting: role of calibration data in conceptual and neural network modeling. *Water Resour Res* 43:1–11. <https://doi.org/10.1029/2006WR005383>
- Vapnik V (1995) The nature of statistical learning theory. Springer, New York
- Varis O, Lahtela V (2002) Integrated water resources management along the Senegal River: introducing an analytical framework. *Int J Water Resour Dev* 18:501–521. <https://doi.org/10.1080/0790062022000017374>
- Wang W, Van Gelder PH, Vrijling JK, Ma J (2006) Forecasting daily stream-flow using hybrid ANN models. *J Hydrol* 324:383–399. <https://doi.org/10.1016/j.jhydrol.2005.09.032>
- Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J Hydrol* 374:294–306. <https://doi.org/10.1016/j.jhydrol.2009.06.019>
- Wen X, Si J, He Z et al (2015) Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resour Manag* 29:3195–3209. <https://doi.org/10.1007/s11269-015-0990-2>
- Willmott CJ (1981) On the validation of models. *Phys Geogr* 2:184–194
- Yaseen ZM, El-Shafie A, Afan HA et al (2015a) RBFNN versus FFNN for daily river flow forecasting at Johor river, Malaysia. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-015-1952-6>
- Yaseen ZM, El-shafie A, Jaafar O et al (2015b) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J Hydrol* 530:829–844. <https://doi.org/10.1016/j.jhydrol.2015.10.038>
- Yaseen ZM, Jaafar O, Deo RC et al (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2016.09.035>