Short-Term Streamflow Forecasting Using the Feature-Enhanced Regression Model

Yun Bai¹ · Nejc Bezak² · Klaudija Sapač² · Mateja Klun² · Jin Zhang³

Received: 12 August 2019 / Accepted: 22 October 2019/
Published online: 25 November 2019
© Springer Nature B.V. 2019

Abstract

Reservoir inflow forecasting is extremely important for the management of a reservoir. In practice, accurate forecasting depends on the feature learning performance. To better address this issue, this paper proposed a feature-enhanced regression model (FER), which combined stack autoencoder (SAE) with long short-term memory (LSTM). This model had two constituents: (1) The SAE was constructed to learn a representation as close as possible to the original inputs. Through deep learning, the enhanced feature could be captured sufficiently. (2) The LSTM was established to simulate the mapping between the enhanced features and the outputs. Under recursive modeling, the patterns of correlation in the short term and dependence in the long term were considered comprehensively. To estimate the performance of the FER model, two historical daily discharge series were investigated, i.e., the Yangtze River in China and the Sava Dolinka River in Slovenia. The proposed model was compared with other machine-learning methods (i.e., the LSTM, SAE-based neural network, and traditional neural network). The results demonstrated that the proposed FER model yields the best forecasting performance in terms of six evaluation criteria. The proposed model integrates the deep learning and recursive modeling, and thus being beneficial to exploring complex features in the reservoir inflow forecasting. Moreover, for smaller catchments with significant torrential characteristics, more data are needed (e.g., at least 20 years) to effectively train the model and to obtain accurate flood-forecasting results.

Keywords Long short-term memory · Stack autoencoder · Feature enhanced · Daily reservoir inflow · Forecast

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11269-019-02399-1>) contains supplementary material, which is available to authorized users.

✉ Yun Bai
yunbai@foxmail.com

Extended author information available on the last page of the article

1 Introduction

River discharge forecasting provides a powerful tool for management of multipurpose reservoirs, optimization of hydropower production, water supply, flood control, and irrigation. In particular, the daily or hourly short-term forecasting models are of special interest to the end users, because they enable efficient decision-making. Hydrological data are complex and strongly subjected to climate variability and anthropological effects, which poses a strong limitation to the development of robust short-term forecast models. River discharge modeling approaches could be divided into three basic groups: conceptual, physically based, and empirical models (Devia et al. 2015). A lot of research work has focused on the use of conceptual models, such as the Soil and Water Assessment Tool (Sun et al. 2015) and the XXX distributed hydrological model (the first X denotes Xinanjiang, the second X denotes hybrid, and the T denotes TOPMODEL) (Zhao et al. 2016). Most of the hydrological processes, however, are characterized by a high degree of spatial and temporal variability, which can be difficult to fully understand and model (Humphrey et al. 2016). Furthermore, some studies suggest that conceptual models have a limited ability to provide short-term predictions and need long-term datasets to calibrate the model (Mosavi et al. 2018). Moreover, some examples show that conceptual models, in some cases, could have limited flood-forecasting ability (Costabile and Macchione 2015). Data-driven models can be divided into statistical—for example, autoregressive-moving-average model (Wang et al. 2015; Myronidis et al. 2018) and black-box models—that are based on various techniques of artificial intelligence (Long et al. 2019). In the literature, it is evident that applications of data-driven techniques have increased in the past decade.

The use of artificial intelligence modeling with adaptive learning capability is a promising alternative to traditional forecasting (Yang et al. 2017a; Su et al. 2019; Qiao et al. 2020). Examples from the literature show that the use models, including the artificial neural network (Badrzadeh et al. 2016), support vector machine (Zhao et al. 2017), decision tree model (Jothiprakash and Kote 2011), fuzzy inference system (Allawi et al. 2018), Bayesian particle filter (Xu et al. 2017), expert system (Zhang et al. 2018), and hybrid model (Bai et al. 2016b), have resulted in satisfactory model performance. A comprehensive review of the machine-learning models for both short-term and long-term predictions was presented by Mosavi et al. (2018). A comparative study of wavelet transform-based machine-learning models for monthly inflow forecasting was addressed by Poul et al. (2019).

