https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

Using LSTM to monitor continuous discharge indirectly with electrical conductivity observations

Yong Chang[1], Benjamin Mewes[2], and Andreas Hartmann[3,4]

[1] School of Earth Science and Engineering, Hohai University, Nanjing 210098, China..

[2] Ruhr-University Bochum, Institute of Hydrology, Water Resources and Environmental Engineering, Bochum, Germany.

[3] Institute of Groundwater Management, Technical University of Dresden, Dresden, Germany.

[4] Chair of Hydrological Modeling and Water Resources, Freiburg University, Freiburg, 79098, Germany.

Corresponding author: Yong Chang (wwwkr@163.com)

Abstract

Due to EC's easy recordability and the existence of a strong correlation between EC (electrical conductivity) and discharge in certain catchments, EC is a potential predictor of discharge. This potential has not yet to be widely addressed. In this paper, we investigate the feasibility of using EC as a proxy for long-term discharge monitoring in a small karst catchment where EC always shows a negative correlation with the spring discharge. Given their complex relationship, a special machine learning architecture, LSTM (Long Short Term Memory), was used to handle the mapping from EC to discharge. LSTM results indicate that the spring discharge can be predicted well with EC, particularly in storms when the dilution dominates the EC dynamic; however, the prediction may have relatively large uncertainties in the small or middle recharge events. A small number of discharge observations are sufficient to obtain a robust LSTM for the long-term discharge prediction from EC, indicating the practicality of recording EC in ungauged catchments for indirect discharge monitoring. Our study also highlights that the random or fixed-interval discharge measurement strategy, which covers various climate conditions, is more informative for LSTM to give robust predictions than other strategies. While our study is implemented in a karst catchment, the method may be also suitable for non-karst catchments where there is a strong correlation between EC and discharge.

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions

1 Introduction

The measurement of streamflow is crucial for hydrologists and hydraulic engineers since it is the fundamental data for estimating the hydrology cycle, water resource management, the design and operation of water projects. For continuous monitoring of streamflow, depth is often recorded continuously by an automatic instrument and translated into discharge based on a defined relationship. The most convenient way is to build a standard hydraulic structure, e.g. weirs or flumes, and the discharge can be easily calculated from the depth based on the theoretical hydraulic equations (*Boiten*, 1993). The establishment of these structures is often laborious and costly, which limits their application. Another common approach is to establish the stage-–discharge curve of the natural channel based on historical observations (*Herschy*, 1995; *Turnipseed and Sauer*, 2010). However, some natural stream beds are not always regular and may change dramatically, especially in mountain areas, due to turbulent erosion and deposition of the sediments (*Weijs et al.*, 2013). This would lead to strong variations in the rating curve and bring a huge uncertainty to discharge estimation.

Instead of depth, electrical conductivity (EC) as a bulk parameter representing overall content of ions in the water may also be a potential discharge predictor. As well as being easy to record, EC has often been observed in many catchments to have a strong correlation with discharge (*Cano-paoli et al.*, 2019; *Dzikowski and Jobard*, 2012; *Gurnell and Fenn*, 1985). Several studies have already discussed the potential of using EC to estimate the spring discharge. For example, *Weijs et al*. (2013) investigated the potential of EC to predict discharge in alpine watersheds and found the EC–streamflow relationship even slightly outperforms the stage–discharge relationship. *Cano-paoli et al.* (2019) presented a preliminary study about the streamflow estimation from EC through calibrated functional EC-Q relationships in a snow-dominated catchment. For the typical karst aquifer without intense human interventions, a strong negative correlation is also observed between EC and discharge (*Goldscheider and Drew*, 2007). Higher discharge often corresponds to lower EC. Therefore, if the EC–discharge relationship can be well established, EC may provide another good proxy for discharge monitoring.

The EC–discharge relationship is more complex than the stage–discharge relationship due to the existence of the hysteresis phenomenon (*Toran and Reisch*, 2012）. A simple empirical formula or regression can hardly describe this complex non-linear relationship. Instead, machine learning methods, which are widely used in the field of hydrology (*Feng et al.*, 2020; *Kratzert et al.*, 2018; *Mewes et al.*, 2020; *Sudriani et al.*, 2019), may be an effective tool to handle their links. Long Short Term Memory (LSTM) architectures, as a special type of current neural networks, are well known for their capabilities to learn long-term dependencies between input and output variables due to the extra consideration of dedicated memory cells and different gates. Its advantage over other machine learning structures to process the long-sequence data has been widely reported (*Gao et al.*, 2020; *Zhang et al.*, 2018). This characteristic makes them an ideal candidate to cope with the hysteresis between discharge and EC.

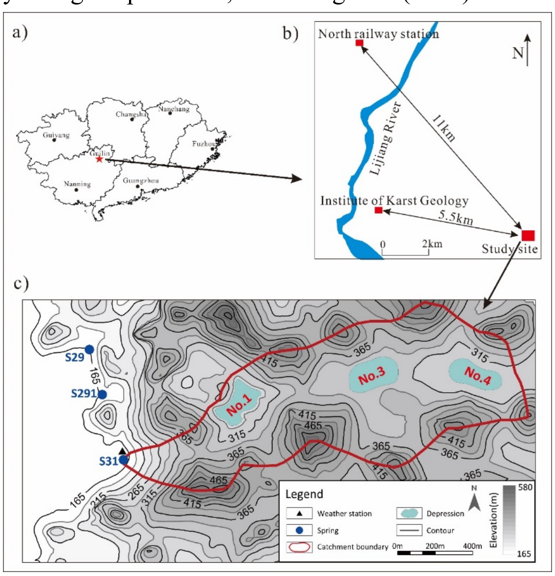
https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

In this paper, we investigate the potential of EC to predict the discharge of a karst spring using LSTM, and whether EC can be used as a proxy for the continuous long-term monitoring of discharge. The purpose of this paper is twofold: (1) to explore the feasibility of discharge prediction with EC; (2) to investigate the optimal strategy of discharge measurement when using EC to indirectly monitor discharge.

