

可见光-近红外光谱的矽卡岩型铁矿反演模型

毛亚纯¹, 温健^{1*}, 付艳华², 曹旺¹, 赵占国³, 丁瑞波¹

1. 东北大学资源与土木工程学院, 辽宁 沈阳 110819

2. 东北大学江河建筑学院, 辽宁 沈阳 110819

3. 中国黄金集团, 北京 100000

摘要 铁矿资源是我国国民经济基础产业中的重要组成要素,在我国经济发展中有举足轻重的地位。铁矿品位的检定效率对铁矿石开采效率有重大影响。目前,铁矿石品位的化学分析检定法,不仅存在成本较高,化验周期长的问题,更主要的是其无法实现铁矿品位原位测定,相对配矿流程存在滞后效应,无法有效降低矿石开采的损失贫化率;基于可见光-近红外光谱分析的铁矿品位原位测定技术是解决这一问题的有效途径。以225个红岭矽卡岩型铁矿测试样本的可见光-近红外光谱数据及化学分析数据为数据源,首先对原始数据进行了平滑处理,并分析了矽卡岩型铁矿可见光-近红外光谱特征,然后利用倒数对数、多元散射校正(MSC)两种预处理方法对平滑后的光谱数据进行处理,再分别以主成分分析法(PCA)、遗传算法(GA)两种降维算法对预处理前后的光谱数据进行了处理,获取了六种不同预处理组合算法处理后的数据源。其中以PCA降维算法所降维数分别为3维、3维、7维;以GA降维算法所降维数分别为477维、489维、509维。最后基于随机森林(RF)和极限学习机(ELM)建立了矽卡岩型矿石金属铁品位的定量反演模型,以决定系数(R^2)、均方根误差(RMSE)和平均相对误差(MRE)三个指标分别对模型的稳定性、精确度、可信度进行评价。结果表明,经MSC处理及PCA降维后的数据基于ELM算法建立的定量反演模型效果最优,其 R^2 可达0.99、RMSE为0.0057、MRE为2.0%,该方法所建模型对红岭矽卡岩型铁矿品位反演精度有明显的提升。对矽卡岩铁矿品位的实时、快速分析提供了一种有效的方法,对实现矽卡岩型铁矿的高效开采具有重要的现实意义。

关键词 可见光-近红外光谱;矽卡岩铁矿;降维算法;预处理组合算法;定量反演模型

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)01-0068-06

引言

铁矿是我国经济发展的重要矿产资源,其中矽卡岩型铁矿是我国重要的铁矿床类型之一,其储量约占全国铁矿床总储量的11%,矿石类型以磁铁矿为主^[1]。传统的矿石品位检定方法以化学分析法为主,检测方法较为准确,但由于成本昂贵、检定周期较长,无法实现矿石品位的即时原位测定^[2],相对配矿流程均存在滞后效应,因此难以有效降低矿石开采的损失贫化率。如今,如何做到低成本,快速、准确确定铁矿品位及合理、高效开采铁矿已成为铁矿开采中亟待解决的关键问题。高光谱遥感由于其高分辨率、波谱连续、信息丰富^[3],已被广泛应用于植被指数反演、土壤含盐量反

演、重金属品位反演等领域^[4-6]。

由于原始高光谱数据存在数据冗余、信噪比低等问题,严重影响其特征分析的准确性和建模反演精度,为此大量学者在高光谱数据的预处理、及降维处理等方面进行了研究^[7-8],这些方法均有效降低了高光谱数据的冗余度,快速准确地提取出有效的光谱信息,达到提升预测精度降低误差的目标。同时国内外很多学者也对数据处理以及建模方法进行了大量的研究。Chudnovsky等基于偏最小二乘法多元分析,证明了仅通过沉积物粉尘的可见光-近红外高光谱数据即可对其进行准确预测^[9];高伟等以铁矿粉的高光谱数据为数据源,建立了多种反演模型,其中SFIM-RFR模型预测结果误差最小,证明了基于高光谱数据预测铁矿粉中全铁品位的可行性和有效性^[10];陈俊英等建立了SNV-SR-ELM模