Deep learning was introduced by Bengio et al. (2006) and has been used widely in different fields, including, for example, environment forecasting (Xie et al. 2018), hydrological modeling (Bai et al. 2016a), and energy forecasting (Wang et al. 2017). A long short-term memory (LSTM) feature presented by Hochreiter and Schmidhuber (1997) has been used to introduce feedback connections that account for the short-term memory cells. The LSTM models, however, cannot capture both long-term and short-term features at the same time. To improve the forecasting ability, a relationship between the long- and short-term memory has to be established (Srivastava and Lessmann 2018; Wei et al. 2019). In this study, we presented a feature-enhanced regression (FER) model with integrated stack autoencoder (SAE) with LSTM for the daily discharge forecasting. To the best of our knowledge, the LSTM has been used successfully only for monthly discharge forecasting (Yuan et al. 2018) and not for daily discharge modeling.

In this work, the SAE was first applied to learn the invariant and abstract features of time series in an unsupervised manner hierarchically (Vincent et al. 2008). The LSTM was then constructed to simulate the mappings between the extracted features and the targets. The

output (i.e., result) is the next day's discharge (i.e., inflow to the reservoir). This work is rooted in the application of a deep learning-based model to improve hydrology forecasting performance. The proposed methodology was applied to two rivers with completely different characteristics (e.g., catchment area, climate characteristics): the Sava Dolinka River in Slovenia and the Yangtze River in China.

Therefore, the main aim of this paper was to investigate whether the proposed FER model could yield similar forecasting performance for a smaller torrential case study as well as for the significant larger case catchment with no torrential properties. The suggested model was compared with some other frequently used data-driven models. In the second step of the study, the impact of the training size on the forecasting performance was evaluated as well.

2 Methodology

This section contains a description of the four tested models—that is, LSTM, SAE, feed-forward neural network (FFNN), and SAE-LSTM—used for the investigation of daily discharge forecasting. This section also presents the proposed SAE-LSTM model, which is also named the constructed FER model. A brief overview of model performance criteria is given at the end of this section. We tested the model, which was proven to be the best, in relation to the length of the dataset used for the training (i.e., 10, 15, and 20 years).

2.1 Long Short-Term Memory Model (LSTM)

The LSTM is an improved version of the recurrent neural network (RNN) model. Compared with the traditional neural network, the LSTM differs from the traditional neural network (e.g., FFNN) in two main properties: (1) the hidden layer of the LSTM has connections between the nodes, and (2) inputs of the hidden layer of LSTM at the present time contain both the outputs of the input layer and the hidden layer at the previous time (Su and Lu 2017). Compared with the basic RNN, the LSTM overcomes the vanishing gradient problem for long-term dependencies modeling (Gers et al. 1999). Compared with other versions of the RNN, the LSTM computes only the memory units using different activation functions and has a simple structure. Therefore, the LSTM is selected for feature regression. A structure of the LSTM unit is shown in Fig. 1c, which is composed of four parts: input gate (IG), output gate (OG), forget gate (FG), and cell.

The modeling process of the LSTM is presented and described in Gers et al. (1999).

- (1) For the t -th time step, the information of the new inputs $\mathbf{x}_t$ is accumulated to the cell if the IG is activated as $\mathbf{I}_t = \sigma(\mathbf{w}_{xi} \mathbf{x}_t + \mathbf{w}_{hi} \mathbf{x}_{t-1} + \mathbf{w}_{ci} \times \mathbf{c}_{t-1} + \mathbf{b}_i)$, where $\sigma(\cdot)$ is a sigmoid function.
- (2) At the same time, the FG is utilized to evaluate, which information is not needed from the previous cell state using $\mathbf{F}_t = \sigma(\mathbf{w}_{xf} \mathbf{x}_t + \mathbf{w}_{hf} \mathbf{x}_{t-1} + \mathbf{w}_{cf} \times \mathbf{c}_{t-1} + \mathbf{b}_f)$.
- (3) The old cell status $\mathbf{c}_{t-1}$ is updated using the new one $\mathbf{c}_t = \mathbf{F}_t \times \mathbf{c}_{t-1} + \mathbf{I}_t \times \tanh(\mathbf{w}_{xc} \mathbf{x}_t + \mathbf{w}_{hc} \mathbf{h}_{t-1} + \mathbf{b}_c)$.
- (4) The updated cell state $\mathbf{c}_t$ is passed through the “tanh” function and multiplied with the sigmoid activation function of the OG to determine the final output from the LSTM unit $\mathbf{h}_t$. The expression is written as $\mathbf{h}_t = \mathbf{O}_t \times \tanh(\mathbf{c}_t)$, where $\mathbf{O}_t = \sigma(\mathbf{w}_{xo} \mathbf{x}_t + \mathbf{w}_{ho} \mathbf{x}_{t-1} + \mathbf{w}_{co} \times \mathbf{c}_t + \mathbf{b}_o)$.