2 Study site and data

The spring S31 is the biggest karst spring in Yaji karst experimental site (Fig.1), which is located in the southwest of Guilin city, China, and it developed in the Devonian pure limestone. This karst catchment belongs to the typical peak-cluster depression landform and only receives the precipitation recharge. The study site has a typical subtropical monsoon climate, with the rainy season from April to August, during which 75% of annual precipitation occurs. Storms are frequent in this season and the highest recording of rainfall is 286 mm/day. The average annual temperature is around 18.8 ℃ and the annual precipitation is 1915 mm. According to the historical record, it seldom snows in the winter.

Due to the abundant rainfall and warm climate, the karstification degree of this karst system is very high, with strong developments of epikarst and conduits. The catchment area of this spring is around 1.0 km^2 and mainly contains three depressions according to the previous tracer tests (Yuan et al., 1996). For each depression, there are several sinkholes at the bottom that connect to the spring directly through a main conduit. During the recharge events, these sinkholes are the main rapid recharge passages that drain the fast lateral flow within the epikarst into the conduit directly. This is also the main reason for the drastic change of the spring discharge during storms. Besides, part of the rainwater could also recharge the karst aquifer slowly through the small fissures to mainly maintain the base flow. For more details about this catchment and its internal hydrological processes, see Chang et al.(2015) and Chang et al. (2019).



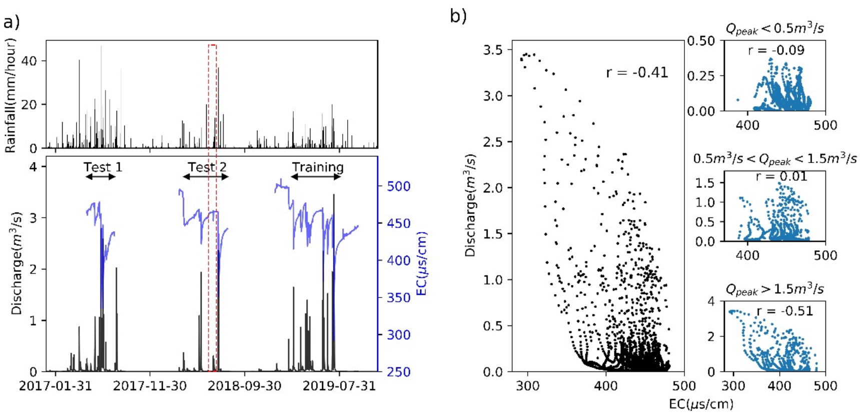
https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

Fig.1 a) Location of study site, b) locations of North railway station and Institute of karst geology relative to the study site, c) catchment area of karst spring S31 (Chang et al., 2021).

The spring discharge is measured by a combination of rectangular weirs and there is a rain gauge located near spring S31 to record the precipitation with a precision of 0.2 mm. A HOBO salt conductivity data logger is used to monitor spring EC continuously at the spring outlet with a temporal resolution of 15 minutes (corrected for 25 ℃). Due to a malfunction of the rain gauge in the study site, there are two recording gaps (14.05.2018–31.07.2018 and 29.04.2019–31.07.2019), which have been filled with information from two nearby climatic stations in North railway station and Institute of Karst Geology with the distance of 11.0 km and 5.5 km (Fig.1b), respectively. According to the previous simulation result of the conceptual rainfall-runoff models driven by these gap-filled data (Chang et al., 2021), most data have a relatively good quality only except the precipitation on June 21, 2018 (red dashed box in Fig.2, labelled as OBGD), which was severely overestimated by the gap-filled data.

The hydrochemical composition of the spring water in the study site is dominated by calcium carbonate equilibria resulting from the dissolution of carbonate rocks. There is limited human intervention in the area. As such, the spring's EC dynamic is mainly controlled by the rock dissolution and the dilution from the low-EC event water during storms (*Liu et al.*, 2004). Figure 2a shows the spring's discharge and EC measurements (corrected for 25℃) from 2017 to 2019. The spring's EC always shows a sharp drop during a storm due to the arrival of unsaturated fast flow, and it then gradually increases after the storm, corresponding to the gradual recession of the spring discharge. For the EC observations in 2018 and 2019, we find that the spring's initial EC after the long dry period is much higher than the following maximum EC in the rainy season. These higher EC observations are mainly caused by the flush of long-stagnant water after a long dry period; as such, we do not include them in the following analysis or simulations. It is worth mentioning that the original observations of the spring's EC in 2017 have a higher maximal EC value than the other two years, which is mainly caused by equipment drift (*Chang et al.*, 2021). Therefore, the EC observations for 2017 were simply adjusted by subtracting a certain value (23 us/cm) to remove the drift and keep the maximum EC consistent with the other two years.



https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

Fig. 2 a) The observed spring's discharge and EC from 2017 to 2019. The missing EC data are due to the drying-out of the spring during the dry period or equipment malfunction. The red-dashed box indicates the severely overestimated precipitation by the gap-filled rainfall data (OBGD). b) The correlation between EC and discharge, further divided into three categories according to the discharge peak (Q_{peak}) in the recharge events: small recharge events (Q_{peak} <0.5 m^3/s), middle recharge events (0.5 $m^3/s \leq Q_{peak} < 1.5$ m^3/s) and storms ($Q_{peak} \geq 1.5$ m^3/s). r is the linear correlation coefficient between EC and discharge.

