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

[4] Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom.

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

Key Points:

- Discharge can be predicted with EC using LSTM machine learning techniques.
- The discharge predictions from EC have relatively large uncertainties in small or middle recharge events.
- The random or fixed-interval discharge measurement strategy is more informative for obtaining a robust LSTM prediction model.

Abstract

Due to EC's easy recordability and the existence of a strong correlation between EC and discharge in certain catchments, EC is a potential predictor of discharge. This potential has 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's 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. The results indicate, based on LSTM, that the spring's 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. While our study is implemented in a karst catchment, the method is also suitable for non-karst catchments where there is a strong correlation between EC and discharge.

Keywords: electrical conductivity, discharge monitoring, LSTM, karst spring

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. There are many ways to measure streamflow, like the current meter method, dilution gauging method, acoustic doppler method and electromagnetic method [*Dobriyal et al.*, 2017]. However, these methods all concentrate on one-time measurements and are not executable for long-term monitoring. For continuous monitoring, 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, 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) is 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]. *Weijs et al*. (2013) investigate the potential of EC to predict discharge in alpine watersheds and find the EC–streamflow relationship even slightly outperforms the stage–discharge relationship. For the typical karst aquifer without intense human interventions, a strong negative correlation is 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.

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 karst catchment of spring S31 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 catchment area is around 1.0 km^2 according to the previous tracer tests [*Yuan et al.*, 1996]. The karstification degree of this karst system is very high, with strong developments of epikarst and conduits. 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. For more details about this catchment, see *Chang et al.*(2015) and *Chang et al.* (2019).

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 1a 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; submitted to Water resources research]. 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.

[Figure 1]

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 nearby climatic stations. According to the previous simulation result of the conceptual rainfall-runoff model [*Chang et al.,* 2021, submitted to Water Resources Research], the precipitation on June 21, 2018 (red dashed box in Fig.2), was severely overestimated by the gap-filled data, which may strongly affect the simulation results.

Figure 1b 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. When the recharge events 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), 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 cell state 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}):

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

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

For comparison, the results of the traditional method are presented (M_P); here the precipitation data were used as the input to predict the spring's discharge. The discharge at time t was simulated just by the previous and current precipitation:

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

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.

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

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 resolution of observations is one hour. Given the random nature of the machine learning algorithm, each model was repeated 10 times to show its uncertainty. Selections of the appropriate hidden layer, input length and neuro number for each model are shown in the supplemental material.

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 coefficient, putting more emphasis on the high flow, was used to compare the performance among different models.

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

Where Q_s and Q_o are simulated and observed 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.

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 total selected data from 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 middle recharge events or 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% 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

Figure 2a 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 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, the performances of M_P and M_{ECP} deteriorate obviously 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. We find that M_{EC} 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. 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).

[Figure 2]

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.2b). 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.

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. 6a and 6b). 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 additional discharge observations.

[Figure 3]

Figure 7 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

[Figure 4]

The results of this paper indicate it is feasible to predict discharge with EC using LSTM. However, it should be noted that EC may provide different accuracies of discharge prediction under different recharge events due to the different correlation between EC and discharge as shown in Fig. 1b. Fig. 4 shows the scatter plot between the observed and simulated discharge with M_P or M_{EC} (one simulation result chosen from ten repeated simulations), which is also divided into the same three groups. Generally, the linear correlation coefficient (r) of M_{EC} is very close to M_P when considering all available data. When further inspecting each group, M_{EC} provides a good simulation result of discharge in storms (r = 0.92), which is even a little better than M_P (r = 0.88). Whereas, for the discharge under the middle rain events, the performance of M_{EC} (r = 0.72) is worse than M_P (r = 0.91). Neither model can reproduce the discharge well under the small recharge events. The different prediction accuracies are probably due to the different control mechanisms of EC behavior under different rainfall conditions. For the typical karst system, the EC dynamic mainly results from the dilution from the fast flow and the dissolution of carbonate rocks. 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 fast flow. However, for the middle recharge events, the EC dynamic may be related to both the dissolution and dilution processes. Because the dissolution process not only depends on discharge, 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.

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 a few 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 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]. 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). In contrast, the EC measurement, like the depth measurement, only needs to focus on the outlet without a spatial observation uncertainty. 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. 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.

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 good 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].

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). All the data and codes are available for download: https://zenodo.org/record/4568702#.YDt9Tdh7rIU

References

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*, 157–166.

Boiten, W. (1993). Flow-measuring structures. *Flow Measurement and Instrumentation*, *4*(1), 17–24.

Cano-paoli, K., Chiogna, G., & Bellin, A. (2019). Convenient use of electrical conductivity measurements to investigate hydrological processes in Alpine headwaters. *Science of the Total Environment*, 685, 37–49.