收稿日期:2020-11-30,修订日期:2021-03-09

基金项目:国家自然科学基金项目(52074064)资助

作者简介:毛亚纯,1966年生,东北大学资源与土木工程学院教授

* 通讯作者 e-mail: 1241042848@qq.com

e-mail: dbdxmyc@163.com

型,对高光谱反演水质模型的优化以及污水水质的快速监测和综合评价提供了有效途径^[11]。虽然国内外学者对高光谱数据处理及建模方法的研究已经取得了一定的进展,但目前基于矽卡岩型铁矿高光谱数据的研究相对较少。

以红岭矽卡岩型铁矿的化学分析与光谱测试数据为数据源,并对其数据进行预处理及降维多种组合算法处理,同时以随机森林算法和极限学习机算法两种算法为建模方法建立矿石铁品位的定量反演模型。结果表明,经 MSC 处理及 PCA 降维后的数据基于 ELM 算法(MSC-PCA-ELM)建立的定量反演模型效果最优,其中决定系数 R^2 为 0.99、均方根误差 RMSE 为 0.005 7、平均相对误差 MRE 为 2.0%,由此可见利用该方法可准确快速反演矽卡岩型铁矿的品位,为我国矽卡岩型铁矿品位的快速原位分析提供了有效手段。

1 实验部分

1.1 研究区与矿石采样

红岭铅锌矿位于我国内蒙古赤峰市,主要开采铁、锌矿体,是以铁、铅、锌为主的矽卡岩型多金属矿床。2019年6月在内蒙古赤峰红岭矿区采集了井下矽卡岩型铁矿样本。为保证所采集的样本具有代表性、多样性及建模的普适性,故采取矿区均匀抽样的方式采集了相应的矽卡岩型铁矿样本,如图 1(a)所示。

为使所建模型在井下具有实际应用性,因此对矽卡岩型铁矿样本进行了钻孔、取芯及切块处理,最终制成共 225 件块状样本,如图 1(b)和(c)所示。

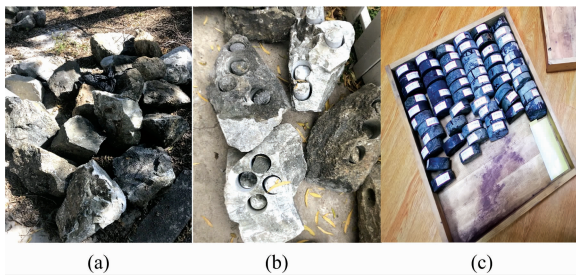


图 1 实验样品

Fig. 1 Experimental samples

1.2 矿石铁品位与光谱测定

使用美国 SVC HR-1024 便携式地物光谱仪采集光谱,波段范围 350~2 500 nm,通道数为 1 024,最小积分时间为 1 s。在观测角度等条件保持不变的情况下,分别以太阳光和卤素灯光作为光源的测试结果基本一致。但考虑到井下应用将以卤素灯作为光源,因此以卤素灯作为测试光源,并在夜间封闭环境条件下对样本进行了测试,测试时使样本观测面保持水平,光谱仪镜头垂直于样品观测面,采样积分时间设置为 2 s,每个样品重复测试 2 次,视场角为 4°。为避免光谱测试出现的偶然性,取两次反射率平均值作为该矿石的实际反射光谱数据。在实验过程中,每隔 10~15 min 进行一次白板测定。

光谱采集完毕后,为了进一步降低噪声的干扰,采用

Savitzky Golay 法对原始光谱数据进行了平滑处理^[12],图 2 是 225 个块状矽卡岩型铁矿样本的光谱曲线。

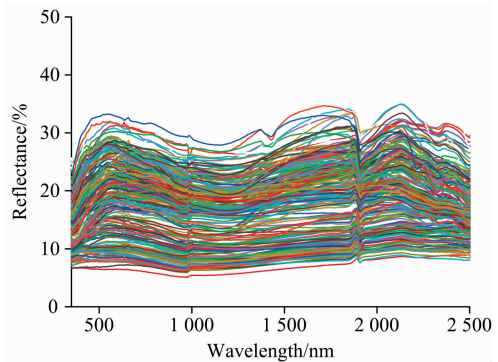


图 2 样品可见光-近红外光谱曲线

Fig. 2 Visible and near infrared spectra of samples

光谱特征如下:

(1)样品的光谱反射率大部分在 10%~30%之间。

(2)在 400~550 nm 反射率为上升趋势,一部分曲线上上升趋势显著,斜率较大,另一部分曲线上上升趋势缓慢,斜率较小,且在 550 nm 附近出现波峰。

(3)在 550~1 180 nm 反射率下降,其中在 870 nm 附近出现微弱波谷。

对上述现象进行分析发现,在 350~550 nm 间的光谱差异与样本铁品位有一定相关性,斜率小的样本普遍铁品位均值高于斜率大的样本铁品位均值。

光谱测试结束后,将全部实验样品进行了研磨化验处理,以确定各个样品的铁品位。由化验结果得出,样本铁品位在 6.75%~66%之间,平均品位为 27.84%。

2 结果与讨论

2.1 光谱数据预处理

由于原始高光谱数据存在数据冗余、信噪比低等问题,为了增强光谱信息,突出光谱特征,提高信噪比,采用倒数对数、多元散射校正两种预处理方法。

(1)倒数对数法

倒数对数法不仅可以有效增强光谱在可见光波段的差异,还可以有效的减弱因测试时光照条件变换所引起的乘性因素的影响。倒数对数法计算公式如式(1)所示

$$\theta' = \log\left(\frac{1}{\theta}\right) \quad (1)$$

式(1)中, θ 为平滑后的原始光谱数据, θ' 为经倒数对数处理后的光谱数据。图 3 是 225 个块状矽卡岩型铁矿样本经过倒数对数处理后的光谱曲线。

(2)多元散射校正

多元散射校正处理可有效降低因散射对光谱数据的影响,在一定程度上增强特征波段的有效信息^[13]。该算法的具体实现过程为:首先由式(2)计算样本的平均光谱作为标准光谱,然后将各种原始光谱与标准光谱作一元回归,如式(3)所示,最后由式(4)计算多元散射校正后的光谱数据。

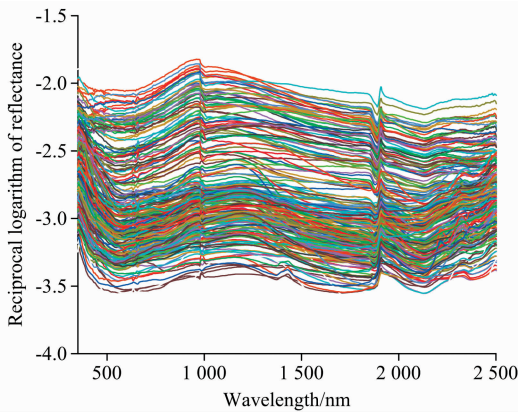


图 3 倒数对数处理后的光谱曲线

Fig. 3 Spectral curves after reciprocal logarithm processing

$$\bar{A}_{m,n} = \frac{\sum_{i=1}^n A_{m,i}}{n} \quad (2)$$

$$A_m = l_m \bar{A} + B_m \quad (3)$$

$$A_{m(msc)} = \frac{A_m - B_m}{l_m} \quad (4)$$

式中, A 为 $i \times \omega$ 维定标光谱矩阵, i 为样品数, ω 为光谱采集时所用波段数, l_m 和 B_m 分别表示经平滑处理后的原始光谱数据和平均光谱数据作为一元线性回归后的相对偏移系数和平移量。图 4 是 225 个块状矽卡岩型铁矿样本经过多元散射校正处理后的光谱曲线。

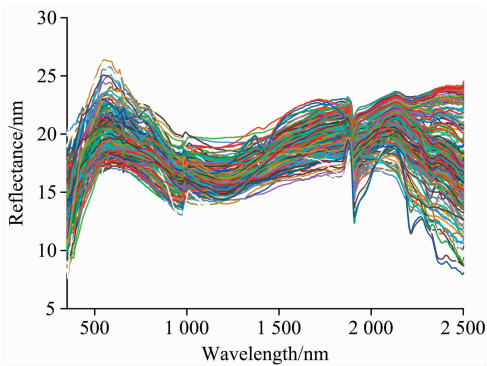


图 4 多元散射校正后的光谱曲线

Fig. 4 Multivariate scatter corrected spectral curves

2.2 降维处理

(1) 遗传算法

遗传算法 (genetic algorithm, GA) 作为一种自适应的全局概率搜索算法, 它的优越性很大程度体现在建模自变量提取方面。利用遗传算法对 225 个样本进行波长筛选时, 先将矿石样本按照铁品位从小到大均匀选出训练集 169 个与测试集 56 个。然后根据高光波段数, 设置染色体长度为 973, 初始种群大小设为 20, 最大繁殖代数设置为 100, 设置交叉概率和变异概率范围为 $[0, 1]$, 然后采用 BP 神经网络法对每一个体计算个体适应度 \hat{W} , 其中训练函数采用 trainlm

