

DOI: 10.19797/j.cnki.1000-0852.20220047

# 典型北方山前岩溶泉历史流量序列重建研究

许亮<sup>1,2</sup>, 郭高轩<sup>1,2</sup>

(1. 北京市地质环境监测所, 北京 100195;  
2. 城市地下水安全防控技术创新基地, 北京 100195)

**摘要:** 泉水流量序列的缺失严重影响对水文地质条件变化的认识, 故对缺失的泉水流量序列重建具有重要意义。以北方某典型山前岩溶大泉群实测流量、流入下游湖水流量和补给区降水量为研究对象, 基于泉水动态的周期性和趋势性, 采用多元线性回归模型、考虑泉流滞后降水的多元线性回归模型、SARIMA模型和SARIMA与多元线性回归相结合模型, 开展了1961—1974年缺失的该典型大泉群流量序列的重建研究。结果表明采用SARIMA和多元线性回归相结合模型重建的流量结果较合理。

**关键字:** 泉水流量; 降水量; 入湖流量; 数据重建

中图分类号: P641.74; P333; TV121

文献标识码: A

文章编号: 1000-0852(2023)03-0088-05

## 0 引言

泉水是地下水的天然露头, 泉水动态承载着重要的水文信息, 其动态流量序列一般采用实测获得<sup>[1-3]</sup>, 在实际中由于自然或者人为的原因, 往往造成实测泉水动态流量序列缺失或不完整, 从而对水文地质条件和规律的认识以及地下水资源的计算与评价造成较大影响。因此, 泉水流量序列的重建对于系统研究区域的水文地质条件, 揭示地下水运移、富集规律和优化泉水利用方案等具有重要意义。

泉水动态数据重建的主要方法有移动平均模型、指数平滑模型、逐步回归模型、多元线性回归和神经网络模型等。如朱燕燕<sup>[4]</sup>分别采用移动平均模型和指数平滑模型还原了2000—2019年娘子关泉流量序列。姜宝良<sup>[5]</sup>采用逐步回归分析方法对河南百泉泉流量进行了拟合。张建芝<sup>[6]</sup>采用多元线性回归模型, 模拟降水和人工开采对地下水动态影响分析, 为泉水保护提供了依据。姜宝良<sup>[7]</sup>根据百泉泉水流量与降水量之间的关系, 构建了BP神经网络模型, 预测了天然状态下的百泉泉水流量。但移动平均方法<sup>[8]</sup>适合模拟流

量较稳定, 季节变化不明显的泉流量。指数平滑模型是使用本期指数平滑值作为下期的预测值, 未考虑相关因素对泉流量的影响。逐步回归模型和多元线性回归模型均未能表达年内周期性的泉水动态特征, 而神经网络模型<sup>[9]</sup>则需要的样本较多。综上, 各研究方法较难拟合年内泉流量的季节变化特征。

本文以北方某山前典型大泉群为实例, 整理了该泉群历史监测原始资料, 根据该泉群实测流量数据、泉群流入下游湖泊的流量数据和补给区降水量数据, 分别采用多元线性回归模型、SARIMA模型、SARIMA和多元线性回归相结合的模型对1961—1974年间缺失的泉群流量数据进行重建研究, 以为该泉群的进一步研究和工程实践提供支撑, 为其他岩溶大泉进行缺失的历史流量序列重建研究提供借鉴。

## 1 典型泉水概况

选取某一北方大型泉, 该泉位于山前, 在地面出露多个泉眼, 属于典型北方山前岩溶泉。该泉补给区位于泉口西南近20 km的大片岩溶裸露区, 含水层主要为奥陶系石灰岩。大气降水和河道渗漏是岩溶水

收稿日期: 2022-01-22

网络首发日期: 2023-06-13

网络首发地址: <https://kns.cnki.net/kcms2/detail/11.1814.P.20230612.1347.010.html>

基金项目: 北京市百千万人才工程(2020A55); 北京市科学技术协会学术示范项目—北京西山岩溶裂隙水环境演变及其生态效应

作者简介: 许亮(1984—), 男, 河北石家庄人, 硕士, 高级工程师, 主要从事水文地质、环境地质及城市地质方面的研究工作。E-mail: 307711670@qq.com

