

doi: 10.6046/gtzyyg.2017.01.24

引用格式: 于峻川, 刘文良, 闫柏琨, 等. 基于野外实测光谱的玄武岩地球化学成分反演[J]. 国土资源遥感, 2017, 29(1): 158 - 163. (Yu J C, Liu W L, Yan B K, et al. Inversion of geochemical compositions of basalts based on field measured spectra[J]. Remote Sensing for Land and Resources, 2017, 29(1): 158 - 163.)

基于野外实测光谱的玄武岩地球化学成分反演

于峻川¹, 刘文良², 闫柏琨¹, 董新丰¹, 王喆¹, 李娜¹

(1. 中国国土资源航空物探遥感中心, 北京 100083; 2. 中国地质大学(北京)地球科学与资源学院
地质过程与矿产资源国家重点实验室, 北京 100083)

摘要: 岩石地球化学成分对岩石分类、成因及演化研究有重要地质意义。利用遥感手段反演岩石地球化学成分是一个较新的课题, 也是遥感相关应用研究的难点。以甘肃省柳园镇南部二叠纪玄武岩带为研究目标, 在进行系统采样、光谱分析、地球化学测试的基础上, 采用偏最小二乘回归(partial least squares regression, PLSR)对拥有 2 150 个波段的实测波谱数据及相应的 6 种主量矿物数据进行反演建模。首先选择有效的预处理方法对目标数据集进行优化, 再利用 k 折交叉检验方法获得最小均方根误差下的最适主成分个数。运算结果表明, PLSR 模型具有较好的稳定性和精度, 在利用遥感数据进行岩石地球化学成分反演方面有很好的应用前景。

关键词: 玄武岩; 地球化学成分; 反演; 实测光谱; 偏最小二乘回归(PLSR); 柳园

中图分类号: TP 79 **文献标志码:** A **文章编号:** 1001-070X(2017)01-0158-06

0 引言

近年来, 遥感技术作为一种重要技术手段从地质解译、矿物信息提取到岩石地球化学成分含量反演, 已广泛应用于地质调查领域。随着遥感技术在光谱分辨率和空间分辨率方面的不断提高, 遥感地质应用由定性分析走向定量分析, 由岩石到矿物再到元素的反演研究已成为遥感地质的发展趋势。目前, 人们已在利用遥感数据进行岩石地球化学成分反演方面做了很多尝试。在地质环境方面, Choe 等^[1]和 Song 等^[2]利用地面光谱或高光谱数据提取重金属元素和稀土元素, 实现了对土壤污染及矿山环境的检测。在月球探测方面, 通过利用环月轨道探测器(clementine)、干涉成像光谱仪(interference imaging spectrometer, IIM)、月球矿物制图仪(moon mineralogy mapper, M3)、“预言者”(diviner)月球辐射计(lunar radiometer)等遥感数据, 基于角度参数、吸收深度、克里斯琴森指数(Christiansenfeature, CF)等方法, 实现了 Fe, Ti 等元素含量的反演^[3-6]。在地表岩石成分反演方面, 部分学者尝试利用热红外数据实现了对 SiO₂ 含量的反演^[7-9], 但对利用其他

遥感数据对 SiO₂ 之外岩石成分反演的研究较少, 有待于进一步尝试。从地质角度讲, 岩浆作用与成矿作用关系密切, 岩浆岩中隐含着其成因、演化及构造环境等重要信息, 而玄武岩更是探测地球深部的“探针”和“窗口”^[10-11], 对岩浆岩地球化学成分的反演对岩石分类、成因及演化研究具有重要意义。

野外岩矿实测光谱数据具有较高的空间分辨率和光谱分辨率, 为地表岩石地球化学成分的定量反演提供了广阔的应用空间。本文以甘肃省柳园镇南部二叠纪玄武岩带为研究目标, 采用偏最小二乘回归(partial least squares regression, PLSR)模型, 对实测光谱数据进行了玄武岩的主量矿物反演, 为高光谱遥感在岩石地球化学成分反演方面的研究和应用提供理论参考。

1 研究区地质概况

研究区位于甘肃省西北部, 隶属于柳园镇, 在大地构造上位于北山南部陆缘活动带上。该区在二叠纪以伸展张作用为主, 地幔活动频繁, 有大量基性-超基性岩发育。研究区的地层以二叠系为主, 主要包括下二叠统哲斯群和上二叠统。下二叠统不整合