函数, 设置学习率为 0.000 1, 训练次数为 1 000 次。每次计算均用遗传算法对神经网络权值和阈值进行优化, 具体步骤如图 5 所示, 其中以测试集铁品位预测值与真实值的误差平方和的倒数构建适应度函数, 如式 (5) 所示

$$g(X) = \frac{1}{SE} = \frac{1}{\text{sse}(\hat{W} - W)} = \frac{1}{\sum_{i=1}^n (\hat{w}_i - w_i)^2} \quad (5)$$

式 (5) 中 $\hat{W} = \{\hat{w}_1, \hat{w}_2, \dots, \hat{w}_n\}$ 为样本铁品位预测值, $W = \{w_1, w_2, \dots, w_n\}$ 为样本铁品位真实值, n 为实验测试样本的数目。

未经处理的数据、经倒数对数变换后的数据和经多元散射校正变换后的数据经遗传算法处理后, 维度由 973 维分别降至 477 维、489 维和 509 维。

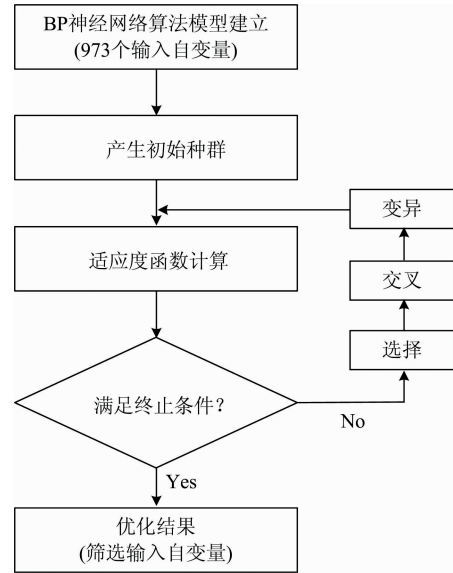


图 5 遗传算法选择最优波段流程图

Fig. 5 Flow chart of genetic algorithm to select the optimal band

(2) 主成分分析法

主成分分析 (principle component analysis, PCA) 是一种线性数据降维分析方法, 其主要思想是通过线性变换提取原始数据中的主要特征, 在减少数据冗余的同时保留原始数据的绝大多数有用信息, 从而解决特征维数过高的问题, 即用压缩后尽可能少的信息来代替原有的信息^[14]。步骤如下:

设原始矩阵为 $X = \{x_0, x_1, \dots, x_n\}$, 其中 $x_0 = x$, 首先计算原始矩阵的协方差阵以及它的最大特征值和对应的特征变量, 并组成投影矩阵 V , $V = [v_1, v_2, \dots, v_w]$ 。其中, w 的值由累积贡献率 $F_{(w)}$ 确定, $F_{(w)}$ 的计算公式如式 (6) 所示, 由式 (7) 得到降维后的向量 $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n]$ 。

$$F_{(w)} = \frac{\sum_{k=1}^w \lambda_k}{\sum_{i=1}^n \lambda_i} \quad (6)$$

$$\tilde{x}_i = V^T (x_i - \bar{x}) \quad (7)$$

累积贡献率设置为 99%, 以累计贡献率为依据, 计算最终维度。如图 6(a, b, c) 所示, 横坐标为依次各主成分, 纵坐标为各主成分贡献率, 未经处理的数据、经倒数对数变换后