通信作者: 郭高轩(1979—), 男, 陕西乾县人, 博士, 教授级高级工程师, 主要从事水文、环境地质及城市地质方面的研究工作。E-mail: ggx2008@qq.com

的主要补给来源,地下水沿地层向处于东北端的隐伏灰岩地区流动,东部受透水性极差的第四系泥砾层和含碎石红粘土层阻挡成泉。

该泉仅有1950—1960年的实测流量序列。1961年以后,该泉泉域干旱少雨,人工大规模开采第四系孔隙水,造成岩溶水顶托补给第四系地下水,改变了该泉泉域裂隙岩溶水的天然排泄方式,使泉水在1974年断流,缺失了1961—1974年流量序列。后仅在丰水年的丰水期有短期复出,下游干涸,湿地也逐渐消失<sup>[10-11]</sup>。

## 2 数据来源

(1)降水量:采用泉域直接补给区的雨量站1959—1974年日降水观测数据,选取月尺度数据。数据来源于北方某气象局补给区附近监测站1959—1974年逐日降水量数据。

(2)诸泉流量:该泉与其他泉水汇合一起出口处有监测点,监测的流量称为“诸泉流量”。采用1950—1960年诸泉流量月尺度数据。数据来源于北方某水文队于1950—1960年监测的月尺度数据。

(3)入湖流量:诸泉排泄后与下游泉流汇合,共同流入下游湖泊,以入湖前监测点的监测流量作为“入湖流量”。数据来源于北方某水文队于1958—1974年监测的入湖流量月尺度数据。

1960年以前,诸泉流量动态较稳定,主要随季节变化,一般在月降水量高峰过后1个月升至最高。年内诸泉流量最大值出现在雨季开始后半个月至2个月内。据1950—1960年观测资料显示,诸泉月平均流量 $1.33 \text{ m}^3/\text{s}$ ,月平均最大流量 $2.59 \text{ m}^3/\text{s}$ (1959年8月),月平均最小流量 $0.558 \text{ m}^3/\text{s}$ (1954年5月)。诸泉入湖月平均流量 $0.86 \text{ m}^3/\text{s}$ ,月平均最大流量 $1.72 \text{ m}^3/\text{s}$ (1969年10月),月平均最小流量 $0.13 \text{ m}^3/\text{s}$ (1970年4月)。流量极值之比为4.64,降水量、诸泉流量及入湖流量变化详见图1。

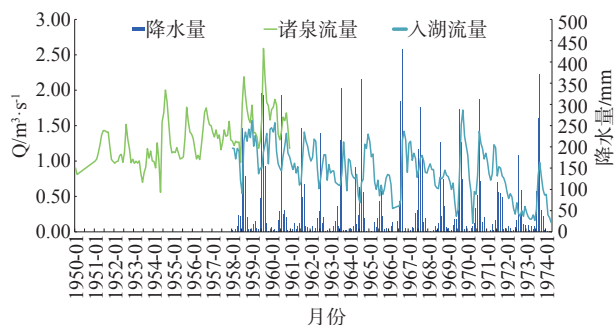


图1 历史观测数据

Fig.1 Historical observation data

## 3 研究方法与研究结果

### 3.1 研究方法

分别采用多元线性回归模型、滞后多元线性回归模型、SARIMA模型、SARIMA和多元线性回归相结合的模型进行数据系列重建。

(1)多元线性回归模型<sup>[6]</sup>。该方法主要用于拟合多种变量之间的线性变化关系,并且使拟合误差最小。在多元线性回归中多元表示为多个变量之间线性关系。

(2)季节性差分自回归滑动平均模型(简称SARIMA)。

SARIMA来源于自回归单整移动平均模型(ARIMA),主要用于分析由于季节性变化包括(年度、季度、月度、周度等变化)或因其他因素引起的具有周期性变化的时间序列,能够利用历史数据对模型进行实时调整<sup>[12]</sup>。它充分考虑了时间序列的周期性和随机波动性。目前SARIMA模型在传染病发病预测<sup>[13]</sup>和交通流量预测<sup>[14-15]</sup>中都取得了较好效果。

