RESEARCH

Application of deep learning approaches to predict monthly stream flows

H. Yildirim Dalkilic · Deepak Kumar ·
Pijush Samui · Barnali Dixon · S. Nur Yesilyurt ·
O. Mert Katipoğlu

Abstract Accurate and reliable flow estimations are of great importance for hydroelectric power generation, flood and drought risk management, and the effective use of water resources. This research carries out a comprehensive study on the application of gated recurrent unit (GRU) neural network, recurrent neural network (RNN), and long short-term memory (LSTM) to predict with river flows at three different streamflow observation stations in Erzincan, Bayburt, and Gümüshane. Monthly streamflow time series covering the years 1978 to 2015 were used to set up artificial intelligence models. During the modeling phase, 70% of the data was divided into training (October 1978–April 2004), 15% validation (May 2004–September 2009), and 15% test set (October 2010–September 2015). Model performances were made according to the correlation coefficient, root mean square error, the ratio of RMSE to the standard deviation, Nash–Sutcliffe efficiency coefficient, index of agreement, and volumetric efficiency values. The calculation results show that GRU leads efficient estimation results for estimating streamflow and can also be used in allied water resources.

Keywords Streamflow forecasting · Gated recurrent unit · Long short-term memory · Recurrent neural network

H. Y. Dalkilic (✉) · S. N. Yesilyurt · O. M. Katipoğlu
Department of Civil Engineering, Erzincan Binali Yıldırım University, Erzincan, Turkey
e-mail: hydalkilic@erzincan.edu.tr

S. N. Yesilyurt
e-mail: sefa.yesilyurt@erzincan.edu.tr

O. M. Katipoğlu
e-mail: okatipoglu@erzincan.edu.tr

D. Kumar
Department of Agriculture, Government of Bihar, Bihar, India
e-mail: decage007@gmail.com

P. Samui
Department of Civil Engineering, National Institute of Technology Patna, Patna, India
e-mail: pijushsamui@gmail.com

B. Dixon
Professor of GIS and Remote Sensing, University of South Florida, 140 Seventh Ave South, St. Petersburg, FL 33701, USA
e-mail: bdixon@mail.usf.edu

Introduction

Effective estimation of streamflow data, one of the basic parameters of the hydrological cycle, is one of the basic steps of effective management of water resources and disaster reduction, early warning, and management (Sharma and Machiwal, 2021). Daily and hourly flow forecasts are of great importance for flood management systems, while monthly and annual flow forecasts are valuable for reservoir operation, irrigation system management, and hydroelectric generation

(Yaseen et al., 2015; Wegayehu and Muluneh, 2022). However, flow data depend on many meteorological parameters such as precipitation, temperature, evaporation, ground moisture, and infiltration, making estimating flow data difficult. Therefore, artificial intelligence technologies that can easily model nonlinear relationships for flow prediction have been in vogue recently (Zhang et al., 2018a, 2018b).

In recent years, many studies have been carried out in many different disciplines about data modeling through artificial neural networks. Many artificial neural networks are used mainly in modeling the hydrological and hydrometeorological data. Quite a few artificial neural networks (ANN) types such as feed forward neural networks (FFNN), generalized regression neural networks (GRNN), radial basis neural networks (RBNN), and Levenberg–Marquardt (LM) algorithms have been used in modeling many hydrometeorological data especially like streamflow, precipitation, and evaporation. On the other hand, the modeling technique through recurrent artificial neural networks has been applied frequently in recent years in polyphonic music, sound and video models, in modeling the signals of speaking, in language models, and in modeling the sequential time series types, and lately this technique has become recommended thanks to the successful results widely that it has led (Sattari et al., 2012; Bahdanau et al., 2014; Cho et al., 2014; Chung et al., 2014; Shoaib et al., 2016; Xiao et al., 2017).

Elman demonstrated the recurrent neural network technique for the first time (1990). Elman applied this technique to divide the sentence structures into noun and verb categories and achieved successful results. Connor et al. (1994) applied this technique to modeling nonlinear load time series and achieved successful results by integrating the robust learning algorithm into recurrent neural networks. Coulibaly et al. (2000) studied the effect of climatic trends on forecasting the annual flow values using the RNN approach. They used wavelet transform when deciding the climatic patterns. They showed that this modeling technique could successfully model the climatic trend effect. Kumar et al. (2004) applied two different networks, feed forward, and recurrent, and demonstrated the use of ANNs to forecast monthly river flows. Coulibaly and Baldwin (2005) used the dynamic RNN technique in forecasting the non-stationary hydrological time series. They compared the results of the RNN-based model with the

multivariate adaptive regression splines (MARS) model and found that the RNN-based model performed better than MARS. Cheng et al. (2008) proposed a three-stage indirect multi-step-ahead prediction model for long-term hydrologic forecasting. Banerjee et al. (2011) evaluate the prospect of (ANN) simulation over mathematical modeling in estimating safe pumping rates to maintain groundwater salinity in island aquifers. Chandra and Zhang (2012) suggested the ANN technique and the use of an alternative approach, real-time recurrent learning (RTRL). They produced various synthetic time series and operated auto-regressive (AR) and moving average (MA) models. Then, they compared RTRL and other models and found that the RTRL model proved more successful results. Sattari et al. (2012) evaluated the performance of the time lag recurrent neural networks (TLRN) model to predict the daily inflow into the Elevian reservoir. Prasad and Prasad (2014) studied the ability of deep networks to extract high-level features and recurrent networks to perform time series inference. Shoaib et al. (2016) explored the potential of wavelets for the first time and they modeled river flows by using coupled time-lagged recurrent neural network (TLRNN). Chang et al. (2018) proposed a deep learning-based model named memory time series networks which is used for time series modeling and prediction. Che et al. (2018) suggested the GRU model, which is a new deep learning approach to model the missing patterns much better. They applied this model to clinic data sets and found that it gave more successful results, especially in modelling the missing patterns. Alizadeh et al. (2021) GRU, LSTM, and SAINA-LSTM methods were compared in four different basins in the USA. As a result of this study, it was seen that SAINA-LSTM gave promising results for the region. In addition, it was stated that LSTM and GRU models performed better than RNN. Hu et al. (2018) used ANN and LSTM network models to model the precipitation-runoff relationship in the Fen River basin. Zhang et al., (2018a, 2018b) aimed to predict and simulate the water level in combined sewer overflow structures using four different neural network models such as MLP, WNN, LSTM, and GRU. Zhao et al., (2021a, 2021b) combined the gray wolf optimizer (IGWO) and GRU method to estimate flow data. The model evaluated the success of the LSSVM

and ELM methods with the model results obtained. Wegayehu and Muluneh (2022) employed stacked-LSTM (S-LSTM), bidirectional LSTM (Bi-LSTM), and GRU with the classical multilayer perceptron (MLP) network for the prediction of daily stream-flow in Awash river basin. MLP and GRU models showed better prediction results than other models.

The present study aims to test and model the performance of recurrent neural network algorithms in terms of high variability. For this purpose, three flow observation stations with a high coefficient of variation of the algorithm were used to avoid the disappearing gradient problem. Model performances were evaluated with various statistical parameters and graphical methods.

