

马从文, 张志才, 陈喜, 等. 基于机器学习的西南岩溶泉流量模拟研究[J]. 中国岩溶, 2024, 43(1): 48-56.

DOI: 10.11932/karst2023y013

基于机器学习的西南岩溶泉流量模拟研究

马从文¹, 张志才¹, 陈喜², 程勤波¹, 彭韬^{3,4}, 张林^{3,4}

(1. 河海大学水文水资源学院, 江苏南京 210098; 2. 天津大学地球系统科学学院, 天津 300072; 3. 中国科学院地球化学研究所, 贵州贵阳 550081; 4. 中国科学院普定喀斯特生态系统观测研究站, 贵州普定 562100)

摘要: 岩溶泉对西南岩溶区生态系统稳定和经济社会发展具有重要意义。受岩溶区独特水文地质结构与多重水流过程控制, 岩溶泉流量具有复杂的动态变化特征, 机器学习模型为其模拟和预测提供了有效手段。然而, 岩溶泉域降雨-泉流量过程及其时空变异特征对机器学习模型结构与模拟精度的影响仍不明晰。本文选取西南典型岩溶泉, 基于长短期记忆网络(LSTM)建立岩溶泉流量模拟模型, 利用泉域实测逐小时降雨与泉流量序列进行模型训练与验证。在此基础上, 分析了不同降雨-泉流量过程对岩溶泉流量模拟精度的影响, 以及岩溶水文地质结构对降雨-泉流量响应滞时的控制作用。研究结果显示, 山坡岩溶泉与流域出口岩溶泉训练期纳什效率系数(NSE)分别为0.942与0.951, 验证期分别为0.831与0.834。对于山坡岩溶泉与流域出口岩溶泉, 利用全年实测序列训练的模型预测雨季泉流量存在较大偏差, NSE分别为0.793与0.798, 而利用雨季实测序列训练的模型预测雨季泉流量, 精度显著提升, NSE分别为0.956与0.962, 且此差异在暴雨频繁的5、6、7月尤为显著。受浅薄土壤与表层岩溶带分布影响, 山坡岩溶泉LSTM模型时序步长显著小于流域出口岩溶泉。

关键词: 机器学习; LSTM; 岩溶泉流量; 响应滞时; 岩溶降雨-泉流量过程

创新点: 将实测水文数据分雨、旱两季分别进行训练与验证, 结果显示将具有不同产汇流特征的数据分类训练可以提高LSTM模型模拟精度; 将LSTM网络时序步长率定过程与物理意义结合, 对比分析了泉流量模拟精度受岩溶水文地质结构的影响。

中图分类号: P641.134 文献标识码: A

文章编号: 1001-4810(2024)01-0048-09

开放科学(资源服务)标识码(OSID):



0 引言

受岩性与风化条件的影响, 西南岩溶裂隙发育特征表现出极强的空间异质性。低渗透裂隙、粗大裂隙、管道及各种裂隙间的相互衔接、嵌套使岩溶裂隙系统具有复杂的水流特征和转化机制^[1], 水文地质空间结构差异对流域地下水补给、径流、排泄特征具有重要影响^[2]。水流在土壤/岩石裂隙-植被-大气连续体内运动且界面水分快速转化, “快速流”

与“慢速流”系统间水文连通性的时空演变^[3], 落水洞/漏斗等岩溶地貌对地表、地下系统水流的连接作用^[4], 形成了岩溶流域独特的产汇流过程。同时, 根据含水介质特征、水动力性质等因素划分的浅循环径流带与深循环径流带具有不同的径流特性^[5-6]。此外, 由于岩石裂隙、管道发育, 导致岩溶流域地下与地表边界不重合, 进而造成岩溶流域降雨-径流计算中的水量不平衡^[7], 这一问题在西南岩溶山区小流域尤为突出。因此, 模拟和预测反映岩溶流域降雨-泉

基金项目: 自然科学基金重点项目(41571130071); 面上项目(41971028, 41571020)

第一作者简介: 马从文(1999-), 男, 硕士研究生, 主要研究方向为岩溶流域水文。E-mail: 756073751@qq.com。

通信作者: 张志才(1980-), 男, 博士, 教授, 研究方向为岩溶水文学。E-mail: zhangzhicai_0@hhu.edu.cn。

收稿日期: 2022-07-27

流量特征的岩溶泉流量动态过程面临巨大的挑战。

机器学习模型从实测数据出发,通过识别隐藏在数据中的信息,建立各类预测因子与响应变量之间的线性或非线形映射关系,进而实现对响应变量的准确预测。该方法无需对复杂的物理过程进行描述,对非线性系统变量模拟预测具有超强的优势。作为一种常用的机器学习方法,神经网络模型(如前馈神经网络 BP,循环神经网络 recent neural network (RNN))已广泛应用于流域降雨-径流过程模拟^[8]。然而,在处理时间序列的数据时,普通的前馈人工神经网络模型缺乏对之前输入信息的记忆^[9-10],而 RNN 容易出现梯度消失和梯度爆炸的问题^[11-13],体现为模拟学习过程停滞不前,陷入局部最优解。在 RNN 的基础上,Hochreiter^[14]提出了带有记忆单元的长短时记忆神经网络(long short-term memory, LSTM)。LSTM 包含特殊的细胞状态和“门”结构,可动态地控制时间序列信息的流动和保存,其细致的内部处理单元,可以更好的处理长序列关联性强的时序数据^[15-16]。因此,与 RNN 相比,LSTM 可有效提高降雨-径流模拟精度^[11,17]。