Figure 2b shows the relationship between discharge and EC using all available observations. In general, two observations show a negative correlation with the linear correlation coefficient of -0.41, but also an obvious hysteresis since the EC peak always lags several hours behind the discharge peak in the study site. To explore the relationship between discharge and EC under different rainfall conditions, the recharge events in the monitoring periods are further divided into small rain events, middle rain events and storms according to the discharge peaks (Q_{peak} <0.5 m^3/s, 0.5 $m^3/s \leq Q_{peak} <$ 1.5 m^3/s, $Q_{peak} \geq 1.5$ m^3/s, respectively). To divide the monitoring discharge series into different recharge events, we first select all the discharge peaks and two adjacent peaks with a time interval lower than one day are considered as a same recharge event. The end point or start point of each recharge event is determined by the lowest discharge point between two selected adjacent peaks. The final correlation degree between discharge and EC in different recharge events is shown in Fig. 1b. We find that a strong relationship between discharge and EC exists mainly in storms, while the relationship is relatively weaker in the small or middle recharge events.

3 Methodology

To explore the feasibility of EC as a proxy for continuous discharge monitoring, we first investigate whether the discharge can be predicted with EC using LSTM. If the prediction is feasible, another fundamental concern is how to establish the stable mapping from EC to discharge in the ungauged catchment. This leads to two questions: (1) How many discharge observations should be measured? (2) What is the optimal discharge measurement strategy? To this end, we further investigate the variations of the model performances trained by a different proportion of randomly selected discharge observations. In addition, the model performances trained by several common strategies of discharge measurement were compared to inspect the potential optimal strategy.

3.1 Modeling approach

LSTM belongs to a special kind of recurrent neural network (RNN), aiming to overcome the weakness of the traditional RNN, i.e. the problem of vanishing or exploding gradients (*Bengio et al*, 1994). Due to the additional consideration of the memory cell in the hidden layer and special gates, LSTM can capture the complex correlation well in both short and long sequences, and was therefore selected to handle the mapping from EC to spring discharge. Because the EC response always lags behind the discharge, the discharge at time t (Q_t) was predicted by the EC observations before and after this time with the same length (M_{EC}):

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

$$Q_t = f(EC_{t+m}, EC_{t+m-1}, \dots, EC_t, EC_{t-1}, EC_{t-2}, \dots, EC_{t-m}) \quad (1)$$

Where EC_{t+m} and EC_{t-m} are the EC values at time t+m and t-m, respectively.

For comparison, the LSTM model was also trained by the precipitation data (M_P) to predict the spring's discharge. The discharge at time t was simulated just by the previous and current precipitation:

$$Q_t = f(P_t, P_{t-1}, \dots, P_{t-n}) \quad (2)$$

Where P_{t-n} is the precipitation at time t-n.

Meanwhile, we also used precipitation and EC data together as the input to predict the spring's discharge (M_{ECP}) to explore whether considering both sets of data in the model can improve discharge prediction.

$$Q_t = f(EC_{t+m}, EC_{t+m-1}, \dots, EC_t, EC_{t-1}, EC_{t-2}, \dots, EC_{t-m}, P_t, P_{t-1}, \dots, P_{t-n}) \quad (3)$$

In addition to these three models, the simple linear regression between discharge and EC involving all observations was used as a benchmark to compare with the results simulated by LSTM. Considering the delay behavior of EC, the best-fitting results with 7 hours forward-shifting of EC were used for comparison. Implementation of LSTM was realized using Python 3.7 based on the Keras library.

For all models, the longest data series from March 1 to August 1 in 2019 was used for model training (training period) and data in the other two periods, May 12 to August 8 in 2017 (test period 1) and March 20 to August 6 in 2018 (test period 2), were used for the model test. The time step in all models is set to one hour. Given the random nature of the machine learning algorithm, each model was repeated 10 times to show its uncertainty.

For each model, the mean squared error (MSE) was used as the objective for model training. According to Fig.1b, EC has a strong negative correlation with discharge mainly in storms, so it is expected that in high-flow periods EC provides better discharge predictions. Therefore, the Nash-Sutcliffe efficiency coefficient, putting more emphasis on the high flow, was used to compare the performance among different models.

$$Nash = 1 - \frac{\sum (Q_s - Q_o)^2}{\sum (Q_o - \overline{Q_o})^2} \quad (4)$$

Where Q_s and Q_o are simulated and measured discharge.

3.2 Different measurement strategies

To investigate how many discharge observations are required for M_P or M_{EC} to obtain a stable prediction, we randomly selected a certain percentage of discharge data in the training period (1%, 2%, 3%, 4%, 5%, 10%, 15%, 20% … 50%) as the available measurements for the model training. The trained LSTM models were then tested in the three periods to analyze prediction performance variations with the amount of available training data.

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

To explore the optimal measurement strategies, the discharge measurements from four different measurement strategies were chosen to train the model, and their performances were compared:

(1) Discharge was measured once in each day randomly during the daytime (9:00 A.M. – 5:00 P.M.). This situation is similar to the sampling strategy at relatively fixed intervals. Given that the training period contains five months, we consider the spring's discharge was measured continuously in the first one month, two months, three months, four months and five months, which accounts for 0.7%, 1.6%, 2.5%, 3.4% and 4.2% of the total data, respectively.

(2) Discharge was measured continuously over a short time. To compare with the results of situation (1), with 4.2% of available data, we randomly selected 4.2% continuous discharge data for the model training. To prevent the selected data all coming from the dry period, the selected data must contain a discharge higher than 1.5 m^3/s, that is, it should contain a certain proportion of discharge in the storms.

(3) Discharge in the largest storm or two largest storms in the training period was measured continuously, which accounted for about 2.9% and 5.0%, respectively, of the total data. In addition, we also considered the situation that the discharge was measured continuously under the largest storm and the rest was measured randomly in the remaining period, which gives 4.2% of total available data.

(4) Discharge was measured randomly in the training period. In contrast to situation (1), the result with 4.0% randomly measured discharge observations for investigating the data requirement was presented for comparison.

For each scenario, the discharge selection was repeated 100 times to consider the uncertainty caused by the random selection.

