

砂岩的近红外光谱特征及其含水量反演

王东升¹, 王海龙^{1,2}, 张芳^{1,3*}, 韩林芳^{1,3}, 李运¹

1. 中国矿业大学(北京)力学与建筑工程学院, 北京 100083

2. 河北省土木工程诊断、改造与抗灾重点实验室, 河北 张家口 075000

3. 中国矿业大学(北京)深部岩土力学与地下工程国家重点实验室, 北京 100083

摘要 沉积岩石的强度往往会受到水的影响,含水量不同其影响程度也不相同。因此,沉积岩石的含水量测定对于后续评估其物理力学特性具有重要的价值。在岩石含水量测定方面,传统的方法往往费时、费力,而且破坏了工程结构的完整性。近红外光谱分析技术是一种非常有潜力的方法,具有实时、无损等优点。基于近红外光谱分析技术对砂岩的光谱特征以及含水量的反演进行了研究。首先,通过室内试验获取了砂岩试样不同饱和度的近红外光谱曲线;然后,对原始光谱曲线进行了异常曲线剔除以及一阶导数预处理,消除噪声、环境等因素的影响;其次,对 R1(1 400 nm)和 R2(1 900 nm)附近的两个吸收峰进行光谱特征变量提取以及归一化处理,消除量纲和域值的影响;接着,基于最大信息系数对提取的光谱特征变量进行分析和筛选;最后,基于筛选后的光谱特征变量采用自行搭建的 BP 神经网络分类器对砂岩的含水量进行了反演。结果表明:(1)含水砂岩的近红外光谱吸收曲线在 1 400 和 1 900 nm 附近有着明显的吸收峰,位于 1 400 nm 附近的吸收峰,谱带比较宽缓,位于 1 900 nm 附近的吸收峰,谱带比较尖锐;随着含水量的增加,近红外光谱曲线在 1 400 和 1 900 nm 附近吸收峰的吸收强度也在增加,具有明显的正相关性,可作为砂岩含水量分析、反演的主要谱段。(2)根据计算的最大信息系数值,1 400 nm 附近的峰高与含水量的相关性最强,同样 1 900 nm 附近的峰高与含水量的相关性最强;1 400 nm 附近的峰面积、峰高和 1 900 nm 附近的峰高、峰面积、半高宽、右肩宽,共 6 个光谱特征变量,其最大信息系数值 >0.9 ,可作为 BP 神经网络反演砂岩含水量的特征变量。(3)利用最大信息系数筛选出 1 400 和 1 900 nm 附近两个吸收峰的特征变量进行 BP 神经网络建模,所建立的砂岩含水量反演模型训练集准确率为 90.3%,测试集的准确率为 83.9%,说明基于近红外光谱分析技术砂岩含水量的反演方法是可行的。

关键词 岩石;含水量;近红外光谱;光谱特征;人工神经网络

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)11-3368-05

引言

水的存在使岩石的强度特性发生了改变,含水量不同其影响程度也不相同。在边坡、地下工程和文物保护领域中,许多灾变和病害都是由水导致岩石强度减弱而诱发的^[1-2]。因此,岩石中含水量的测定对于评估岩石的物理力学特性以及工程结构的稳定性具有重要的意义。

传统的岩石含水量测定方法需要现场原位取样,破坏了工程结构的完整性,尤其是在文物保护领域。近红外光谱分

析技术(near infrared spectroscopy, NIRS)测定岩石中的含水量是近些年研究的新思路,通过测定含水岩石中 OH 基团的光谱吸收强度,建立含水量与近红外光谱特征之间的相关关系,从而实现含水量测定的目的,与传统方法相比具有实时、无损的优点^[3]。本文基于近红外光谱分析技术对砂岩的光谱特征以及其含水量反演的可行性进行了研究。首先,通过室内试验获取砂岩试样不同饱和度的近红外光谱曲线;其次,基于最大信息系数(maximal information coefficient, MIC)对试样的近红外光谱特征进行了分析和筛选;最后,采用搭建的 BP 神经网络分类器对岩石的含水量进行了反演。