近年来,LSTM 及其改进方法已应用于不同流域特征与时空尺度的降雨-径流模拟。如顾逸^[18]提出的基于年、月径流特性的 Simple-LSTM 模型,在流域径流模拟中表现出较高的模拟精度和计算效率。Kratzert^[10]等使用 LSTM 模型在美国 241 个流域进行了降雨-径流模拟,结果表明 LSTM 方法可应用于多种类型流域,其模拟精度与具有物理基础的流域

水文模型相似。研究发现,传统水文模型与 LSTM 方法结合,可有效提高降雨-径流过程模拟精度^[19]。同时,利用代表性强的短时间观测序列进行模型训练,也可得到较好的模型训练和径流预测效果^[20]。然而,将 LSTM 模型应用于岩溶流域降雨-径流过程,模拟高度空间异质的水文地质结构以及受干湿条件控制的产汇流过程,对其模型结构与模拟精度的要求仍有待进一步研究。

论文选取西南典型岩溶泉为研究对象,分别建立流域出口与山坡岩溶泉降雨-泉流量 LSTM 计算模型,利用实测气象、水文数据进行模型训练与验证。基于模拟结果,对比分析不同降雨-泉流量形成过程对岩溶泉流量模拟精度的影响,以及岩溶水文地质结构对降雨-泉流量响应滞时的影响。

1 研究区概况与数据

陈旗流域(26°15'36"~26°15'56"N, 105°43'30"~105°44'42"E)位于贵州省普定县的岩溶盆地,流域面积为 0.9 km²。流域地势东南高,西北低,三面环山,海拔 1 309~1 521 m。流域具有西南岩溶地区典型的峰丛山体-洼地地貌及岩溶水文地质特征,土壤层与表层岩溶带沿山坡向洼地逐渐增厚^[21]。流域内山体植被覆盖以灌草与林地为主,洼地主要为水田。流域岩性主要为以灰岩为主的碳酸盐岩,包括泥灰岩、厚层灰岩与薄层灰岩,少量白云岩出露地表,土层主要为碳酸盐岩土(图 1)。流域地理气候属于典型的

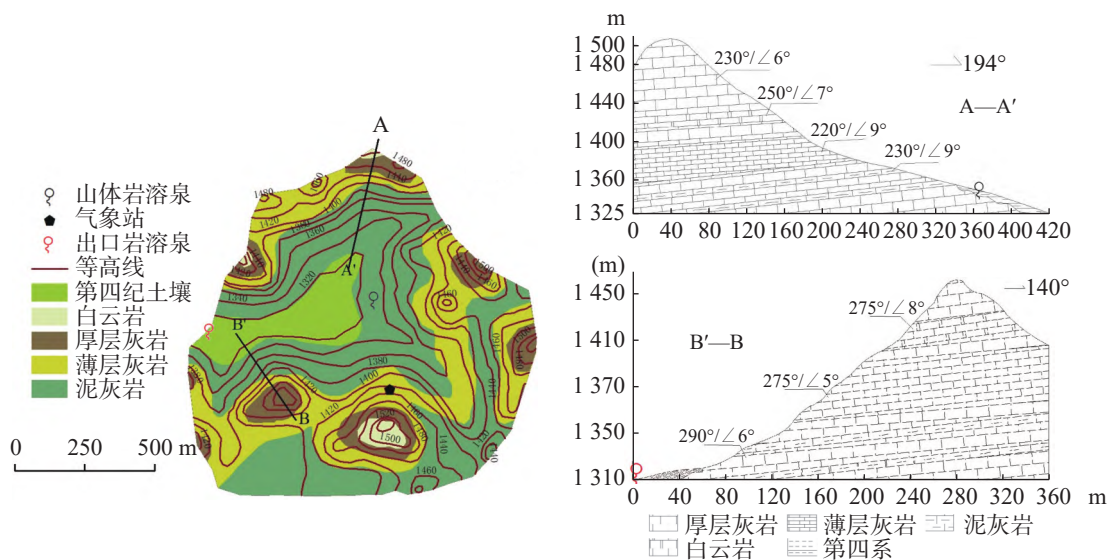


图 1 陈旗流域地形、岩性及水文气象观测

Fig. 1 Topography, lithology and hydrometeorological observations of the Chenqi watershed

东南亚热带季风气候,年平均气温在 15 ℃ 左右,雨季为 4 月到 10 月,枯季为 11 月到次年 3 月。该地区降雨时空分布不均,多年平均降雨量 1 336 mm,全年降雨最为集中的时段为 6—8 月。本文的研究对象为位于流域出口与山坡坡脚处的两眼岩溶泉(图 1),下文分别称为出口岩溶泉与山坡岩溶泉。

两处岩溶泉分别装有三角堰,利用水位计(HOBO U20-001-01 自计式水位采集自动记录仪)进

行堰前水位连续观测,并通过堰流公式进行泉流量换算。在流域内的山坡上设有小型气象站用于观测降雨(图 1),泉流量与降雨观测时间步长均为 5 分钟,分别通过平均与求和得到泉流量与降雨量的小时实测序列。本研究选取 2017 年 1 月 1 日 0 时—2019 年 12 月 31 日 24 时实测序列进行模拟研究。研究期内的降雨及岩溶泉流量统计特征如表 1 所示。

表 1 研究期降雨、泉流量统计特征

Table 1 Statistical characteristics of rainfall and spring flow during the study period

站点	降雨量		泉流量/m ³ ·s ⁻¹		
	时段总量/mm	最大值/mm·h ⁻¹	最大值	最小值	均值
山坡岩溶泉	3 584.2	59.9	0.002 3	0	0.000 2
出口岩溶泉			0.164 6	0	0.006 8