4 Results

4.1 Discharge predictions by different inputs

The selection of appropriate hidden layer, input length (m or n) and neuro number is crucial to apply the LSTM model to avoid overfitting or underfitting problem. In this paper, their appropriate values are determined through the performance comparison among models with different layers, input lengths or neuro numbers in the training and test period 1. The detail information about the selection procedures is shown in the appendix. Finally, the hidden layer, m and n are set to 1, 10 and 6 in three models, respectively.

Figure 3a shows the model performances of three models (M_P, M_{EC} and M_{ECP}). For the training period, all three models have excellent simulation results, with Nash coefficients larger than 0.90. Their performances become a little worse in test period 1 (Fig. 3b) and the median Nash values of M_P, M_{EC} and M_{ECP} are 0.78, 0.61 and 0.76, respectively. However, for test period 2 (Fig. 3c), the performances of M_P and M_{ECP} deteriorate obviously probably due to the large error of precipitation observations, whereas M_{EC} still has a relatively stable performance with a median Nash value of 0.47.

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

If the OBGD recharge event is removed, the median Nash values of three models increase to 0.07, 0.53 and 0.16, respectively. The performance of M_{EC} is still better than M_P and M_{ECP}. This indicates the gap-filled precipitation in test period 2 except OBGD may still have some errors that affect the simulation results of M_P and M_{ECP}. Comparing M_{ECP} to the other two models, except for the training period, M_{ECP} always presents the in-between Nash value. This implies the additional integration of EC into M_P can, to some degree, avoid a severe deterioration in model performance caused by the precipitation error (test period 2), but it cannot effectively improve the discharge prediction (test period 1). The Nash value of the benchmark model is 0.20. M_{EC} always has much better prediction results than the benchmark model in all three different periods, which indicates the excellent capability of LSTM to handle the complex nonlinear relationship between EC and discharge.

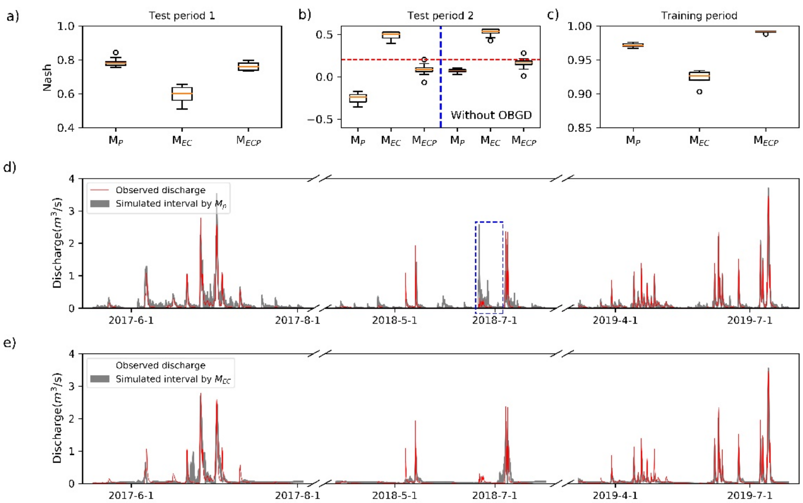


Fig. 3 a), b) and c) Performance comparison of three LSTM models with different input data (M_P: Rainfall, M_{EC}: EC, M_{ECP}: Rainfall + EC) in test period 1, test period 2 and train period, respectively. The red-dashed line in Fig.2b represents the Nash value of the benchmark model, which just considers the simple linear regression using all available data. d) and e) The simulation results of the spring's discharge by M_P and M_{EC}. The simulated interval was obtained from ten repeating simulations of each model. The blue-dashed box in Fig.2d indicates the severely overestimated discharge by M_P caused by the gap-filled precipitation data.

When further inspecting the simulated hydrographs in the three periods, we find M_P can capture the most discharge dynamics, except the severe overestimation in test period 2 caused by the precipitation error (blue dashed box in Fig.2d). Meanwhile, the simulated hydrograph by M_P contains many small discharge peaks in the dry period that are not observed. In contrast, while M_{EC} can also reproduce the spring's discharge, especially under storms, it cannot capture small discharge peaks lower than 0.50 m^3/s and the recession curve in the dry period (Fig. 3e).

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

Given the different correlations between discharge and EC in different recharge events as in Fig.1b, we further compare the performance of M_P and M_{EC} in three recharge events separately and the results are shown in Figure 4. M_{EC} shows a little better performance in storm events than M_P, whereas its performance in the middle recharge events is worse than M_P with the mean Nash value of 0.52 and 0.82, respectively. For the small recharge events, both models show very bad results and the Nash values are lower than 0. Even though the OBGD event is removed, their performances are still not good and only M_{EC} has a Nash value a little higher than zero.

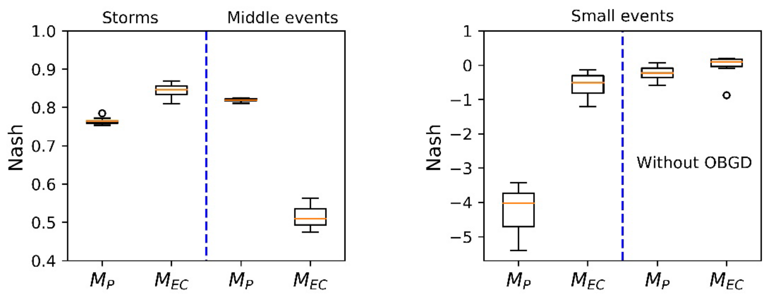


Fig. 4 Performance comparison of two models (M_P and M_{EC}) in different recharge events.

4.2 Discharge predictions under different monitoring strategies

To investigate the data requirement of discharge observations to obtain a stable prediction, we compare the performances of M_P and M_{EC} trained by different proportions of random selections (Fig. 5a and 5b). Our results show that the Nash coefficients of the two models gradually increase with available observations except for M_P in test period 2 (precipitation error). For both models, when the percentage of selected observations is higher than 20%, their performances tend to be stable and the consideration of extra observations would not highly improve the model performance. Meanwhile, in contrast to M_P driven by precipitation, M_{EC} does not need more additional discharge observations.