### 3.2 研究结果

采用1950—1960年数据,构建不同模型,重建1961—1974年缺失的诸泉流量序列,并对不同模型重建结果进行比较。

(1)多元线性回归模型。考虑补给区大气降水量因素和入湖流量因素,以1958—1960年的诸泉水流量作为因变量,以入湖流量和降水量作为自变量,使用多元线性回归模型进行拟合,得到公式如下:

$$y = 0.412 + 0.874 \times Q_{\text{入湖}} + 0.002 \times P \quad (1)$$

式中: $Q_{\text{入湖}}$ 为入湖流量( $\text{m}^3/\text{s}$ ); $P$ 为降水量( $\text{mm}$ )。

拟合结果显示标准误差为0.223,相关系数为0.734。根据式(2)重建1961—1974年诸泉历史流量结果见图2。

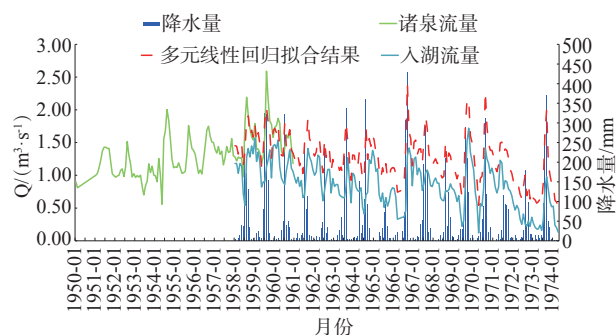


图2 多元线性回归模型重建泉水数据

Fig.2 Reconstruction of spring data by linear regression model

(2)考虑滞后的多元线性回归模型。由于大气降水入渗补给区距离泉口较远,泉流量对降水的响应有一定滞后效应,故在自变量中添加滞后信息。通过观察降水量和诸泉流量,发现诸泉流量的峰值较降水量峰值晚1~2个月,故分别采用滞后1个月和2个月的情况进行重建,考虑滞后时间为1个月的多元线性回归方程为:

$$y = 0.466 + 0.803 \times Q_{入湖} + 0.002 \times P + 0.039 \times Q_{入湖+1} \quad (2)$$

式中: $Q_{入湖}$ 为入湖流量, ( $m^3/s$ );  $Q_{入湖+1}$ 为延迟1个月入湖流量( $m^3/s$ )。经计算标准误差为0.222,相关系数为0.737,较不考虑滞后效应情况,精度略有提升。根据公式(2)重建诸泉历史流量序列见图3。

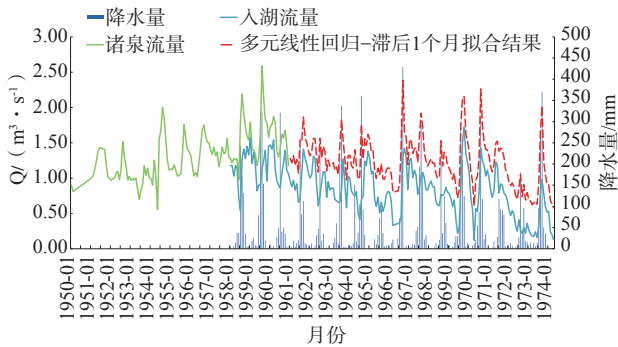


图3 考虑滞后性的多元线性回归模型重建泉水数据  
Fig.3 Reconstruction of spring water data by multiple linear regression model considering hysteresis

考虑滞后时间为2个月的多元线性回归模型为:

$$y = 0.775 + 0.675 \times Q_{入湖} + 0.002 \times P + 0.02 \times Q_{入湖+1} - 0.072 \times Q_{入湖+2} - 0.001 \times P_2 \quad (3)$$

式中: $Q_{入湖+1}$ 为滞后1个月时间的入湖流量( $m^3/s$ );  $Q_{入湖+2}$ 为滞后2个月时间的入湖流量( $m^3/s$ )。  $P_2$ 为延迟2个月降水量数据(mm)。经计算标准误差为0.220,相关系数为0.744。滞后2个月的多元线性回归拟合精度有提升,但从拟合方程中发现滞后两个月入湖流量与预测结果呈负相关关系,与实际不符,故不考虑2个月的滞后效应。