Material and method

Study area and data

In the present study, the data belonging to the water years between 1978 and 2015 were obtained from three current observation stations, numbered E23A004, E14A022, and E21A019, located in Erzincan, Bayburt, and Gümüshane provinces were used. Data statistics and station details are given in Table 1. These three selected gauging stations have approximately the same climatic conditions. And, the schematic view of the locations of these stations are shown in Fig. 1.

Recurrent neural networks (RNN)

To understand the RNN, it is advisable to remember the artificial neural networks that work feed forward. The operating logic of these two techniques is similar. In other words, it can be said that they are two structures that produce outputs by applying a set of mathematical operations to the information that comes to the neurons in the networks (Coulibaly & Baldwin, 2005; Coulibaly et al., 2000; Kumar et al., 2004).

The information in the feed forward network is processed forward only and cannot be returned back to any point. In this structure, input data is simply passed through the network and output data is obtained. Feed forward neural network structure is shown below in Fig. 2.

Also, as shown in Fig. 3, besides the input, the content units that refer to the previous output also affect the network in a RNN structure. For example, the decision for the information at (t-1) also affects the decision to be made at t. In a word, the inputs in such networks produce outputs by combining the existing and the previous information (Chandra & Zhang, 2012; Connor et al., 1994; Coulibaly et al., 2000; Donate & Cortez, 2014; Elman, 1990; Prasad & Prasad, 2014).

The main aim of recurrent neural networks is to use sequential information. The main reason for naming it "recurrent" is that the output always depends on the previous calculation steps. In other words, the RNNs store and make use of information about the steps that are so far calculated. Therefore, they work like a memory (Chang et al., 2002; Cheng et al., 2008; Bahdanau et al., 2014; Smith & Yin 2014; Che et al., 2018).

Gated recurrent neural networks

Long short-term memory unit

LSTM is a RNN structure that remembers the values at random intervals. This specific RNN type can learn long-term dependencies and is widely used and applied to various problems of many different disciplines. LSTM, considering the unknown size and time delays between important events, is a very convenient method to sort, process, and foresee the time series. Moreover, LSTM's relative insensitivity to gap length provides a significant advantage when compared to alternative RNNs, concealed Markov models, and many other learning methods.

Table 1 Station details and data information

Station No	Station name	Location	Time period	Mean (m³/s)	Max (m³/s)	Min (m³/s)	S_x (m³/s)	C_v
E21A019	Erzincan	39°23′37″N 39°41′02″E	1978–2015	87.16	469	14.9	82.09	0.942
E14A022	Gümüshane	39°18′42″N 40°06′45″E	1978–2015	8.01	70.7	0.08	11.22	1.402
E23A004	Bayburt	40°13′36″N 40°15′32″E	1978–2015	16.31	94.8	1.4	26.78	1.642

Fig. 1 The schematic locations of the gauging stations

The recurrent module in standard RNNs has a very simple structure, like a single tanh layer. On the other hand, a LSTM network contains LSTM units instead of the other network units. The LSTM unit remembers the long or short periods. The key to this capability is that it does not use any activation functions in its recurrent components. Therefore, the stored value is not recursively changed and the gradient does not vanish over time with backprop (Hu et al., 2018; Kim et al., 2018; Zhang et al., 2018a, 2018b).
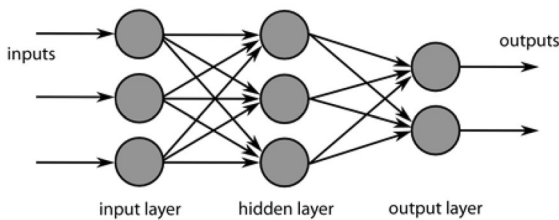

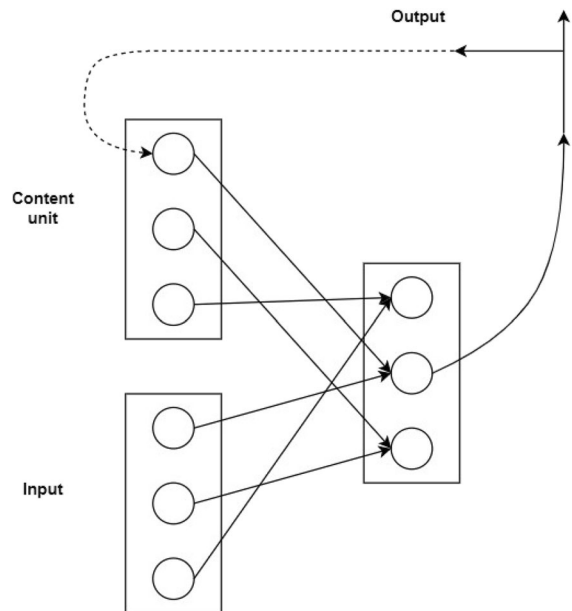**Fig. 2** Structure of feed forward neural networks

Fig. 3 Recurrent neural network cycle

Gated recurrent unit

The recurrent cells or the gated recurrent cells proposed by Cho et al. (2014) are a gate mechanism in recurrent neural networks. It is seen that the performance of these cells is similar to LSTM in many areas, and sometimes even better. GRUs have fewer parameters than LSTM because they have no exit gates. The LSTM unit inspired the GRU, but it is considered simpler to compute and implement. They also have a memory mechanism but with significantly fewer parameters than LSTM. GRU is often used when there is fewer data available and is faster to compute (Chang et al., 2018, Hu et al., 2018, Kim et al., 2018, Zhang et al., 2018a, 2018b).

Evaluation of model performance

This research follows the basic guideline to assess the goodness of fit for the developed models. The guideline uses the correlation coefficient (*r*), root mean square error (*RMSE*), ratio of *RMSE* to the standard deviation (RSR), Nash–Sutcliffe efficiency coefficient (*NSE*), index of agreement (*d*), and volumetric efficiency (*VE*) as fitness indices to evaluate the model performance. All these fitness indices are calculated using the Eqs. (1–6).

$$r = \left(\frac{\sum_{i=1}^{l} \left(Q_{E_i} - Q_{\overline{E}_i}\right)\left(Q_{O_i} - Q_{\overline{O}_i}\right)}{\sqrt{\sum_{i=1}^{l} \left(Q_{E_i} - Q_{\overline{E}_i}\right)^2 \sum_{i=1}^{n} \left(Q_{O_i} - Q_{\overline{O}i}\right)^2}} \right) \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{l} \left(Q_{E_i} - Q_{O_i}\right)^2}{l}} \quad (2)$$

$$RSR = \frac{RMSE}{STDEV^{obs}} = \frac{\sqrt{\sum_{i=1}^{l} \left(Q_{O_i} - Q_{E_i}\right)^2}}{\sqrt{\sum_{i=1}^{l} \left(Q_{O_i} - \overline{Q}_{E_i}\right)^2}} \quad (3)$$

$$NSE = \left(1 - \frac{\sum_{i=1}^{l} \left(Q_{Ei} - Q_{O_i}\right)^2}{\sum_{i=1}^{l} \left(Q_{O_i} - Q_{\overline{O}_i}\right)^2} \right) \quad (4)$$

$$d = 1 - \left[\frac{\sum_{i=1}^{l} \left(Q_{E_i} - Q_{O_i}\right)^2}{\sum_{i=1}^{l} \left(\left|Q_{E_i}\right| - \left|Q_{O_i}\right|\right)^2} \right], 0 \le d \le 1 \quad (5)$$

$$VE = 1 - \left(sum\left(Q_{O_i} - \overline{Q}_{E_i}\right) / sum\left(Q_{O_i}\right) \right) \quad (6)$$

where Q_{E_i} is the ith estimated monthly streamflow discharge using models; Q_{O_i} is the ith observed monthly streamflow discharge; $Q_{\overline{E}_i}$ is the average of the estimated monthly streamflow discharge; $Q_{\overline{O}_i}$ is the average of the observed monthly streamflow discharge; and *l* is the number of observations.