Chang, Y., J. Wu, & G. Jiang (2015), Modeling the hydrological behavior of a karst spring using a nonlinear reservoir-pipe model, *Hydrogeology Journal*, *23*, 901–914.

Chang, Y., J. Wu, G. Jiang, L. Liu, T. Reimann, & M. Sauter (2019), Modelling spring discharge and solute transport in conduits by coupling CFPv2 to an epikarst reservoir for a karst aquifer, *Journal of Hydrology*, *569*, 587–599.

Chang,Y., A. Hartmann, L. Liu, G. Jiang, & J. Wu (2021), Identifying more realistic model structures by electrical conductivity observations of the karst spring. (submitted to Water Resources Research).

Dobriyal, P., Badola, R., Tuboi, C., & Hussain, S. A. (2017). A review of methods for monitoring streamflow for sustainable water resource management. *Applied Water Science*, *7*(6), 2617–2628.

Dzikowski, M., & Jobard, S. (2012). Mixing law versus discharge and electrical conductivity relationships: application to an alpine proglacial stream. *Hydrological Processes*, *26*(18), 2724–2732.

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 Resources Research*, *56*, e2019WR026793.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology*, *589*, 125188.

Goldscheider, N., & Drew, D. (2007). *Methods in Karst Hydrogeology: IAH: International Contributions to Hydrogeology, 26*. CRC Press.

Gurnell, A. M., & Fenn, C. R. (1985). Spatial and Temporal Variations in Electrical Conductivity in a Pro-Glacial Stream System. *Journal of Glaciology*, *31*(108), 108–114.

Herschy, R. W. (1995). *Streamflow measurement*. CRC press.

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

Juston, J., Seibert, J., & Johansson, P.-O. (2009). Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrological Processes*, *23*(21), 3093–3109.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall – runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrology and Earth System Sciences, 22*(11), 6005–6022.

Liu, Z., Groves, C., Yuan, D., & Meiman, J. (2004). South China Karst Aquifer Storm-Scale Hydrochemistry. *Ground Water*, *42*(4), 491–499.

Liu, Z., Li, Q., Sun, H., & Wang, J. (2007). Seasonal, diurnal and storm-scale hydrochemical variations of typical epikarst springs in subtropical karst areas of SW China: Soil CO_2 and dilution effects. *Journal of Hydrology*, *337*(1–2), 207–223.

McIntyre, N. R., & Wheater, H. S. (2004). Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty. *Journal of Hydrology*, *290*(1), 100–116.

Mewes, B., Oppel, H., Marx, V., & Hartmann, A. (2020). Information-Based Machine Learning for Tracer Signature Prediction in Karstic Environments. *Water Resources Research*, *56*, e2018WR024558.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., & Michel, C. (2006). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology*, *320*(1), 62–83.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., & Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological Sciences Journal*, *52*(1), 131–151.

Pool, S., Viviroli, D., & Seibert, J. (2017). Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *Journal of Hydrology*, *554*, 613–622.

Seibert, J., & Beven, K. J. (2009). Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences*, *13*(6), 883–892.

Seibert, J., & McDonnell, J. J. (2015). Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *Journal of Hydrologic Engineering*, *20*(1), A4014004.

Singh, S. K., & Bárdossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources*, *38*, 81–91.

Sudriani, Y., Ridwansyah, I., & A Rustini, H. (2019). Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *299*(1), 012037.

Toran, L., & Reisch, C. E. (2012). Using Stormwater Hysteresis to Characterize Karst Spring Discharge. *Ground Water*, *54*(4), 575–587.

Turnipseed, D. P., & Sauer, V. B. (2010). *Discharge measurements at gauging stations*. *U.S. Geological Survey Techniques and Methods book 3, chap. A8, 87p.*

Weijs, S. V, Mutzner, R., & Parlange, M. B. (2013). Could electrical conductivity replace water level in rating curves for alpine streams? *Water Resources Research*, *49*(1), 343–351.

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 (1996). 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. (in Chinese)

Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, *561*, 918–929.