(3)SARIMA模型。采用SARIMA模型模拟1950—1957年诸泉流量,使用1958年1月—1960年12月数据作为验证测试。其中标准误差为0.159,相关系数为0.871。其模拟结果见图4。从图4可发现泉流量时序具有明显的周期性(年际)和明显的上升趋势,根据历史资料,与实际情况不符合。

(4)SARIMA和多元线性回归相结合模型。

①对1958—1960年流量序列采用SARIMA模型、多元线性回归序列和滞后1个月的多元线性回归预测序列相结合的模型进行拟合。所得到的线性回归方程如下:

$$y = -0.114 + 0.915 \times y_{SARIMA预测结果} - 2.137 \times y_{考虑滞后1月的多元线性回归结果} + 2.508 \times y_{多元线性回归结果} \quad (4)$$

其中标准误差为0.149,相关系数为0.889,所得拟合结果较仅采用多元线性回归方法精度有明显提升。

②对①中重建的1961—1974年曲线进行线性拟合,可得线性拟合方程:

$$y = 0.0035t + 1.4881 \quad (5)$$

式中: $t$ 为时间(月)。根据区域水文地质资料,补给源中永定河上游1958年修建官厅水库,河道来水量减少,泉域内城市规模逐步扩大,人口逐渐增多,用水量增加,大规模开采开始于1961年,该泉于1974年4月断流,故采用线性拟合1961年1月—1974年4月期间诸泉流量下降趋势,则诸泉流量时序的线性拟合为(设 $t=35, y=1.5088; t=193, y=0$ ):

$$y = -0.0078t + 1.3669 \quad (6)$$

令式(5)与式(6)相减,再叠加①中预测模型,得诸泉流量重建公式:

$$y = -0.114 + 0.915 \times y_{SARIMA预测结果} - 2.137 \times y_{考虑滞后1月的多元线性回归结果} + 2.508 \times y_{多元线性回归结果} - 0.0113 \times t - 0.1212 \quad (7)$$

根据式(7)重建诸泉历史流量见图5。

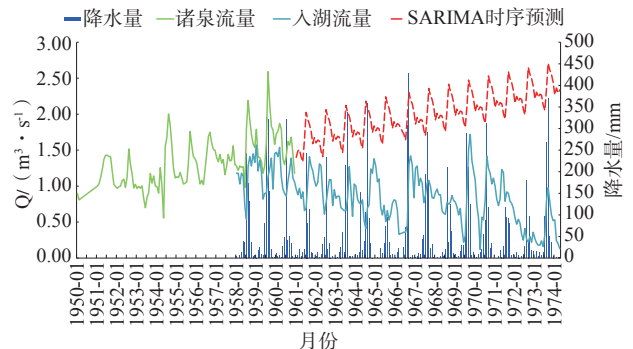


图4 采用SARIMA模型的泉水流量数据重建曲线  
Fig.4 Reconstruction curve of spring water flow data using SARIMA model

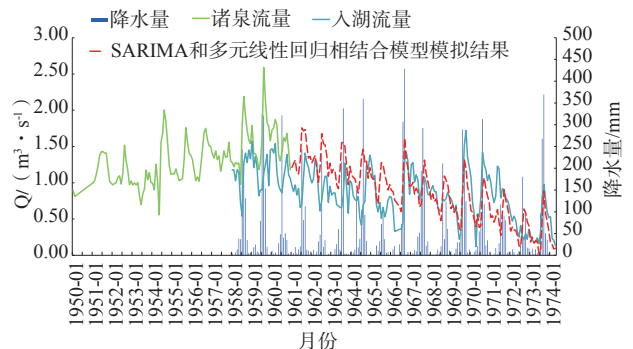


图5 采用SARIMA和多元线性回归相结合模型的泉水流量重建曲线  
Fig.5 Reconstruction curve of water flow using SARIMA and multiple linear regression model

### 3.3 结果对比分析

通过图6和表1可看出,四种模型对1958—1960年的拟合结果与实测值相比,采用SARIMA模型以及SARIMA方法和多元线性回归相结合模型的平均值更接近实测平均值。多元回归模型重建的诸泉流量相关系数最小。采用SARIMA和多元线性回归模型重建的泉水流量序列标准误差最小,相关系数和 $R^2$ 最高。

