



Daily Runoff Forecasting Using a Cascade Long Short-Term Memory Model that Considers Different Variables

Yun Bai^{1,2} · Nejc Bezak³ · Bo Zeng¹ · Chuan Li¹ · Kloudija Sapač³ · Jin Zhang⁴

Received: 8 July 2020 / Accepted: 29 December 2020 /

Published online: 19 February 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Accurate forecasts of daily runoff are essential for facilitating efficient resource planning and management of a hydrological system. In practice, daily runoff is needed for various practical applications and can be predicted using precipitation and evapotranspiration data. To this end, a long short-term memory (LSTM) under a cascade framework (C-LSTM) approach is proposed for forecasting daily runoff. This C-LSTM model is composed of a 2-level forecasting process. (1) In the first level, an LSTM is established to learn the relationship between the precipitation and evapotranspiration at present and to learn several meteorological variables one day in advance. (2) In the second level, an LSTM is constructed to forecast the daily runoff using the historical and simulated precipitation and evapotranspiration data produced by the first LSTM. Through cascade modeling, the complex features of the numerous targets in the different stages can be sufficiently extracted and learned by multiple models in a single framework. In order to evaluate the performance of the C-LSTM approach, four mesoscale sub-catchments of the Ljubljana River in Slovenia were investigated. The results indicate that based on the root-mean-square error, the Pearson correlation coefficient, and the Nash-Sutcliffe model efficiency coefficient, the proposed model yields better results than two other tested models, including the normal LSTM and other neural network approaches. Based on the results of this study, we conclude that the LSTM under the cascade architecture is a valuable approach and can be regarded as a promising model for forecasting daily runoff.

Keywords Long short-term memory · Cascade framework · Meteorological conditions · Precipitation–evapotranspiration pattern · Daily runoff forecast

✉ Yun Bai
yunbai@foxmail.com

1 Introduction

Runoff forecasting is extremely important for water resource management. Due to the influences of multiple factors, such as precipitation and evapotranspiration, runoff time series exhibit nonlinearity, time-varying characteristics, and indeterminacy, thus making it difficult to accurately forecast runoff (Lei et al. 2014). In order to address this issue, many models, such as data-driven methods, have been established and reported in the literature.

The statistical forecasting model (i.e., the Box–Jenkins model) is a classical method for time series forecasting (Box et al. 2013). This technique can yield a good performance when the forecasting conditions are within the scope of the modeling conditions, but it produces poor results when the conditions are outside or near the limits of the past observations incorporated into the model. Furthermore, the Box-Jenkins model is assumed to follow the assumptions for a stationary univariate process (Wang et al. 2015). Accordingly, it is not suitable for runoff time series modeling since actual hydrological systems are multivariate and non-stationary. Artificial intelligence (AI)-based modeling (Badrzadeh et al. 2015) can be regarded as a promising approach, with characteristics such as adaptive learning, a non-fixed mapping structure, and rapid convergence. For instance, in a comparative study of these methods, AI modeling produced better daily forecasting results than the traditional statistical methods (Sudhishri et al. 2016). Moreover, AI models such as the artificial neural network (Nourani 2017), support vector machine (Zhao et al. 2017), and deep learning model (Bai et al. 2016; Li et al. 2016) have been used on several occasions to simulate the complex characteristics of hydrological systems. These AI models achieved better forecasting results, but time series forecasting remains a bottleneck due to the long-term dependencies. To address this issue, long short-term memory (LSTM), which is one of the most popular types of recurrent neural networks (RNNs), has recently been proposed and employed in various fields (Karim et al. 2018; Srivastava and Lessmann 2018; Xu and Niu 2018). These studies indicate that the LSTM performs better in long time-horizon forecasting than other AI methods. However, only a few studies have investigated the use of LSTM for runoff forecasting (Yuan et al. 2018; Feng et al. 2020; Xiang et al. 2020). For daily rainfall-runoff modeling, different types of models can be applied, such as physically-based models and conceptual models. These types of models usually take into account the main characteristics of rainfall-runoff processes. However, a recent study by Sezen et al. (2019) indicated that selected AI models yield results comparable to those of the tested lumped conceptual model in particular catchments.

Precipitation (P) and evapotranspiration (E) are more closely related to runoff than other variables (Coulibaly et al. 2015) such as wind speed or sunshine duration since P and E are the main processes affecting runoff formation at the catchment scale, i.e., P and E have a dynamic hydrological balance (Berghuijs et al. 2017). Furthermore, it should be noted that P and E are also related. Moreover, several other variables such as soil moisture that can be remotely measured can affect runoff generation. In forecasting, P and E are unknowns, as is daily runoff (Q). Furthermore, P and E forecasts are usually used as inputs in calibrated hydrological models for forecasting Q (Jain et al. 2018).

In many catchments, discharge data are not available and/or where the precipitation stations are sparse the forecasting performance is limited. Even with detailed hydro-meteorological measurements, due to the complexity of catchment processes, the performance of rainfall-runoff models can always be improved and new methods should be tested to enhance the performance of rainfall-runoff modeling. Thus, the cascade long short-term memory model (C-LSTM) is proposed in this study. Using the cascade framework, different feature mappings can

be constructed and transferred into the final target model. Using the LSTM model, the features (i.e., the lags of unknown duration between important events in a time series) hidden in a long-term time series can be captured and learned by the RNN architecture. In this study, the first-level LSTM was established to forecast P and E at time step $t + 1$ ($P(t + 1)$, $E(t + 1)$) using several meteorological variables at time step t . Then, the second-level LSTM was used to forecast $Q(t + 1)$ using $P(t + 1)$ and $E(t + 1)$ values and $P(t)$ and $E(t)$. Moreover, $Q(t)$ was also used in the second-level LSTM. Thus, the short-term correlations and long-term dependencies can be merged in a single cascade forecasting model, which is the main contribution of this study.