Rank analysis Determining the best model is a complex task in models in which many statistical indicators are used together. For this reason, the rank values of the statistical indicators used in this study were determined separately, and then, the most effective model was determined according to the total rank values. A rank is assigned to each performance parameter when performing rank analysis. Rankings were arranged from the maximum value equal to the number of models, which was three in our study, to the minimum value equal to one. Here, the best-performing model is assigned the third rank, and the lowest-performing model is appointed the first rank. The model with the highest total rank shows the best, while the model with the lowest shows the worst (Zhang et al., 2020).

Results and discussion

This work aims to develop a forecasting model to predict streamflow using the GRU model. In addition, two similar configured sequential type models (RNN and LSTM) were also tested to find its applicability to forecasting the streamflow. Furthermore, to check its robustness, three different stations, namely Erzincan, Gümüşhane, and Bayburt were considered.

The dataset was split into three parts; training (from Oct 1978 to April 2004), validation (from May 2004 to Sep 2009), and testing set (from Oct 2010 to Sep 20,015), i.e., data split is 0.7/0.15/0.15. The training data is used to develop the models, and validation data is used to tune and to select the best-performing models. At the same time, the test data is used to evaluate the performance of the developed models. As stated above, obtaining final model configuration or hyper-parameter tuning is big topic and requires skill. Therefore, multiple scenarios with different time horizons of trail and the final architecture for each station is selected based on the highest correlation coefficient (r).

The algorithm for sequential deep learning models (RNN, LSTM, and GRU) were developed under TensorFlow using "Keras deep Learning library." The major problem in time series analysis is selection of random components. Therefore, this study considered monthly streamflow discharge as a random component to developing the models. The appropriate architecture of sequential models consists of model input (i.e., previous time steps), number of the memory cell (memory block), and model output. This research trails different combination of inputs and memory cells for the three stations (Erzincan, Gümüshane, Bayburt). Initially, the previous time steps (i.e., look back) are varied between 1 and 20 and the final architecture of models are selected based on the highest correlation coefficient (r). The input selection (i.e., look back) is one of the unwieldy and important tasks during model development. Based on the critical appraisal, different researchers considered different looks back for the model development. Ouyang and Lu (2018) have considered 12-month previous time steps for the development of ANN model and multi-gene genetic programming and support vector machine. Qin et al. (2019) have also tested the LSTM model for hydrological time series analysis by adopting different batch sizes and the number of the

memory cell. Furthermore, Kumar et al. (2019) have tested different time steps to check their effect on the performance of the LSTM model. The final model has complied by adopting the loss function, i.e., mean square error (MSE) and optimizer *ReLu*. The final model configuration was discovered by trial and error based on skill. This leaves the door open to explore new and possibly better configurations. Table 2 shows the final configuration of the models at the three selected stations.

Model performance for Erzincan station (E21A019)

Table 3 shows the performance indices of models at the different stations. The results show that during the training phase, the RNN model outperformed, followed by LSTM and GRU. The correlation coefficient (r) for RNN and LSTM model is approximate, showing that the models were well trained during the training phase, whereas, for GRU, the performance of slightly less as compared to the RNN and LSTM (Figs. 4 and 5). While the validation period, the correlation coefficient (r) was found to be approximate similar for all the models (RNN ($r = 0.906$), LSTM ($r = 0.904$), and GRU ($r = 0.904$)). In addition, other fitness parameter was also calculated. In general, the lowest the error better the model. The lowest RMSE was found for the GRU ($RMSE = 0.121$) followed by the RNN ($RMSE = 0.125$) and LSTM ($RMSE = 0.127$). The other fitness index confirmed that the GRU model is well developed and can give significant results to forecast the streamflow at the Erzincan station. Furthermore, all three models were also tested on remained 15% test dataset to check the robustness of the models. From the analysis of results, it was found that GRU model well capable and performance is similar to RNN and LSTM in terms of fitness indices. The result signifies that the GRU model can be used as a replacement of

Table 2 The final configuration of the models at different stations

	Erzincan			Gümüshane			Bayburt		
	RNN	LSTM	GRU	RNN	LSTM	GRU	RNN	LSTM	GRU
Memory cell	40	20	12	40	10	9	40	10	5
Look back	20	20	20	10	10	10	12	12	12
Batch size	10	10	10	10	10	10	10	10	10
Epoch	200	200	200	200	200	200	200	200	200