收稿日期: 2015-07-27; 修订日期: 2015-08-28

基金项目: 国土资源部公益性行业科研基金项目“月球试验场遥感探测研究”(编号: 201211071-2)和中国地质调查局地质调查项目“高光谱地质调查技术方法研究”(编号: 12120115040801)共同资助。

第一作者: 于峻川(1984-), 男, 博士, 工程师, 主要从事高光谱及相关地质研究工作。Email: jasonyu@live.cn。

于下古生界地层之上,与上二叠统也呈不整合接触。下二叠统哲斯群分为上、下 2 组,上组发育玄武岩,下组则以凝灰质砂岩及砂砾岩为主。哲斯群上组的底部为深绿色微晶质细粒玄武岩和灰绿色中细粒玄武岩夹少量枕状玄武岩^[12],向上为灰绿色中粒枕状玄武岩和微晶细粒枕状玄武岩,总厚度约 2 600 m。玄武岩表面呈墨绿色,块状、枕状构造;镜下呈斑状结构,斑晶为粒状辉石和长板状斜长石(含量约 10%),基质由长石及辉石微晶组成。根据玄武岩的岩石类型及展布特征共设计 6 条采样路线,对整个玄武岩带进行了较好的控制;并开展了细致的采样工作,共采集样品 120 个(图 1),用于后期分析测试。

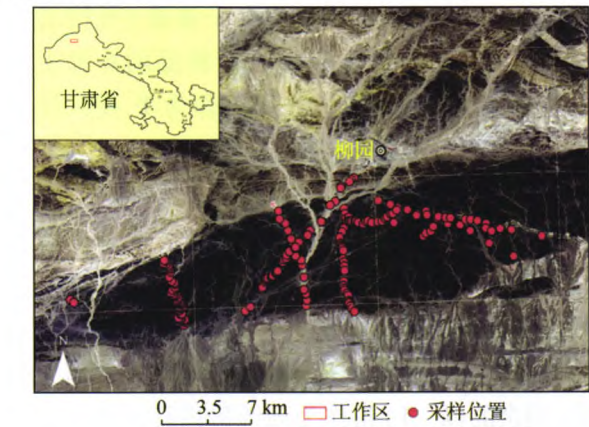


图 1 甘肃柳园玄武岩采样位置

Fig. 1 Sampling location of Liuyuan basalts in Gansu Province

2 研究方法

2.1 PLSR 理论

偏最小二乘回归(PLSR)是一种多因变量对多自变量的新型的多元统计数据分析方法,其优势在于对数据建模的同时能消除数据之间的冗余信息,达到降维的目的,且具简单稳定、计算量小等优点,可较好地解决许多普通多元回归方法无法解决的问题。该方法于 1983 年由 Wold 等^[13]首次提出;近 10 a 来,在理论和方法上得到了迅速的发展,并在各研究领域均有应用。其中,PLSR 方法结合可见-近红外光谱在医药检验、食品安全等方面的应用均取得了较好效果^[14]。为此,本文将作为研究的理论基础。

PLSR 的基本原理是将 n 个样品的 m 个组分含量矩阵 $Y = (y_{ij})_{n \times m}$ 和测定的 n 个样品的 p 个波段处的反射率值矩阵 $X = (x_{ij})_{n \times p}$ 分解成特征向量的形式,即

$$Y = U \cdot Q + F, \tag{1}$$

$$X = T \cdot P + E, \tag{2}$$

式中: U 和 T 分别为 n 行 d 列的组分含量特征因子矩阵和反射率载荷矩阵; Q 为 $d \times m$ 阶组分含量载荷矩阵; P 为 $d \times p$ 阶反射率载荷矩阵; F 和 E 分别为 $n \times m$ 和 $n \times p$ 阶组分含量残差矩阵和反射率残差矩阵^[15]。

PLSR 根据特征向量的相关性分解 Y 和 X ,建立回归模型,即

$$U = T \cdot B + E_d, \tag{3}$$

式中: E_d 为随机误差矩阵; B 为 d 维对角回归系数矩阵。若反射率向量为 x ,则组分含量为 $y = x(U \cdot X)'B \cdot Q$ 。