To verify the performance of the proposed model, four mesoscale catchments located in the non-homogenous karst Ljubljana River catchment (Slovenia) with different geological characteristics were investigated. It is a well-known fact that model performance decreases with increasing lead time (i.e. forward forecast time) (Jain et al. 2018). The relationship between the decrease in the model's performance and the lead time depends on the catchment characteristics such as size, land use, geological structure, and hydro-meteorological network quality. For example, better modeling results can be obtained for longer lead times in larger catchments than in smaller catchments with torrential characteristics (Bai et al. 2019). In the case of mesoscale catchments, one- or two-day lead times are important since the forecasting performance for longer lead times can be questionable. The most relevant lead time can be determined based on the autocorrelation analysis purely from the time series perspective. Furthermore, the forecasting lead time should be long enough to allow the authorities and affected population to respond to a potential disaster. Thus, the lead time basically represents the forecasting forward time. In the case of mesoscale catchments, use of hourly data would be more meaningful for effective flood forecasting. Such data are often unavailable and the daily time step can be regarded as sufficient for applications such as reservoir inflow modeling, water supply modeling, and climate change impact modeling (Sapač et al. 2019; Sezen et al. 2020).

2 Methodology

2.1 Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)

A recurrent neural network (RNN) is a type of artificial neural network that consists of an input layer, a hidden layer, and an output layer. There are two differences between an RNN and a traditional network such as feedforward neural network (FFNN) (Su and Lu 2017). (1) In the same hidden layer there are connections between the nodes in an RNN, whereas in the FFNN there are none. (2) The inputs of the hidden layer at the current time contain both the input layer at the current time and the hidden layer at the previous time. The special structure of the RNN (Fig. 1(a)) allows for a better description of the temporal dynamic behavior, because it uses the previous information it has learned to model the pattern of the current step, which is beneficial for sufficiently exploring the features of the current time series. Therefore, in this study, an RNN with a memory function was investigated and applied in the time series forecasting.

In practice, an RNN cannot maintain a good memory if the time interval is large and has a vanishing gradient problem (Gers et al. 2000). Therefore, various improved RNN models have been introduced such as an LSTM with the simple structure (Fig. 1(b)) that has been widely

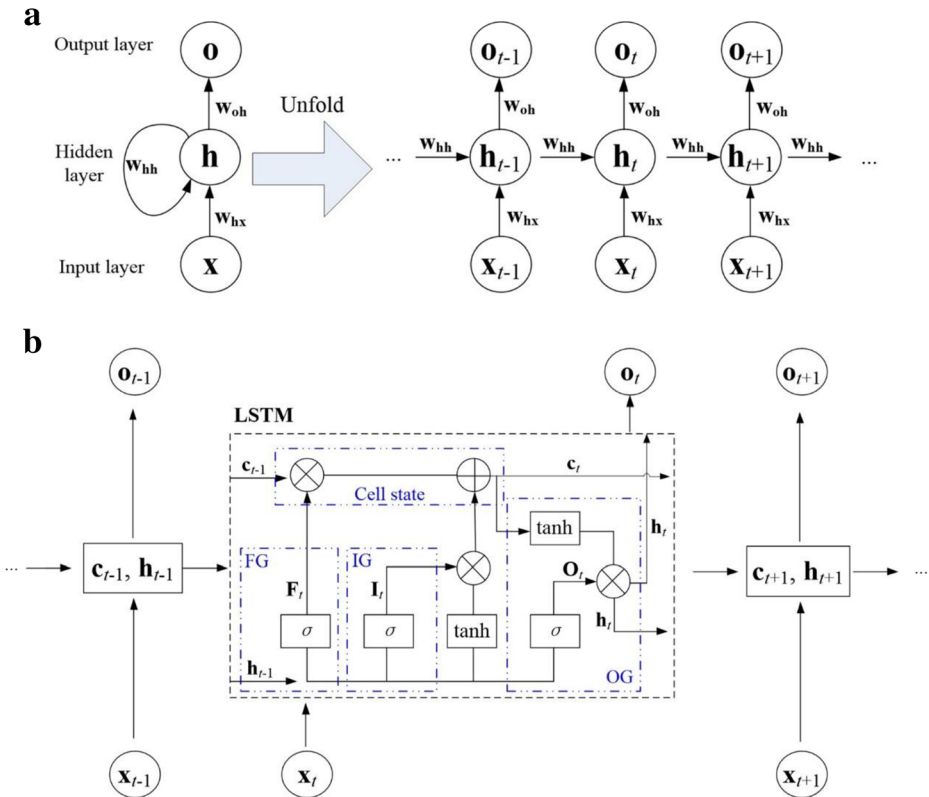


Fig. 1 Unfolded basic RNN structure (a) RNN: x : input nodes; h : hidden nodes; o : output nodes; w : shared parameters in each layer and LSTM unit with peephole connection (b) LSTM

utilized for time series forecasting in various fields. The LSTM computes its memory units using different activation functions. However, thus far, its application to hydrological data has been limited.

The LSTM unit with peephole connections includes an input gate (IG), an output gate (OG), and a forget gate (FG) (Fig. 1(b)). By using this special interactive operation among three gates with a memory cell (c in Fig. 1 (b)), which serves as an accumulator of the cell state, the LSTM can mitigate the vanishing gradient effect of long-term dependencies. The computing process of the LSTM with peephole connections is briefly described below (Gers et al. 2000).

- (1) For the t -th time step, the information of the new inputs \mathbf{x}_t will be accumulated to the cell if the IG is activated as a sigmoid function $\mathbf{I}_t = \sigma(\mathbf{w}_{xi}\mathbf{x}_t + \mathbf{w}_{hi}\mathbf{h}_{t-1} + \mathbf{w}_{ci} \times \mathbf{c}_{t-1} + \mathbf{b}_i)$.
- (2) At the same time, the FG evaluates which information to eliminate from the previous cell state, using $\mathbf{F}_t = \sigma(\mathbf{w}_{xf}\mathbf{x}_t + \mathbf{w}_{hf}\mathbf{h}_{t-1} + \mathbf{w}_{cf} \times \mathbf{c}_{t-1} + \mathbf{b}_f)$.
- (3) The old cell status \mathbf{c}_{t-1} will be updated to the new state $\mathbf{c}_t = \mathbf{F}_t \times \mathbf{c}_{t-1} + \mathbf{I}_t \times \tanh(\mathbf{w}_{xc}\mathbf{x}_t + \mathbf{w}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c)$.
- (4) The updated cell state \mathbf{c}_t is passed through “tanh” function and multiplied by the sigmoid activation function of the OG to determine the final output from LSTM unit \mathbf{h}_t . This is expressed as $\mathbf{h}_t = \mathbf{O}_t \times \tanh(\mathbf{c}_t)$, where $\mathbf{O}_t = \sigma(\mathbf{w}_{xo}\mathbf{x}_t + \mathbf{w}_{ho}\mathbf{h}_{t-1} + \mathbf{w}_{co} \times \mathbf{c}_t + \mathbf{b}_o)$.