2 研究方法

2.1 LSTM 模型

虽然 LSTM 与 RNN 一样都具有重复的链状结构单元(图 2a)用来记住历史信息,但 LSTM 的链状结构单元中包含更加复杂的运算。LSTM 在隐藏层中于加入了类似“传送带”的细胞状态,并由遗忘门、输入门、输出门来调整细胞状态^[11](图 2b)。其中,遗忘门决定了 t 时刻细胞状态需要移除的 t-1 时刻细胞状态的信息,输入门决定了 t 时刻细胞状态需要存储的新信息,输出门决定了 t 时刻细胞状态需要输出的信息,而 t 时刻的细胞状态则记录了 t 时刻的输入、门结构信息及 t-1 时刻隐藏层状态、t-1 时刻细胞状态。这种结构能够有效解决时间序列数据的长期运

输和记忆问题,比传统神经网络算法更适合用在径流模拟与预报上。

泉流量的变化表征了泉域蓄量动态特征^[19],加之模型计算步长为 1 h,蒸散发对当前时段内的泉流量影响较小,因此本研究中,将降雨量和泉流量作为模型输入变量。通过输入层的当前时刻输入变量:降雨量(P_(t))与前时段泉流量(Q_(t-1)),以及前一时段的隐藏层状态变量(h_(t-1))和细胞状态变量(S_(t-1)),经由遗忘门、输入门、输出门三个内部控制机制,得到本时段隐藏层状态变量(h_(t))和细胞状态变量(S_(t)),再传递至下一时段运算,同时,输出本时段泉流量(Q_(t))至输出层。LSTM 模型的关键是细胞状态,细胞状态公式表达如下:

$$f_i = \sigma(W_f \cdot [h_{t-1}, i_t] + b_f) \quad (1)$$

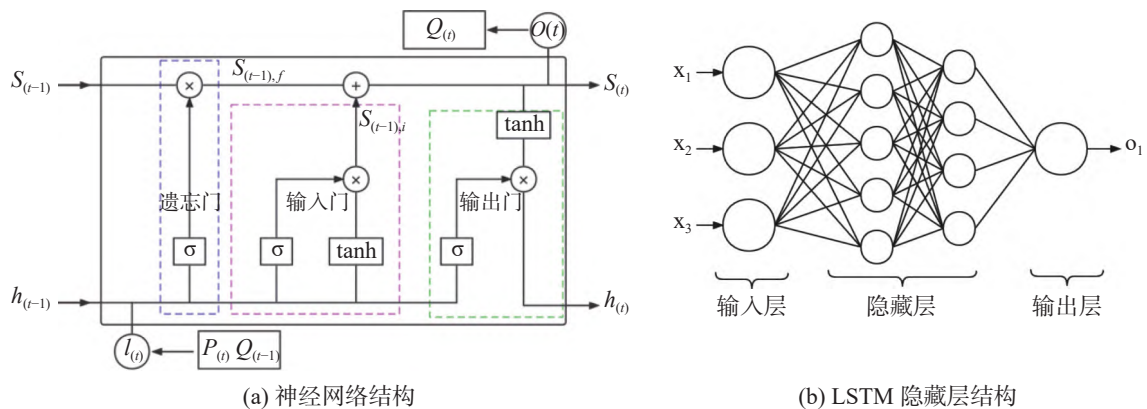


图 2 LSTM 模型原理示意图

Fig. 2 Schematic diagram of LSTM model: neural network structure (a); LSTM hidden layer structure (b)

$$x_t = \sigma(W_i \cdot [h_{t-1}, i_t] + b_i) \quad (2)$$

$$\tilde{S}_t = \tanh(W_s \cdot [h_{t-1}, i_t] + b_s) \quad (3)$$

$$S_t = f_t \cdot S_{t-1} + x_t \cdot \tilde{S}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, i_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(S_t) \quad (6)$$

式中: W_f 、 W_i 、 W_o 分别为遗忘门、输入门和输出门中关于输入向量的权重矩阵; W_s 为 CEC 单元状态权重矩阵; f_t 、 x_t 、 o_t 分别为 t 时刻遗忘门、输入门和输出门关于 CEC 单元的权重系数; \tilde{S}_t 为暂存变量; S_t 为细胞状态变量; b_f 、 b_i 、 b_s 、 b_o 为偏置向量; h_t 为 t 时刻隐含层输出向量; i_t 为 t 时刻输入向量; σ 和 \tanh 为激活函数^[11]。

利用 LSTM 模型模拟泉流量过程,模型从训练样本中训练得到细胞状态、遗忘门、输入门和输出门相关参数。此外,模型自带的隐藏层数、神经元数量、训练轮数、批处理量、学习率、时序步长等超参数无法通过训练获得,需在训练前预先定义或设置^[9]。

2.2 模型评价指标

为了量化 LSTM 模型模拟的精度,本研究采用两种量化指标,分别为 Nash-Sutcliffe 效率系数(NSE)和拟合优度(R^2),具体计算公式如下:

$$NSE = 1 - \frac{\sum_{t=1}^T (Q'_t - Q'_m)^2}{\sum_{t=1}^T (Q'_t - \bar{Q}_o)^2} \quad (7)$$

$$R^2 = \frac{\left[\sum_{t=1}^T (Q'_t - \bar{Q}_o)(Q'_m - \bar{Q}_m) \right]^2}{\sum_{t=1}^T (Q'_t - \bar{Q}_o)^2 (Q'_m - \bar{Q}_m)^2} \quad (8)$$

式中: Q'_t 为 t 时刻实测泉流量; Q'_m 为 t 时刻模拟泉流量; \bar{Q}_o 为实测泉流量平均值; \bar{Q}_m 为模拟泉流量平均值; T 为模拟开始后第 T 时刻。

