

DOI: 10.20040/j.cnki.1000-7709.2025.20242020

# 基于可解释机器学习的岩溶介质渗透系数估算模型

李红星<sup>1</sup>, 白伟<sup>1</sup>, 李傲<sup>2a</sup>, 杨艳娜<sup>2a,2b</sup>, 王之正<sup>1</sup>

(1. 长江三峡勘测研究院有限公司(武汉), 湖北 武汉 430073; 2. 成都理工大学 a. 地下水科学与工程系;  
b. 地质灾害防治与地质环境保护国家重点实验室, 四川 成都 610059)

**摘要:** 岩溶介质的渗透性能受到可溶岩地质特征、岩溶发育程度、流体性质等多重因素的影响, 裂隙岩体的渗透系数估算模型难以反映岩溶发育的复杂性, 使其适用性较差。为此, 采用随机森林(RF)、支持向量回归(SVR)、CatBoost机器学习算法结合贝叶斯优化算法分别构建岩溶介质渗透系数估算模型, 利用均方根误差( $R_{RMSE}$ )、均方误差( $M_{MSE}$ )、决定系数( $R^2$ )开展估算模型的评价精度验证, 并通过SHAP算法分析机器学习模型中岩溶介质渗透系数的主导影响因子, 阐明各影响因子对岩溶介质渗透系数的影响程度。结果表明, 优化后的支持SVR模型 $R_{RMSE}$ 为0.1288,  $M_{MSE}$ 为0.0166,  $R^2$ 为0.74, 均优于随机森林和CatBoost模型, 能够较好地估算岩溶介质渗透系数。SHAP图揭示了各主导因子不同特征值对岩溶介质渗透系数有明显差异, 钻孔线岩溶率(BK)、深度(Z)和充填物含量(AFC)为岩溶介质渗透系数的主要影响因子, 对岩溶介质渗透系数影响显著。可见SVR模型具有较高的估算精度, 模型可解释性强, 为岩溶地区工程应用提供一定参考价值。

**关键词:** 随机森林; 支持向量回归; CatBoost; 岩溶介质; 渗透系数估算

**中图分类号:** [TV221.2]

**文献标志码:** A

**文章编号:** 1000-7709(2025)09-0069-04

## 1 引言

岩体的渗透系数对于工程防渗设计和水文地质评价至关重要。岩溶介质的渗透性受岩溶发育特征和岩体地应力等多种因素的影响, 通常表现出显著的各向异性和空间变异性。渗透系数在自然环境中的空间分布及在工程扰动条件下的变化规律, 已成为水文地质研究的前沿课题和渗流场分析的核心内容<sup>[1]</sup>。确定岩体渗透系数或渗透张量的方法主要有现场水文地质试验、裂隙测量法、反演法及渗透系数估算法<sup>[2-6]</sup>。渗透系数估算模型大多以裂隙岩体为研究对象, 而针对岩溶介质的渗透系数估算模型则相对较少。在机器学习领域, 通过数据和算法训练模型使其能够自主识别模式和做出决策这一技术已广泛应用于金融、医疗等行业<sup>[7-8]</sup>。在地质领域, 邵良杉等<sup>[9-12]</sup>已将机器学习应用于边坡稳定性、隧道涌突水等灾害方面的预测, 但未见在岩体渗透性估算中的应用。机器学习算法能够解决岩溶介质渗透系数估算中的影响因素选取局限性、数据预处理不足及非线性

性问题。为此, 本文基于随机森林(RF)、CatBoost(Categorical Boosting)和支持向量回归(SVR)算法构建岩溶介质渗透系数估算模型并进行验证, 借助SHAP算法对模型估算结果进行解释, 以探讨不同影响指标的重要性并进行量化, 从而直观展示不同特征值对岩溶介质渗透系数的影响, 以提高复杂模型的可解释性和可信度。