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

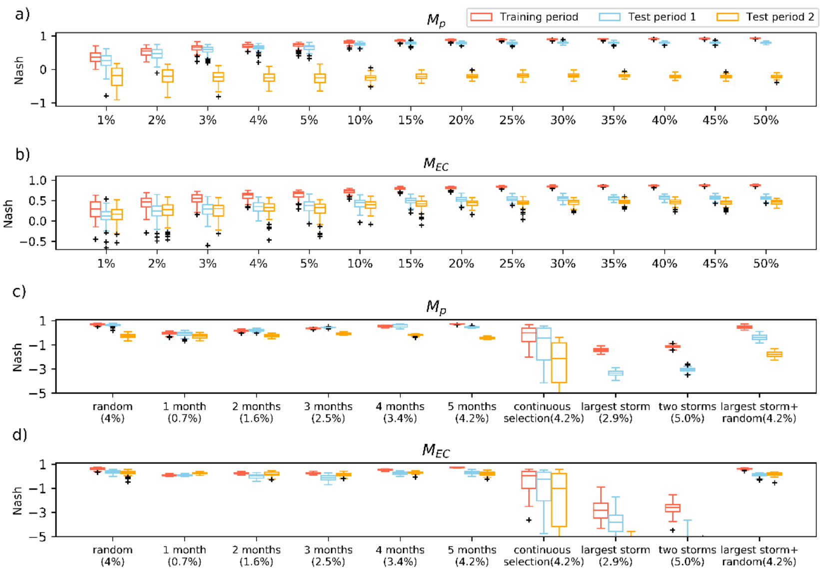


Fig. 5 a) and b) Model performances in the three periods when the available discharge data is randomly selected from the training period with a certain percentage (1%, 2%, 3%, 4%, 5%, 10%, 15%, …, 50%). c) and d) Model performances with different measurement strategies of discharge in the training period. Random corresponds to random discharge measurements. 1 month, 2 months, 3 months, 4 months indicate that one discharge was randomly selected on one day during the daytime from one month, two months, three months and four months, respectively. Continuous selection means the discharge data were selected in a continuous way. Largest storm and two storms indicate that only the discharge data under the largest storm or the two largest storms were selected to train the model. Largest storm + random denotes that the discharge data under the largest storm was used along with a random selection of data, together accounting for 4.2% of the total data. The number in brackets shows the proportion of the randomly selected data in the training period.

Figure 5c and 5d shows the performances of two models (M_P and M_{EC}) in the three periods trained by different discharge observations relating to different measurement strategies. Generally, no matter which variable is used to predict the discharge (precipitation or EC), the optimal discharge measurement strategy for obtaining the best prediction results is consistent. The model trained by the random or relatively fixed-interval observations gives the best prediction results, while the one trained by the observations under one or two largest storms has the worst performance. However, if the observations in the largest storm are combined with some random measurements to train the model, the model performance will be highly improved, but is still worse than the best prediction. This result further demonstrates the superiority of considering random observations to train the model to get a better prediction result. For the model trained by the continuous discharge observations, the model performance shows wide ranges indicating its strong dependence on the measurement period.

4 Discussion

The results of this paper indicate it is feasible to predict discharge with EC using LSTM in the study catchment. However, as shown in Fig.4, EC provides different accuracies of discharge prediction under different recharge events due to the different correlation between EC and discharge (Fig. 2b). These different prediction accuracies are probably due to the different control mechanisms of EC behavior under different rainfall conditions. For the typical karst system without strong human interventions, the EC dynamic mainly results from the dilution from the new fresh rainwater and the dissolution of carbonate rocks within the catchment (*Goldscheider and Drew*, 2007; *Chang et al.*, 2021). During storms, the EC dynamic is mainly dominated by dilution, which leads to the close dependence of EC reduction and discharge because larger discharge always means more new fresh rainwater. However, for the middle recharge events, the EC dynamic may be related to both the dissolution and dilution processes. The dissolution process not only depends on discharge, but also relates to the rainfall style and internal hydrological path. For example, for a same spring discharge, higher rainfall intensity may lead to more fast flow and lower EC. Therefore, the effect of dissolution on EC, to some degree, can reduce the correlation between EC and discharge and increase the prediction uncertainty of discharge. For small recharge events, the dissolution process dominates EC behavior. At the study site under small rainfall conditions, the spring's EC always shows a very limited fluctuation or even does not change, indicating that the dissolution of carbonate rock almost reaches the equilibrium at the outlet. Therefore, under such conditions, there is a very weak correlation between EC and discharge, and large uncertainties in discharge predictions. From this aspect, EC is more suitable to be used for the large discharge monitoring when the dilution effect dominates the EC dynamic.

Several studies have investigated how many discharge measurements are needed to obtain robust predictions in ungauged catchments, although most concentrate on the conceptual rainfall-runoff model. *Perrin et al*. (2007) find that 350 random observations sampled out of a 39 year recorded period (around 2.5% of full data), including dry and wet conditions, are sufficient to get similar calibrations to those of a full calibration based on 12 basins in the USA. *Seibert and Beven* (2009) report that 32 random selections from each hydrological year (around 8.7%) can provide robust runoff simulations based on 11 catchments in Sweden. In contrast, our study indicates that more discharge observations are needed (around 20% of full data) for M_P or M_{EC} to reach similar discharge predictions to those predicted by the model trained using all data. This requirement is probably because LSTM is a hyperparameter model that contains many more calibrated parameters than the traditional conceptual model since a more complex model often needs more calibration data to reach a stable performance (Perrin et al., 2007).