3 结果与分析

3.1 模型训练与验证

本文以 2017 年 1 月 1 日 0 时—2018 年 12 月 31 日 24 时为模型训练期,共 17 520 个时段;2019 年 1 月 1 日 0 时—2019 年 12 月 31 日 24 时为模型验证期,共 8 760 个时段。模型超参数根据经验值与手动调参确定,取值如表 2 所示。

模型训练期与验证期泉流量模拟结果如图 3 所示。结果显示,训练期山坡岩溶泉 R^2 与 NSE 分别为 0.988 和 0.942,出口岩溶泉 R^2 与 NSE 分别为 0.989 和 0.951。验证期模型结果总体可以反映泉流量动态过程,山坡岩溶泉 R^2 与 NSE 分别为 0.943 和 0.831,出口岩溶泉 R^2 与 NSE 分别为 0.954 和 0.834,但在洪水过程模拟中,尤其对雨季暴雨洪水过程中洪峰模拟存在低估(图 3)。

3.2 降雨-泉流量过程形成特征对泉流量模拟结果的影响

针对模型结果中验证期枯季与雨季泉流量过程模拟精度存在较大差异的情况,本文将验证期划分为枯季(2019 年 1 月 1 日—2019 年 3 月 31 日,2019 年 11 月 1 日—2019 年 12 月 31 日)和雨季(2019 年 4 月 1 日—2019 年 10 月 31 日)时段,通过模型模拟,分别对比分析模型对不同季节泉流量过程的模拟效果。结果显示,山坡岩溶泉枯季 NSE 与 R^2 为 0.973 与 0.987,雨季为 0.793 与 0.942,出口岩溶泉枯季 NSE 与 R^2 为 0.982 与 0.993,雨季为 0.798 与 0.935(表 3)。枯季山坡岩溶泉与出口岩溶泉流量模拟精度均显著高于雨季,此结果与 Cho 等^[19]的研究结论相一致。LSTM 的核心是训练建立输入和输出之间

表 2 超参数取值

Table 2 Hyperparameter values

超参数名称	含义	取值
隐藏层数(layer)	神经网络结构层数	2
单个隐藏层神经元数(hidden size)	隐藏层节点的数量	10
训练轮数(epoch)	模型进行完整训练的次数	100
批量大小(batch size)	一次训练所选取的样本数	32
学习率(learning rate)	神经网络的学习速度	0.001
时序步长(time step)	山坡岩溶泉	15
	出口岩溶泉	28

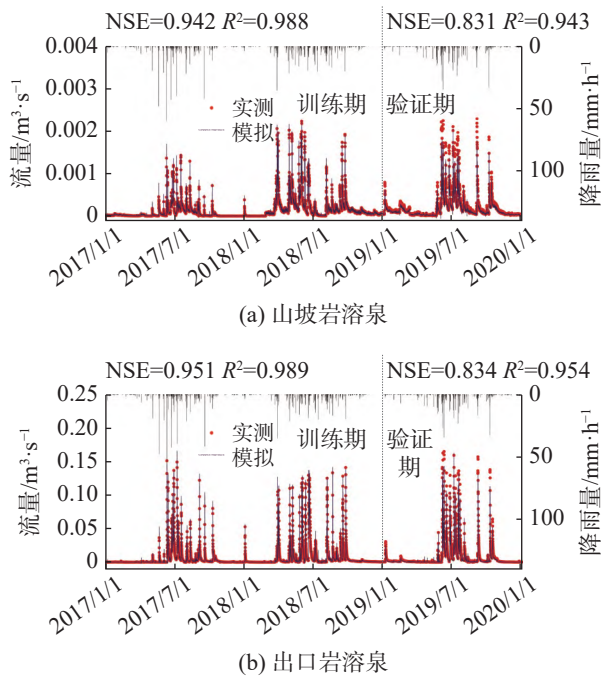


图3 山坡岩溶泉与出口岩溶泉 LSTM 模型训练期与验证期泉流量模拟结果

Fig. 3 Simulation results of spring flow in the training and validation period of LSTM model: hillside karst spring (a); outlet karst spring (b)

表3 基于全年数据训练的验证期雨、枯季 LSTM 模拟结果
Table 3 LSTM simulation results for the rainy and dry seasons in validation period trained based on annual data

时间	山坡岩溶泉		出口岩溶泉	
	NSE	R ²	NSE	R ²
训练期	0.942	0.988	0.951	0.989
验证期	枯季	0.973	0.987	0.993
	雨季	0.793	0.942	0.798

的函数关系,并以此进行预测,其函数关系的建立总体划分为分类问题与回归问题^[14],以数值型数据为输入的泉流量模拟即属于回归问题。LSTM 在训练过程中,当局部数据与整体数据差异性较大时,局部回归建立的函数关系会被整体所影响甚至忽略,进而产生局部数据预测误差。

为研究流域不同降雨-泉流量特征对泉流量过程模拟的影响,论文以原训练期(2017年1月1日—2018年12月31日)中雨季实测序列(2017年4月1日—2017年10月31日,2018年4月1日—2018年10月31日)作为训练样本进行模型训练,并以此模型模拟雨季(2019年4月1日—2019年10月31日)岩溶泉流量。结果显示:利用雨季实测序列训

练的模型模拟山坡岩溶泉雨季泉流量 NSE 为 0.956, R² 为 0.975, 出口岩溶泉 NSE 和 R² 分别为 0.962 与 0.977, 显著高于利用全年实测序列训练的模型(表 3, 图 4)。