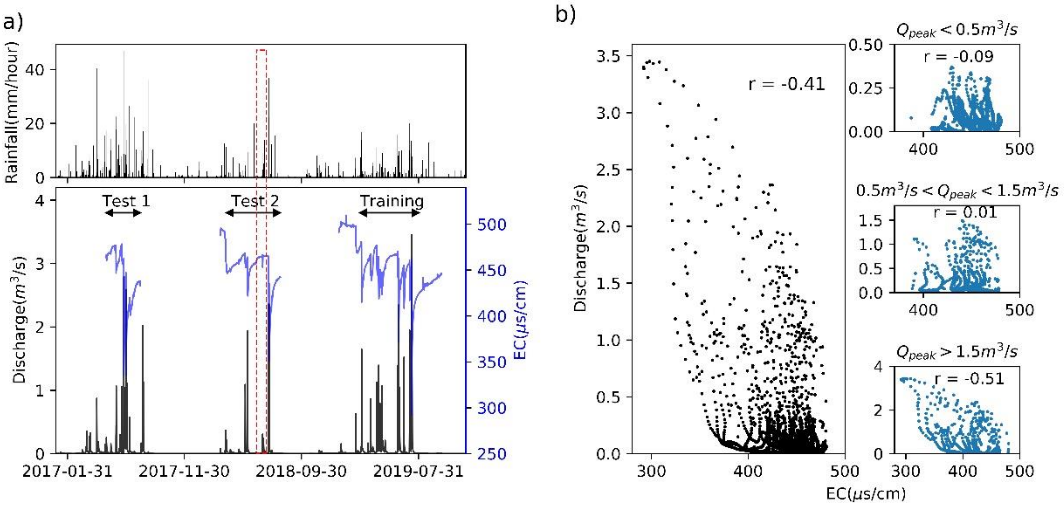

Fig. 1 a) The observed spring's discharge and EC from 2017 to 2019. The missing EC data are due to the drying-out of the spring during the dry period or equipment malfunction. The red-dashed box indicates the severely overestimated precipitation by the gap-filled rainfall data. b) The correlation between 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.

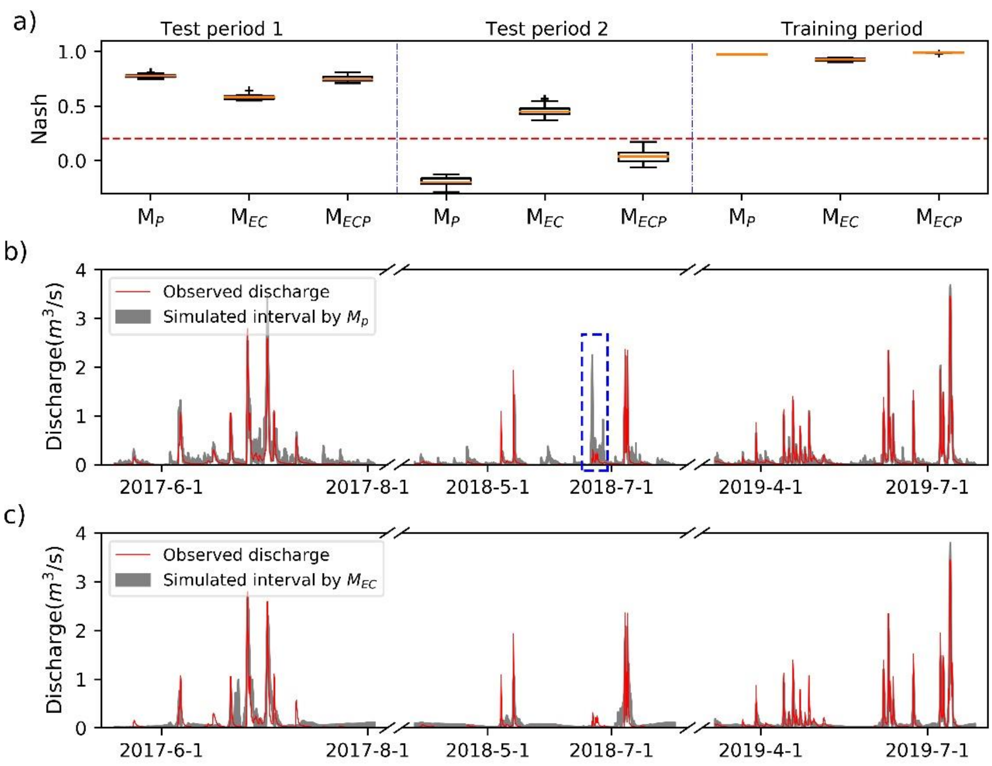

Fig. 2 a) Performance comparison of three LSTM models with different input data (M_P: Rainfall, M_{EC}: EC, M_{ECP}: Rainfall + EC). The red-dashed line represents the Nash value of the benchmark model, which just considers the simple linear regression using all available data. b) and c) The simulation 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 indicates the severely overestimated discharge by M_P caused by the gap-filled precipitation data.