Fig. 1 Architecture of the proposed FER model (left), a) structure of single AE, b) stack AEs, and c) LSTM unit

In the steps above, w_{xi} , w_{xf} , w_{xo} are the input weights, w_{hi} , w_{hf} , w_{ho} denote the recurrent weights, w_{ci} , w_{cf} , w_{co} represent the peephole weights, b_i , b_f , b_c , b_o are the bias vectors, and “ $\times$ ” means point-wise multiplication.

2.2 Stack Auto-Encoder Neural Network (SAE-NN)

In the process of machine-learning application, it is essential to select a representative set of features before building the model. Regarding the feature extraction, deep learning has been regarded as a promising approach because of its layer-wise learning framework (Li et al. 2016). That is, the higher levels are specified by the lower levels, where the representation of the low levels may define several different features of the high levels, which makes the data representation more abstract and nonlinear for the higher levels (Hinton and Osindero 2006). Therefore, the SAE, as a typical approach in deep learning, enjoys all of the benefits of any deep network of greater expressive power.

A structure of the SAE is shown in Fig. 1b. The SAE is a neural network method with multiple layers of the AE in which the outputs of each layer are wired to the inputs of the successive layer. As one unit in the SAE shown in Fig. S1a, the single AE aims to reconstruct its input vectors h^{l-1} (when $l = 1$, $h^0 = x$, where x represents the original time series) at the outputs (reconstruction results r) through unsupervised learning (Vincent et al. 2008). That is, $r = f(h^{l-1}|\xi) \approx h^{l-1}$, where ξ means a parameter set, including w (weights) and b (bias values) between neighboring layers. The process of the SAE is described in detail in Hinton and Salakhutdinov (2006).

- (1) Train the first AE on the raw inputs x to learn primary features h^1 according to $h^1 = f_e(x) = \text{Sigm}(wx + b)$, where $f_e(\cdot)$ means encoder. Use $x' = f_d(h^1) = \text{Sigm}(w^T h^1 + b^T)$ to reconstruct the input information, where $f_d(\cdot)$ means decoder. Evaluate the reconstruction error $(x - x')$ to determine the primary feature h^1 (minimize the error to make the as close as possible), which gives a representation of the input in terms of higher-order features.

- (2) Use the primary feature $\mathbf{h}^1$ as the “raw input” to the second AE to learn secondary features $\mathbf{h}^2$. Following the previous step (i.e., encoder $\mathbf{h}^1 = f_e(\mathbf{h}^{l-1}) = \text{Sigm}(\mathbf{w}\mathbf{h}^{l-1} + \mathbf{b})$, decoder $\mathbf{r} = f_d(\mathbf{h}^1) = \text{Sigm}(\mathbf{w}^T\mathbf{h}^1 + \mathbf{b}^T)$, and error evaluation), the l -th feature $\mathbf{h}^l$ can be obtained.
- (3) Combine $(l + 1)$ layers together to form a SAE with l hidden layers and one input layer. It is noted that the SAE in this paper is just used for enhanced feature extraction, and thus the decoder function in the last layer is removed.

2.3 Feed Forward Neural Network

FFNNs are known as relatively simple but powerful artificial neural networks. As evident from the name, the information propagates in the forward direction (i.e., from the input layer through one or more hidden layers to the output layer, without loops). FFNNs commonly are used for supervised learning. The layers are fully connected, but there are no connections among the perceptrons in the same layer. Encoders and decoders, part of the SAE described in the previous section, are examples of the FFNN. More details about FFNN can be found in a review paper by Sazli (2006).

2.4 Feature Enhanced Regression (FER) Model

This section introduces details of the tested FER model. The goal of this model is to predict the daily discharge data based on the discharge of the previous n days. Deep learning with the multilayer NN reduces the dimensionality of data and accelerates the learning capabilities of the network (Hinton and Osindero 2006; Hinton and Salakhutdinov 2006). In the first step, an unsupervised learning is used on raw data to produce useful intermediate correlations. For further training the primary result is used as raw input, which makes it less sensitive to the error in the input data (Vincent et al. 2008). Note that the SAE is used only for enhanced feature extraction—that is, the decoder function in the last layer is removed (Hinton and Salakhutdinov 2006). Forecasting and feature regression is done using the LSTM.

The architecture of the proposed FER model is presented in Fig. 1 (left). The modeling steps are as follows.

- Step 1. Collect historical daily data and normalize all the data into $[0, 1]$.
- Step 2. Divide data into two categories: training and testing sets.
- Step 3. Train the FER model using training set.
 - (a) Build SAE to obtain the enhanced feature of inputs, and output $\mathbf{h}^l$.
 - (b) Build LSTM to simulate the mappings between $\mathbf{h}^l$ and desired value of daily inflow.
 - (c) Evaluate the performance to determine the FER model structure.
- Step 4. Forecast the target variable using the trained FER model with the testing data.
- Step 5. Output results.