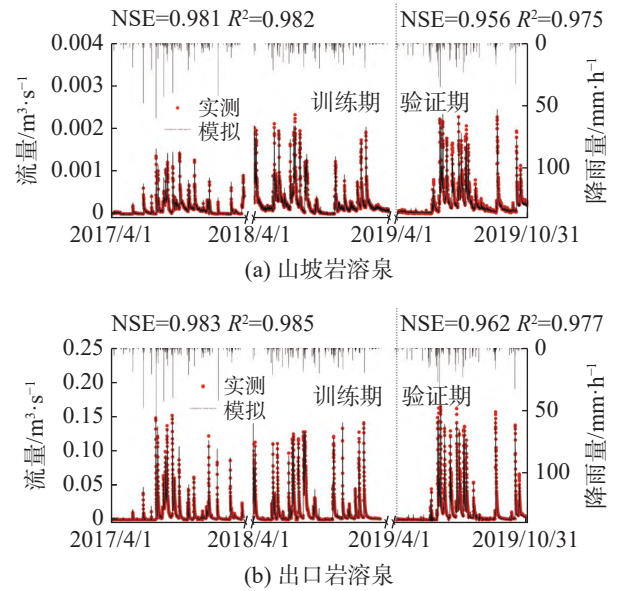


图4 基于雨季实测序列训练的模型模拟结果

Fig. 4 Simulation results from the model trained based on the sequence in the rainy season: hillside karst spring (a); outlet karst spring (b)

岩溶流域水力连通性受流域干湿条件的影响显著,对流域产汇流过程及其变异具有控制作用。雨季山坡与洼地地貌单元,地表与地下管道水力连通性显著增强。暴雨过程中,通过大裂隙、落水洞等快速水流通道的的水流在泉流量中的比例迅速增大^[22]。暴雨洪水后期,随着含水层水位下降,泉流量中快速流比重快速下降;枯季土壤含水率较低,岩溶含水层水位相对较低,泉流量主要为来自于储存在土壤与细小裂隙中的慢速水流^[23]。此外,岩溶流域降雨-径流过程与植被生理、生态过程密切相关^[24],植被变化对产汇流过程与流域调蓄能力产生重要影响^[25-27]。基于全年实测序列进行模型训练,模型结构反映了雨季与枯季岩溶泉域综合的产汇流特征,其对雨季降雨-泉流量过程,特别是暴雨过后陡涨陡落的洪水过程模拟精度较低。雨季暴雨过程中泉流量峰值存在显著低估现象,表明模型没有反映出暴雨过程中快速水流对岩溶泉的补给作用(图 3)。枯季泉流量较小且相对稳定,表明模型对其模拟仍然具有较高精度,与之对应,利用汛期实测序列进行模型训练,

模型较好地反映了雨季岩溶泉流量形成特征,模型对雨季暴雨-岩溶泉流量过程的模拟精度显著提高。

为定量分析泉域干湿条件对岩溶泉流量 LSTM 模型模拟精度的影响,本文以 2017—2018 年枯季实测序列(2017 年 4 月 1 日—2017 年 10 月 31 日,2018 年 4 月 1 日—2018 年 10 月 31 日)训练模型,则模型反映枯季降雨-泉流量形成特征。然后,以此模型模拟 2019 年雨季岩溶泉流量(2019 年 4 月 1 日—2019 年 10 月 31 日),则模拟结果反映了雨季降雨在枯季降雨-泉流量过程模式下所形成的岩溶泉流量过程,将此模拟结果与实测泉流量的差值记为 R1。根据雨季序列训练的模型模拟所得到的模拟结果与实测泉流量的差值记为 R2,则可用 R1-R2 表征干湿条件改变(即降雨-泉流量形成特征变化)对岩溶泉流量模拟结果的影响。图 5 表示不同降雨-泉流量形成特征下 LSTM 模型的雨季泉流量模拟结果。结果表明,基于枯季实测序列训练的模型对雨季泉流量的模拟误差比雨季序列训练模型高 27.5%~76.2%,5、6、7 月差异尤为显著。该地区暴雨洪水过程主要集中在

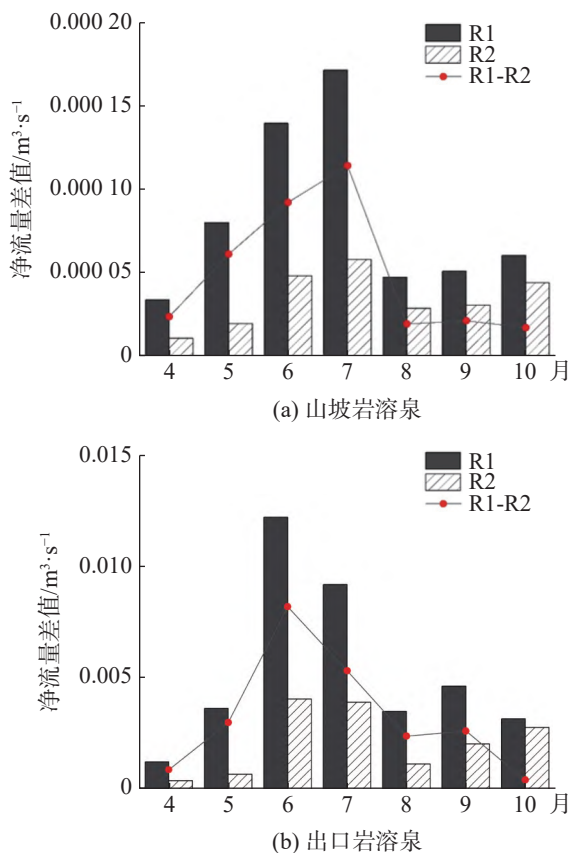


图 5 基于不同降雨-泉流量形成特征模型的泉流量模拟结果
Fig. 5 Simulation results of spring flow based on different models: hillside karst spring (a); outlet karst spring (b)