In the steps above, \mathbf{w}_{xi} , \mathbf{w}_{yf} , \mathbf{w}_{xo} are the input weights; \mathbf{w}_{hi} , \mathbf{w}_{hf} , \mathbf{w}_{ho} denote the recurrent weights; \mathbf{w}_{ci} , \mathbf{w}_{cf} , \mathbf{w}_{co} represent the peephole weights; \mathbf{b}_i , \mathbf{b}_f , \mathbf{b}_c , \mathbf{b}_o are the bias vectors; and “ \times ” represents point-wise multiplication. In this way (the cell state accumulating activities over time), the LSTM can discover the long-term features.

2.2 Present Model: Cascade LSTM (C-LSTM)

Recently, the LSTM approach has been demonstrated to be effective at handling temporal correlations (Li et al. 2017; Liang et al. 2020), although in practice some limitations inevitably exist. For instance, the target value at the current time t is not only related to the variables at previous times (e.g., $t-1$), but it is also connected to the variables at the current time t . However, the variables at the current time are nonexistent in reality since they coexist with the target value to be forecasted (Q in this case study). To address this issue, cascade modeling was utilized in this investigation. The cascade model is composed of many sub-models, which are independent and complementary in feature extraction and mapping. The C-LSTM has k levels, and the LSTM model is applied for the time series forecasting in each level using the different input variables, which are composed of the learning results from the previous level and the corresponding new inputs (Fig. 2). Using the cascade architecture, the mixed time series characteristics are identified at different levels, which may effectively reject vague patterns.

By combining this theory and Fig. 2, the proposed method can be summarized as follows:

- (1) Collect and rescale the original data.
- (2) Divide the data into training (90%) and testing (10%) data.
- (3) Train the C-LSTM model. The training data are divided into k categories according to pattern identification based on the research target. For example, first the $[P, E]$ are forecasted through other variables (LSTM 1), and then the forecasted values ($[P, E]$) are used to forecast Q (LSTM 2). Because there are two sub-targets, the dataset identifies two patterns.
- (4) Test the trained C-LSTM model by dividing the testing data as in the previous step and forecasting the target variable.

3 Case Study

3.1 Study Area and Dataset

The Ljubljana River catchment is part of the Sava River catchment, which drains into the Danube River. In this study, four mesoscale sub-catchments located in the larger Ljubljana River catchment were examined (Fig. S1). The basic properties of the investigated catchments and a list of the stations used for the specific catchments are presented in Table 1. Some of these catchments have already been investigated in previous studies (Bezak et al. 2017; Sezen et al. 2019; Rusjan et al. 2019).

In this study, the following variables were used and a daily time step was selected: runoff (Q), precipitation (P), air temperature (T), evapotranspiration (E), wind speed (WS), sunshine duration (SD), saturation vapor pressure deficit (SVP), and relative humidity (RH). These variables were selected based on the data availability and because they have at least a minor

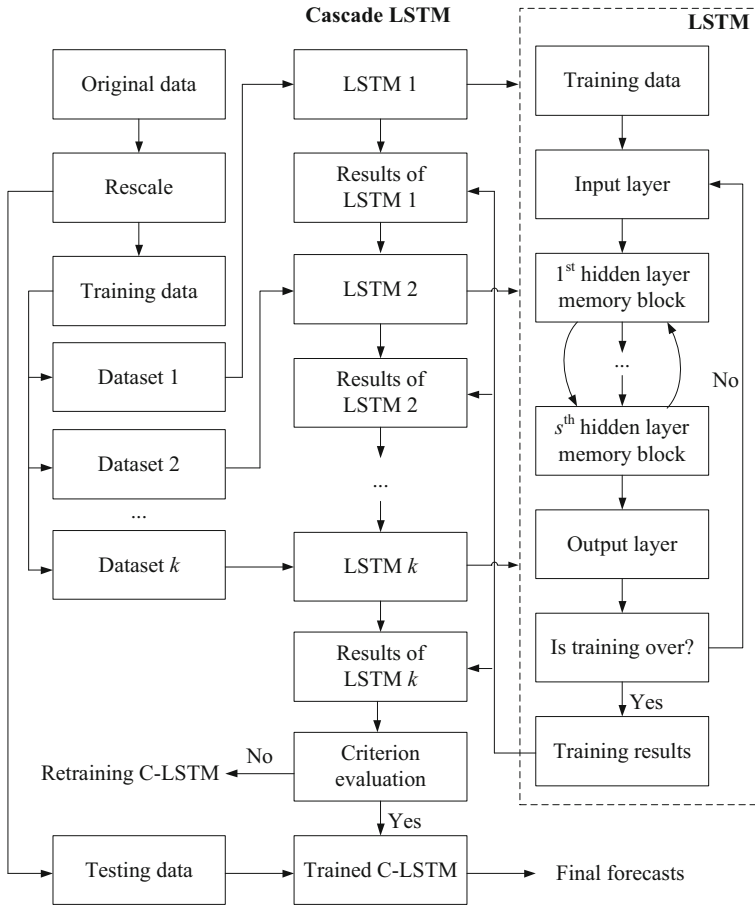


Fig. 2 Flow chart of the proposed Cascade LSTM approach (C-LSTM)

connection to runoff generation. For the meteorological variables (T , E , WS , SD , SVP , and RH) Ljubljana and Postojna stations were used since the necessary data were only available from these stations. More information about the SVP calculations and the impact of the investigated meteorological variables on the evapotranspiration can be found in Maček et al. (2018). A total of 17 years’ of data (2000–2016) were utilized for the model calibration. The statistical

Table 1 Main properties of the investigated catchments and selected discharge stations (located at the outlet of the catchment as shown in Fig. S1), precipitation, and meteorological stations (evapotranspiration, temperature, wind speed, sunshine duration, relative humidity, and saturation vapour pressure deficit data)

| River | Discharge gauging station | Catchment area [km ²] | Min/Mean/Max catchment elevation [m.a.s.l.] | Mean catchment slope [°] | Precipitation station | Meteorological station |
|-------------|---------------------------|-----------------------------------|---|--------------------------|-----------------------|------------------------|
| Cerkniščica | Cerknica | 50 | 562/712/1113 | ~9 | Cerknica | Postojna |
| Gradaščica | Dvor | 79 | 343/616/1020 | ~16 | Postojna | Postojna |
| Nanoščica | Mali Otok | 51 | 516/578/993 | ~5 | Črni vrh | Ljubljana |
| Šujica | Razori | 47 | 299/425/880 | ~11 | Šentjošt | Ljubljana |

properties of these data are summarized in Table 2 in order to present the main hydro-meteorological characteristics of the investigated area.