Fig. 4 Scatter plot displaying the models (RNN, LSTM, and GRU) performance at stations Erzincan during the training phase
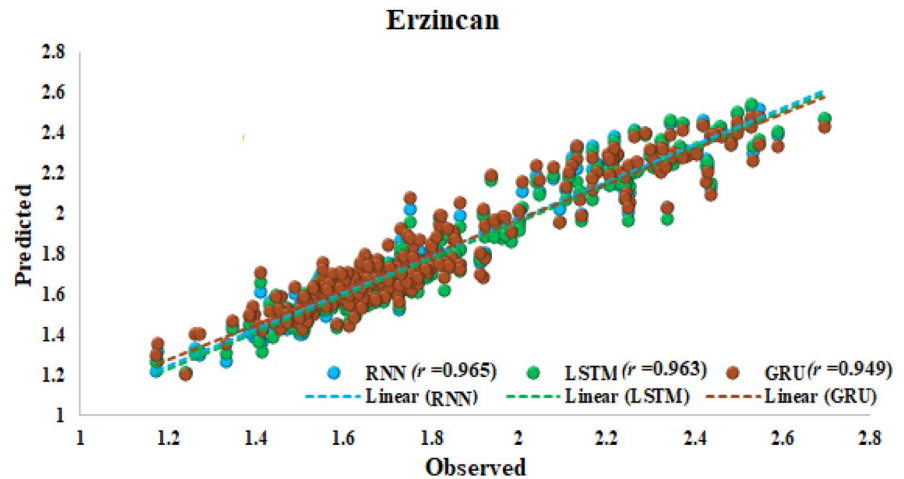
LSTM to predict the streamflow for this station. For better representation, a graphical Taylor diagram has been drawn to visualize the performance of the models in terms of correlation coefficient (r), standard deviation and root mean square error (blue line). Figure 6a shows that the GRU model is clearly better than the RNN and LSTM as it gives smaller standard deviation without changing much in the correlation coefficient. In addition, box wisher plot (Fig. 6b) is also drawn to map the
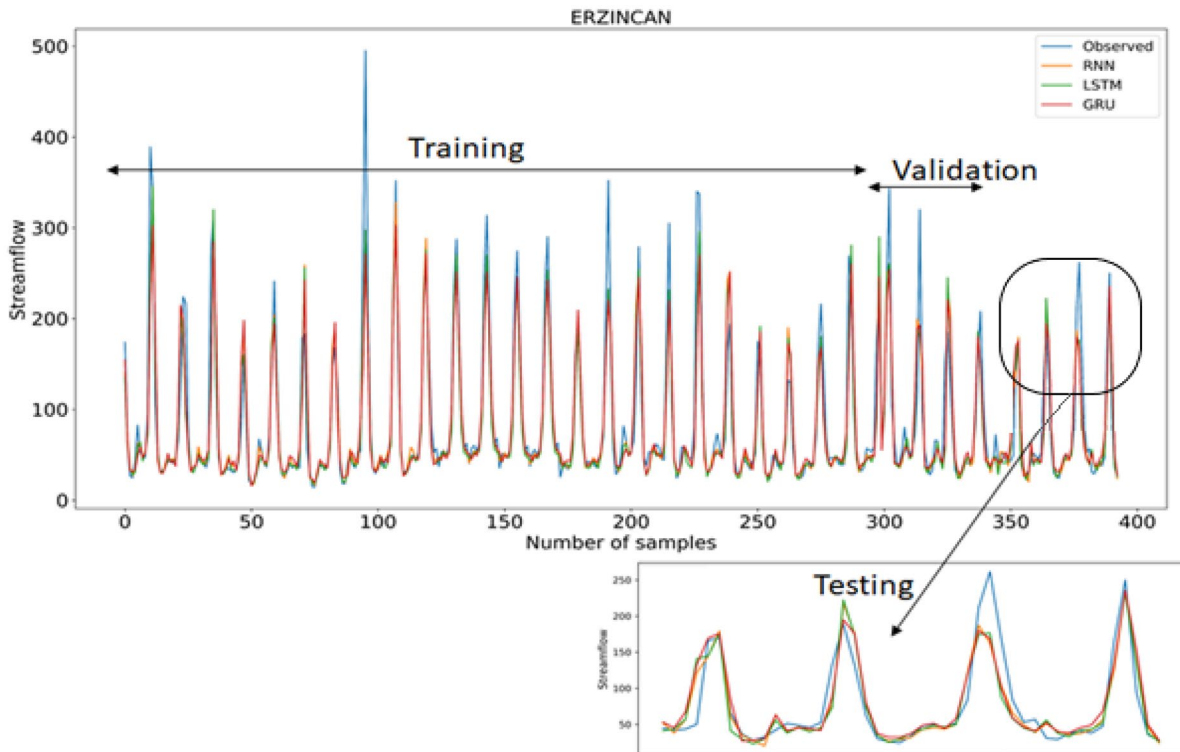
Fig. 5 Comparison of observed and forecasted streamflow using the proposed model (GRU) and LSTM and RNN models at Erzincan station
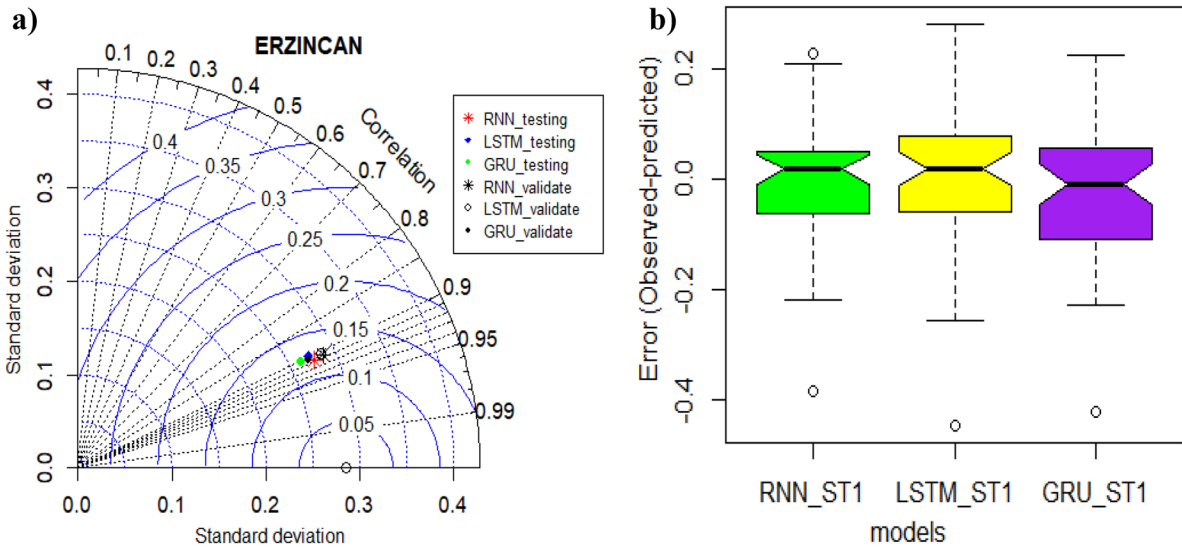
Fig. 6 a Taylor plot displaying the models (RNN, LSTM, and GRU) performance at Erzincan station during the validation and testing phase. **b** Box whisker plot displaying testing error (observed-predicted) at Erzincan station

spread of the error (observed-predicted) to visualize the range, median of the error, that occurred during the testing phase by the different models.

Again, the GRU model shows better performance as the error range is lesser than the LSTM and the median of error occurred towards the center of the box. Whereas in the case of RNN, the median of the error shifted to the upper interquartile range. Therefore, analysis of the results showed than the GRU model performed better than the LSTM at this station.

Model performance for Gümüshane station (E14A022)

In order to test the robustness of the GRU model, it is further tested on the Gümüshane station. During the training phase RNN model ($r=0.931$) outperformed followed by the GRU ($r=0.914$) and LSTM ($r=0.908$) (Fig. 7). But during the validation period, the performance of the RNN model is reduced significantly and showed the underfit condition and found LSTM model ($r=0.901$) is better than the RNN model ($r=0.895$) (Table 3). Whereas during the testing