5、6、7 月,降雨量占全年的 61.3%,其产汇流特征与枯季差异显著,因此反映两种产汇流特征的模型模拟结果精度具有明显差异。

3.3 岩溶水文地质结构对降雨-泉流量响应滞时的影响

时序步长是区分 LSTM 与一般神经网络的关键,其在 LSTM 模型中表征历史信息传递长短。LSTM 由常量误差传递单元(CEC)和三个特殊的“门”结构实现历史信息的储存与更新^[3],将 LSTM 应用于流域径流模拟时,可将 CEC 理解为流域当前状态。其由前 N 天的流域降雨与径流共同决定,N 为时序步长,即降雨径流响应滞时(前期影响时长)。根据建立的 LSTM 模型,分别改变模型的时序步长,对比分析不同时序步长条件的泉流量模拟结果(图 6)。结果显示,山坡岩溶泉 LSTM 模型的时序步长低于出口岩溶泉,尤其当模型具有较高的模拟精度时(如 $NSE > 0.78$),二者的时序步长差异显著。山坡岩溶泉与出口岩溶泉时序步长分别为 15 h 与 28 h 时,各自的泉流量模拟精度最高(图 6)。

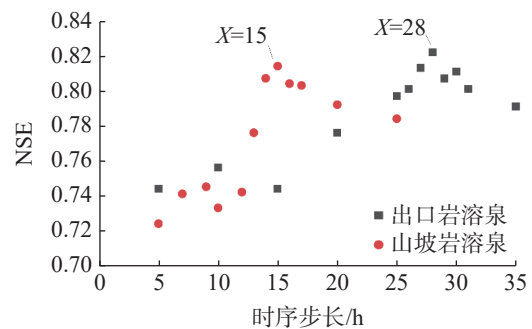


图 6 时序步长与 LSTM 模型精度的关系

Fig. 6 Relationship between timing steps and LSTM

表层岩溶带厚度对岩溶流域流量降雨响应特征具有重要影响^[28]。岩溶区,表层岩溶带发育程度与地形具有密切关系,其厚度总体随坡度的增加而降低,一般陡峭山坡表层岩溶带厚度小于平坦的洼地,且沿山坡由上至下逐步增大^[21]。与洼地相比,受浅薄表层岩溶带控制,山坡单元含水层中水流迅速,更易形成以较大裂隙为通道的快速水流,进而表现为山坡岩溶泉快速的流量降雨响应特征。因此,受表层岩溶带发育特征影响,山坡地貌单元水流对降雨的响应滞时小于洼地,即其岩溶泉流量 LSTM 模型时序步长显著低于出口岩溶泉。因此,岩溶水文地质结构特征为 LSTM 模型时序步长的确定提供了参

考依据与物理解释。

4 结论

(1) 论文针对岩溶降雨-泉流量过程, 选取了西南地区不同地貌单元的典型岩溶泉, 采用双隐层, 双输入的 LSTM 模型, 构建了典型岩溶泉流量机器学习模型, 利用研究区实测序列进行了模型训练与验证。山坡岩溶泉与出口岩溶泉训练期纳什效率系数 (NSE) 分别为 0.943、0.951, 验证期分别为 0.831、0.834。结果表明模型能较好地模拟不同地貌单元岩溶泉流量的整体动态过程。

(2) 岩溶泉域降雨-泉流量形成过程及其变异对机器学习模型的模拟精度具有重要影响。与利用全年实测序列训练模型相比, 利用雨季实测序列训练的模型可显著提高对雨季岩溶泉流量的模拟精度, 山坡岩溶泉 NSE 由 0.793 提高至 0.956, 出口岩溶泉由 0.798 升高至 0.962, 且此差异在暴雨-洪水集中的 5、6、7 月最为显著。

(3) 岩溶水文地质结构对降雨-泉流量响应滞时有影响。山坡岩溶泉与出口岩溶泉时序步长分别为 15 h 与 28 h 时各自的模拟精度最高, 表明受表层岩溶带发育浅薄的影响, 相较于洼地区域的岩溶泉, 山坡泉域具有更小的降雨-泉流量响应滞时。