3.2 Experiment Design

According to the modeling procedure, all of the data were rescaled into $[0, 1]$ and were divided into training (2000–2014, total of 5479 days) and testing (2015–2016, total of 731 days) subsets.

The daily Q is closely linked to the daily P and E , but the forecasted value and the values of these variables for the same day are unknown. Hence, P and E should be estimated before Q modeling. Furthermore, other meteorological variables affect P and E (Maček et al. 2018). It should be noted that the correlation between P and E is negative and rather weak with a Pearson correlation coefficient of -0.11 to 0.13 . Therefore, in this study, the proposed C-LSTM model was composed of two forecasting system levels ($k=2$) (Fig. 2). Specifically, the mapping of the first LSTM level was between the meteorological variables (T , WS , RH , SD , and SVP) and P and E , while the input-output structure of the second LSTM level was the pattern $([P, E] \rightarrow Q)$. Table 3 presents the input-output structure of the C-LSTM according to Fig. 2. For comparison, the basic LSTM model and a typical network (i.e., FFNN) are also listed in Table 3. Two input scenarios were designed for a single application of the LSTM to investigate the influence of the other meteorological variables on the model's performance. That is, LSTM (I) considers all of the variables at t time which is in line with the variables of the C-LSTM, whereas, the LSTM (II) only considers Q , P , and E at t time which is in line with the variables of the second level of C-LSTM (LSTM2).

In addition, a convolution LSTM network was constructed in this study (i.e., $s=2$, Fig. 2). The two hidden layers contain LSTM neurons for exploring temporal dependencies. The other layer consists of normal neurons for regression, and it makes use of the temporal features calculated in the previous layers and provides the final forecast. Based on the pre-experiments (i.e., trial and error method, in which the number of neurons in each hidden layer is set to be 5, 10, and 20), the first and second LSTM layers contained 20 and 10 neurons, respectively. The other computation parameters were set as mini-batches with 30 size, 100 epochs, and training rate of 0.05 at the beginning (scaling ratio of 0.1 with a drop period of 30).

The root-mean-square error (RMSE), the Pearson correlation coefficient (R), and the Nash-Sutcliffe model efficiency coefficient (NSE) were employed to evaluate the models' performance.

Table 2 Statistical properties of the variables used in this study

| Variables | Maximum | Mean | Minimum |
|-------------|-------------------------|---------------------|---------------------|
| Q (mm) | 47.2/61.6/107.8/38.4 | 1.6/2.5/2.2/2.4 | 0/0/0/0 |
| P (mm) | 144.2/178.6/130.8/132.4 | 4.2/4.4/4.6/4.0 | 0/0/0/0 |
| E (mm) | 7.3/7.7/7.7/7.3 | 2.2/2.2/2.2/2.2 | 0/0/0/0 |
| T (°C) | 27.7/29.6/29.6/27.7 | 10.0/11.6/11.6/10.0 | -11/-10.8/-10.8/-11 |
| WS (m/s) | 11.8/7.7/7.7/11.8 | 2.5/1.3/1.3/2.5 | 0/0/0/0 |
| RH (%) | 99/100/100/99 | 76.3/74.7/74.6/76.2 | 29/31/31/29 |
| SD (h) | 14/15/15/14 | 5.4/5.4/5.4/5.4 | 0/0/0/0 |
| SVP (kPa) | 2.3/2.6/2.6/2.3 | 0.5/0.5/0.5/0.5 | 0.01/0.02/0.02/0.01 |

Label “/” is used to separate stations Cerknjšica, Gradašica, Šujica and Nanošica respectively

Table 3 Input-output structure of different models. The variable with the caret “^” means the forecasted value

| Model | Input variables | Output |
|-----------|---|---|
| C-LSTM | LSTM1: $T(t)$, $WS(t)$, $RH(t)$, $SD(t)$, $SVP(t)$, $P(t)$, $E(t)$ LSTM2: $\hat{P}(t+1)$, $\hat{E}(t+1)$, $Q(t)$, $P(t)$, $E(t)$ | $\hat{P}(t+1)$, $\hat{E}(t+1)$ $\hat{Q}(t+1)$ |
| LSTM (I) | $T(t)$, $WS(t)$, $RH(t)$, $SD(t)$, $SVP(t)$, $P(t)$, $E(t)$, $Q(t)$ | $\hat{Q}(t+1)$ |
| LSTM (II) | $Q(t)$, $P(t)$, $E(t)$ | $\hat{Q}(t+1)$ |
| FFNN | $T(t)$, $WS(t)$, $RH(t)$, $SD(t)$, $SVP(t)$, $P(t)$, $E(t)$, $Q(t)$ | $\hat{Q}(t+1)$ |

4 Results and Discussion

Based on the experimental design (Table 3), P and E were modeled using the first-level LSTM. The validation results (Fig. 3) suggest that the LSTM approach was able to model the P and E dynamics. Moreover, the R value between the modelled and observed precipitation ranged from 0.60 to 0.67 while R between the modelled and measured evapotranspiration data ranged from 0.90 to 0.92. Thus, a relatively suitable model performance can be obtained. Due to the precipitation generation characteristics, some deviations between the modeled and observed values did occur in the precipitation forecasts, meaning that the next-day P often had a limited connection to the previous-day P . This is especially evident in areas with frequent thunderstorms, which is also the case in the studied area with a temperate continental climate where thunderstorms with high erosive power frequently occur. This sort of modeling results was expected because evapotranspiration is influenced by the selected input variables such as T or SVP and thus, it generally exhibits more significant seasonal characteristics, with higher values in the summer and lower values in winter, while the day-to-day variations are smaller (Fig. 4).