Our study also highlights the significance of the measurement strategy in model performance. The random observations are more informative for model calibration than the continuous dataset of the same length, which is consistent with previous studies (*Perrin et al.*, 2007; *Seibert and Beven*, 2009; *Seibert and McDonnell*, 2015). In

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.

contrast to several reports (*Juston et al.*, 2009; *McIntyre and Wheater*, 2004; *Singh and Bárdossy*, 2012), we find that the event-based sampling strategy results in much worse model performance than sampling at relatively fixed intervals. This mainly depends on the characteristic of LTSM that belongs to a pure data-driven model and has a limited extrapolation capability. Therefore, to obtain stable prediction results, LTSM should be trained by the dataset covering various climate conditions. The model trained only by event-based observations would provide large prediction uncertainties when used to predict discharge beyond the training condition. This is also the main reason that the random or relative fixed measurement strategy performs better than others. Hence, in practical applications, we should measure discharge under a variety of rainfall conditions, particularly extreme conditions as much as possible so as to obtain a robust LSTM model.

Although depth is commonly used for continuous discharge monitoring based on the stage–discharge rating curve, this method is only suitable for the relatively regular channel, where the channel geometry should not change during the monitoring period (*Weijs et al.*, 2013). In contrast, our method to use EC to substitute for discharge monitoring is independent of the channel geometry and can be applied in any channel condition. Therefore, it is more stable than the stage–discharge method when applied in a channel where the geometry may change obviously with time. In addition, the rainfall-runoff model calibrated by limited random measurements also has a huge potential to obtain long-term discharge series (*Perrin et al.*, 2007; *Pool et al.*, 2017; *Seibert and Beven*, 2009). Our study shows the LSTM trained by precipitation (M_p) always show a better prediction performance than that trained by EC (M_{EC}) in test period 1. However, these models need accurate precipitation measurements, which often exhibit a strong spatial variability. Measuring precipitation with a sparse gauge network may produce large errors that can result in large uncertainties of discharge predictions (*Oudin et al.,* 2006), as our study shows (M_P in the test period 2, Fig. 2). Although now the satellite precipitation product can provide the precipitation data at reatively high resolutions in the ungauged site, these products still suffer from systematic, random and detection errors, which are more pronounced in mountain regions (Maggioni and Massari, 2018l; Beck et al., 2019). In contrast, the EC measurement, like the depth measurement, only needs to focus on the outlet without a spatial observation uncertainty. Therefore, compared to the rainfall-runoff model, our method still has an obvious advantage in mountain regions where the precipitation have a large spatial variability. Despite these advantages, our method also has obvious drawbacks. Firstly, the application of our method is restricted to catchments where EC has a strong relationship with discharge. Meanwhile, the EC sensor also needs a periodic calibration to avoid a strong drifting. Secondly, as discussed before, predicting discharge with EC may have large uncertainties in the small recharge events, during which the EC dynamic is strongly affected by mineral dissolution. In our study catchment, higher values of EC after a long dry period due to the flush of old water within the conduit are discarded in our model. Further research is needed to test whether more complex neuro networks can handle this situation.

https://doi.org/10.5194/hess-2022-77
Preprint. Discussion started: 21 March 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences Discussions

5 Conclusions

In this paper, we evaluate the feasibility of using EC as a proxy for the long-term discharge monitoring based on a machine learning architecture LSTM in a small karst catchment where EC exhibits a strong negative correlation with discharge. The results indicate the huge potential of EC to predict discharge and it is feasible to train a robust LSTM with just a small number of discharge observations; however, in some recharge events the prediction uncertainty is relatively large. The random or fixed-interval measurement strategy can give more informative values for LSTM training. Our study provides useful guidance for the application of our method in other ungauged catchments where the installation of gauging weirs or representative rainfall stations is prohibited. Furthermore, at the study site, the EC dynamic of the karst spring is relatively simple without obvious seasonal variations (*Liu et al.*, 2007) or 'piston effects' (a temporal EC peak before it drops during storms) (*Hess and White*, 1993), further investigations are required to evaluate whether LSTM could handle more complex situations. It should also be noted that although our work was conducted in a karst region, our method and conclusion may also be useful in non-karst catchments where a strong correlation between EC and streamflow exists (*Cano-paoli et al.*, 2019; *Weijs et al.* 2013). For example, *Cano-paoli et al.* (2019) used several empirical equations to estimate the river discharge by EC in a snow-dominated non-karst catchment and obtained relatively good prediction results. Compared to the empirical equation, LSTM show a more flexible capability to handle the relationship between discharge and EC, and therefore it is expected to get more robust results.

Appendix

A1 Selection of appropriate hidden layer, input length and neuro number in LSTM models

The fundamental problem when using LSTM to simulate discharge is to determine the appropriate hidden layer, input length (m or n) and neuro number. The dropout technique was applied in LSTM to avoid overfitting and the value is set to 0.4 (Srivastava et al., 2014). In many cases, one hidden layer is often found to be sufficient for discharge simulation (Campolo et al., 1999; Gao et al., 2020). In our study, we also found one hidden layer in LSTM is enough to simulate the discharge by precipitation, EC or both after a series of experiments. To choose the appropriate input length of precipitation or EC to simulate the discharge, we first used a relatively large neuro number (40) and then gradually increased the input length of EC (m) and precipitation (n) from 1 to 20 to compare their performances in the training and test period 1. The appropriate value of m or n was obtained based on the relative balance between the performances in these two periods. The comparison results of models using different input length of EC or precipitation were shown in Fig. 1A and Fig. 2A, and the optimal values of m and n were found to be about 10 days and 6 days.