Fig.7 Scatter plot displaying the models (RNN, LSTM, and GRU) performance at stations Gümüshane during the training phase
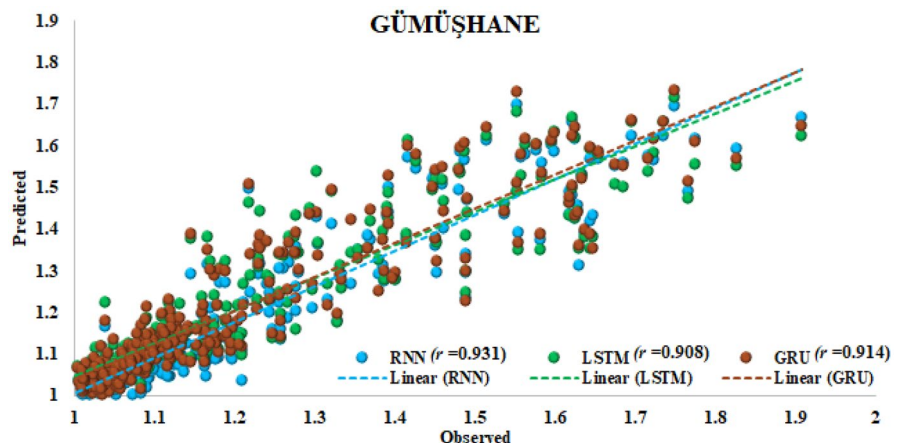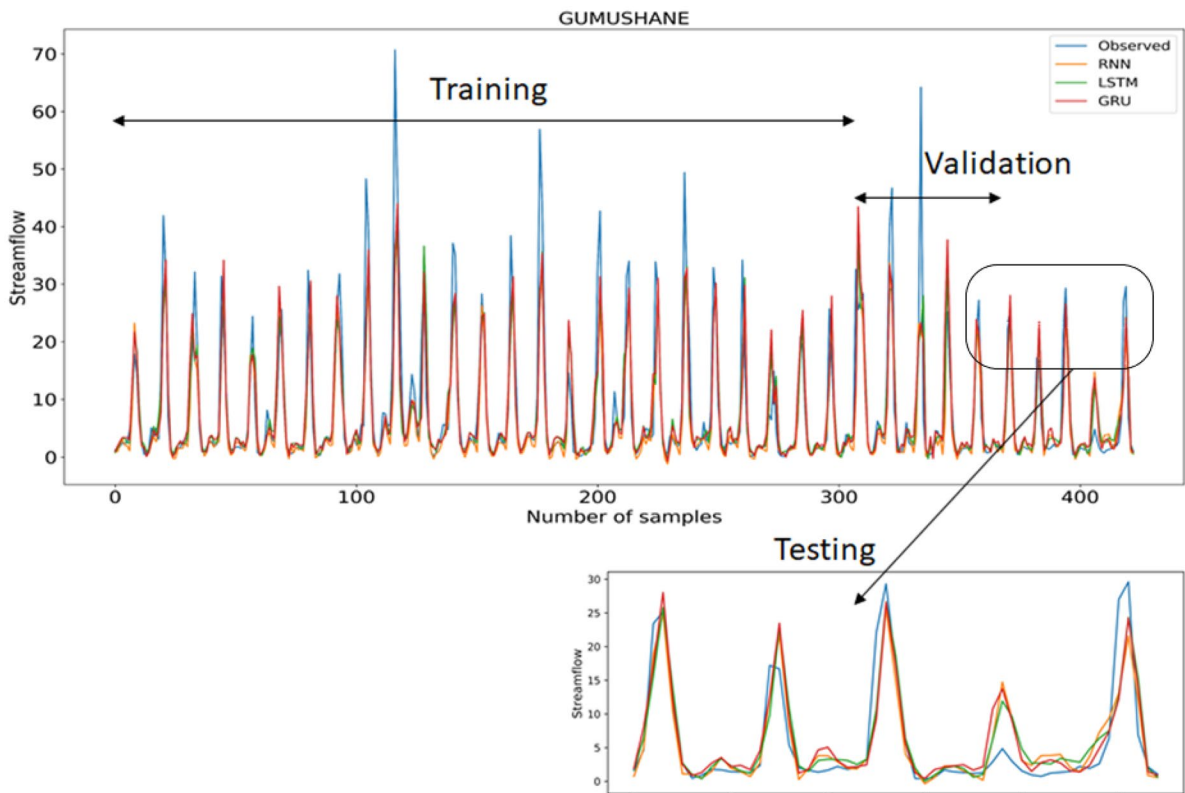
Fig. 8 Comparison of observed and forecasted streamflow using the proposed model (GRU) and LSTM and RNN models at Gümüshane station
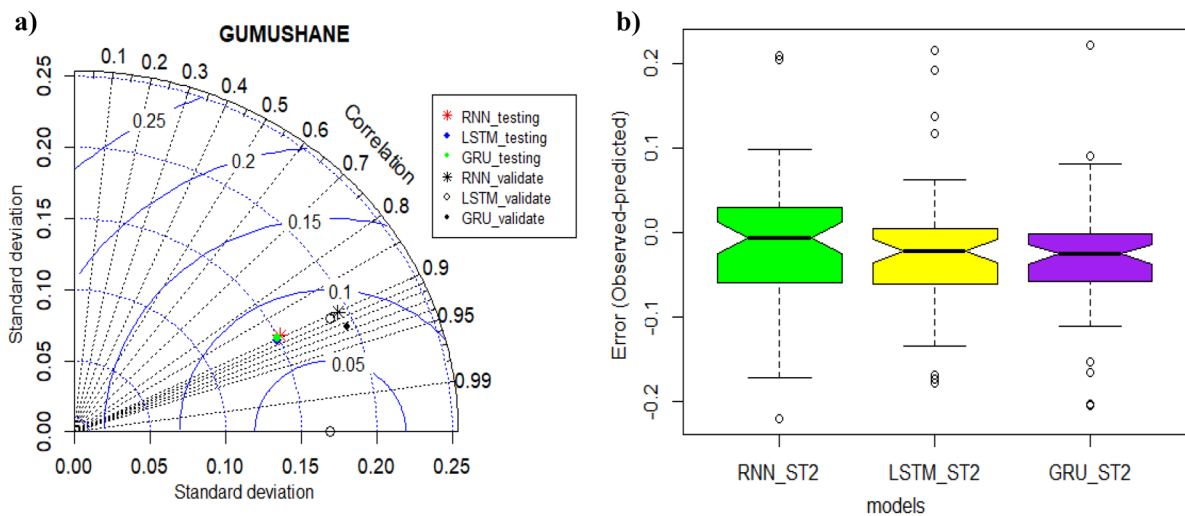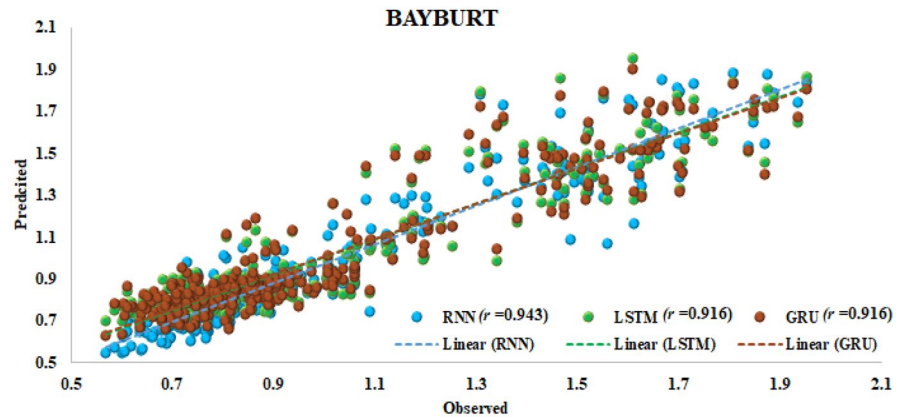