评价 PLSR 模型性能的基本指标有决策系数(R^2)和均方根误差(root mean square error, RMSE)。

R^2 反映了测量数据与真实值的相关程度,即

$$R^2 = \frac{\sum_{i=1}^n (\hat{y} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \tag{4}$$

式中: $\hat{y}$ 为测量数据; $\bar{y}$ 为测量数据的均值; y_i 为真实值。 R^2 越接近 1,模型的拟合度越高。

RSME 反映了测量数据偏离真实值的程度,即

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}, \tag{5}$$

式中: y_i 为测量数据; $\bar{y}$ 为真实数据。 $RSME$ 越小,模型的拟合效果越好,预测精度越高。

2.2 光谱数据优化

可见光-近红外波段(350 ~ 2 500 μm)主要由含氢基团的倍频和组频吸收峰组成,吸收强度弱,灵敏度相对较低,且吸收带宽有一定叠加现象。光谱数据优化的目的是通过对光谱的适当预处理和变换,减弱乃至消除各种非目标因素对光谱的影响,增强和突出自变量光谱的吸收特征,使算法能够更准确地捕捉到这些特征并与因变量建立起对应关系,从而提高模型的预测能力和稳定性。常见的优化处理方法有标准化、归一化、平滑处理、一阶和二阶导数、标准正态变量(standard normal variable, SNV)变换及多元散射校正(multiplicative scatter correction, MSC)等^[16-17]。其中,对光谱数据求导方法有直接差分法和 Savitzky-Golay 求导法。直接差分法一般用于对分辨率高、波长采样点多的光谱数据进行操作,不适合稀疏波长采样点的光谱数据;而 Savitzky-Golay 卷积平滑方法适用于各种条件下的光谱数据的优化。

SNV 主要用于消除固体颗粒大小、表面散射以及光程变化对可见光 - 近红外反射光谱的影响。SNV 算法与标准化算法的不同之处在于其只针对一条光谱进行处理,其计算公式为

$$X_{i,SNV} = \frac{X_{i,k} - \Delta X_i}{\sqrt{\frac{\sum_{k=1}^m (X_{i,k} - \Delta X_i)^2}{(m-1)}}}, \quad (6)$$

式中: ΔX_i 为第 i 个样品光谱的平均值; $k=1,2,\cdots,m$; m 为波长点数; $i=1,2,\cdots,n$; n 为校正集样品数。

3 PLSR 建模及结果

3.1 数据集获取

对采集的 120 个样品进行反射率光谱测试,测试仪器采用美国 ASD 公司生产的 FieldSpec 光谱仪。先将光谱仪预热 10 ~ 30 min 后再进行测定,并用标准白板进行校正;视场角为 10°,传感器探头垂直向下,波段范围为 350 ~ 2 500 nm,光谱采样间隔为 1 nm。挑选其中具有代表性的 110 件岩石样品进行全岩地球化学分析,获得样品的 SiO_2 , TiO_2 , Fe_2O_3 , MgO , CaO 和 Al_2O_3 等主量矿物数据。

3.2 数据处理

在进行 PLSR 建模之前,需对光谱数据进行一

些预处理,并剔除异常数据。首先利用多项式卷积 (Savitzky - Golay 多项式法) 平滑处理消除测量过程中产生的随机高频噪声。分别选用 3,5,7,9 点平滑处理后的光谱进行建模,通过观测其 $RMSE$,经过反复试验认为 5 点平滑处理对后续建模的效果最好。然后按照同样的原理,分别对一阶导数、二阶导数、MSC 和 SNV 变换进行试验,通过对 R^2 、最佳主成分及 $RMSE$ 等指标的观察,最终确定“卷积平滑 + 一阶导数 + SNV 校正”是适用于本次研究的最佳处理方法。对挑选的 110 个样品预处理后的光谱数据进行投图,观测其波谱形态,剔除与整体趋势反差较大的 10 个样品的异常波谱,将剩余的 100 个样品的光谱数据 (350 ~ 2 500 nm, 共 2 150 个波段) 所构成的 $[2\,150 \times 100]$ 矩阵作为 X ,将所有样品的 6 种主量矿物含量值组成的 $[6 \times 100]$ 矩阵作为 Y ,带入建立的 PLSR 模型 (式 (3)) 中进行运算。