## 2 研究方法

### 2.1 机器学习算法简介

随机森林回归(RF)是一种通过构建多个决策树(这些树形结构用于进行决策)来实现回归预测的集成学习方法。在每棵树的构建过程中算法对训练数据进行随机采样, 并从特征集中随机选择部分特征进行分裂决策。这种“袋装法”策略提高了模型的泛化能力, 有效减少了单个决策树的高方差问题。随机森林能够捕捉复杂的非线性关系, 且对高维数据和异质性数据具有鲁棒性, 广泛用于分类和回归任务中<sup>[9]</sup>。

支持向量回归(SVR)是一种强大的监督学习算法, 广泛应用于回归任务。其核心思想是通过

收稿日期: 2024-10-28, 修回日期: 2024-11-29

作者简介: 李红星(1985-), 男, 教授级高级工程师, 研究方向为水文地质与工程地质, E-mail: 63135319@qq.com

通讯作者: 白伟(1984-), 男, 高级工程师, 研究方向为水文地质与工程地质, E-mail: 277165372@qq.com

构建一个决策边界(即超平面),将不同类别的数据点分隔开,并最大化边界与最近样本点(支持向量)之间的间距。SVR 能够借助核函数将数据映射到更高的维度,从而应对那些线性不可分的问题。由于其良好的泛化能力和有效性,SVR 在多个领域得到了广泛应用<sup>[10]</sup>。

CatBoost(Categorical Boosting)是开源梯度提升决策树(GBDT)框架,专门优化用于处理类别特征。相较于传统的 GBDT 算法,CatBoost 通过一系列独特的技术,减少了过拟合的风险,并提升了模型性能。它不仅能够处理大规模数据集,还支持多种模式,包括分类、回归和排序任务。由于其优异的表现和易用性,CatBoost 逐渐成为数据科学和机器学习领域的重要工具之一<sup>[7]</sup>。

贝叶斯优化是一种基于概率模型的全局优化算法,主要用于寻找具有高计算成本或黑盒性质的目标函数的最优解。在机器学习中,贝叶斯优化被广泛应用于超参数调优。该方法通过构建一个基于高斯过程(GP)的代理模型来近似目标函数的行为,并根据此模型选择后续的探索点。贝叶斯优化通过平衡探索与利用,提高了对参数空间的高效搜索能力,特别适合计算资源有限的情境下进行模型调优。

### 2.2 SHAP 算法

SHAP(Shapley Additive Explanations)是一种用于解释机器学习模型输出的技术<sup>[13]</sup>。SHAP 值提供了一种评估特征的重要性标准,使得可量化每个特征对预测结果的贡献。这一方法为理解各个特征对模型输出的影响提供了严谨的理论框架,并在有效解释复杂机器学习模型方面表现出显著的效果。对于特征集合  $S$  中的特征  $i$ ,Shapley 值计算公式为:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} \cdot [v(S \cup \{i\}) - v(S)] \quad (1)$$

式中, $\Phi_i$  为特征  $i$  在所有可能子集上的平均边际贡献; $N$  为所有特征的集合; $S$  为不包含特征  $i$  的任何一个特征子集; $|S|$  为集合  $S$  中的特征数; $v(S)$  为特征集  $S$  对模型预测输出的贡献; $v(S \cup \{i\})$  为包含特征  $i$  的特征集  $S \cup \{i\}$  对模型预测输出的贡献。

通过对文献<sup>[14]</sup>与工程资料搜集,获取了岩溶地区有效地质信息点 110 份。模型构建时将总样本按照 8 : 2 划分训练集与测试集用于模型训练及测试,并利用贝叶斯优化算法进行模型超参数优化,通过模型输出的均方根误差( $R_{RMSE}$ )、均

方误差( $M_{MSE}$ )、决定系数( $R^2$ )评价模型精度,进一步选取适宜的模型。最后通过 SHAP(Shapley Additive Explanations)算法进一步解释影响因子对渗透系数的影响。研究思路见图 1。