参考文献

- [1] Hartmann A, Goldscheider N, Wagener T, Lange J, Weiler M. Karst water resources in a changing world: Review of hydrological modeling approaches[J]. *Reviews of Geophysics*, 2014, 52(3): 218-242.
- [2] 李潇, 漆继红, 许模. 西南典型狭窄褶皱小尺度浅层岩溶水系统特征及隧道涌水分析[J]. *中国岩溶*, 2020, 39(3): 375-383. LI Xiao, QI Jihong, XU Mo. Analysis on the characteristics of small-scale shallow karst water systems in typical tight-narrow folds and tunnel water inrush in Southwestern China[J]. *Carsologica Sinica*, 2020, 39(3): 375-383.
- [3] Zhang Zhicai, Chen Xi, Soulsby Chris. Catchment-scale conceptual modelling of water and solute transport in the dual flow system of the karst critical zone[J]. *Hydrological Processes*, 2017, 31(19): 3421-3436. doi:10.1002/hyp.11268.
- [4] 蒙海花, 王腊春, 苏维词, 霍雨. 基于落水洞的岩溶半分布式水文模型的构建及其应用[J]. *地理科学*, 2009, 29(4): 550-554. MENG Haihua, WANG Lachun, SU Weici, HUO Yu. Development of a karst sinkhole-based semi-distributed hydrological model and its application[J]. *Scientia Geographica Sinica*, 2009, 29(4): 550-554.
- [5] 王宇. 西南岩溶区岩溶水系统分类、特征及勘查评价要点[J]. *中国岩溶*, 2002, 21(2): 114-119. WANG Yu. Classification, features of karst water system and key point for the evaluation to karst water exploration in Southwest China karst area[J]. *Carsologica Sinica*, 2002, 21(2): 114-119.
- [6] 王宇. 岩溶高原地下水径流系统垂向分带[J]. *中国岩溶*, 2018, 37(1): 1-8. WANG Yu. Vertical zoning of groundwater runoff system in karst plateau[J]. *Carsologica Sinica*, 2018, 37(1): 1-8.
- [7] 王在高, 徐萍莉. 浅析喀斯特流域水文地貌过程—响应系统[J]. *贵州师范大学学报(自然科学版)*, 2002, 20(1): 36-39. WANG Zaigao, XU Pingli. An elementary analysis of hydrogeomorphological process—response system[J]. *Journal of Guizhou Normal University (Natural Sciences)*, 2002, 20(1): 36-39.
- [8] Hu Caihong, Wu Qiang, Li Hui, Jian Shengqi, Li Nan, Lou Zhengzheng. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation[J]. *Water*, 2018, 10(11): 1543.
- [9] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge, USA: The MIT Press, 2016.
- [10] Kratzert F, Klotz D, Brenner C, Schulz K. Rainfall-runoff modelling using Long Short-Term Memory(LSTM) networks[J]. *Hydrology and Earth System Sciences*, 2018, 22(11): 6005-6022.
- [11] 党池恒, 张洪波, 陈克宇, 支童, 卫星辰. 长短期记忆神经网络在季节性融雪流域降水-径流模拟中的应用[J]. *华北水利水电大学学报(自然科学版)*, 2020, 41(5): 10-18, 33. DANG Chiheng, ZHANG Hongbo, CHEN Keyu, ZHI Tong, WEI Xingchen. Application of the long-short-term memory neural network for rainfall-runoff simulation in seasonal snowmelt basin[J]. *Journal of North China University of Water Resources and Electric Power (Natural Science Edition)*, 2020, 41(5): 10-18, 33.
- [12] Y Bengio, P Simard, P Frasconi. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 157-166.
- [13] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions[J]. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1998, 6(2): 107-116.
- [14] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [15] Gao Shuai, Huang Yuefei, Zhang Shuo, Han Jingcheng, Wang Guangqian, Zhang Meixin, Lin Qingsheng. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation[J]. *Journal of Hydrology*, 2020, 589: 125188.
- [16] Zhang Di, Lin Junqiang, Peng Qidong, Wang Dongsheng, Yang Tiantian, Soroosh Sorooshian, Liu Xuefei, Zhuang Jiangbo.

- Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm[J]. *Journal of Hydrology*, 2018, 565: 720-736.
- [17] Tennant C, Larsen L, Bellugi D, Moges E, Zhang Liang, Ma Hongxu. The utility of information flow in formulating discharge forecast models: A case study from an arid snow-dominated catchment[J]. *Water Resources Research*, 2020, 56(8): e2019WR024908.
- [18] 顾逸. 基于长短期记忆循环神经网络及其结构约减变体的中长期径流预报研究[D]. 武汉: 华中科技大学, 2018.
GU Yi. Research on mid-long-term runoff forecasting based on long-short-term memory recurrent networks and its structural reduction variant[D]. Wuhan: Huazhong University of Science and Technology, 2018.
- [19] Cho K, Kim Y. Improving streamflow prediction in the WRF-Hydro model with LSTM networks[J]. *Journal of Hydrology*, 2022, 605: 127297.
- [20] Lu Dan, Goutam Konapala, Scott L Painter, Shih Chieh Kao, Sudershan Gangrade. Streamflow simulation in data-scarce basins using bayesian and physics-Informed machine learning models[J]. *Journal of Hydrometeorology*, 2021, 22(6): 1421-1438.
- [21] 张志才, 陈喜, 刘金涛, 彭韬, 石朋, 严小龙. 喀斯特山体地形对表层岩溶带发育的影响: 以陈旗小流域为例[J]. *地球与环境*, 2012, 40(2): 137-143.
ZHANG Zhicai, CHEN Xi, LIU Jintao, PENG Tao, SHI Peng, YAN Xiaolong. Influence of topography on epikarst in karst mountain areas: A case study of Chenqi catchment[J]. *Earth and Environment*, 2012, 40(2): 137-143.
- [22] Zhang Zhicai, Chen Xi, Cheng Qinbo, Soulsby Chris. Using Storage Selection(SAS) functions to understand flow paths and age distributions in contrasting karst groundwater systems[J]. *Journal of Hydrology*, 2021, 602: 126785.
- [23] Zhang Zhicai, Chen Xi, Cheng Qinbo, Soulsby Chris. Storage dynamics, hydrological connectivity and flux ages in a karst catchment: Conceptual modelling using stable isotopes[J]. *Hydrology and Earth System Sciences*, 2019, 23(1): 51-71.
- [24] Cai Lianbin, Chen Xi, Huang Richao, Keith Smettem. Runoff change induced by vegetation recovery and climate change over carbonate and non-carbonate areas in the karst region of Southwest China[J]. *Journal of Hydrology*, 2022, 604: 127231.
- [25] 侯怡萍. 流域生态水文敏感度及其影响因子分析[D]. 成都: 电子科技大学, 2019.
HOU Yiping. Research on ecohydrological sensitivity of watershed and its influencing factors analysis[D]. Chengdu: University of Electronic Science and Technology of China, 2019.
- [26] 刘炜, 焦树林, 李银久, 莫跃爽, 赵宗权, 张洁, 赵梦. 喀斯特地表植被覆盖变化及其与气候因子相关性分析[J]. *水土保持研究*, 2021, 28(3): 203-215.
LIU Wei, JIAO Shulin, LI Yinjiu, MO Yueshuang, ZHAO Zongquan, ZHANG Jie, ZHAO Meng. Analysis on the correlation between vegetation cover of land surface and climatic factors in karst area[J]. *Soil and Water Conservation Research*, 2021, 28(3): 203-215.
- [27] 李林立. 西南典型岩溶区生态环境对表层岩溶水调蓄功能的影响研究[D]. 重庆: 西南大学, 2009.
LI Linli. Study of effects of ecological environment on regulated function of epikarst water in typical karst area of Southwest, China[D]. Chongqing: Southwest University, 2009.
- [28] Zhang Zhicai, Chen Xi, Chen Xunhong, Shi Peng. Quantifying time lag of epikarst-spring hydrograph response to rainfall using correlation and spectral analyses[J]. *Hydrogeology Journal*, 2013, 21(7): 1619-1631.