3.3 主成分个数选择

PLSR 主成分的选择需同时考虑主成分个数对 Y 的累计方差贡献率以及模型的 $RMSE$ 。本文采用 k 折 (k -fold) 交叉验证法进行主成分个数的选取,即将数据集随机分为 k 份,每次将其中 1 份作为测试集,剩余的 ($k-1$) 份作为训练集进行训练。通过计算,得到主成分个数与对应 $RMSE$ 及 Y 的累计方差贡献率之间的关系图,如图 2 所示。

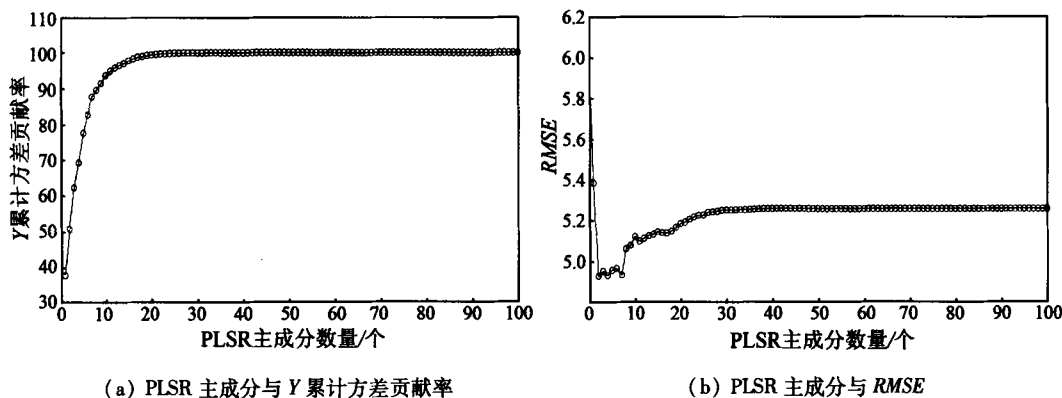


图 2 PLSR 主成分个数与 Y 累计方差贡献率及 $RMSE$ 的关系

Fig. 2 Relationship between number of PLSR components and percent variance explained in Y as well as $RMSE$

从图 2(a)可以看出,当主成分个数大于 6 时, Y 的累计方差贡献率接近 85%。而从图 2(b)可以看出,在主成分个数为 7 时,模型具有最小的 $RMSE$ 。综上所述,选择主成分个数为 7 较为合理。

3.4 结果与讨论

3.4.1 实验结果

通过 R^2 和 $RMSE$ 这 2 个指标对 PLSR 模型的性能进行评价,表 1 给出了 6 种主量矿物对应的 $RMSE$ 和 R^2 。

表 1 PLSR 模型计算 6 种主量矿物的 $RMSE$ 及 R^2

Tab. 1 $RMSE$ and R^2 of six oxides calculated by PLSR model

评价指标	主量矿物						均值
	Al_2O_3	Fe_2O_3	MgO	SiO_2	TiO_2	CaO	
$RMSE$	0.380	0.522	0.441	0.550	0.148	0.495	0.423
R^2	0.905	0.894	0.851	0.829	0.911	0.879	0.878