2.5 Parameter Definition Process and Performance Criteria

In this work, one week’s data (i.e. 7 days before the forecasting day) is considered as an input and used for the forecasting of the next day’s discharge value (Li et al. 2016). Regarding the SAE modelling, its deep structure is determined by the trial and error method. Considering the

input-output structure, the structure of the SAE is tested with the sizes of hidden layers from 1 to 3 (interval 1) and the number of nodes in each hidden layer from 5 to 15 (interval 5). Regarding the LSTM modelling, its structure is affected by the outputs from the SAE. In addition, the optimal FER model structure is determined by the mean square error (MSE) after running two cases (the SAE and the LSTM). Through the experiments using the training dataset (described in the section 3), the optimal structures and parameters of the proposed model in different steps are listed in Table S1. All the algorithms are computed using Matlab 2018a software (The MathWorks 2018).

In this paper, six criteria are used to estimate the forecasting performance: mean absolute percentage error (MAPE), root-mean-square error (RMSE), Pearson correlation coefficient (R), Nash-Sutcliffe model efficiency coefficient (NSE), Kling-Gupta efficiency (KGE) (Gupta et al. 2009), and non-parametric modified Kling-Gupta efficiency (RNP) (Pool et al. 2018).

3 Case Study and Data

3.1 Yangtze River, China, Asia

The Yangtze River is 6300 km long and is the longest river in Asia (Fig. S1). Yangtze River originates in the Tibetan Plateau and outflows to the East China Sea. Its flow regime is driven by the Asian monsoon climate, with wet season and peak discharges in the summer months. As the most important water conservancy project on the Yangtze River, the Three Gorges reservoir, which is located between Chongqing and Hubei Province of China (106°50'-110°50'E and 29°16'-31°25' N), was selected as the study area. The reservoir has an area of 1084 km² with full capacity of 39.3 km³, and normal storage water level of 175 m. The Three Gorges reservoir region is characterized by a subtropical monsoon climate with an average annual precipitation of 1100 mm.

3.2 Sava Dolinka River, Slovenia, Europe

Sava Dolinka is a river in the northwest part of Slovenia, Europe (Fig. S1). It is one of the upper branches of the Sava River, which is a tributary of Danube River, the second largest river in Europe. Sava Dolinka is 47 km long with the drainage area of 521 km². Mean discharge at the Jesenice gauging station (catchment area of 258 km²) is 9.99 m³/s with peaks of more than 100 m³/s (Bat et al. 2008). The flow regime is Alpine snow-rain with primary peak in spring and secondary peak in autumn. Although autumn peak is a consequence of the higher amount of precipitation, the spring peak is related to snow melting in the Alpine region (Bat et al. 2008). The Jesenice gauging station was chosen because it is the last station before the Moste dam, which influences the flow downstream. From the spring at Zelenci to the confluence with Sava Bohinjka at Radovljica, Sava Dolinka has many short tributaries that are characterized as torrents (Sodnik and Mikoš 2006). The Sava Dolinka River catchment is characterized by its Alpine climate with an average air temperature 6.6 °C as well as by an average annual precipitation of 1459 mm (ARSO 2019a).

3.3 Data

The historical daily discharge data of the Yangtze River (records of Three Gorges reservoir) and the Sava Dolinka River were obtained from China Three Gorges Corporation (CTG 2019) and Slovenian Environment Agency (ARSO 2019b), respectively. Location of both cases is

shown in Fig. S1. For both stations, we acquired the latest continuous datasets (without gaps) with a length of at least 25 years; that is, between January 1, 1992, and December 31, 2017, for the Yangtze River, and between January 1, 1988, and December 31, 2013, for the Sava Dolinka River (Fig. 2). Note the significant differences in the seasonal dynamics of both tested rivers: the Yangtze River had a clear summer peak, and low discharge values were measured in the winter; conversely, the Sava Dolinka River had significant torrential characteristics, and, compared with the Yangtze River, was more sensitive to the input rainfall. Before modeling, the data from both stations were divided into a training set (accounting for about 90%) and a testing set (accounting for about 10%) (Table S2).