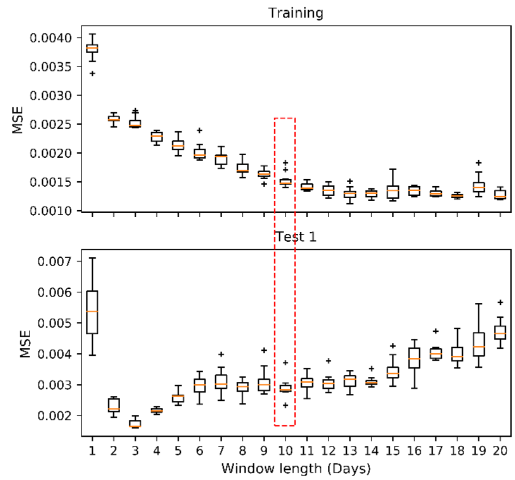


Fig. A1 MSE variations with the input length of EC (m) increasing from 1 to 20 when the EC is used to simulate the discharge (the neuro number is 40). Based on the performance comparison in the training and test 1 period, 10-days is regarded as the best window length for EC to predict the discharge.

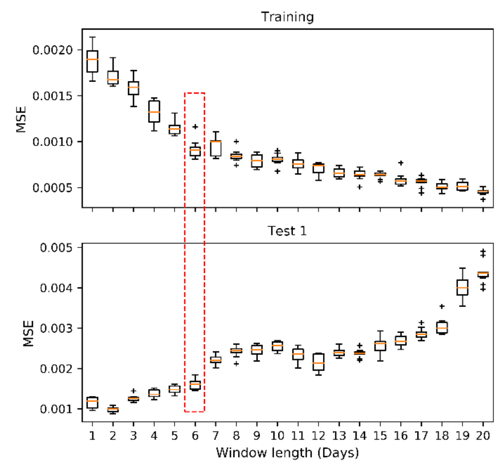


Fig. 2A MSE variations with the input length of precipitation (n) increasing from 1 to 20 when the precipitation is used to simulate the discharge (the neuro number is set to 40). Based on the performance comparison in the training and test 1 period, 6-days is regarded as the best window length for precipitation to predict the discharge.

Finally, we gradually reduced the neuro number from 40 to 1 to compare the model performance with different neuron numbers and the results were shown in Fig. 3A and 4A. The simulation results just exhibit pronounced deteriorations when the neuro number is smaller than 10, while the larger neuro number does not bring noticeable overfitting (performance deterioration in test period 1) which may mainly due to the consideration of dropout technique. Therefore, for three models (M_{EC}, M_P and M_{ECP}), the neuro number is set to 10.

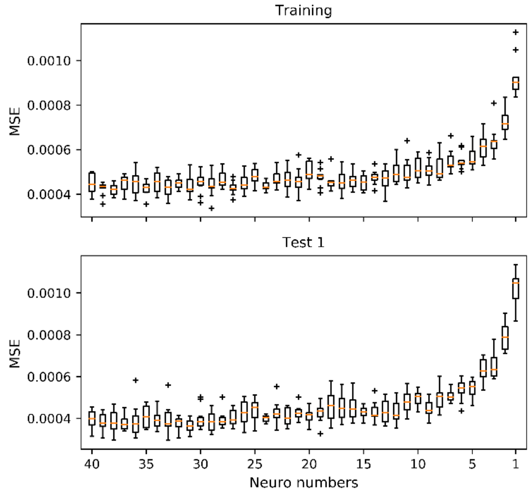


Fig. 3A MSE variations with the neuro number decreasing from 40 to 1 gradually in M_{EC}.

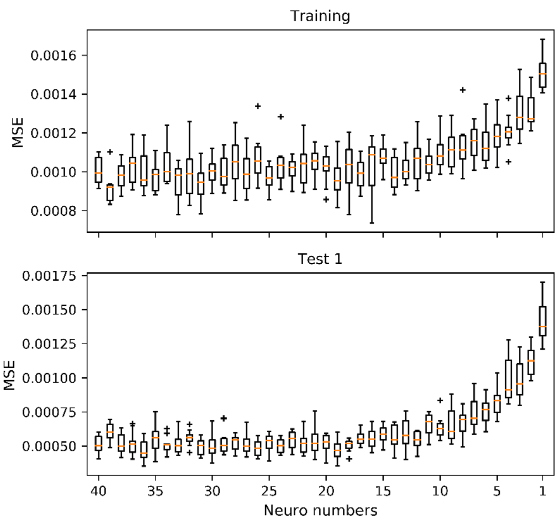


Fig.4A MSE variations with the neuro number decreasing from 40 to 1 gradually in M_P.

Code and data availability. All data and simulation results are available from the corresponding author upon request.

© Author(s) 2022. CC BY 4.0 License.

Author contributions. All authors designed this study. YC carried out all analysis, model simulations and wrote the initial manuscript. BM and AH contributed to discussing the results and improving the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. Yong Chang was supported by the China Scholarship Council (ID: 201906195028). Andreas Hartmann was supported by the Emmy-Noether-Programme of the German Research Foundation (DFG, Grant Nos. HA 8113/1- 1).

© Author(s) 2022. CC BY 4.0 License.

References

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R. and Adler, R. F.: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, Bull. Am. Meteorol. Soc., 100(3), 473–500, doi:10.1175/BAMS-D-17-0138.1, 2019.

Bengio, Y., Simard, P. and Frasconi, P.: Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw., 5, 157–166, doi:10.1109/72.279181, 1994.

Boiten, W.: Flow-measuring structures, Flow Meas. Instrum., 4(1), 17–24, doi: 10.1016/0955-5986(93)90006-5, 1993.

Campolo, M., Andreussi, P. and Soldati, A.: River flood forecasting with a neural network model, Water Resour. Res., 35(4), 1191–1197, doi: 10.1029/1998WR900086, 1999.

Cano-paoli, K., Chiogna, G. and Bellin, A.: Convenient use of electrical conductivity measurements to investigate hydrological processes in Alpine headwaters, Sci. Total Environ., 685, 37–49, doi:10.1016/j.scitotenv.2019.05.166, 2019.

Chang, Y., Wu, J. and Liu, L.: Effects of the conduit network on the spring hydrograph of the karst aquifer, J. Hydrol., 527, 517–530, doi:10.1016/j.jhydrol.2015.05.006, 2015.