从表 1 可以看出,所有矿物对应的 $RMSE$ 均小于 1,均值为 0.423,表明该模型计算出的预测值偏离实测值的程度非常小,模型精度很高。

由 PLSR 模型计算出各主量矿物预测值与实际值的比值,如图 3 所示。

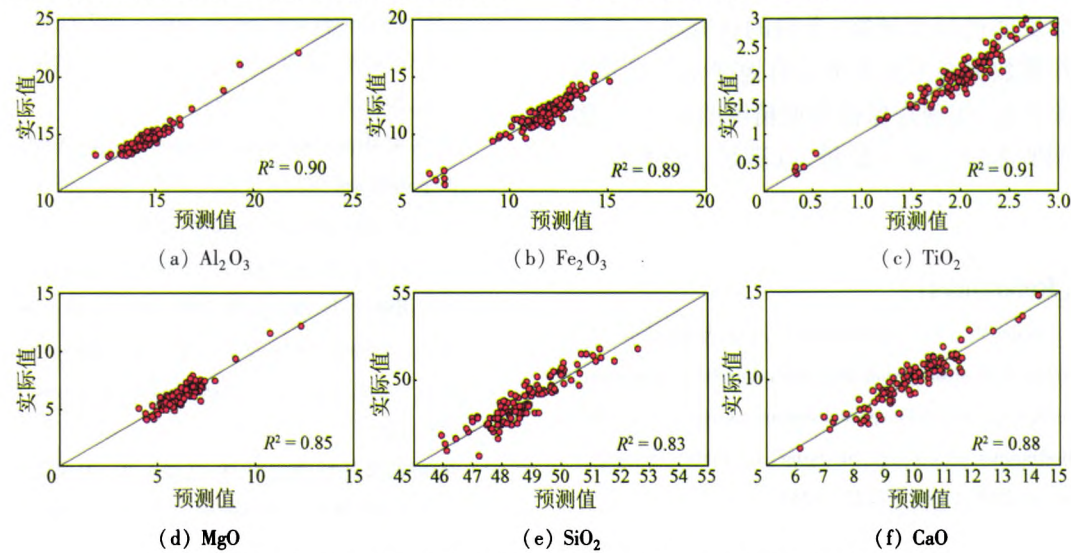


图 3 由 PLSR 模型计算出的 6 种主量矿物预测值与实际值对比

Fig.3 Comparison between prediction values of 6 oxides calculated by PLSR model and observed reflectance

从图 3 可以看出,PLSR 模型的模拟效果整体上较好,各主量矿物的预测值与实际值均沿 $y=x$ 直线分布,其 R^2 均高于 0.80,平均 R^2 达 0.88,其中 R^2 最高的为 TiO_2 ($R^2=0.91$), R^2 最低的为 SiO_2 ($R^2=0.83$),可见基于 PLSR 模型、结合实测光谱数据,可以实现对岩石地球化学成分的高精度反演。除此之外,该模型还给出了各主量矿物对应各波段的载荷系数(图 4),这些系数反映出各主量矿物与各波长位置上反射率值间的相关关系。

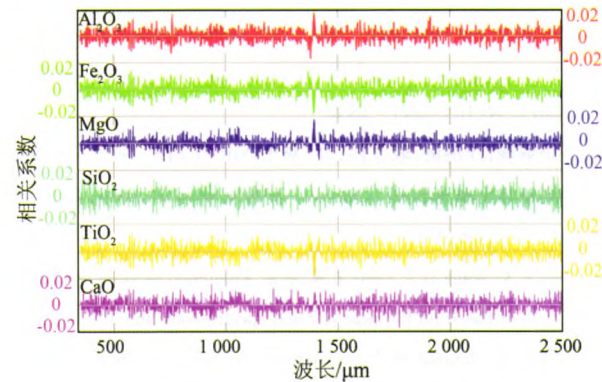


图 4 由 PLSR 模型计算出的 6 种主量矿物对应各波段值的相关系数

Fig.4 Correlation coefficients of 6 oxides calculated by PLSR model to relative bands

3.4.2 讨论

1)根据用 PLSR 获得的各主量矿物的拟合方程,结合该地区同类型岩石光谱数据,可实现对其主量矿物含量的反演。如果针对不同时代、不同地区的玄武岩利用 PLSR 进行建模,由于研究对象在特征上的差异,所得到的结果也会有一定的差异;但主量矿物含量与光谱吸收特征之间关系的趋势是不变的。因此,该方法在研究岩石地球化学成分反演

方面具有一定的推广意义。

2)研究对象也不仅限于对玄武岩主量矿物含量的反演。通过实验证明,PLSR 模型对于花岗岩等岩石也同样适用,且根据该模型获得的载荷矩阵可以进一步精简模型的输入变量,利用最少的波段实现较好的反演效果。利用实测数据保证了光谱数据的质量;而从应用的角度来讲,该模型也适用于高光谱遥感数据,但需要考虑大气等因素造成的影像波谱与实测波谱的差异对实验结果的影响。相信通过对该方法的逐步优化,有望实现利用航空、航天高光谱数据对地面岩石地球化学成分进行区域性填图的目标。

4 结论

本文利用偏最小二乘回归(PLSR)方法,结合实测光谱数据和地球化学分析数据,实现了对甘肃柳园地区二叠纪玄武岩的岩石地球化学成分的反演。得出如下结论:

1)PLSR 可以在实现数据建模的同时消除数据之间的冗余信息,达到降维的目的;该方法具有诸多优势,可用于实现地表岩石地球化学成分的反演,弥补遥感在该方面应用的不足。

2)本文利用地面实测光谱数据、结合岩石地球化学分析数据,以柳园南部玄武岩带为例,实现了对玄武岩中 SiO_2 , TiO_2 , Fe_2O_3 , MgO , CaO 和 Al_2O_3 这 6 种主量矿物信息的提取。

3)实验结果表明,PLSR 模型具有较好的稳定性和精度, R^2 均在 0.8 以上,对利用高光谱数据进行岩石地球化学成分反演具有推广意义。

4) 本文仅利用地面实测光谱数据对玄武岩地球化学成分反演进行了试验, 今后的研究重点是利用高光谱影像数据对不同类型岩石地球化学成分反演开展实验分析, 实现在保证模型精度的前提下, 逐步精简模型所需自变量的数量, 以满足工程化应用的需求。

参考文献 (References):

- [1] Choe E, van der Meer F, van Ruitenbeek F, et al. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain [J]. *Remote Sensing of Environment*, 2008, 112(7): 3222 - 3233.
- [2] Song L, Jian J, Tan D J, et al. Estimate of heavy metals in soil and streams using combined geochemistry and field spectroscopy in Wan - sheng mining area, Chongqing, China [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, 34: 1 - 9.
- [3] Lucey P G, Blewett D T, Jolliff B L. Lunar iron and titanium abundance algorithms based on final processing of Clementine ultraviolet - visible images [J]. *Journal of Geophysical Research*, 2000, 105(E8): 20297 - 20305.
- [4] Le Mouélic S, Lucey P G, Langevin Y, et al. Calculating iron contents of lunar highland materials surrounding Tycho crater from integrated Clementine UV - visible and near - infrared data [J]. *Journal of Geophysical Research*, 2002, 107(E10): 5074.
- [5] Greenhagen B T, Lucey P G, Wyatt M B, et al. Global silicate mineralogy of the Moon from the Diviner lunar radiometer [J]. *Science*, 2010, 329(5998): 1507 - 1509.
- [6] Yan B K, Xiong S Q, Wu Y Z, et al. Mapping lunar global chemical composition from Chang' E - 1 IIM data [J]. *Planetary and Space Science*, 2012, 67(1): 119 - 129.
- [7] 闫柏琨, 刘圣伟, 王润生, 等. 热红外遥感定量反演地表岩石的 SiO_2 含量 [J]. *地质通报*, 2006, 25(5): 639 - 643.
Yan B K, Liu S W, Wang R S, et al. Quantitative inversion of the SiO_2 content in surface rocks using thermal infrared remote sensing [J]. *Geological Bulletin of China*, 2006, 25(5): 639 - 643.
- [8] 王俊虎, 张杰林, 张静波. 基于 ASTER 热红外数据的 SiO_2 含量反演在某矿田信息提取中的应用 [J]. *铀矿地质*, 2010, 26(5): 306 - 312.
Wang J H, Zhang J L, Zhang J B. SiO_2 content retrieving based on ASTER thermal data and its application in substrating metallogenic factor of a uranium deposit [J]. *Uranium Geology*, 2010, 26(5): 306 - 312.
- [9] 杨杭, 张立福, 黄照强, 等. 基于热红外发射光谱的岩石 SiO_2 定量反演模型研究 [J]. *光谱学与光谱分析*, 2012, 32(6): 1611 - 1615.
Yang H, Zhang L F, Huang Z Q, et al. Quantitative inversion of rock SiO_2 content based on thermal infrared emissivity spectrum [J]. *Spectroscopy and Spectral Analysis*, 2012, 32(6): 1611 - 1615.
- [10] 莫宜学. 岩浆与岩浆岩: 地球深部“探针”与演化记录 [J]. *自然杂志*, 2011, 33(5): 255 - 259.
Mo X X. Magma and magmatic/igneous rocks: A lithoprobe into the deep earth and records of the Earth's evolution [J]. *Chinese Journal of Nature*, 2011, 33(5): 255 - 259.
- [11] 莫宜学, 赵志丹, 邓晋福, 等. 印度 - 亚洲大陆主碰撞过程的火山作用响应 [J]. *地学前缘*, 2003, 10(3): 135 - 148.
Mo X X, Zhao Z D, Deng J F, et al. Response of volcanism to the India - Asia collision [J]. *Earth Science Frontiers*, 2003, 10(3): 135 - 148.
- [12] 赵泽辉, 郭召杰, 韩宝福, 等. 新疆东部 - 甘肃北山地区二叠纪玄武岩对比研究及其构造意义 [J]. *岩石学报*, 2006, 22(5): 1279 - 1293.
Zhao Z H, Guo Z J, Han B F, et al. Comparative study on Permian basalts from eastern Xinjiang - Beishan area of Gansu Province and its tectonic implications [J]. *Acta Petrologica Sinica*, 2006, 22(5): 1279 - 1293.
- [13] Wold S, Albano C, Dunn III W J, et al. Pattern recognition: Finding and using regularities in multivariate data [C]//*Proceedings of Food Research and Data Analysis*. London, UK: Applied Science Pub, 1983: 146 - 189.
- [14] 陈洁梅, 肖青青, 潘涛, 等. NIR 光谱结合稳定等效 MW - PLS 方法用于高脂血症指标分析 [J]. *光谱学与光谱分析*, 2014, 34(10): 2827 - 2832.
Chen J M, Xiao Q Q, Pan T, et al. NIR spectroscopy combined with stability and equivalence MW - PLS method applied to analysis of hyperlipidemia indexes [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(10): 2827 - 2832.
- [15] 李耀翔, 汪洪涛, 耿志伟, 等. 基于 NIR 及 PLS - PCR - SVR 预测森林土壤有机碳含量 [J]. *安徽农业科学*, 2014, 42(15): 4702 - 4706, 4742.
Li Y X, Wang H T, Geng Z W, et al. Prediction of forest soil carbon content based on the near infrared spectroscopy and PLS - PCR - SVR [J]. *Journal of Anhui Agricultural Sciences*, 2014, 42(15): 4702 - 4706, 4742.
- [16] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用 [J]. *化学进展*, 2004, 16(4): 528 - 542.
Chu X L, Yuan H F, Lu W Z. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique [J]. *Progress in Chemistry*, 2004, 16(4): 528 - 542.
- [17] 黎文兵, 药林桃, 刘木华, 等. 光谱预处理对 LIBS 检测脐橙中 Cu 的偏最小二乘定量模型影响 [J]. *光谱学与光谱分析*, 2015, 35(5): 1392 - 1397.
Li W B, Yao L T, Liu M H, et al. Influence of spectral pre - processing on PLS quantitative model of detecting Cu in navel orange by LIBS [J]. *Spectroscopy and Spectral Analysis*, 2015, 35(5): 1392 - 1397.