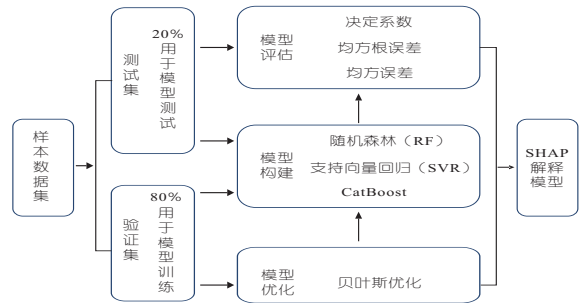


图 1 岩溶介质渗透系数预测模型研究路线图

Fig. 1 Roadmap of the prediction model of permeability coefficient of karst media

## 3 岩溶介质渗透系数估算模型训练

### 3.1 指标选取

采用易获得、且能直观反映岩溶发育程度的岩溶层组类型(TK)、钻孔线岩溶率(BK)<sup>[15]</sup>、地质构造特征(GSC)及地下水的运动特征(GMC)作为岩溶指标,结合已有裂隙介质的渗透系数估算模型中深度指标(Z)<sup>[16]</sup>及泥质充填物含量(AFC)<sup>[17]</sup>作为地质指标来建立岩溶介质的渗透系数估算模型<sup>[14]</sup>,具体指标划分见表 1,部分样本数据见表 2。

表 1 岩溶介质渗透系数预测模型指标的量化取值划分结果表

Tab. 1 Quantitative value division results of karst media permeability coefficient prediction model index

种类	类型 / 特征	岩溶化强度	取值		
			TK	GSC	GMC
岩溶层	均匀状灰岩	强	0.75~1		
	纯碳酸岩与不纯碳酸岩互层	较强	0.5~0.75		
	灰岩与不纯碳酸岩	弱	0.25~0.5		
	不纯碳酸岩与碎屑岩互层	极弱	0~0.25		
地质构造	断裂带、褶皱核部	较强—强		0.5~1	
	断裂影响带、褶皱翼部	弱		0.25~0.5	
	平缓岩层	极弱		0~0.25	
地下水运动	水平循环带	较强—强			0.75~1
	垂直入渗带	弱			0.25~0.75
	深部缓流带	极弱			0~0.25

### 3.2 模型对比研究

RF、SVR 和 CatBoost 算法经贝叶斯优化算法调参后,3 种模型主要参数设置见表 3。

通过均方根误差( $R_{RMSE}$ )、决定系数( $R^2$ )及均方误差( $M_{MSE}$ )对 3 种经过贝叶斯优化的机器学习算法随机森林(RF)、支持向量回归(SVR)和 CatBoost 进行性能评估,结果见表 4。其中  $M_{MSE}$  和  $R_{RMSE}$  用于评估模型的估算误差,帮助理解模

表 2 部分样本数据表

Tab. 2 A partial sample data table

序号	TK	BK	GSC	GMC	Z	AFC	K
1	0.75	1	0.80	0.25	85	10	0.007
2	0.75	2	0.80	0.25	87	10	0.007
3	0.75	9	0.25	0.25	29	40	0.054
4	0.75	12	0.25	0.25	31	10	0.025
5	0.75	4	0.25	0.25	35	30	0.183
6	0.75	32	0.25	0.75	37	20	0.437
7	0.75	28	0.25	0.75	39	10	0.670
...	...	...	...	...	...	...	...
106	0.75	6	0.25	0.25	75	30	0.008
107	0.75	4	0.25	0.25	77	30	0.009
108	0.75	3	0.25	0.25	79	30	0.010
109	0.75	2	0.25	0.25	81	10	0.009
110	0.75	9	0.25	0.25	83	30	0.008

表 3 各模型参数设置

Tab. 3 Parameter settings for each model

算法	超参数	搜索空间	数值
RF	N estimators(树的数量)	[50,200]	100
	Max depth(树的最大深度)	[5,20]	5
	Min samples split(一个节点必须包含的最小样本数才能被分割)	[2,20]	20
	Min samples leaf(一个叶子节点必须包含的最小样本数)	[1,10]	1
	SVR	C(惩罚项系数)	[0.1,1 000]
	Epsilon(容忍度)	[0.01,0.5]	0.018 7
	Gamma	[0.000 1,10]	0.000 18
	(单个训练样本的影响范围)		
	Kernel(核函数)	线性(linear)、 径向基(rbf)	径向基(rbf)
CatBoost	Depth(树的深度)	[5, 10]	5
	Learning rate(学习率)	[0.001,0.2]	0.002 3
	Iterations(迭代次数)	[100, 2 000]	1 958