Using the forecasted P and E , the second-level LSTM was used to forecast the daily Q . (Table 3, Figs. 4, S2-S4).

Figure 4 suggests that the second-level LSTM model can sufficiently learn the relationship between the inputs ($Q(t)$, $P(t)$, $E(t)$, $P(t+1)$, and $E(t+1)$) and the output ($Q(t+1)$). As can be seen from Fig. 4, the model was able to satisfactorily reproduce the discharge dynamics for all four catchments. In addition, the scatter plots exhibited a high fitting degree ($R=0.92-0.95$) between the observations and the forecasts, but Šujica (Fig. S3) displayed deviations between the 200th and 350th day. It should be noted that the investigated catchments are geologically nonhomogeneous. The geology of the Gradašćica and Šujica catchments is dolomite and incomplete karst. However, the Nanošćica and Cerknjšćica catchments are located in areas with impermeable surfaces and underground flow through the karst edges and the high karst area in Notranjska (Sezen et al. 2019). In spite of these differences, the C-LSTM model was able to adequately forecast the daily runoff.

Furthermore, in order to study the distributions of the residual error (absolute error), violin plots of the absolute errors are plotted (Fig. 5). The absolute errors of C-LSTM model were characterized by a quasi-normal distribution with a mean near 0. The ranges of the main errors were $[-5, 7]$, $[-7, 6]$, $[-9, 7]$, and $[-6, 8]$ for the Cerknjšćica, Gradašćica, Šujica, and Nanošćica catchments, respectively. The margin of the cumulative contributions beyond 0.9 was $[-2, 2]$.

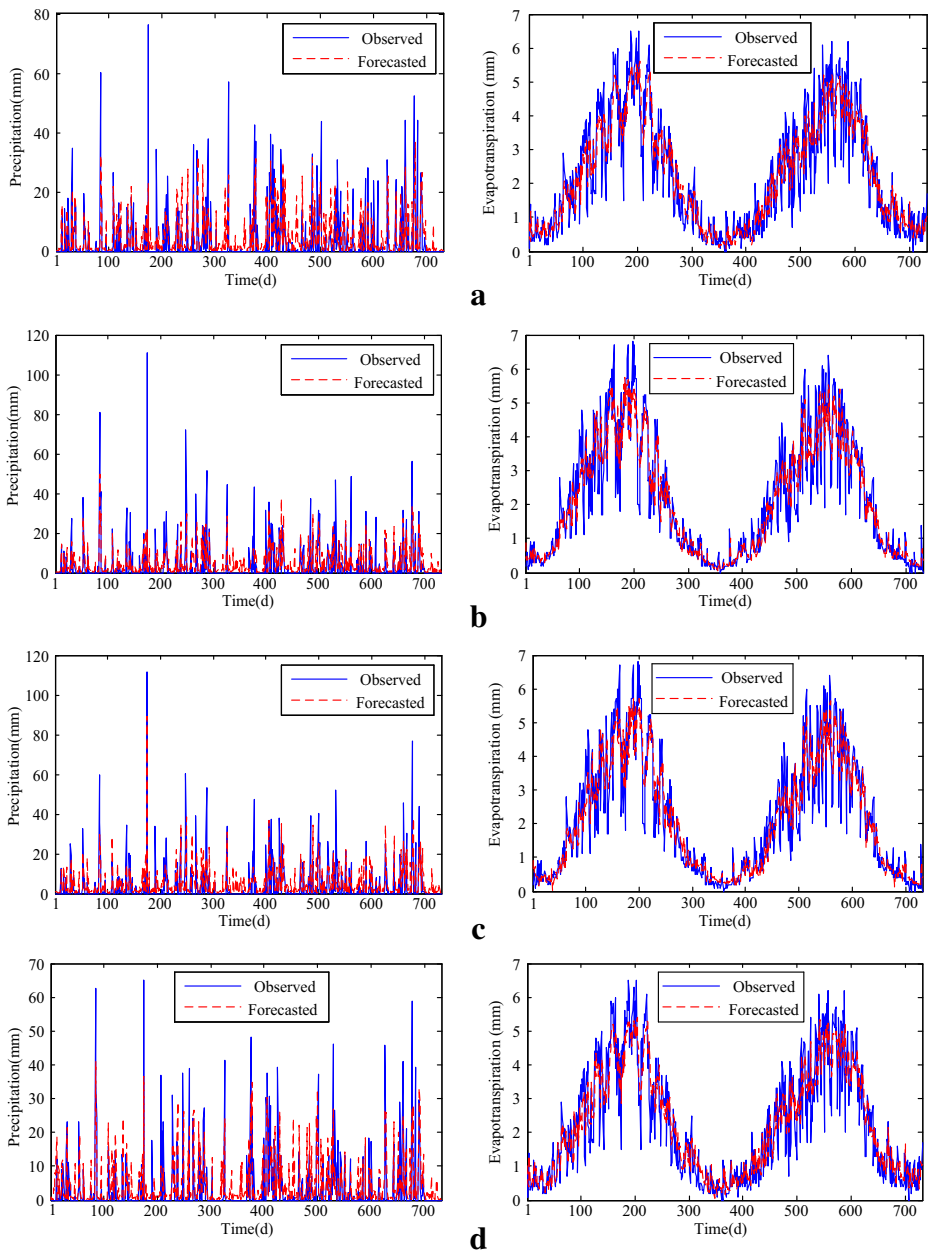


Fig. 3 Results of the precipitation (left) and evapotranspiration (right) forecasted using the first level LSTM for the validation period for four stations. **a** Cerknjiščica **(b)** Gradaščica **(c)** Šujica **(d)** Nanoščica

The qualitative analysis results are supplemented by the quantitative evaluation results. The RMSE of C-LSTM model for four stations range from 0.79 to 1.47 mm, and the NSE ranges from 0.85 to 0.90 (Table 4). The comparison of the efficiency criteria results of the C-LSTM model and the lumped conceptual rainfall-runoff model reveals that in most cases the C-LSTM model yielded better results. However, it should be noted that different data lengths were used