收稿日期: 2021-07-28, 修订日期: 2022-04-05

基金项目: 国家自然科学基金青年科学基金项目(51604276)资助

作者简介: 王东升, 1992 年生, 中国矿业大学(北京)力学与建筑工程学院博士研究生 e-mail: wds0301@outlook.com

* 通讯作者 e-mail: zhangf76@126.com

1 实验部分

1.1 光谱曲线采集

砂岩试样取自中国陕西省榆林市神木县柠条塔煤矿，经加工后制作成标准试样，试样及监测点位置见图 1。数据采集系统采用瑞士万通的 XDS Smart Probe 近红外光谱分析仪，见图 2。该仪器采集光谱范围为 400~2 500 nm，数据采集间隔为 0.5 nm，采用漫反射方式采集砂岩试样的光谱信号。试验中砂岩试样不同饱和度的近红外光谱曲线采集步骤如下^[4]：

(1)将砂岩试样放入干燥箱中干燥 24 h，待冷却至室温后取出并称重，测量饱和度为 0% 的近红外光谱曲线；

(2)将砂岩试样放入恒温水箱中煮沸 8 h，待冷却至室温后取出擦去表面自由水分，测量饱和度为 100% 的近红外光谱曲线；

(3)将饱和度为 100% 的砂岩试样放到天平上进行蒸发试验，观察含水量的变化，当达到理论计算饱和度时(99%，98%，97%，...，1%)，放入密封袋中待其内部含水分布均匀后测量其近红外光谱曲线；

(4)重复步骤(3)即可采集到砂岩不同饱和度的近红外光谱曲线。

试验过程中为了保证采集到的近红外光谱曲线能够较准确的包含砂岩含水饱和度的信息，测点选取位置为试样中部并将光纤探头垂直接触试样表面。每块试样采集的饱和度位于 0~100% 之间，整个试验共采集到 120 条近红外光谱曲线。

1.2 近红外光谱预处理

近红外光谱仪在采集光谱数据的时候受机器、试样以及外界的干扰会存在各种噪声，比如高频随机噪声、基线漂移和光散射等^[5-6]。因此，想要通过近红外光谱曲线特征准确的分析试样的物质成分，光谱的预处理显得格外重要。首先，利用马氏距离法剔除试验中采集到的异常光谱曲线^[7]；

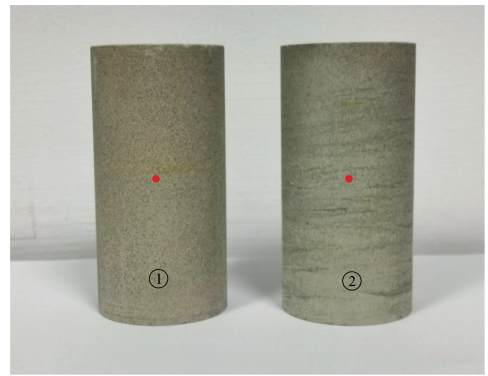


图 1 岩样及其测试点位置(红点)

Fig. 1 Rock sample and its test point location (red dot)

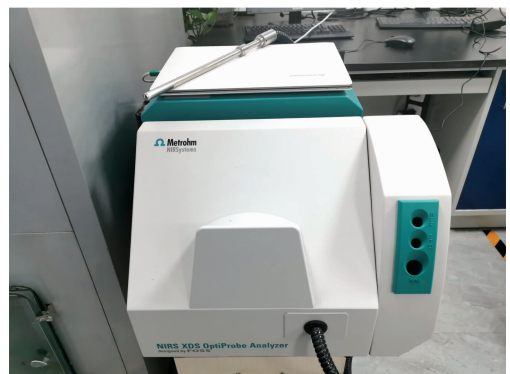


图 2 XDS Smart Probe 近红外光谱分析仪

Fig. 2 XDS Smart Probe near infrared spectrum analyzer

其次，对剔除后剩余的光谱曲线进行一阶导数预处理，消除仪器背景或漂移对信号的影响，提高光谱信号的分辨率和灵敏度^[8-9]。图 3(a)是砂岩的原始近红外光谱曲线，图 3(b)是一阶导数预处理后的近红外光谱曲线，受篇幅的影响只展示了部分原始和一阶导数预处理后的近红外光谱曲线。