表 4 模型性能对比

Tab. 4 Comparison of model performance

算法	$R_{RMSE}$	$R^2$	$M_{MSE}$
RF	0.139 2	0.700	0.019 4
SVR	0.128 8	0.742	0.016 6
CatBoost	0.171 1	0.546	0.029 3

型的实际表现; $R^2$  用于评估模型对数据方差的解释能力,了解模型的适用性。由表 4 可看出,SVR 在所有评估指标中均表现出色,其  $R_{RMSE}$  为 0.128 8,  $M_{MSE}$  为 0.016 6,均为 3 种算法中最低值,说明该模型具备处理此类特征数据的较强拟合能力和估算准确性。 $R^2$  为 0.742,说明该模型能够解释目标变量约 74.2%的变异,具有较好的适用性。RF 模型的表现也相对良好,其  $R_{RMSE}$  为 0.139 2,  $M_{MSE}$  为 0.019 4,  $R^2$  值为 0.700。这些结果表明 RF 模型在捕捉数据趋势方面相对可靠,尽管其整体性能稍低于 SVM,但仍展现出一定的估算能力。CatBoost 模型的表现逊色于前两者,其  $R_{RMSE}$  为 0.171 1,  $M_{MSE}$  为 0.029 3,且决

定系数仅为 0.546,表明该模型仅能够解释约 54.6%的数据变异性。综合来看,贝叶斯优化的 SVR 模型处理此类任务具有更好的性能,可用于讨论各影响因子对岩溶介质渗透系数的影响。CatBoost 在处理本数据集时面临一定挑战,可能由于 CatBoost 更适合处理类别特征,但本文收集的数据集中类别特征(TK、GSC、GMC)区别并不明显。

## 4 基于 SVR 模型的岩溶介质渗透系数影响因素分析

图 2 为 SVR 模型 SHAP 全局解释图,图 2 中黄色点表示特征值在这个观察模型中对模型估算产生了正面影响;深紫色点表示该特征值在这个观察中对模型估算产生负面影响。垂直轴(特征排列)的特征按影响力从上到下进行排序(BK、Z、AFC、GMC、GSC、TK),上方的特征对模型输出的总影响更大,而下方的特征影响较小。水平轴(SHAP 值)显示每个特征对估算结果的影响大小,点越远离中心线(零点),表示该特征对模型输出的影响越大。由表 2 可看出,岩溶介质渗透系数最重要的影响因子为钻孔线岩溶率(BK),钻孔线岩溶率越大,岩体渗漏通道越大,SHAP 值越大,渗透系数就越大。深度指标(Z)对渗透系数也有着显著影响,SHAP 值随着深度指标增加(埋深越大)而减少,与钻孔线岩溶率正好相反。泥质充填物含量(AFC)与深度指标相同,泥质充填物含量越高,水流通过的有效空间减小,渗透系数越小。地下水的运动特征(GMC)、地质构造特征(GSC)和岩溶层组类型(TK)对渗透系数影响较小。可见,岩溶介质渗透系数的主要影响因素为钻孔线岩溶率(BK)、深度指标(Z)、泥质充填物含量(AFC)。

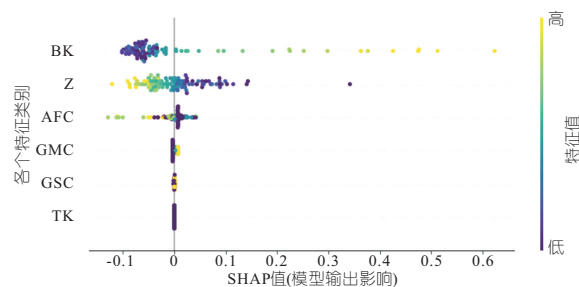


图 2 SVR 模型 SHAP 全局解释图(训练集与测试集)  
Fig. 2 Global interpretation of SHAP of SVR model (training set vs. test set)