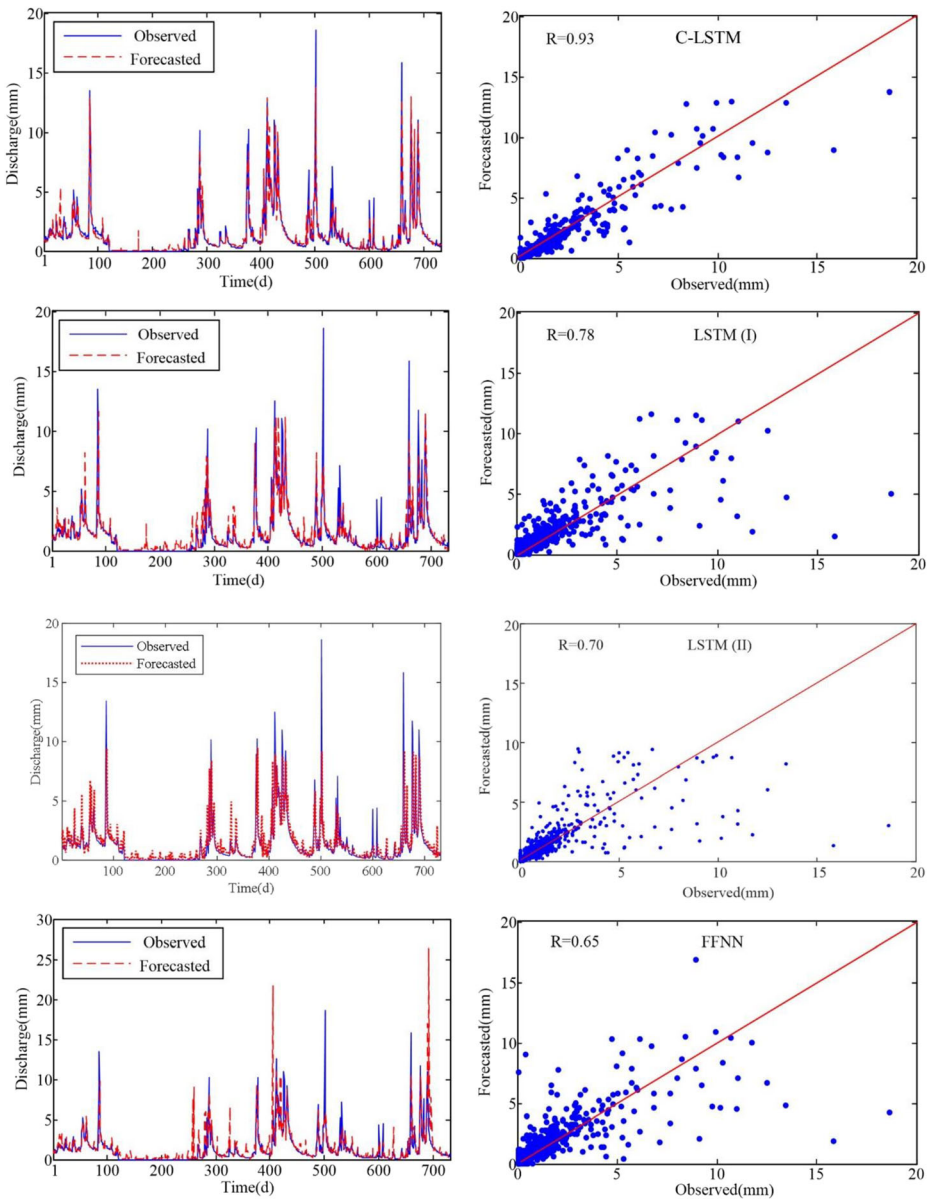


Fig. 4 Results of the runoff forecasted using C-LSTM, LSTM (I), LSTM (II), and FFNN models for the Cerknjišica station. Results of other stations can be found in the Supplement Fig. S2–S4

for the calibration and validation in these studies (e.g., Sezen et al. (2019) used 5 years' of data for validation). The rainfall-runoff process is characterized by a lag time (Jain et al. 2018), which in the case of mesoscale catchments, can vary from a few hours to 1 or 2 days. Therefore, in terms of $Q(t+1)$ forecasting, in some cases $P(t)$ is more important than $P(t+1)$. In such cases the autocorrelation analysis (e.g., applying the autocorrelation test before the modelling) can reveal the most suitable time steps (i.e., lag times) that need to be considered in

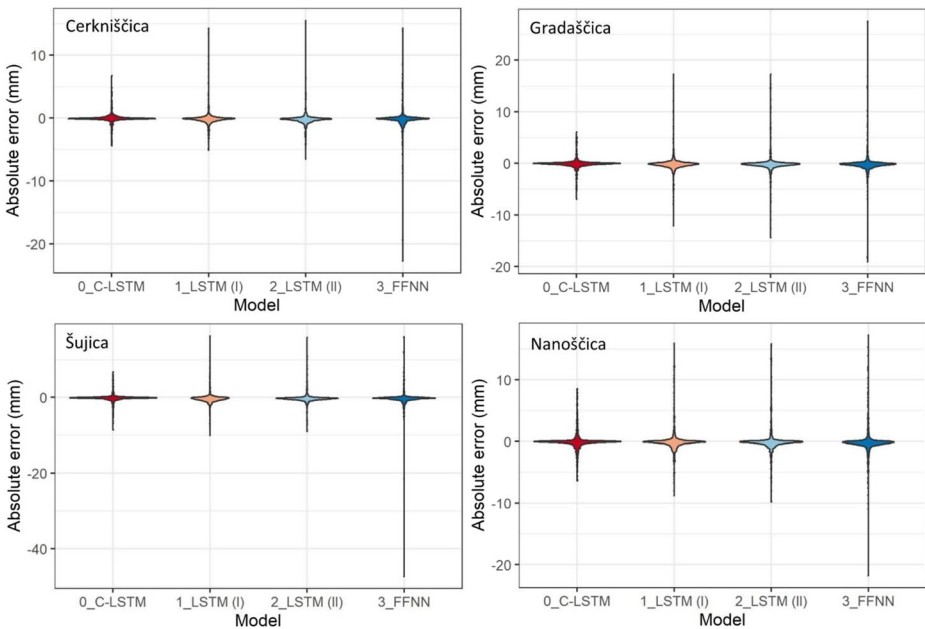


Fig. 5 Distribution of absolute errors of C-LSTM, LSTM (I), LSTM (II), and FFNN models for four stations

such models to obtain as optimal a model performance as possible. Moreover, other variables can improve the modeling performance to some extent, and therefore, they were included in the scope of this study.