The Three Gorges dam project was completed in 2012, although it had started to influence the water level in 2003 (Yang et al. 2017b). The Moste dam was constructed in 1952. Thus, construction of the Moste dam did not affect the measured hydrological data used in this study because we used data from 1988 to 2017. For the Three Gorges dam, the selected discharge gauging station was located upstream from the dam (before the reservoir), and the dam did not have a significant impact on the discharge series (this was evident in Fig. 2). We conducted additional tests using the Mann-Kendall test, which can be used to detect trends in the environmental data. The Mann-Kendall test result for the Yangtze River case study was 0.0429 with a p value of 2.22×10^{-16} , which meant that there was no trend present in the discharge data. In case that dam had a significant impact on the discharge, we would have expected it to have a statistically significant trend. This, however, was not the case in the presented study. Moreover, the Mann-Kendall test result for the Sava Dolinka case study was 0.106×10^{-16} , which also meant that no significant trend was present in the data. Furthermore, we regarded the discharge data as a representation of the different conditions in the investigated watersheds, such as climate, hydrology, geomorphology, and engineering structures.

The case studies and, consequently, the stations were selected, based on the prior knowledge of their hydrological characteristics. Moreover, the stations were located upstream of the inflow to the hydropower plant reservoir. Rivers, however, had some significant differences that may have influenced the model performance. For example, the Yangtze River had a much higher autocorrelation of the discharge series than the Sava Dolinka River. A significant decrease in the ACF was visible after a lag time of 2 days for Sava Dolinka River. However, the Yangtze River had an autocorrelation that was higher than 0.5 for the 40 days' time lag (Fig. 2). Therefore, we could assume that models described in section 2 could yield better performance for the Yangtze River case study compared with the Sava Dolinka River.

Moreover, there was a significant difference in the discharge scale (i.e., mean, min, max) between the selected stations. Compared with the Jesenice station, the median and mean discharges at the Three Gorges reservoir were more than 1000 times higher (Table S2). When comparing specific discharges of the two stations through the entire year, however, the Sava Dolinka River had a higher specific discharge than the Yangtze River (Fig. S2). Fig. S2 also shows the differences between the average seasonal dynamics of the two tested cases.

After we evaluated the model with the best performance of daily discharge forecasting, we investigated the influence of the shortened length of data for training the model. We used 10, 15, and 20 years of data for the training of the model, and the length of the validation period remained the same (i.e., 3 years). The distribution of discharge data used for the calibration and validation is shown in Fig. S2. For the Sava Dolinka River, four calibration periods and one validation period had similar properties. The same can be said for the Yangtze River. Differences between the validation period and the four calibration periods for the Yangtze River were a bit larger than they were for the Sava Dolinka River case study.

Fig. 2 Measured discharge series for the Three Gorges-Yangtze (inflow to the reservoir) and Jesenice-Sava Dolinka gauging stations (above plots). Below plots show autocorrelation of discharge series for the Yangtze River and Sava Dolinka River. Lag units are days. Dotted blue lines indicate 95% confidence intervals

4 Results and Discussion

4.1 Comparison of Different Models

To evaluate the forecasting performance of the FER model that was proposed for the daily discharge forecasting in this study, three models were selected for comparison: the LSTM, the SAE-NN, and the FFNN. A detailed description of these models was given in section 3, and the main characteristics are summarized in Table S1. Figure 3 compares the four tested models for the 3-year testing period for the Yangtze and Sava Dolinka Rivers. On the basis of the presented results, one can notice that in terms of seasonal dynamics for the Yangtze River, all four tested models yielded similar performance. The largest differences could be detected in the summer period (i.e., high-flow period). In the case of the Sava Dolinka, differences in the low-flow period were larger compared with the Yangtze River. A comparison between the

Fig. 3 Evaluation of tested models using daily data (a), median monthly values (b), flow duration curve (c) and scatter plot (d) for the Three Gorges reservoir on the Yangtze and using daily data (e), median monthly values (f), flow duration curve (g) and scatter plot (h) for the Jesenje station on the Sava Dolinka River

Table 1 Results of six model efficiency criteria for four different models and for two case studies

Case	Yangtze River				Sava Dolinka River			
Evaluation Model	FER	LSTM	SAE-NN	FFNN	FER	LSTM	SAE-NN	FFNN
PBIAS [%]	-0.8	0.1	1	1.9	-0.9	2.5	2.3	0.3
RMSE [m ³ /s]	1634	2316	2232	2617	1.76	2.95	2.95	3.55
R	0.97	0.95	0.95	0.93	0.98	0.93	0.93	0.87
NSE	0.95	0.89	0.9	0.87	0.94	0.83	0.83	0.76
KGE	0.97	0.95	0.94	0.93	0.83	0.76	0.75	0.83
RNP	0.99	0.98	0.98	0.96	0.98	0.94	0.95	0.96

observed and modeled flow duration curves confirmed these findings. For the Sava Dolinka River, we detected larger differences among modeled and observed data for low-flows, and the FER model yielded the closest fit to the observed data. Scatter diagrams between simulated and observed data (Fig. 3d and h) indicated that for both tested case studies, the FFNN yielded the worse results. For all four models, the differences between modeled and observed data were normally distributed.