基于 SHAP 可视化影响因子重要性排序前 3 个主导因子钻孔线岩溶率、深度指标、充填物含量的单因子依赖图(图 3),来直观揭示影响因子的

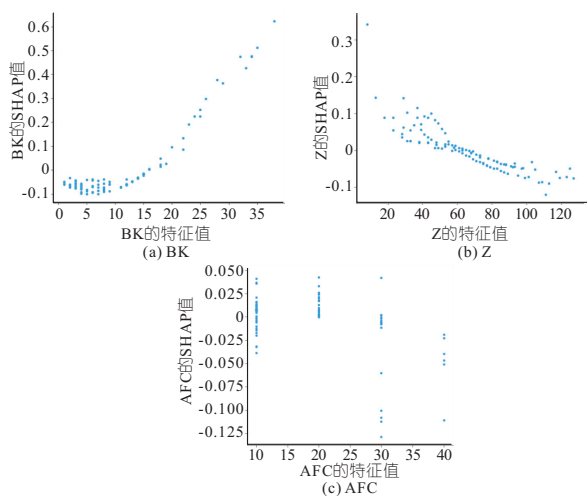


图 3 主导因子与模型预测结果的依赖关系图

Fig. 3 Dependence between the dominant factor and the prediction results of the model

不同特征值对岩溶介质渗透系数估算结果的影响,以增加模型的可信度。由图 3 可看出,钻孔线岩溶率 10% 以下时,对岩溶介质渗透系数起到的促进作用还不是特别强烈,在达到钻孔线岩溶率 10% 以上后开始作用强烈。这是因为随着岩溶率的增加,岩体内的溶蚀孔隙和裂隙数量增多,形成的裂隙系统为地下水提供了更为通畅的流动通道,使水流能够更快地穿过岩体,从而提高了岩体的渗透性。随着埋深的增加,岩体所承受的压力增大,导致孔隙结构愈发紧密,从而降低了渗透性。同时岩溶介质中流动的地下水溶解度逐渐降低,从而减少了其对碳酸盐岩的侵蚀能力,这进一步影响了水流的通道和流动能力,最终导致岩体的渗透性降低。随着泥质充填物含量的增加,SHAP 值对模型造成的负面影响也开始增大。这是因为泥质充填物填补岩石中的大孔隙,使得颗粒间的间隙变得更加紧凑,这使得水流通过的有效空间减小;同时这些泥质充填物可能具有较高的吸水性,进一步影响水的流动,从而降低岩体渗透性。

## 5 结论

a. 经贝叶斯优化后的 SVR 模型的决定系数最高且误差小,表明 SVR 模型处理此类任务性能相对更好,估算效果更好,可用于岩溶地区的渗透系数估算。

b. 通过 SHAP 算法对 SVR 模型估算结果进行解释,钻孔线岩溶率(BK)、深度指标(Z)、泥质充填物含量(AFC)对岩溶介质渗透系数的影响较大,为主导因子。地下水的运动特征(GMC)、地质构造特征(GSC)和岩溶层组类型(TK)对渗

透系数影响较小。

c. SHAP 算法主要用于对模型输出结果的解释上,有助于理解模型的决策成因,提高复杂模型的可信度。后续可将 SHAP 的结果运用于模型的重构及指标的进一步优化,以提高模型的估算能力。