Fig. 9 **a** Taylor plot displaying the models (RNN, LSTM, and GRU) performance at Gümüshane station during the validation and testing phase. **b** Box whisker plot displaying testing error (observed-predicted) at Gümüshane station

Fig. 10 Scatter plot displaying the models (RNN, LSTM, and GRU) performance at Bayburt stations during the training phase

period, the LSTM model reached the condition of overfitting ($r=0.913$) and predict higher value than the observed, but same time GRU model showed consistent performance to forecast the streamflow (Fig. 11). In terms of other performance indices, the GRU model ($RMSE=0.073$, $RSR=0.439$) showed approximately close RMSE error than the LSTM ($RMSE=0.076$, $RSR=0.435$). The volumetric efficiency of the GRU model ($VE=0.953$) is approximately equal to the LSTM model ($VE=0.956$). The Taylor plot showed that GRU and LSTM model gives smaller standard deviation during testing in comparison to RNN, whereas accuracy of the GRU is approximately similar to LSTM (Fig. 9). In addition, the interquartile range of box wisher plot shows that the LSTM model has more outlier in the lower quantile which is opposite in the case of GRU. This signifies that the LSTM model frequently predicts less value than the observed value.
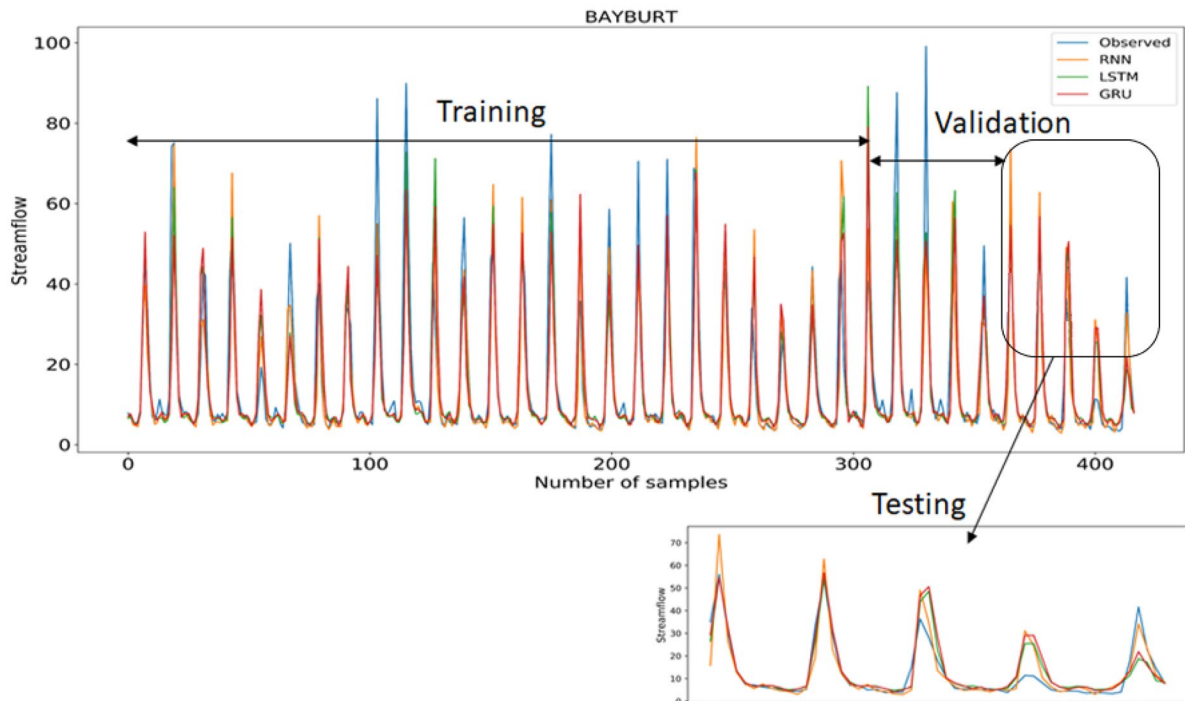

Fig. 11 Comparison of observed and forecasted streamflow using the proposed model (GRU) and LSTM and RNN models at Bayburt station
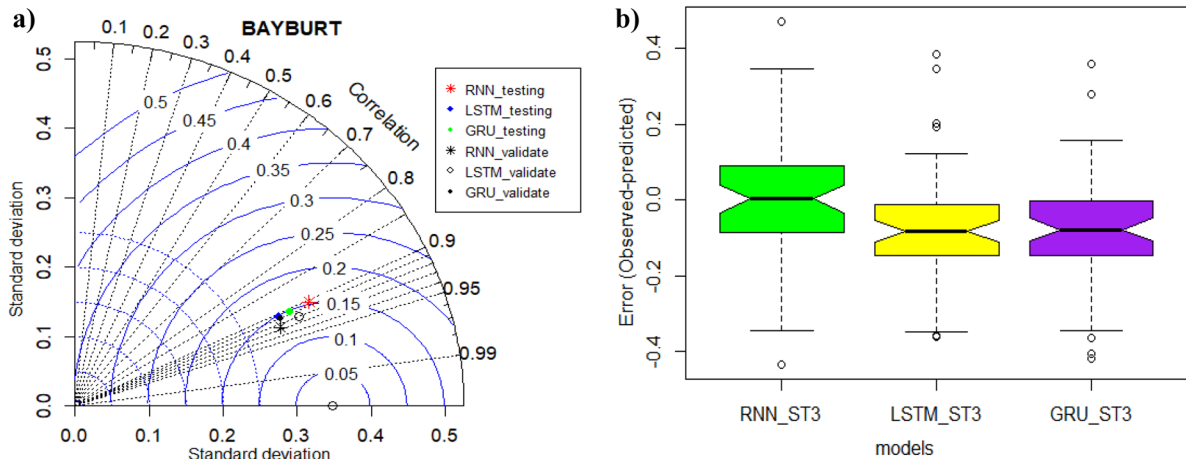
Fig. 12 **a** Taylor plot displaying the models (RNN, LSTM, and GRU) performance at Bayburt station during the validation and testing phase. **b** Box whisker plot displaying testing error (observed-predicted) at Bayburt station

This is also observed in Fig. 8 as the LSTM model cannot predict the streamflow peak. This analysis of the results showed that the GRU model is better than the LSTM model to forecast the peak streamflow (Fig. 9).

Model performance for Bayburt station (E23A004)

GRU model is again established on the Bayburt station to check it performance. During the training phase RNN model ($r = 0.943$) outperformed the other models. But during the validation period, the performance of the RNN model ($r = 0.925$) is reduced significantly and showed the underfit condition. This underfit condition was observed consistently for the three stations during model development. This concludes that the RNN model fails to maintain accuracy efficiently (Table 3). The performance of the GRU ($r = 0.916$) and LSTM ($r = 0.916$) are same during the training phase (Fig. 10). During the validation period, the performance of the LSTM model ($r = 0.914$) is more consistent than the GRU model ($r = 0.902$), but it failed during the testing period. During the testing period, the LSTM model overfit and predicted a slightly higher value than the observed. But same time GRU model showed the consistent performance to forecast the streamflow (Fig. 11). In terms of other performance indices, the GRU model ($RMSE = 0.159$, $RSR = 0.459$) showed approximately close RMSE error than the LSTM ($RMSE = 0.154$, $RSR = 0.451$) during testing. The volumetric efficiency of the GRU model ($VE = 0.869$) is approximately equal to

the LSTM model ($VE = 0.872$). The Taylor diagram further confirms this fact as the model coincided (Fig. 12a). In addition, box whisker also shows a similar range of error and median for the GRU and LSTM (Fig. 12b). This analysis concludes that the GRU model can replace the LSTM model for this station.