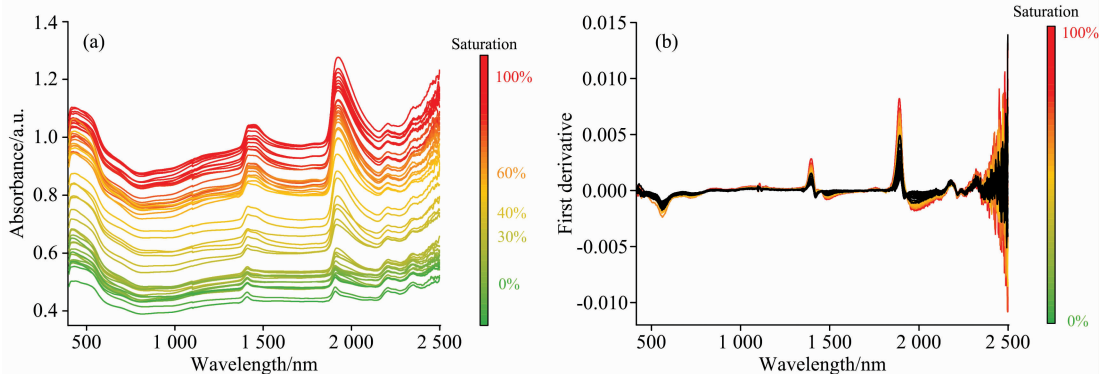


图 3 不同饱和度砂岩近红外光谱

(a): 原始光谱; (b): 一阶导数预处理后光谱

Fig. 3 Near-infrared spectra of sandstone with different degrees of saturation

(a): Original spectra; (b): First derivative pre-processed spectra

含水砂岩在 400~2 500 nm 波长范围内主要有 2 个明显的吸收峰, 1 400 nm 处 OH 和 $\cdot n\text{H}_2\text{O}$ 的倍频、合频, 以及 1 900 nm 处 H_2O 的振动谱带, 依次将这 2 个吸收峰命名为 R1 和 R2。结晶水吸收峰位于 1 400 nm 附近, 谱带比较宽缓; 吸附水的吸收峰位于 1 900 nm 附近, 谱带比较尖锐。在砂岩的原始光谱曲线中, 两个吸收峰 R1 和 R2 随着含水量的增加, 整体吸收强度也在增加, 具有明显的正相关性; 砂岩饱和度在 30%~40% 附近原始光谱曲线发生了整体跳跃, 其可能是砂岩内部水的赋存状态发生了改变。

2 结果与讨论

2.1 光谱特征变量提取及归一化处理

原始光谱曲线经过一阶导数去燥处理后提取其光谱初始特征, 其初始特征变量提取见图 4, 分别为: 峰高(height)、峰面积(area)、左肩宽(left half width)、右肩宽(right half width)、半高宽(full width at half maximum), 共计 5 个初始特征参数, 分别用首字母 H , A , L , R 和 F 表示。各个初始特征变量之间存在量纲和阈值的差异, 可能导致在后续的分析计算过程中, 一些数量级较小的特征变量所占的权重较小其作用无法体现。因此, 对提取的初始特征参数进行归一化处理, 消除量纲和域值差异带来的影响。归一化的方法是将原始数据各元素值减去数据集的最小值后再除以该数据集的极差, 见式(1)

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

受篇幅的限制, 只展示了部分饱和度砂岩试样的归一化结果, 见表 1。

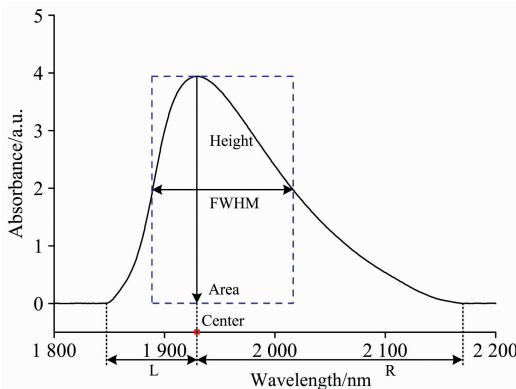


图 4 近红外光谱特征变量的示意图

Fig. 4 Schematic diagram of characteristic variables of near infrared spectrum

2.2 基于 MIC 光谱特征分析

模式识别系统中相关性弱的特征会影响分类的准确性, 因而特征选择就显得尤为重要。特征选择是从系统中挑选出一些相关性强的特征并降低特征空间维数的过程, 其选择结果的好坏将直接影响着模式识别中分类器的分类精度和泛化性能^[10-11]。利用 MIC 进行光谱特征分析和筛选, 其具体计算过程如式(2)和式(3)

$$I_{\max}(C, f, x_i, y_j) = \max I((C, f) |_{G_{x_i, y_j}(j)}), \quad j = 1, 2, \dots, n \quad (2)$$

$$\text{MIC} = \max \left\{ \frac{I_{\max}(C, f, x_i, y_j)}{\log \min(x_i, y_j)} \right\} \quad (3)$$