## 参考文献:

- [1] 张人权,梁杏,靳孟贵,等.当代水文地质学发展趋势与对策[J].水文地质工程地质,2005,32(1):51-56.
- [2] 孙蓉琳,梁杏,靳孟贵.裂隙岩体渗透系数确定方法综述[J].水文地质工程地质,2006,33(6):120-123.
- [3] 景来红,段世超,杨顺群.渗流反演分析在工程设计中的应用[J].岩石力学与工程学报,2007,26(增刊2):4503-4509.
- [4] 丁立丰,郭啟良,王成虎.工程岩体裂隙渗透性试验方法研究及应用[J].岩土力学,2009,30(9):2599-2604.
- [5] 宋琨,晏鄂川,陈刚.地下水封洞库岩体渗透系数估算研究[J].岩石力学与工程学报,2014,33(3):575-580.
- [6] 陈君,刘明明,李星,等.一种基于地质指标的裂隙岩体渗透系数估算模型[J].岩土力学,2016,37(6):1706-1714.
- [7] 薛博元.基于改进 CatBoost 算法的糖尿病回归和分类预测模型研究[D].银川:北方民族大学,2024.
- [8] 王萌.信用违约风险视角下我国房地产行业中心企业的识别与应用[D].呼和浩特:内蒙古财经大学,2024.
- [9] 邵良杉,马寒,温廷新.基于 RF-ELM 模型的边坡稳定性预测研究[J].中国安全生产科学技术,2015,11(3):93-98.
- [10] 王璐.基于遗传算法和支持向量机的西南岩溶越岭隧道涌水量预测[D].成都:成都理工大学,2019.
- [11] 汪长重,韩旭,赵鑫,等.多种机器学习方法在岩质类型快速判别中的可靠性分析[J].水电能源科学,2024,42(6):88-92.
- [12] 杨小平,段生锐,蒋力,等.基于机器学习算法的滑坡土壤含水率预测方法研究[J].水电能源科学,2024,42(3):73-77.
- [13] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions [C]//NIPS Long Beach, 2017.
- [14] 范全忠.川东隔挡式背斜区隧道涌水量预测方法的适宜性研究[D].成都:成都理工大学,2021.
- [15] 中国地质调查局.水文地质手册(第二版)[M].北京:地质出版社,2012.
- [16] 万力,蒋小伟,王旭升.含水层的一种普遍规律:渗透系数随深度衰减[J].高校地质学报,2010,16(1):7-12.
- [17] 蒋小伟,万力,王旭升.自重应力和岩性对砂泥岩裂隙岩体渗透性空间分布的控制[J].工程勘察,2008(增刊1):210-216.

[10] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 气象干旱等级: GB/T 20481-2017[S]. 北京: 中国标准出版社, 2017.

[11] KENDALL M, GIBBONS J D. Rank correlation methods[M]. 5th ed. London: Oxford University Press, 1990

[12] 包为民, 沈丹丹, 倪鹏, 等. 滑动平均差检测法的提出及验证[J]. 地理学报, 2018, 73(11): 2075-2085.

[13] 张璇, 许杨, 郝芳华, 等. 滦河流域气象干旱向水文干旱传播特征及风险分析[J]. 水利学报, 2022, 53(2): 165-175.

[14] 王水寒, 邱建秀, 王大刚. 1960—2014 年广东省干旱时空演变特征[J]. 热带地理, 2020, 40(2): 357-366.

[15] 王文, 王靖淑, 陶奕源, 等. 人类活动对水文干旱形成与发展的影响研究进展[J]. 水文, 2020, 40(3): 1-8.

[16] TIJDEMAN E, HANNAFORD J, STAHL K. Human influences on streamflow drought characteristics in England and Wales[J]. Hydrology and earth system sciences, 2018, 22: 1051-1064.

[17] 陈晶, 张天力, 顾世祥. 滇中引水工程水源区与受水区径流丰枯遭遇及水文干旱研究 [J]. 水电能源科学, 2024, 42(6): 23-27.

### Study on Characteristics of Hydrologic and Meteorological Droughts and Responses Mechanism in Leizhou Peninsula

FENG De-zeng<sup>1</sup>, SUN Jing-jing<sup>2</sup>, PENG Ying<sup>1</sup>, ZHENG Jing-hua<sup>1</sup>, WANG Bin<sup>3</sup>, WEI Ling-na<sup>4a,4b</sup>