In Table 3, the success of the LSTM, GRU, deep-learning models predict monthly flows at a satisfactory level and close to each statistical parameters. At the end of the analysis, the RNN model at Erzincan and Bayburt stations and the RNN and GRU models at Gümüshane station showed the most successful estimation results. In addition, it has been determined that all the established deep learning models predict monthly flows at a satisfactory level and close to each other.

De melo et al. (2019) determined that the models of GRU and LSTM networks perform more effectively than the MLP and ARIMA models. Sahoo et al. (2019) India used LTSM and RNN models to estimate daily discharge data in the Mahanadi River basin. He found that the LSTM-RNN model showed more successful prediction outputs than the RNN model. Zhao et al., (2021a, 2021b) and Shu et al. (2021) found that deep learning models outperform machine learning models such as ANN, ELM, and SVR for monthly flow predictions. When the study is compared with the existing literature, it overlaps with the literature in that deep learning algorithms produce effective results from stream estimation.

Table 3 Evaluation of model performance

Fitness Parameters	Training			Validation			Testing		
RNN	LSTM	GRU	RNN	LSTM	GRU	RNN	LSTM	GRU	
	Station 1 (Erzincan)								
r	0.965	0.963	0.949	0.906	0.904	0.904	0.908	0.897	0.899
Rank	3	2	1	3	2	1	3	1	2
RMSE	0.082	0.086	0.099	0.125	0.127	0.121	0.119	0.126	0.126
Rank	3	2	1	2	1	3	3	2	1
RSR	0.264	0.276	0.316	0.443	0.448	0.429	0.418	0.443	0.443
Rank	3	2	1	2	1	3	3	2	1
NSE	0.930	0.923	0.900	0.800	0.795	0.812	0.821	0.799	0.800
Rank	3	2	1	2	1	3	3	1	2
d	0.981	0.979	0.972	0.949	0.948	0.949	0.952	0.946	0.944
Rank	3	2	1	3	1	2	3	2	1
VE	0.965	0.963	0.958	0.946	0.946	0.948	0.949	0.947	0.946
Rank	3	2	1	2	1	3	3	2	1
Total rank	**18**	12	6	14	7	**15**	**18**	10	6
Total rank	**RNN:50,** LSTM:29, GRU:27								
	Station 2 (Gümüşhane)								
r	0.931	0.908	0.914	0.895	0.901	0.922	0.902	0.913	0.902
Rank	3	1	2	1	2	3	2	3	1
RMSE	0.077	0.084	0.081	0.101	0.094	0.084	0.07	0.07	0.072
Rank	3	1	2	1	2	3	3	2	1
RSR	0.384	0.419	0.405	0.463	0.430	0.384	0.438	0.435	0.432
Rank	3	1	2	1	2	3	1	2	3
NSE	0.852	0.823	0.836	0.781	0.811	0.849	0.804	0.806	0.770
Rank	3	1	2	1	2	3	2	3	1
d	0.959	0.948	0.952	0.938	0.944	0.957	0.945	0.946	0.936
Rank	3	1	2	1	2	3	2	3	1
VE	0.955	0.953	0.953	0.945	0.954	0.956	0.957	0.956	0.953
Rank	3	2	1	1	2	3	3	2	1
Total rank	**18**	7	11	6	12	**18**	13	**15**	8
Total rank	**RNN:37,** LSTM:34, **GRU:37**								
	Station 3 (Bayburt)								
r	0.943	0.916	0.916	0.925	0.914	0.902	0.907	0.927	0.920
Rank	3	2	1	3	2	1	3	3	2
RMSE	0.123	0.143	0.143	0.141	0.143	0.148	0.150	0.154	0.159
Rank	3	2	1	3	2	1	3	2	1
RSR	0.346	0.400	0.401	0.419	0.425	0.439	0.441	0.451	0.459
Rank	3	2	1	3	2	1	3	2	1
NSE	0.880	0.839	0.839	0.821	0.816	0.803	0.802	0.792	0.764
Rank	3	2	1	3	2	1	3	2	1
d	0.967	0.953	0.953	0.950	0.953	0.946	0.951	0.945	0.942
Rank	3	2	1	2	3	1	3	2	1
VE	0.916	0.899	0.899	0.904	0.903	0.900	0.882	0.872	0.869
Rank	3	2	1	3	2	1	3	2	1
Total rank	**18**	12	6	**17**	11	6	**18**	13	7
Total rank	**RNN:53,** LSTM:36, GRU:19								

Bold characters indicate the best models

Conclusion

Application of these deep learning techniques is encouraged because this type of model offers the possibility to take advantage of the sequential nature that can help achieve higher accuracy. This article discusses the feasibility of deep learning approaches such as GRU, RNN, and LSTM for estimating monthly stream flow. In this, a new generation of sequential deep learning models is tested to forecast the streamflow, which is always challenging in water resources. As a result of the rank analysis, it was determined that the most successful prediction model was RNN.

Multiple benchmarks are comprehensively tested and compared to the proposed GRU to forecast framework on a real-world dataset. Considering this dataset, the proposed GRU framework achieved significant forecasting accuracy like the LSTM model. The inconsistency in the monthly streamflow profile at different stations generally affects the predictability. The higher the inconsistency, the more the GRU can contribute to the forecasting improvement compared to the simple RNN and LSTM. As for future work, methodologies for parameter tuning can be developed to further increase forecasting accuracy on different types of stations, especially for stations with varying flow regimes. In this study, parameter optimization was done simply. However, in the future, the parameters can be adjusted thanks to hybrid structures. Moreover, although individual streamflow forecast is far from accurate, aggregating all individual predictions yields a better prediction for the aggregation level than the conventional strategy of directly forecasting the streamflow.

Data availability   All data used in this study is available in a public repository, namely the Climate Data Store.

Declarations

Ethical approval   The manuscript complies with all the ethical requirements. The paper was not published in any journal.

Consent to publish   The authors confirm that the work described has not been published before, and it is not under consideration for publication elsewhere.

Conflict of interest   The authors declare no competing interests.

References

Alizadeh, B., Bafti, A. G., Kamangir, H., Zhang, Y., Wright, D. B., & Franz, K. J. (2021). A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *Journal of Hydrology, 601*, 126526.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. Technical report, arXiv preprint arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473

Banerjee, P., Singh, V. S., Chatttopadhyay, K., Chandra, P. C., & Singh, B. (2011). Artificial neural network model as a potential alternative for groundwater salinity forecasting. *Journal of Hydrology, 398*(3–4), 212–220.