的数据和经多元散射校正变换后的数据利用 PCA 算法处理后维度由 973 维分别降至 3 维、3 维和 7 维。

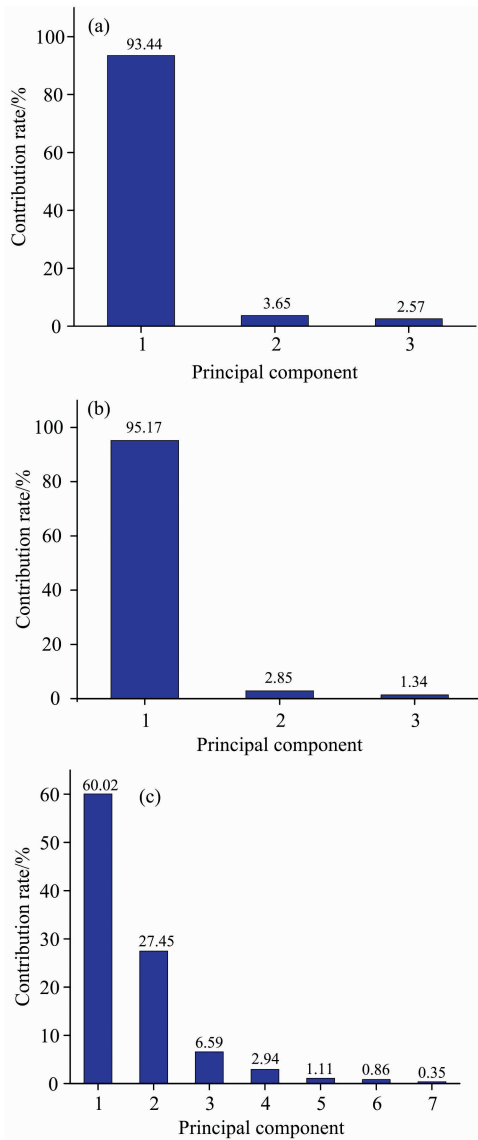


图 6 主成分分析结果

Fig. 6 Principal component analysis results

2.3 模型建立与验证

随机森林(random forests, RF)是由 Leo Breiman 提出的一种基于统计学的非线性组合智能机器学习算法。对于回归问题, RF 模型通过多次 bootstrap 抽样获得随机样本, 然后通过学习样本特征分别建立相应的决策树, 最后基于投票和平均的方法输出多个不同功能决策树的最终结果。在算法模型建立过程中, 设置决策树的数量为 500。

极限学习机(extreme learning machine, ELM), 是黄广赋等依据广义逆矩阵理论提出的一种性能优良的单隐含层前馈神经网络模型。它不仅具有较强的非线性拟合能力, 还因较其他算法模型速度更快、精度更高、参数调整简单而被广泛应用于多个领域。ELM 在执行过程中随机产生输入层和隐含层间的连接权值及隐含层的神经元的阈值且在训练中无

需调整, 可以获得最优解^[15]。模型建立过程中将隐含层节点数设置为 30。

模型的稳定性、精确度、可信度分别由决定系数 R^2 、均方根误差 RMSE 以及平均相对误差 MRE 来检验。

选择 169 个样品为训练样本和 56 个样品为测试样本分别建立随机森林算法模型和极限学习机算法模型。表 1 和表 2 分别为未经处理的数据以及对数据进行不同预处理后使用随机森林算法和极限学习机算法建模的结果。

表 1 RF 反演模型反演结果评价

Table 1 Assessment of RF inversion results

预处理结果	R^2	RMSE	MRE
未经处理的数据	0.91	0.054	0.16
GA	0.88	0.054	0.17
倒数对数-GA	0.88	0.068	0.20
MSC-GA	0.89	0.059	0.18
PCA	0.91	0.053	0.16
倒数对数-PCA	0.89	0.061	0.21
MSC-PCA	0.91	0.051	0.16

表 2 ELM 反演模型反演结果评价

Table 2 Assessment of ELM inversion results

预处理结果	R^2	RMSE	MRE
未经处理的数据	0.90	0.057	0.20
GA	0.91	0.056	0.20
倒数对数-GA	0.95	0.044	0.16
MSC-GA	0.88	0.073	0.22
PCA	0.94	0.042	0.16
倒数对数-PCA	0.91	0.055	0.20
MSC-PCA	0.99	0.0057	0.02