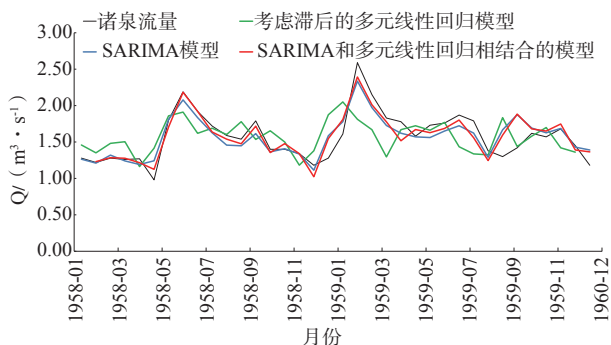


图6 各种模型逐月拟合流量图(1958—1960年)  
Fig.6 Monthly fitting flow chart of various models

表1 不同模型重建流量序列均值和方差(1958—1960年)

Table1 Mean and variance of discharge series rebuilt by different models

重建模型	平均值/ ( $m^3 \cdot s^{-1}$ )	标准 误差	相关 系数	$R^2$
实测值	1.58			
多元线性回归模型	1.553	0.223	0.734	0.527
考虑滞后的多元线性回归模型	1.573	0.222	0.737	0.534
SARIMA模型	1.575	0.159	0.871	0.754
SARIMA方法和多元线性回归相结合模型	1.585	0.149	0.889	0.791

通过图5可以看出,采用SARIMA和多元线性回归模型重建的诸泉流量在后期出现了断流现象,与实际的变化情况相符。采用SARIMA和多元线性回归模型对泉水流量的重建从物理模型上来讲,泉流量波动幅度相对较小,且从20世纪60年代开始出现大量开采第四系地下水,引起第四系地下水位降低,压力降低,引起岩溶水顶托补给第四系孔隙水,进而造成诸泉流量的降低,故采用SARIMA和多元线性回归相结合的模型对泉流量序列重建有较好作用。

自1950年监测以来,该泉诸泉月平均流量最大出现在1959年8月为 $2.59 m^3/s$ ,月平均最小流量出现在1954年5月为 $0.558 m^3/s$ 。自1960年后泉流量逐年减少,直至1974年断流。通过重建1961—1974年的流量从而组成完整的数据序列。1950—1974年多年平均流量为 $1.064 m^3/s$ 。

### 4 结论与建议

本次研究表明,岩溶大泉的流量动态特征常常具有周期性和趋势性,基于完整的监测数据,采用季节性差分自回归滑动平均模型(SARIMA)拟合缺失的北方典型泉水的诸泉流量序列是可行的。

岩溶泉水流量的模拟重建是一个十分复杂的问题。通过对典型岩溶泉群补给区降水量、已有诸泉监测数据和流入下游湖水流量数据,采用多种模型模拟结果表明,季节性差分自回归滑动平均模型(SARIMA)和多元线性回归相结合的模型比多元线性回归模型、考虑滞后的多元线性回归模型以及SARIMA模型的模拟精度更高。

本次重建北方典型岩溶大泉流量序列是一次探索性的尝试,忽略了1961—1974年下垫面条件和土地利用方式等因素的变化,模型还是比较粗糙,希望为后续研究起到抛砖引玉的作用。