**Inversion of geochemical compositions of basalts based
on field measured spectra**

YU Junchuan¹, LIU Wenliang², YAN Bokun¹, DONG Xinfeng¹, WANG Zhe¹, LI Na¹

(1. *China Aero Geophysical Survey and Remote Sensing Center for Land and Resources, Beijing 100083, China*; 2. *State Key Laboratory of Geological Processes and Mineral Resources, School of Earth Sciences and Mineral Resources, China University of Geosciences(Beijing), Beijing 100083, China*)

Abstract: Geochemical compositions have significant implications for rock classification, identification of the petrogenesis and evolution of the rocks. The utilization of remote sensing method to estimate the geochemical compositions of the rocks is a new subject, and is also a difficult point in remote sensing related researches due to its relatively immature applications. In this study, the Permian basalts were chosen as the study object. Based on systematical sampling, spectral analysis and geochemical test, the authors constructed a mathematical model between field measured spectra data (2 150 bands) and available data of six representative major elements by using partial least squares regression (PLSR). It is essential to initially choose proper preprocessing method to optimize the spectra data, and then search for the optimal number of principal components with minimum root – mean – square error through *k* – fold cross – validation. The results show that the PLSR model yields higher stability and precision, and plays a significant role in applications of geochemical composition inversion using remote sensing data.

Keywords: basalt; geochemical compositions; inversion; field measured spectra; partial least squares regression (PLSR); Liuyuan

(责任编辑: 刘心季)