Modelling karst spring flow in Southwest China based on machine learning

MA Congwen¹, ZHANG Zhicai¹, CHEN Xi², CHENG Qinbo¹, PENG Tao^{3,4}, ZHANG Lin^{3,4}

(1. College of Hydrology and Water Resources, Hohai University, Nanjing, Jiangsu 210098, China; 2. School of Earth System Science, Tianjin University, Tianjin 300072, China; 3. Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, Guizhou 550081, China; 4. Puding Karst Ecosystem Observation and Research Station, Chinese Academy of Sciences, Puding, Guizhou 562100, China)

Abstract Karst springs are important for ecosystem and economic development in Southwest China. Controlled by the unique karst hydrogeological structure and multiple water flow processes, the karst spring flow has complex dynamic characteristics, thus posing a great challenge to simulate and predict the dynamic process of karst spring flow which can reflect the characteristics of rainfall–spring flow in the karst basin. As a data-driven model, the machine learning model omits the necessity of considering complex physical processes, showing its significant advantages in the simulation and prediction of nonlinear system variables. Therefore, it provides an effective approach for simulation and prediction of karst spring discharge. However, the influence of the flow processes and hydrogeological conditions on the structure and simulation accuracy of machine learning model is still unclear. Among the machine learning algorithms, LSTM, as the most popular algorithm in recent years, is widely used in the simulation and prediction of

various long-time series data. LSTM adds a cell state similar to "conveyor belt" in the hidden layer, and the cell state is adjusted by forgetting gate, input gate and output gate. This structure can effectively solve the long-term transportation and memory problems of time series data, and is more suitable for runoff simulation and prediction than the traditional neural network algorithm.

In this study, for the processes of rainfall–spring flow in karst areas, two typical karst springs (hillside karst spring and outlet karst spring) representing different geomorphic units in Southwest China are selected. Through hyper parameter optimization, a double hidden layer and double input LSTM model are adopted to build a machine learning model of typical karst spring flow. The measured meteorological and hydrological data from 0:00 on January 1, 2017 to 24:00 on December 31, 2019 are used. 2017–2018 is the training period and 2019 is the validation period. The model has been trained and verified. Based on the simulation results, the influence of different rainfall–spring flow forming processes on the simulation accuracy of karst spring flow and the influence of karst hydrogeological structure on the response time lag of rainfall–spring flow are compared and analyzed. The results show that the Nash efficiency coefficients (NSE) for the hillside karst spring and the outlet karst spring were 0.942 and 0.951 in the training period, and 0.831 and 0.834 in the validation period, respectively. The model can well simulate the whole dynamic process of karst spring discharge in different geomorphic units, but there are significant errors in the simulation of flood peak in the rainy season. The formation process and variation of rainfall–spring discharge in the karst spring area have an important influence on the simulation accuracy of machine learning model. Compared with the model trained by the annual measured sequence, the model trained by the measured sequence in the rainy season can significantly improve the simulation accuracy of the karst spring flow in this season. The NSE of the hillside karst spring increases from 0.793 to 0.956, and the outlet karst spring increases from 0.798 to 0.962. The difference is most significant in May, June and July when rainstorms and flood are concentrated. Under the same precision of simulating spring flow, the time step of LSTM model of hillside karst spring is obviously smaller than that of outlet karst spring. When the simulation precision of the two types of spring flow is the highest, the time step of the model is 15 h and 28 h, respectively. This result can be combined with the actual geological structure. In karst areas, the development degree of the epikarst zone is closely related to the terrain, and its thickness usually decreases with the increase of the slope. Generally, the thickness of the epikarst zone on steep hillsides is less than that in flat depressions, and gradually increases from top to bottom along the hillsides. Compared with the water flow in the depression, the water flow in the aquifer of the hillside unit is rapid because it is controlled by the shallow epikarst zone, and it is easier to form the fast water flow with the large fissure as the channel, which may further show the rapid flow and rainfall response characteristics of the hillside karst spring. Therefore, affected by the development characteristics of the epikarst zone, the response lag time of rainfall–spring flow in the hillside karst spring area is less than that in the depression, and the time step of LSTM model of hillside karst spring flow is significantly smaller than that of the karst spring at the outlet of the basin.

Key words machine learning, LSTM, karst spring flow, flow hysteresis to rainfall, rainfall–spring flow process in the karst area

(编辑 杨杨)