#### 参考文献:

- [1] 张宗祜,李烈荣.中国地下水资源[M].北京:中国地图出版社,2004.
- [2] 梁永平,韩行瑞,王维泰,等.中国北方岩溶地下水环境问题与保护[M].北京:地质出版社,2013.
- [3] 罗丹,杨勇,邢国章.基于水化学和同位素的白浮泉补径排条件分析[J].水文,2021,41(6):13-17,90.
- [4] 朱燕燕.基于时间序列分析的娘子关泉流量预测[J].山西水利,2021,37(7):41-43.
- [5] 姜宝良,付北锋,赵延涛.泉水动态分析预测和资源评价:以辉县百泉为例[J].水文地质工程地质,2002(3):43-46.
- [6] 张建芝,邢立亭.回归分析法在地下水动态分析中的应用[J].地下水,2010,32(4):88-90.
- [7] 姜宝良,许来慧,崔江利,等.新乡市百泉泉水流量动态预测与资源评价[J].人民黄河,2013,35(12):71-72,76.
- [8] 郭艺,王枫,甘甫平,闫柏琨.基于移动平均模型和指数平滑模型的岩溶泉流量预测[J].河北地质大学学报,2020,43(4):19-25.
- [9] 姜宝良,李林晓,李腾超.基于BP神经网络的新乡百泉逐月泉水流量动态分析[J].矿产勘查,2018,9(3):516-521.
- [10] 王南.古都北京[M].北京:清华大学出版社,2012.
- [11] 蔡向民,郭高轩,张磊,等.北京城湖泊的成因[J].中国地质,2013,40(4):1092-1098.
- [12] 张辉,刘嘉琨,柳湘月,等.交通流的季节ARIMA模型与预报[J].天津大学学报,2005,38(9):838-841.
- [13] 杨仁东,胡世雄,邓志红,等.湖南省手足口病发病趋势SARIMA模型预测[J].中国公共卫生,2016,32(1):48-52.
- [14] 王莹,韩宝明,张琦,等.基于SARIMA模型的北京地铁进站客流量预测[J].交通运输系统工程与信息,2015,15(6):205-211.
- [15] 李晓磊,肖进丽,刘明俊.基于SARIMA模型的船舶交通流量预测研究[J].武汉理工大学学报(交通科学与工程版),2017,41(2):329-332,337.

## Study on Reconstruction of Historical Flow Sequence of Typical Karst Springs in Northern Piedmont

XU Liang<sup>1,2</sup>, GUO Gaoxuan<sup>1,2</sup>

(1. *Beijing Institute of Geological Environment Monitoring, Beijing 100195, China;*

2. *Urban Groundwater Safety Prevention and Control Technology Innovation Base, Beijing 100195, China*)

**Abstract:** The lack of spring flow series seriously affects the understanding of the changes of hydrogeological conditions, so it is of great significance to reconstruct the missing spring flow sequence. Based on the periodicity and trend of spring water dynamics, the reconstruction of the flow series of a typical karst spring group missing from 1961 to 1974 was carried out by using multiple linear regression model, multiple linear regression model considering spring lag precipitation, SARIMA model and the combination of SARIMA and multiple linear regression model. The results show that the flow reconstructed by the model combining SARIMA and multiple linear regression is reasonable.

**Keywords:** spring flow; rainfall; flow data into the lake; data reconstruction

(上接第87页)

## Research on Watershed Flood Forecasting Model Based on Deep Learning

XU Gang<sup>1,2</sup>, HU Tingting<sup>3</sup>, WANG Chenni<sup>1</sup>

(1. *College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang 443002, China;*

2. *Hubei Provincial Collaborative Innovation Center for Water Resources Security, Wuhan 430072, China;*

3. *Immigration Management Consulting Center of the Ministry of Water Resources, Yichang 443000, China*)

**Abstract:** With the popularization of hydrological information systems, the process of collecting data has become more and more convenient. Traditionally, various hydrological models and parameter calibration are used to describe the complex relationship between hydrological data and prediction results. How to achieve hydrological data-driven Automatic Hydrological Prediction is a problem at present. In recent years, artificial intelligence technology has begun to flourish, and deep learning technology has begun to be applied in the field of hydrology, trying to solve this barrier. Taking the watershed from the West Branch of Juzhang River to Yuan'an as the research object, combined with the principle of distributed hydrological model and deep learning theory, this paper extracts runoff characteristics of the watershed based the multi-dimensional data, uses the characteristic network to predict the watershed runoff, and constructs a convolutional neural networks (CNN) Watershed rainfall runoff prediction model. The results show that the simulation accuracy of the basin rainfall-runoff forecast model based on the convolutional neural network is 90%, the model is reliable, and it could meet most of the rainfall runoff forecasts, and it is the rainfall runoff forecast of the basin. Provide a new method.

**Keywords:** artificial intelligence; rainfall and runoff forecast; deep learning; convolutional neural network; parameter calibration