To estimate the forecasting performance of the C-LSTM approach, the LSTM with two input scenes and the FFNN model (Table 3) were used for a comparison study using the same dataset. The computing parameters of both LSTM models (LSTM (I), LSTM (II)) were the same as the C-LSTM settings. The FFNN was defined as follows: hidden layers = 1, hidden nodes = 10, learning rate = 0.02, epochs = 500, and goal = 0.0001. Based on these parameter settings, the forecasting results for Cerknjščica catchment are displayed in Fig. 5 and for other stations in Fig. S2–S4 for all models, and their absolute error distributions are shown in Fig. 5.

The results for the four cases reveal that LSTM (I) can capture the dynamics of the daily runoff for the validation period, but it failed at some peak discharge values (e.g., Cerknjščica on the 500th day). Moreover, the scatter plots indicate that the forecasted values did not fit the

Table 4 Comparison of the forecasting performances using different models

| Model | C-LSTM | | | LSTM (I) | | |
|-------------|-----------|------|------|-----------|------|------|
| Case | RMSE (mm) | R | NSE | RMSE (mm) | R | NSE |
| Cerknjščica | 0.79 | 0.93 | 0.85 | 1.31 | 0.78 | 0.60 |
| Gradašćica | 1.02 | 0.93 | 0.87 | 1.89 | 0.75 | 0.54 |
| Šujica | 1.07 | 0.92 | 0.85 | 1.53 | 0.84 | 0.68 |
| Nanošćica | 1.47 | 0.95 | 0.90 | 2.04 | 0.90 | 0.80 |
| Model | LSTM (II) | | | FFNN | | |
| Case | RMSE (mm) | R | NSE | RMSE (mm) | R | NSE |
| Cerknjščica | 1.49 | 0.70 | 0.49 | 1.80 | 0.65 | 0.24 |
| Gradašćica | 2.13 | 0.68 | 0.42 | 2.40 | 0.60 | 0.26 |
| Šujica | 1.60 | 0.82 | 0.66 | 2.43 | 0.65 | 0.20 |
| Nanošćica | 2.34 | 0.86 | 0.74 | 2.46 | 0.84 | 0.71 |

observations for values >10 mm well (e.g., underestimation of the LSTM model for high flows occurred for the Cerknjšica catchment). The results of LSTM (II) show that although it gets good results, its performance decreases as the input variables decrease. The comparison with LSTM (I) reveals that other meteorological factors (T , WS , RH , SD , and SVP) should also be considered in the modeling. The FFNN was also unable to successfully model the peak discharge values. The dispersion degree shown in the scatter plot of the observed and modeled values was even greater than that of the LSTM structure, illustrating that the shallow structure has lower capacity than the LSTM model.

All of the tested models reproduced the runoff dynamics (Figs. 4, S2–S4), although the FFNN model had issues with the peak discharge modeling. The dispersion degree for the LSTM (I) and LSTM (II) models was also larger than that of the C-LSTM model, especially for the peak values. Additionally, the C-LSTM model had smaller errors compared to other models, for which the distribution of the main errors extended to greater than 10 mm (Fig. 5). Thus, the C-LSTM model also exhibits a superior performance for daily runoff forecasting, demonstrating that it has the potential to enhance runoff modeling performances.

According to the quantitative evaluations (Table 4), the deep learning architecture methods (C-LSTM and LSTM (I, II)) had lower RMSE and higher R and NSE values than those of the shallow learning method (FFNN), illustrating the superiority of the deep networks. The LSTM (I) that considered all of the meteorological factors had a better performance than the LSTM (II) model, demonstrating that the extra inputs did not decrease the forecasting performance. Moreover, the proposed C-LSTM model outperformed the LSTM (I), LSTM (II), and FFNN models, as indicated by its lower RMSE and higher R and NSE values. The relative performance ranking is C-LSTM (best), followed by LSTM (I), LSTM (II), and FFNN (worst), revealing that the LSTM under a cascade framework can maximize the data using different mappings in a single model, that is, the information is used to its fullest potential and the features can be deeply explored. Therefore, the C-LSTM is beneficial for synchronously learning the sophisticated features of the target variable and its influencing factors, and thus, it exhibits a better capacity for daily runoff forecasting. Moreover, according to Morisai et al. (2015), all NSE values >0.75 can be regarded as “excellent” modeling results for daily time steps, which was the case in this study. This supports the arguments above regarding the performance of the C-LSTM.

5 Conclusions

Considering the synchronous effects of precipitation and evapotranspiration, in this study, two-leveled cascade long short-term memory (C-LSTM) model was applied for daily runoff forecasting using the data from four mesoscale sub-catchments. The first LSTM level was used to simulate the precipitation and evapotranspiration on the current day. Then, these values were used as inputs for the second LSTM level, which was established for the daily runoff forecasting. The C-LSTM model was demonstrated to have a powerful feature learning ability, and it achieved a high forecasting accuracy relative to the results of other methods and in terms of the three quantitative indices. In summary, the C-LSTM model integrates the fractional modeling capacity of the cascade framework with the recurrent deep learning ability of the LSTM, providing it with the capacity to extract and learn coupled features affected by multiple factors, thus improving its forecasting performance.

The novelty of this study was the use of a cascade framework for different sub-task modeling, which synchronously considers the dynamic changes in the variables and their effects on the daily runoff. In addition, the LSTM considers both short- and long-term dependencies, and thus, it has strong feature learning and time series modeling capabilities, making this a novel attempt in the field of daily runoff forecasting. The results of this investigation indicate that the C-LSTM model, which has not yet been frequently applied in hydrology, demonstrates the ability to reproduce the hydrological characteristics (i.e., runoff) on a daily time scale. Thus, the proposed model can be applied to other hydrological applications in order to improve the modeling performance.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11269-020-02759-2>.