首先, 计算最大信息值 I_{\max} , 见式(2), 然后, 计算最大信息系数, 见式(3)。 x_i 为行数, y_j 为列数, $x_i, y_j < B(n)$, $B(n)$ 为网格分割细度。 $G_{x_i, y_j}(j)$ 为含水量 C 和特征参数 f 的散点网格, j 表示在确定行数 x_i 和列数 y_j 情况下的划分种类, 其结果见表 2。

表 1 归一化后特征变量

Table 1 Normalized characteristic variable

Saturation	A	H	F	L	R
99%	0.865 9	0.842 6	0.218 2	0.259 5	0.206 4
83%	0.588 9	0.594 6	0.285 8	0.197 7	0.354 1
75%	0.731 0	0.661 9	0.301 6	0.400 3	0.186 0
67%	0.588 4	0.651 8	0.279 2	0.259 1	0.288 0
59%	0.674 8	0.656 6	0.318 1	0.285 7	0.315 0
51%	0.694 3	0.446 7	0.457 4	0.482 1	0.317 1
43%	0.201 1	0.227 3	0.431 1	0.547 0	0.221 5
35%	0.118 6	0.175 0	0.536 6	0.839 7	0.088 6
27%	0.358 2	0.284 5	0.565 8	0.498 8	0.445 8
19%	0.169 1	0.132 1	0.641 2	1.00 0	0.078 3
11%	0.204 0	0.179 6	0.552 9	0.548 1	0.382 6
3%	0.230 0	0.160 5	0.917 0	0.534 3	0.880 3
0%	0.304 8	0.251 1	0.366 0	0.387 5	0.283 6

表 2 饱和度与光谱特征的 MIC 值

Table 2 MIC value calculated to get the correlation degree between saturation and spectral characteristics

Peak	MIC	A	H	F	L	R
R1 (1 400 nm)	S	0.946 0	0.980 4	0.642 3	0.619 4	0.456 5
R2 (1 900 nm)	S	0.960 2	0.980 4	0.906 6	0.620 3	0.952 4

注: S 表示饱和度

Note: S represents degree of saturation

根据表 2 所计算的 MIC 值可知, R1 峰(1400 nm)中饱和度与特征变量的相关性为: $H > A > F > L > R$; 其中 A , H 与饱和度 S 的相关性较高, 分别为 0.946 0 和 0.980 4, 而 F , L , R 与饱和度 S 的相关性较低, 分别为 0.642 3, 0.619 4 和 0.456 5。R2 峰(1 900 nm)中饱和度与特征变量的相关性为: $H > A > R > F > L$; 其中 A , H , F , R 与饱和度 S 的相关性较高, 分别为 0.960 2, 0.980 4, 0.906 6 和 0.952 5, 而 L 与饱和度 S 的相关性较低, 为 0.620 3。整体上来看, 峰 R2 与砂岩饱和度的相关性大于峰 R1, 是含水量反演的主要吸收峰; 分析原因可能是峰 R1 是结晶水的吸收峰, 在饱和度较低的情况下具有较好的相关性, 而在整个饱和度区间, 其与水的相关性就大大减弱。由于 R1 峰的 A , H 和 R2 峰的 A , H , F , R 六个特征与岩石含水量的 MIC 值大于 0.9, 具有较强的相关性, 本文将其作为砂岩含水量识别的特征变量。

2.3 基于 BP 神经网络砂岩含水量的分级识别

为了量化砂岩的含水量, 将砂岩的含水量分为 5 个等

级,即 0~20%, 21%~40%, 41%~60%, 61%~80%, 81%~100%, 并采用自主搭建的 BP 神经网络进行砂岩含水量的识别。整个数据集共采集到 120 条近红外光谱曲线,经过筛选后剩余 102 条,将其按 7:3 的比例随机分为训练集(71 条)和测试集(31 条)。