A quantitative estimation for the models' performance for both case studies is presented in Table 1. The feature-enhanced models (i.e., FER and SAE-NN) have lower RMSE and higher R and NSE than the other two tested models (i.e., LSTM and FFNN), which illustrated that the feature learning can improve the forecasting performance. Moreover, KGE and RNP criteria indicated that the FER model yielded better modeling performance than the other three tested models for both case studies. However, note that differences among models were relatively small, which is in accordance with the graphic comparison of results shown in Fig. 3. Additionally, differences in percentage bias among the models were low, but according to the criterion, the LSTM and FFNN models gave the best results for the Yangtze River and the Sava Dolinka River, respectively. In general, the FER model slightly underestimated measured data, and the other three models slightly overestimated observed data for both case studies. More specifically, compared with the SAE-NN model, the RMSE for the FER model decreased for 597.9 m³/s. On the basis of the graphic comparison of the tested models and the numerical values of the different efficiency criteria, we ranked the tested models from the best to the worst: FER, SAE-NN, LSTM, and FFNN. This indicated that the proposed model integrating the deep feature learning with the dynamic regression model could be beneficial to sufficiently learning sophisticated features and thus exhibited a better capacity for the daily reservoir inflow forecasting. Moreover, note that tested models yielded similar performance for both case studies despite some relatively large differences between the Yangtze and Sava Dolinka Rivers, such as catchment area, hydrological response, and autocorrelation. This indicated that the tested methods were suitable for a variety of different catchments.

4.2 Impact of Training Size on the Prediction Results

It was found that the FER model, for which 23 years of discharge data were used for the model training, most successfully forecasted the discharge for both rivers. Therefore, in the next step of this study, we asked the question if and how the shorter length of the training period (i.e., 10, 15, and 20 years) influenced the results of the FER model. Figure 4 shows a graphic comparison for modeling performance in case of using a different length of training set (shown in Fig. S2) for both tested case studies. For the Yangtze River, the shortening of the

Fig. 4 Evaluation of the training size impact on the modelling results using daily data (a), median monthly values (b), flow duration curve (c) and scatter plot (d) for the Three Gorges reservoir on the Yangtze River and using daily data (e), median monthly values (f), flow duration curve (g) and scatter plot (h) for the Jesenice station on the Sava Dolinka River

Table 2 Results of FER model efficiency for three different lengths of training set and two case studies

Case	Yangtze River Sava			Dolinka River		
Evaluation	Length of the training set					
	10 years	15 years	20 years	10 years	15 years	20 years
PBIAS [%]	1	−0.1	1	−1.2	0.2	1.5
RMSE [m³/s]	2409	2258	2197	4.98	4.68	4.17
R	0.94	0.95	0.95	0.72	0.76	0.82
NSE	0.89	0.9	0.91	0.52	0.58	0.67
KGE	0.94	0.95	0.95	0.6	0.68	0.74
RNP	0.98	0.98	0.98	0.92	0.93	0.94

training size did not have a significant impact on the FER model performance. This meant that we obtained similar results in case that model was trained with 10 or 20 years of data. Conversely, differences were larger for the Sava Dolinka case study that had significant torrential characteristics. In this case, more data are needed to train the proposed FER model to obtain good modeling results. The same conclusion could be made based on the numerical results shown in Table 2. Perrin et al. (2007) indicated that 350 calibration days that included both wet and dry conditions are sufficient to obtain robust estimates of model parameters. They also stated that more simple models (i.e., less parameter) generally required less data. Moreover, Juston et al. (2009) pointed out that, in some cases, calibration using a smaller percentage of data could provide similar results as using the entire dataset. This study, however, focused on areas with a catchment area between 5 and 6 km² that had different hydrological behavior than the catchment areas investigated in this study. In this paper, for the torrential Sava Dolinka case study, if we used 10 years of data for calibration instead of 20 years of data, this yielded nearly 25% worse performance according to some of the criteria in Table 2. Additionally, a calibration period should include the most relevant hydrometeorological processes, such as the monsoon period (Islam et al. 2018), which was also the case in this study for which we were able to select enough data to capture the variability in the hydrological regimes.