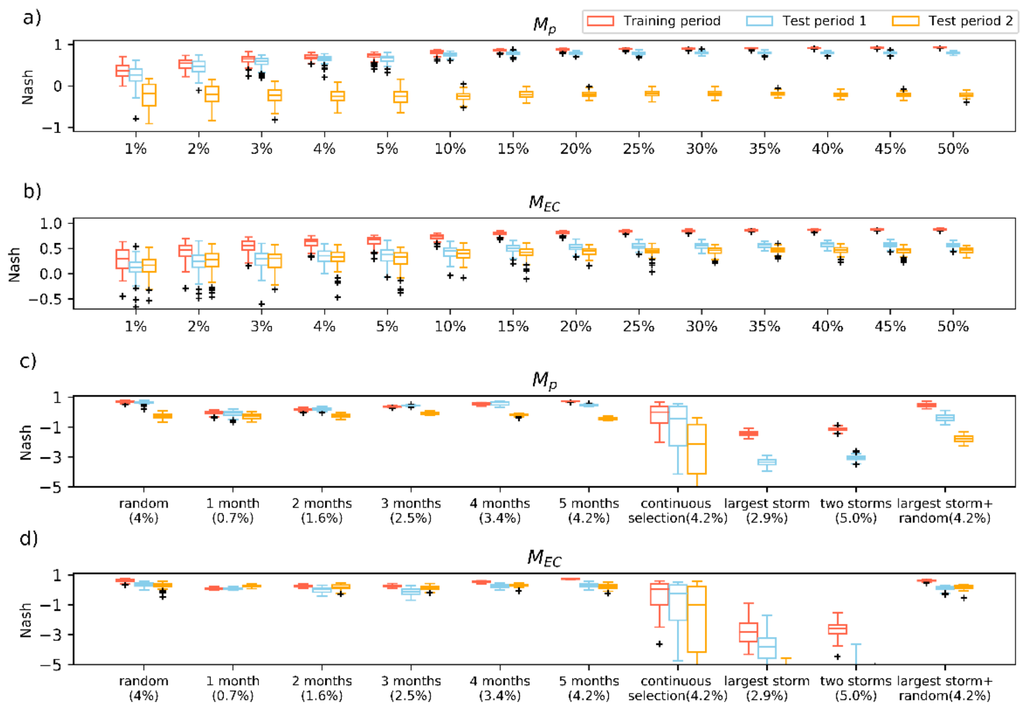

Fig. 3 a) and b) Model performances in the three periods when the available discharge data is randomly selected from the training period with a certain percentage (1%, 2%, 3%, 4%, 5%, 10%, 15%, …, 50%). c) and d) Model performances with different measurement strategies of discharge in the training period. 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, togetheraccounting for 4.2% of the total data. The number in brackets shows the proportion of the total available data.

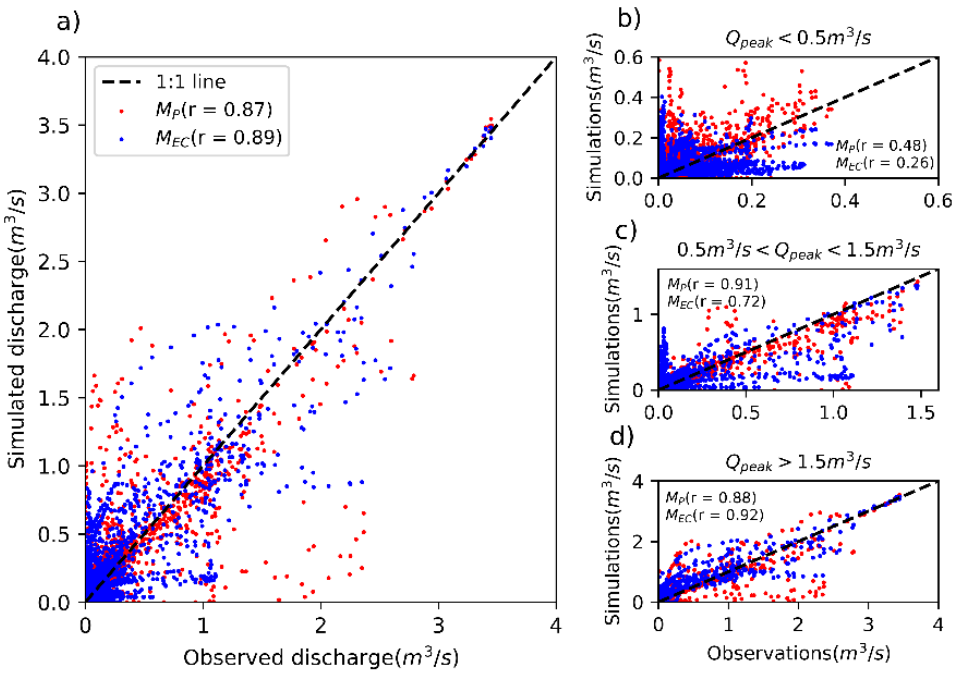

Fig.4 a) Scatter plots between the observed and simulated discharge with M_P and M_{EC} in the three periods, which was trained by all available data in the training period. b) data in the small recharge events with the observed discharge peak (Q_{peak}) lower than 0.5 m^3/s, c) data in the middle recharge events with observed Q_{peak} between 0.5 m^3/s and 1.5 m^3/s, d) data in the storms with observed Q_{peak} larger than 1.5 m^3/s. r is the linear correlation coefficient between observed and simulated discharge.