BP 神经网络是一种按照误差反向传播算法训练的多层前馈神经网络,一般由输入层、隐含层、输出层组成。该网络结构可以很好地将多个特征变量组合起来,自动增强相关性强的特征变量权重,降低相关性弱的特征变量权重。目前,理论上已经表明单隐含层的 BP 神经网络可以拟合出任意的非线性函数,而且在实际的应用中单隐含层的 BP 神经网络已经可以满足工程上精度的需求。

基于近红外光谱砂岩含水量反演的数学思想是综合多个光谱特征变量信息,突出相关性强的变量权重。BP 神经网络采用单隐含层,由于基于 MIC 值筛选后的光谱特征变量共 6 个,所以输入层神经元个数 6(R1-H, R1-A, R2-H, R2-A, R2-F, R2-R);隐含层神经元个数经过计算后最优个数为 10,参见表 3;由于含水量共分为 5 个等级,所以输出层神经元个数为 5。

表 3 隐含层不同神经元个数训练结果

Table 3 Training results of different number of neurons in hidden layer

隐含层神经元个数	准确率/%	
	训练集	测试集
5	72.2	64.5
6	86.1	74.2
7	94.4	74.2
8	94.4	71.0
9	93.1	77.4
10	90.3	83.9
11	95.8	71.0
12	100.0	77.4
13	100.0	74.2
14	100.0	71.0
15	97.2	77.4

将训练集数据输入到 BP 神经网络模型中进行训练,训练完成后输入测试集数据进行测试,训练和测试结果见表 4。

References

- [1] YANG Xiao-jie, WANG Jia-min, ZHANG Min, et al(杨晓杰,王嘉敏,张民,等). Journal of Mining Science and Technology(矿业科学学报), 2017, 2(5): 432.
- [2] WU Bao-yang, LIU Kang, GUO Dong-ming(吴宝杨,刘康,郭东明). Journal of Mining Science and Technology(矿业科学学报), 2020, 5(6): 632.
- [3] ZHANG Fang, ZHANG Xiu-lian, ZHOU Nuan, et al(张芳,张秀莲,周暖,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2021, 41(7): 2028.
- [4] Zhang F, Zhang X L, Hu C, et al. Geotechnical Testing Journal, 2021, 44(3): 564.
- [5] WANG Tao, BAI Tie-cheng, ZHU Cai-die, et al(王涛,白铁成,朱彩蝶,等). Journal of Northwest Forestry University(西北林学院学报), 2020, 35(5): 173.
- [6] LI Shang-ke, LI Pao, DU Guo-rong, et al(李尚科,李跑,杜国荣,等). Journal of Food Safety & Quality(食品安全质量检测学报), 2019, 10(24): 8204.

训练集准确率为 90.3%, 测试集的准确率为 83.9%, 反演效果整体较好,说明基于近红外光谱砂岩含水量反演是可行的。

表 4 识别结果

Table 4 Identification results

类型	准确率/%
0~20%	83.3
21%~40%	83.3
41%~60%	83.3
61%~80%	66.7
81%~100%	100
Training	90.3
Test	83.9

3 结论

基于近红外光谱分析技术对砂岩的光谱特征以及其含水量反演的可行性进行了研究。首先,通过室内试验获取不同饱和度的砂岩试样近红外光谱曲线;其次,基于 MIC 对试样的近红外光谱特征进行了分析和筛选;最后,采用搭建的 BP 神经网络分类器对砂岩的含水量进行了反演。其具体结论如下:

(1)含水砂岩的近红外光谱在 1 400 和 1 900 nm 附近有明显的吸收峰,随着含水量增加,吸收强度也在增加,两者有明显的相关性。吸收峰位于 1 400 nm 附近,谱带比较宽缓,吸附水的吸收峰在 1 900 nm 附近,谱带比较尖锐。

(2)根据计算的 MIC 值,1 400 nm 附近的峰高与含水量的相关性最强,1 900 nm 附近的峰高与含水量的相关性也最强;1 400 nm 附近的峰面积、峰高,1 900 nm 附近的峰面积、峰高、半高宽、右肩宽共 6 个特征变量其 MIC 值 > 0.9,可作为 BP 神经网络反演砂岩含水量的特征变量。