5 Conclusions

In this paper, the FER was proposed for daily discharge forecasting and compared using three methods: SAE-NN, LSTM, and FFNN. To estimate the model performance, two historical daily discharge series were collected from the Yangtze River in China and from the Sava Dolinka River in Slovenia. Moreover, we also investigated the impact of the training size on the forecasting performance. On the basis of the presented results, the following conclusions can be made.

- The proposed FER model exhibited a powerful feature learning ability and achieved high forecasting accuracy. Compared with other models, the proposed model also displayed the best performance in terms of six quantitative criteria and a graphic comparison between modeled and observed data. Through feature-enhanced extraction and recursive modeling, the proposed model could improve forecasting performance significantly.
- Similar performance was obtained for the Yangtze River and Sava Dolinka River despite some significant differences between the two cases.

- For the Yangtze River, the shortening of the training size did not have a significant impact on the FER model performance. For the Sava Dolinka River, however, the shorter training size significantly reduced FER model performance, which indicated that in the case of a river with torrential characteristics, more data were needed to adequately train the model.

This paper addressed a method that integrated a deep framework with recursive modeling, which was useful for exploring sophisticated features of the discharge series. The deep-learning approach (i.e., SAE in this paper) had a strong capacity for feature-enhanced extraction, the performance of which directly affected the regression accuracy. The recursive model (i.e., LSTM in this paper) had good fitting ability for the short-term relationship and the long-term dependence. In other words, the two specific techniques could be replaced with other techniques that had the same functions (e.g., one type of deep learning or recursive model could replace the SAE and the LSTM) following the present method. This demonstrated that the proposed model had a certain generalization and a good basis for the daily discharge forecasting in a variety of catchment areas with different characteristics because we obtained relatively good modeling results in this study.

Acknowledgements This work is supported in part by the National Natural Science Foundation of China (71801044), the international cooperation of the Ministry of Science and Technology of China (12-24: bilateral project between China and Slovenia entitled: “Evaluation of intelligent learning techniques for prediction of hydrological data: useful case studies in China and Slovenia”), and the Natural Science Foundation of Chongqing (cstc2018jcyjAX0436). This work was also partially supported by the Slovenian Research Agency (ARRS) through grants J2-7322 and P2-0180.

Compliance with Ethical Standards

Conflict of Interest No conflict of interest.

References

- Allawi MF, Jaafar O, Hamzah FM et al (2018) Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system : a case study for a semi-arid region. *Theor Appl Climatol* 134:545–563
- ARSO (2019a) Klimatološka povprečja 1981–2010. http://www.meteo.si/met/sl/climate/tables/n-ormals_81_10/. Accessed 24 May 2019
- ARSO (2019b) Arhiv hidroloških podatkov. Arhiv površinskih voda. http://vode.arso.gov.si/hi-darhiv/pov_arhiv_tab.php. Accessed 24 May 2019
- Badrzadeh H, Sarukkalige R, Jayawardena AW (2016) Improving Ann-based short-term and long-term seasonal river flow forecasting with signal processing techniques. *River Res Appl* 32:245–256
- Bai Y, Chen Z, Xie J et al (2016a) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *J Hydrol* 532:193–206
- Bai Y, Xie J, Wang X et al (2016b) Model fusion approach for monthly reservoir inflow forecasting. *J Hydroinf* 18:634–650
- Bat M, Dolinar M, Frantar P, et al (2008) Water balance of Slovenia 1971–2000. Ljubljana, p 119
- Bengio Y, Lamblin P, Popovici D, et al (2006) Greedy layer-wise training of deep networks. In: *NIPS’06 proceedings of the 19th international conference on neural information processing systems*. Canada, pp 153–160
- Costabile P, Macchione F (2015) Enhancing river model set-up for 2-D dynamic flood modelling. *Environ Model Softw* 67:89–107
- CTG (2019) China Three Gorges corporation. <https://www.ctg.com.cn/english/>. Accessed 15 Mar 2019
- Devia GK, Ganasri BP, Dwarakish GS (2015) A review on hydrological models. *Aquatic Procedia* 4:1001–1007
- Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM. In: *Proc. ICANN’99 Int. Conf. On Artificial neural networks*, Edinburgh, Scotland. London, UK, pp 850–855