综合两个表中的数据, 利用经 MSC 处理、PCA 降维后的数据基于 ELM 算法建立的品位定量反演模型效果最优。如图 7 所示, 经该方法处理后使用 ELM 预测的预测值和真实值作拟合曲线, 预测值与真实值的决定系数达到 0.99, 均方根误差为 0.005 7, 平均相对误差为 2.0%, 与未经处理的数据建立的模型相比较, 有较大的提升, 预测效果更为精确。

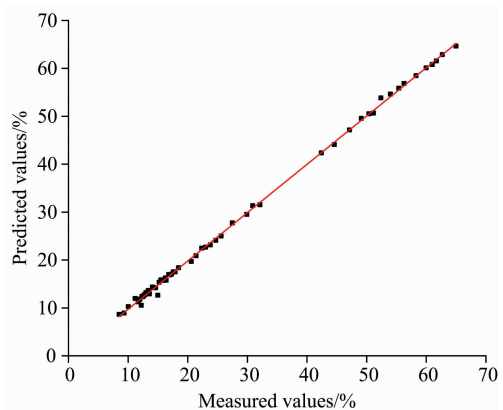


图 7 预测值与真实值差异图

Fig. 7 The difference between the predicated value and the true value

综合分析上述不同方法处理之后的结果,其中经 MSC 处理、PCA 降维后的数据基于 ELM 算法建立的品位定量反演模型效果最优。未经处理的数据可能受到设备的局限性以及实验环境的影响,产生基线平移偏移等现象,对建模造成负面影响,而通过 MSC 处理之后的数据能很大程度上消除这种影响,突出光谱特征信息的同时降低信噪比,有助于反演模型的精度提升,而遗传算法在路径寻优上具有偶然性,且容易收敛到局部最优解,因此满足不了精度要求。而主成分分析能最大程度的提取经过 MSC 处理之后光谱数据主要信息,因此经过该方法建立的模型反演结果精度最优。由于随机森林是一种集成算法,因此经过预处理的数据对其反演模型精度没有产生太大变化;而极限学习机算法,学习速度极快,泛化能力强,预测结果精确,但容易受到噪声以及无用信息的干扰,导致对未经处理的数据反演建模精度较低,而选择合适的预处理方法能最大程度消除负面影响。

经 MSC 处理、PCA 降维后的光谱数据,以 ELM 为极限学习机模型对红岭矽卡岩型铁矿品味反演不仅能大幅度提升模型反演速度,而且满足高精度、高效率的品位反演需求。

3 结 论

以 225 个赤峰红岭矽卡岩型铁矿的化学分析与可见光-近红外光谱测试数据为数据源,深入研究了测试数据的预处理方法以及定量反演模型,结论如下:

(1) 利用 MSC 算法对矽卡岩型铁矿光谱数据进行处理可有效降低散射对数据的影响。利用 PCA 算法对矽卡岩型铁矿光谱数据进行降维处理可有效降低原始数据冗余,提升建模速度。

(2) 对于矽卡岩型铁矿,使用经 MSC 处理、PCA 降维后的数据,以极限学习机为建模方法,反演结果最优。其中 R^2 由 0.94 提升至 0.99, MSE 由 0.042 0 降低至 0.005 7, MRE 从 16% 降低至 2%, 预测精度较高。

针对矽卡岩型铁矿的原位快速品位分析提供了一种有效方法。但由于不同类型矿体的光谱测试结果会存在不同程度的差异,因此数据处理方法、所建模型精度也会不同,对此尚需开展进一步深入研究。