(3)利用 MIC 值筛选出 1 400 和 1 900 nm 两个吸收峰的特征变量进行 BP 神经网络建模,其所建立的砂岩含水量反演模型训练集准确率为 90.3%,测试集的准确率为 83.9%,说明基于近红外光谱分析技术砂岩含水量反演的方法是可行的。

- [7] XIE Yue, ZHOU Cheng, TU Cong, et al(谢 越, 周 成, 涂 从, 等). Chinese Journal of Analytical Chemistry(分析化学), 2017, 45(3): 363.
- [8] TIAN Jing, LI Qiao-ling(田 晶, 李巧玲). Food Science(食品科学), 2018, 39(2): 293.
- [9] Obregon-Cano S, Moreno-Rojas R, Jurado-Millan A M, et al. Foods, 2019, 8(9): 364.
- [10] ZHANG Mu-xing, SHEN Xiao-hong, HE Lei, et al(张牧行, 申晓红, 何 磊, 等). Journal of Northwestern Polytechnical University(西北工业大学学报), 2020, 38(3): 471.
- [11] SUN Guang-lu, SONG Zhi-chao, LIU Jin-lai, et al(孙广路, 宋智超, 刘金来, 等). Acta Automatica Sinica(自动化学报), 2017, 43(5): 795.

Near-Infrared Spectral Characteristics of Sandstone and Inversion of Water Content

WANG Dong-sheng¹, WANG Hai-long^{1, 2}, ZHANG Fang^{1, 3*}, HAN Lin-fang^{1, 3}, LI Yun¹

1. School of Mechanics and Civil Engineering, China University of Mining & Technology, Beijing 100083, China

2. Hebei Key Laboratory of Diagnosis, Reconstruction and Anti-Disaster of Civil Engineering, Zhangjiakou 075000, China

3. State Key Laboratory for Geomechanics & Deep Underground Engineering, China University of Mining & Technology, Beijing 100083, China

Abstract The strength of sedimentary rocks is often affected by water, and the influence degree is different with different water content. Therefore, the measurement of the water content of the rock has important value for the subsequent evaluation of its physical and mechanical properties. In measuring rock water content, traditional methods are often time-consuming and laborious, destroying the integrity of the structure. At present, near-infrared spectroscopy is a very potential method, with real-time and nondestructive advantages. In this paper, sandstone's spectral characteristics and the water content prediction are studied based on near infrared spectroscopy. Firstly, near-infrared spectrum curves of sandstone samples with different saturations were obtained through laboratory tests. Secondly, the first derivative of the original spectral curve is preprocessed to eliminate the influence of noise, environment, and other factors. Thirdly, the spectral characteristic variables of R1 (1 400 nm) and R2 (1 900 nm) were extracted and normalized to eliminate the influence of dimension and domain value. Fourthly, the extracted spectral characteristic variables are analyzed and screened based on the maximum information coefficient; Finally, the self-built BP neural network classifier is used to predict the water content of sandstone. The conclusions are as follows: (1) The near-infrared absorption curve of water-bearing sandstone has obvious absorption peaks near 1 400 and 1 900 nm, the absorption peak is near 1400 nm, the band is relatively broad, the absorption peak is near 1 900 nm, and the band is relatively sharp. As the water content increases, the absorption intensity of each absorption peak is also increasing, which has an obvious correlation and can be used as the main spectrum band for sandstone water content analysis and prediction. (2) According to the calculated maximum information coefficient value, the peak height near 1 400 nm has the strongest correlation with water content, and the peak height near 1900 nm has the strongest correlation with water content. Peak area and peak height near 1 400 nm, peak area, peak height, half-height width, and right shoulder width near 1 900 nm are 6 characteristic variables. The maximal information coefficient value is greater than 0.9, which can be used as characteristic variable for BP neural network to predict sandstone water content. (3) Using the maximum information coefficient to screen out the characteristic variables of the two absorption peaks at 1 400 and 1 900 nm for BP neural network modeling, the accuracy of the training set of the sandstone water content prediction model established by it was 90.3%, and the accuracy of the test set was 83.9%. The method based on near-infrared spectroscopy analysis technology to predict the water content of sand and gravel is feasible.

Keywords Rock; Water content; Near-infrared spectroscopy; Spectral characteristics; Artificial neural network

* Corresponding author

(Received Jul. 28, 2021; accepted Apr. 5, 2022)