- Gupta HV, Kling H, Yilmaz KK et al (2009) Decomposition of the mean squared error and NSE performance criteria : implications for improving hydrological modelling. *J Hydrol* 377:80–91
- Hinton GE, Osindero S (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(80):504–507
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
- Humphrey GB, Gibbs MS, Dandy GC et al (2016) A hybrid approach to monthly streamflow forecasting : integrating hydrological model outputs into a Bayesian artificial neural network. *J Hydrol* 540:623–640
- Islam MMM, Hofstra N, Sokolova E (2018) Modelling the present and future water level and discharge of the tidal Betna River. *Geosciences* 8:271
- Jothiprakash V, Kote AS (2011) Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. *J Hydrol Eng* 16:563–574
- Juston J, Seibert J, Johansson P (2009) Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrol Process* 23(21):3093–3109
- Li C, Bai Y, Zeng B (2016) Deep feature learning architectures for daily reservoir inflow forecasting. *Water Resour Manag* 30(14):5145–5161
- Long JY, Sun Z, Pardalos PM et al (2019) A hybrid multi-objective genetic local search algorithm for the prize-collecting vehicle routing problem. *Inf Sci* 478:40–61
- Mosavi A, Ozturk P, Chau K (2018) Flood prediction using machine learning models: literature review. *Water* 10: 1536
- Myronidis D, Ioannou K, Fotakis D et al (2018) Streamflow and hydrological drought trend analysis and forecasting in Cyprus. *Water Resour Manag* 32(5):1759–1776
- Perrin C, Oudin L, Andreassian V et al (2007) Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. *Hydrol Sci J* 52:131–151
- Pool S, Vis M, Seibert J et al (2018) Evaluating model performance : towards a non- parametric variant of the Kling-Gupta efficiency Kling-Gupta efficiency. *Hydrol Sci J* 63:1941–1953
- Poul AK, Shourian M, Ebrahimi H (2019) A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resour Manag* 33(8):2907–2923
- Qiao W, Lu H, Zhou G et al (2020) A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J Clean Prod* 244:118612
- Sazli MH (2006) A brief review of feed-forward neural networks. *Commun Fac Sci Univ Ankara Ser* 50:11–17
- Sodnik J, Mikoš M (2006) Estimation of magnitudes of debris flows in selected torrential watersheds in Slovenia. *Acta geographica Slov* 46:93–123
- Srivastava S, Lessmann S (2018) A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Sol Energy* 162:232–247
- Su B, Lu S (2017) Accurate recognition of words in scenes without character segmentation using. *Pattern Recogn* 63:397–405
- Su J, Li C, Zeng Q et al (2019) A green closed-loop supply chain coordination mechanism based on third-party recycling. *Sustainability* 11:5335
- Sun L, Nistor I, Seidou O (2015) Streamflow data assimilation in SWAT model using extended Kalman filter. *J Hydrol* 531:671–684
- The MathWorks (2018) MATLAB release 2018a. Natick, Massachusetts, USA
- Vincent P, Larochelle H, Bengio Y, et al (2008) Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25 th international conference on machine learning. Helsinki, Finland, pp 1096–1103
- Wang H, Li G, Wang G et al (2017) Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl Energy* 188:56–70
- Wang W, Chau K, Xu D et al (2015) Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour Manag* 29(8):2655–2675
- Wei N, Li C, Peng X et al (2019) Daily natural gas consumption forecasting via the application of a novel hybrid model. *Appl Energy* 250:358–368
- Xie J, Wang X, Liu Y et al (2018) Autoencoder-based deep belief regression network for air particulate matter concentration forecasting. *J Intell Fuzzy Syst* 34:3475–3486
- Xu X, Zhang X, Fang H et al (2017) A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. *Environ Model Softw* 88:151–167
- Yang T, Asanjan AA, Welles E et al (2017a) Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resour Res* 53:2786–2812
- Yang Y, Zhang M, Zhu L, Liu W, Han J, Yang Y (2017b) Influence of large reservoir operation on water-levels and flows in reaches below dam: case study of the three gorges reservoir. *Sci Rep* 7:15640

- Yuan X, Chen C, Lei X et al (2018) Monthly runoff forecasting based on LSTM–ALO model. *Stoch Environ Res Risk Assess* 32:2199–2212
- Zhang X, Moynihan GP, Ernest ANS et al (2018) Evaluation of the benefits of using a backward chaining decision support expert system for local flood forecasting and warning. *Expert Syst* 35:e12261
- Zhao J, Xu J, Xie X, Lu H (2016) Drought monitoring based on TIGGE and distributed hydrological model in Huaihe River basin, China. *Sci Total Environ* 553:358–365
- Zhao X, Chen X, Xu Y et al (2017) An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* 9:153

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Yun Bai¹ · Nejc Bezak² · Klaudija Sapač² · Mateja Klun² · Jin Zhang³

¹ National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China

² Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, Ljubljana, Slovenia

³ Institute of Groundwater and Earth Sciences, Jinan University, Guangzhou 510632, China