Chandra, R., & Zhang, M. (2012). Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction. *Neurocomputing, 86*, 116–123. https://doi.org/10.1016/j.neucom.2012.01.014

Chang, F. J., Chang, L. C., & Huang, H. L. (2002). Real-time recurrent learning neural network for streamflow forecasting. *Hyd. Processes., 16*, 2577–2588. https://doi.org/10.1002/hyp.1015

Chang, Y. Y., Sun, F. Y., Wu, Y. H., & Lin, S. D. (2018). A memory-network based solution for multivariate time-series forecasting. Cornell Uni. arXiv:1809.02105. https://doi.org/10.48550/arXiv.1809.02105

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Science and Reports, 8*, 6085. https://doi.org/10.1038/s41598-018-24271-9.10.1038/s41598-018-24271-9

Cheng, C. T., Xie, J. X., Chau, K. W., & Layeghifard, M. (2008). A new indirect multi-step-ahead prediction model for a long-term hydrologic prediction. *Journal of Hydology., 361*, 118–130. https://doi.org/10.1016/j.jhydrol.2008.07.040

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259. https://doi.org/10.48550/arXiv.1409.1259

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*, 1251–1258.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. Cornell Uni. arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks, 5*(2), 240–254. https://doi.org/10.1109/72.279188

Coulibaly, P., & Baldwin, C. K. (2005). Non-stationary hydrological time series forecasting using nonlinear dynamic

methods. *Journal of Hydology., 307*, 164–174. https://doi.org/10.1016/j.jhydrol.2004.10.008

Coulibaly, P., Anctil, F., Rasmussen, P., & Bobee, B. (2000). A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrological Processes, 14*, 2755–2777. https://doi.org/10.1002/1099-1085(20001030)14:15%3c2755::AID-HYP90%3e3.0.CO;2-9

De Melo, G. A., Sugimoto, D. N., Tasinaffo, P. M., Santos, A. H. M., Cunha, A. M., & Dias, L. A. V. (2019). A new approach to river flow forecasting: LSTM and GRU multi-variate models. *IEEE Latin America Transactions, 17*(12), 1978–1986. https://doi.org/10.1109/TLA.2019.9011542

Donate, J. P., & Cortez, P. (2014). Evolutionary optimization of sparsely connected and time-lagged neural networks for time series forecasting. *Applied Soft Computing, 23*, 432–443. https://doi.org/10.1016/j.asoc.2014.06.041

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water, 10*, 1542–1557. https://doi.org/10.3390/w10111543

Kim, K., Kim, D. K., Noh, J., & Kim, M. (2018). Stable forecasting of environmental time series via long short term memory recurrent neural network. *IEEE Access., 6*, 75216–75228. https://doi.org/10.1109/ACCESS.2018.2884827

Kumar, D. N., Raju, K. S., & Sathish, T. (2004). River flow forecasting using recurrent neural networks. *Water Resources Management, 18*, 143–161. https://doi.org/10.1023/B:WARM.0000024727.94701.12

Kumar, A., Son, H. L., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access, 7*, 23319–23328. https://doi.org/10.1109/ACCESS.2019.2899260

Ouyang, Q., & Lu, W. (2018). Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. *Water Resources Management, 32*(2), 659–674. https://doi.org/10.1007/s11269-017-1832-1

Prasad, S. C., & Prasad, P. (2014). Deep recurrent neural networks for time-series prediction. IEEE. 1–19. https://doi.org/10.48550/arXiv.1407.5949.

Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., & Zhang, B. (2019). A novel combined prediction scheme based on CNN and LSTM for urban PM2.5 concentration. *IEEE Access, 7*, 20050–20059. https://doi.org/10.1109/ACCESS.2019.2897028

Sahoo, B. B., Jha, R., Singh, A., & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica, 67*(5), 1471–1481. https://doi.org/10.1007/s11600-019-00330-1

Sattari, M. T., Yurekli, K., & Mahesh, P. (2012). Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. *Applied Mathematical Modelling, 36*, 2649–2657. https://doi.org/10.1016/j.apm.2011.09.048

Sharma, P., & Machiwal, D. (Eds.). (2021). *Advances in streamflow forecasting: from traditional to modern approaches*. Amsterdam: Elsevier. https://doi.org/10.1016/C2019-0-02163-2

Shoaib, M., Shamseldin, A. Y., Melville, B. W., & Khan, M. M. (2016). A comparison between wavelet based static and dynamic neural network approaches for runoff prediction.

Journal of Hydology, 535, 211–225. https://doi.org/10.1016/j.jhydrol.2016.01.076

Shu, X. S., Ding, W., Peng, Y., Wang, Z. R., Wu, J., & Li, M. (2021). Monthly streamflow forecasting using convolutional neural network. *Water Resources Management, 35*(15), 5089–5104. https://doi.org/10.1007/s11269-021-02961-w

Smith, C., & Jin, Y. (2014). Evolutionary multi-objective generation of recurrent neural network ensembles for time series prediction. *Neurocomputing, 143*, 302–311. https://doi.org/10.1016/j.neucom.2014.05.062

Wegayehu, E. B., & Muluneh, F. B. (2022). Short-Term daily univariate streamflow forecasting using deep learning models. *Advances in Meteorology*, *2022*. https://doi.org/10.1155/2022/1860460

Wu, C. L., Chau, K. W., & Li, Y. S. (2011) River stage prediction based on a distributed support vector regression. Journal of Hydrology, 358 (1–2), 96–111. https://doi.org/10.1016/j.jhydrol.2008.05.028.

Xiao, S., Yan, J., Yang, X., Zha, H., & Chu, S. M. (2017). Modeling the intensity function of point process via recurrent neural networks. *Conference on Artificial Intelligence*, 1597–1603. https://doi.org/10.1609/aaai.v31i1.10724.

Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A., & Sayl, K. N. (2015). Artificial intelligence based models for streamflow forecasting: 2000–2015. *Journal of Hydrology, 530*, 829–844. https://doi.org/10.1016/j.jhydrol.2015.10.038

Zhang, D., Lindholm, G., & Ratnaweera, H. (2018a). Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *Journal of Hydrology., 556*, 409–418. https://doi.org/10.1016/j.jhydrol.2017.11.018

Zhang, Z., Zhang, Q., & Singh, V. P. (2018b). Univariate streamflow forecasting using commonly used data-driven models: Literature review and case study. *Hydrological Sciences Journal, 63*(7), 1091–1111. https://doi.org/10.1080/02626667.2018.1469756

Zhang, H., Zhou, J., JahedArmaghani, D., Tahir, M., Pham, B., & Huynh, V. (2020). A combination of feature selection and random forest techniques to solve a problem related to blast-induced ground vibration. *Applied Sciences, 10*(3), 869. https://doi.org/10.3390/app10030869

Zhao, X. H., Lv, H. F., Lv, S. J., Sang, Y. T., Wei, Y. Z., & Zhu, X. P. (2021). Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer. *Journal of Hydrology, 601*, 126607. https://doi.org/10.1016/j.jhydrol.2021.126607

Zhao, X. H., Lv, H. F., Wei, Y. Z., Lv, S. J., & Zhu, X. P. (2021). Streamflow forecasting via two types of predictive structure-based gated recurrent unit models. *Water, 13*(1), 91. https://doi.org/10.3390/w13010091