References

- [1] LI Hou-min, WANG Deng-hong(李厚民,王登红). *Geology in China(中国地质)*, 2012, 39(3): 559.
- [2] Hu H, Tang Y, Ying, et al. *Talanta*, 2014, 125: 425.
- [3] Meer F V D. *International Journal of Applied Earth Observation & Geoinformation*, 2014, 5(1): 55.
- [4] CHEN Zhong-xin, REN Jian-qiang(陈仲新,任建强). *Journal of Remote Sensing(遥感学报)*, 2016, 20(5): 748.
- [5] Ridel F, Denk M, Müller L, et al. *Geoderma*, 2018, 315: 188.
- [6] Kemper T, Somme S. *Environmental. Science & Technology*, 2002, 36(12): 2742.
- [7] Meng S, Huang L T, Wang W Q. *IEEE Geoscience and Remote Sensing Letters*, 2017, 13(7): 897.
- [8] Serpico S B, D'Inca M, Melgani F, et al. *Proc SPIE*, 2003, 4(10): 17.
- [9] Chudnovsky A, Ben-Dor E, Paz E. *Journal of Near Infrared Spectroscopy*, 2007, 15(1): 59.
- [10] GAO Wei, YANG Ke-ming, LI Meng-qian, et al(高伟,杨可明,李孟倩,等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(8): 2546.
- [11] CHEN Jun-ying, XING Zheng, ZHANG Zhi-tao, et al(陈俊英,邢正,张智韬,等). *Transactions of The Chinese Society for Agricultural Machinery(农业机械学报)*, 2019, 50(11): 200.
- [12] Roth A. *Computers in Physics*, 2018, 4(6): 669.
- [13] Wu Y, Peng S, Xie Q, et al. *Chemometrics and Intelligent Laboratory Systems*, 2019, 185: 114.
- [14] Cui M S, Prasad S, Li W, et al. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2013, 6(3): 1688.
- [15] Deng C W, Huang G B, Jia X U, et al. *Science China (Information Sciences)*, 2015, 58(2): 20301.

Quantitative Inversion Model Based on the Visible and Near-Infrared Spectrum for Skarn-Type Iron Ore

MAO Ya-chun¹, WEN Jian^{1*}, FU Yan-hua², CAO Wang¹, ZHAO Zhan-guo³, DING Rui-bo¹

1. School of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China

2. Northeastern University JangHo Architecture, Shenyang 110819, China

3. China National Gold Group Co., Ltd., Beijing 100000, China

Abstract Iron ore resources are an important component of the basic industries of China's national economy and play a pivotal role in China's economic development. In particular, the efficiency of iron ore grade determination has a significant impact on the efficiency of iron ore mining. At present, the method of iron ore grade determination is mainly based on chemical analysis. However, it not only has the problems of high cost and long assay cycle but also cannot achieve the in-situ determination of iron ore grade, which has a lag effect relative to the ore allocation process and cannot effectively reduce the loss depletion rate of ore mining, so the in-situ determination of iron ore grade based on visible and near-infrared spectral analysis is an effective way to solve this problem. This paper uses visible and near-infrared spectral data and chemical analysis data from 225 test samples of Hongling Skarn-Type iron ores as data sources. First, the original data were smoothed and analyzed for visible and near-infrared spectral characteristics of Skarn-Type iron ores, and then the smoothed spectral data were processed by using two pre-processing methods, including logarithm of reciprocal and multiple scattering correction (MSC). -Before and after pre-processing, the spectral data were processed using two-dimensionality reduction algorithms, including genetic algorithm (GA) and principal component analysis (PCA), and obtain the data sources were processed by six different pre-processing combination algorithms. The PCA dimensionality reduction algorithm was used to reduce the dimensionality of the spectral data before and after the pre-processing, and the dimensionality reduced were 3, 3 and 7 dimensions respectively; The GA dimensionality reduction algorithm was used to reduce the dimensionality of the spectral data before and after the pre-processing, and the dimensionality reduced were 477, 489 and 509 dimensions respectively. Finally, based on Random Forest (RF) and Extreme Learning Machine (ELM), a quantitative inversion model of iron grades in skarn-type ores was established, and the stability, accuracy and credibility of the model were evaluated in terms of coefficient of determination (R^2), root means square error (RMSE) and mean relative error (MRE). The results show that the quantitative inversion model based on the ELM algorithm, using MSC-processed and PCA-dimensioned data, is the most effective, with R^2 of 0.99, RMSE of 0.005 7 and MRE of 2.0%. The accuracy of the model for inversion of HongLing skarn-type iron ore grades has been significantly improved. This research provides an effective method for the real-time and rapid analysis of skarn-type iron ore grade, which is of great practical importance for efficient skarn-type iron ore mining.

Keywords Visible and near-infrared spectroscopy; Skarn-type iron ore; Dimensionality reduction algorithm; Preprocessing combination algorithm; Quantitative inversion model

(Received Nov. 30, 2020; accepted Mar. 9, 2021)

* Corresponding author