Acknowledgements This work is supported in part by the National Natural Science Foundation of China (71801044, 71771033), the international cooperation of the Ministry of Science and Technology of China (12-24: bilateral project between China and Slovenia), the China Scholarship Council (201908500020), and the Natural Science Foundation of Chongqing (cstc2018jcyjAX0436). N. Bezak gratefully acknowledges funding by the Slovenian Research Agency (grants J2-7322 and P2-0180). We also would like to thank the Slovenian Environment Agency for data provision (<http://www.arso.gov.si/>) and LetPub (www.letpub.com) for its linguistic assistance of this manuscript.

Authors Contributions Conceptualization: Y.B.; Methodology: Y.B. and C.L.; Formal analysis and investigation, Y.B. and N.B.; Writing - original draft preparation: Y.B. and N.B.; Writing - review and editing: K.S. and J.Z.; Funding acquisition: Y.B., N.B., and B.Z.; Resources: N.B.

Funding National Natural Science Foundation of China (71801044, 71771033), International Cooperation of the Ministry of Science and Technology of China (12-24: bilateral project between China and Slovenia), China Scholarship Council (201908500020), Natural Science Foundation of Chongqing (cstc2018jcyjAX0436), and Slovenian Research Agency (grants J2-7322 and P2-0180).

Data Availability Slovenian Environment Agency (<http://www.arso.gov.si/>).

Compliance with Ethical Standards

Ethical Approval and Consent to Participate This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to Publish Not applicable.

Competing Interests None.

References

- Badrzadeh H, Sarukkalige R, Jayawardena AW (2015) Hourly runoff forecasting for flood risk management: application of various computational intelligence models. *J Hydrol* 529:1633–1643
- Bai Y, Chen Z, Xie J, Li C (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *J Hydrol* 532:193–206
- Bai Y, Bezak N, Sapač K, Klun M, Zhang J (2019) Short-term streamflow forecasting using the feature-enhanced regression model. *Water Resour Manag* 33(14):4783–4797
- Berghuijs WR, Larsen JR, Emmerik T, Woods RA (2017) A global assessment of runoff sensitivity to changes in precipitation, potential evaporation, and other factors. *Water Resour Res* 53(10):8475–8486

- Bezak N, Grigillo D, Urbančič T, Mikoš M, Petrovič D, Rusjan S (2017) Geomorphic response detection and quantification in a steep forested torrent. *Geomorphology* 291:33–44
- Box GEP, Jenkins GM, Reinsel GC (2013) *Time series analysis: forecasting and control*. Wiley, Hoboken
- Coulibaly P, Anctil F, Rasmussen P, Bobée B (2015) A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrol Process* 14(15):2755–2777
- Feng D, Fang K, Shen C (2020) Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resour Res* 56(9). <https://doi.org/10.1029/2019WR026793>
- Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. *Neural Comput* 12(10):2451–2471
- Jain S, Mani S, Jain SK, Prakash P, Singh VP, Tullos D, Kumar S, Agarwal SP, Dimri AP (2018) A brief review of flood forecasting techniques and their applications. *Int J River Basin Manage* 16(3):329–344
- Karim F, Majumdar S, Darabi H, Chen S (2018) LSTM fully convolutional networks for time series classification. *IEEE Access* 6:1662–1669
- Lei F, Huang C, Shen H, Li X (2014) Improving the estimation of hydrological states in the SWAT model via the ensemble Kalman smoother: synthetic experiments for the Heihe River basin in Northwest China. *Adv Water Resour* 67:32–45
- Li C, Bai Y, Zeng B (2016) Deep feature learning architectures for daily reservoir inflow forecasting. *Water Resour Manag* 30(14):5145–5161
- Li X, Peng L, Yao X, Cui S, Hu Y, You C, Chi T (2017) Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. *Environ Pollut* 231:997–1004
- Liang Z, Zou R, Chen X, Ren T, Su H, Liu Y (2020) Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *J Hydrol* 581:124432
- Maček U, Bezak N, Šraj M (2018) Reference evapotranspiration changes in Slovenia, Europe. *Agric For Meteorol* 260–261:183–192
- Morisai DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation criteria. *Trans ASABE* 58:1763–1785
- Nourani V (2017) An emotional ANN (EANN) approach to modeling rainfall-runoff process. *J Hydrol* 544:267–277
- Rusjan S, Sapač K, Petrič M, Lojen S, Bezak N (2019) Identifying the hydrological behavior of a complex karst system using stable isotopes. *J Hydro* 577:123956
- Sapač K, Medved A, Rusjan S, Bezak N (2019) Investigation of low- and high-flow characteristics of karst catchments under climate change. *Water* 11:925
- Sezen C, Bezak N, Bai Y, Šraj M (2019) Hydrological modelling of karst catchment using lumped conceptual and data mining models. *J Hydrol* 576:98–110
- Sezen C, Šraj M, Medved A, Bezak N (2020) Investigation of rain-on-snow floods under climate change. *Appl Sci* 10:1242
- 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 recurrent neural network. *Pattern Recogn* 63:397–405
- Sudhishri S, Kumar A, Singh JK (2016) Comparative evaluation of neural network and regression based models to simulate runoff and sediment yield in an outer Himalayan watershed. *J Agric Sci Technol* 18(3):681–694
- Wang WC, Chau KW, Qiu L, Chen Y (2015) Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environ Res* 139:46–54
- Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour Res* 56:e2019WR025326
- Xu S, Niu R (2018) Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in three gorges area, China. *Comput Geosci* 111:87–96
- Yuan X, Chen C, Lei X, Yuan Y, Adnan RM (2018) Monthly runoff forecasting based on LSTM-ALO model. *Stoch Environ Res Risk Assess* 32(8):2199–2212
- Zhao X, Chen X, Xu Y, Xi D, Zhang Y, Zheng X (2017) An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* 9(3):153

Affiliations

Yun Bai^{1,2} · Nejc Bezak³ · Bo Zeng¹ · Chuan Li¹ · Kladija Sapač³ · Jin Zhang⁴

¹ School of Management Science and Engineering, Chongqing Technology and Business University, Chongqing 400067, China

² Faculty of Science and Technology, University of Algarve, 8005139 Faro, Portugal

³ Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, Ljubljana, Slovenia

⁴ Institute of Groundwater and Earth Sciences, Jinan University, Guangzhou 510632, China