(1. China Water Resources Pearl River Planning, Surveying & Designing Co., Ltd., Guangzhou 510000, China; 2. Water Resources Research Institute of Shandong Province, Jinan 250014, China; 3. Pearl River Water Resources Research Institute, Guangzhou 510000, China; 4a. Key Laboratory of Hydro-Meteorological Disasters Mechanism and Warning, Ministry of Water Resources; 4b. College of Hydrology and Water Resources, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** The Leizhou Peninsula exhibits one of the highest incidences of drought within the Pearl River Basin. Research on the characteristics and response mechanisms of meteorological and hydrological droughts in this area holds significant theoretical and practical importance for understanding drought patterns in regions with abundant rainfall yet fragile water resource systems. Based on the monthly meteorological and hydrological data from the representative station in Leizhou Peninsula, the standardized precipitation-evapotranspiration index ( $I_{SPEI}$ ) and standard runoff index ( $I_{SRI}$ ) were used to analyze the variation patterns of meteorological drought and hydrological drought in this region from 1970 to 2016. The Mann-Kendall trend test method, moving average method and Morlet wavelet analysis method were adopted to analyze the changing trends and cycles of meteorological drought and hydrological drought in this region at different time scales (3 and 12 months). Propagation characteristics of meteorological drought to hydrological drought were examined using the run-length theory. The results show that the meteorological drought in Leizhou Peninsula is increasing during the study period. Particularly, the 12-month  $I_{SPEI}$  obtained from Zhanjiang station and Xuwen station decreased significantly, at the rate of  $-0.014/\text{month}$  and  $-0.008/\text{month}$ , respectively. However, no obvious trend of hydrological drought was detected. Both drought types showed multi-scale periodicity, sharing a primary first-order period of 20 years in the north and 14 years in the south. The main cycles of meteorological drought and hydrological drought in the south and north are consistent, and the variation patterns are relatively synchronous. Hydrological drought in the southern peninsula demonstrated stronger responsiveness to meteorological drought than the northern region.

**Key words:** Leizhou Peninsula; meteorological drought; hydrologic drought; climate change

\*\*\*\*\*  
(上接第 72 页)

### Permeability Coefficient Estimation Model of Karst Media Based on Interpretable Machine Learning

LI Hong-xing<sup>1</sup>, BAI Wei<sup>1</sup>, LI Ao<sup>2a</sup>, YANG Yan-na<sup>2a,2b</sup>, WANG Zhi-zheng<sup>1</sup>

(1. Yangtze River Three Gorges Survey and Research Institute Company Limited (Wuhan), Wuhan 430073, China; 2a. Department of Groundwater Science and Engineering; 2b. State Key Laboratory of Geohazard Prevention and Geoenvironmental Protection, Chengdu University of Technology, Chengdu 610059, China)

**Abstract:** The permeability of karst medium is affected by multiple factors such as the geological characteristics of soluble rocks, the degree of karst development, and the fluid properties. The permeability coefficient estimation model of fractured rock mass is difficult to reflect the complexity of karst development, which makes it poorly applicable. Random Forest (RF), Support Vector Regression (SVR), CatBoost machine learning algorithm combined with Bayesian optimization algorithm were used to construct the permeability coefficient estimation model of karst media. The root mean square error ( $R_{RMSE}$ ), mean square error ( $M_{MSE}$ ) and coefficient of determination ( $R^2$ ) were used to verify the evaluation accuracy of the estimation model. The SHAP algorithm was used to analyze the dominant influencing factors of the permeability coefficient of karst media in machine learning model, and the influence of each influencing factor on the permeability coefficient of karst medium was clarified. The results show that the  $R_{RMSE}$  of the optimized SVR model is 0.128 8,  $M_{MSE}$  is 0.016 6 and  $R^2$  is 0.74, which are better than the random forest and CatBoost models, and can better estimate the permeability coefficient of karst media. The SHAP diagram revealed that there were obvious differences in the permeability coefficient of karst media between different eigenvalues of each dominant factor, and the karst rate (BK), depth (Z) and filling content (AFC) of the borehole line were the main influencing factors of the permeability coefficient of karst media, and had a significant impact on the permeability coefficient of karst media. The SVR model has high estimation accuracy and strong interpretability, which provides a certain reference value for engineering applications in karst areas.

**Key words:** random forest; support vector regression; CatBoost; karst media; permeability coefficient estimation