Chang, Y., Wu, J., Jiang, G., Liu, L., Reimann, T. and Sauter, M.: Modelling spring discharge and solute transport in conduits by coupling CFPv2 to an epikarst reservoir for a karst aquifer, J. Hydrol., 569, 587–599, doi: 10.1016/j.jhydrol.2018.11.075, 2019.

Chang, Y., Hartmann, A., Liu, L., Jiang, G. and Wu, J.: Identifying More Realistic Model Structures by Electrical Conductivity Observations of the Karst Spring, Water Resour. Res., 57(4), e2020WR028587, doi: 10.1029/2020WR028587, 2021.

Dzikowski, M. and Jobard, S.: Mixing law versus discharge and electrical conductivity relationships: application to an alpine proglacial stream, Hydrol. Process., 26(18), 2724–2732, doi: 10.1002/hyp.8366, 2012.

Feng, D., Fang, K. and Shen, C.: Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales, Water Resour. Res., e2019WR026793, doi:10.1029/2019WR026793, 2020.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M. and Lin, Q.: Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation, J. Hydrol., 589, 125188, doi:10.1016/j.jhydrol.2020.125188, 2020.

Goldscheider, N. and Drew, D.: Methods in Karst hydrogeology: IAH: international contributions to hydrogeology, 26, Crc Press., 2014.

Gurnell, A. M. and Fenn, C. R.: Spatial and Temporal Variations in Electrical Conductivity in a Pro-Glacial Stream System, J. Glaciol., 31(108), 108–114, doi: 10.3189/S002214300000633X, 1985.

© Author(s) 2022. CC BY 4.0 License.

Herschy, R. W. : Streamflow measurement. CRC press, 1995.

Hess, J. W. and White, W. B.: Groundwater geochemistry of the carbonate karst aquifer, southcentral Kentucky, U.S.A, Appl. Geochemistry, 8(2), 189–204, 1993.

Juston, J., Seibert, J. and Johansson, P.-O.: Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, Hydrol. Process., 23(21), 3093–3109, doi: 10.1002/hyp.7421, 2009.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M.: Rainfall – runoff modelling using Long Short-Term Memory (LSTM) networks, , 6005–6022, 2018.

Liu, Z., Groves, C., Yuan, D. and Meiman, J.: South China Karst Aquifer Storm-Scale Hydrochemistry, Ground Water, 42(4), 491–499, doi:10.1111/j.1745-6584.2004.tb02617.x, 2004.

Liu, Z., Li, Q., Sun, H. and Wang, J.: Seasonal, diurnal and storm-scale hydrochemical variations of typical epikarst springs in subtropical karst areas of SW China: Soil CO2 and dilution effects, J. Hydrol., 337(1–2), 207–223, doi:10.1016/j.jhydrol.2007.01.034, 2007.

Maggioni, V. and Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review, J. Hydrol., 558, 214–224, doi: 10.1016/j.jhydrol.2018.01.039, 2018.

McIntyre, N. R. and Wheater, H. S.: Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty, J. Hydrol., 290(1), 100–116, doi:https://doi.org/10.1016/j.jhydrol.2003.12.003, 2004.

Mewes, B., Oppel, H., Marx, V. and Hartmann, A.: Information-Based Machine Learning for Tracer Signature Prediction in Karstic Environments, Water Resour. Res., 56(2), 1–20, doi:10.1029/2018WR024558, 2020.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. and Michel, C.: Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, J. Hydrol., 320(1), 62–83, doi: 10.1016/j.jhydrol.2005.07.016, 2006.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C. and Mathevet, T.: Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models, Hydrol. Sci. J., 52(1), 131–151, doi:10.1623/hysj.52.1.131, 2007.

Pool, S., Viviroli, D. and Seibert, J.: Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration?, J. Hydrol., 554, 613–622, doi:10.1016/j.jhydrol.2017.09.037, 2017.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, Hydrol. Earth Syst. Sci., 13(6), 883–892, doi:10.5194/hess-13-883-2009, 2009.

Seibert, J. and McDonnell, J. J.: Gauging the Ungauged Basin: Relative Value of Soft and Hard Data, J. Hydrol. Eng., 20(1), doi:10.1061/(asce)he.1943-5584.0000861, 2015.

Hydrology and Earth System Sciences
Open Access
Discussions

Singh, S. K. and Bárdossy, A.: Calibration of hydrological models on hydrologically unusual events, Adv. Water Resour., 38, 81–91, doi: 10.1016/j.advwatres.2011.12.006, 2012.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, J. Mach. Learn. Res., 15, 1929–1958, 2014.

Sudriani, Y., Ridwansyah, I. and A Rustini, H.: Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia, IOP Conf. Ser. Earth Environ. Sci., 299(1), doi:10.1088/1755-1315/299/1/012037, 2019.

Toran, L. and Reisch, C. E.: Using Stormwater Hysteresis to Characterize Karst Spring Discharge, Ground Water, 51(4), 1745-6584, doi:10.1111/j.1745-6584.2012.00984.x, 2012.

Turnipseed, D. P. and Sauer, V. B.: Discharge measurements at gaging stations, Reston, VA., 2010.

Weijs, S. V, Mutzner, R. and Parlange, M. B.: Could electrical conductivity replace water level in rating curves for alpine streams?, Water Resour. Res., 49(1), 343–351, doi:10.1029/2012wr012181, 2013.

Yuan, D.X., A.D. Dai, W.T. Cai, Z.H. Liu, S.Y. He, X.P. Mo, S.Y. Zhou, and W.K. Lao: Karst water system of a peak cluster catchment in South China's bare Karst region and its mathematic model. Guangxi Normal University Publishing House, Guilin, China, 1996. (in Chinese)

Zhang, J., Zhu, Y., Zhang, X., Ye, M. and Yang, J.: Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas, J. Hydrol., 561, 918–929, doi:10.1016/j.jhydrol.2018.04.065, 2018.