

LIBS 与反向传播算法结合的橄榄石成分分析

袁汝俊^{1, 2, 3}, 万雄^{1, 2*}, 何强^{1, 2, 3}, 王泓鹏^{1, 2}

1. 中国科学院上海技术物理研究所, 上海 200083
2. 中国科学院空间主动光电技术重点实验室(上海技术物理研究所), 上海 200083
3. 中国科学院大学, 北京 100049

摘要 LIBS 是一项用于分析物质成分的有力手段, 但是定量分析时存在结果不准确、重复性低等缺点。为了准确预测自然界中橄榄石的成分信息, 通过按照自然界中的橄榄石的成分信息制作 15 组橄榄石样品, 将其中的 11 组作为标准样品, 另外 4 组作为测试样品进行 LIBS 定量分析。每个样本采 50 条 LIBS 光谱建立橄榄石的 LIBS 数据库。然后采用多元线性回归算法和反向传播算法对样本的 50 组数据进行分析, 有效的降低了由于随机误差造成的测试结果的不准确。最终结果表明, 使用激光诱导击穿光谱和反向传播算法对橄榄石中镁橄榄石与铁橄榄石含量进行检测, 预测结果的决定系数为 0.901, 接近常规的多元线性回归算法得到的 0.911, 这说明反向传播算法对橄榄石含量的预测精度接近多元线性回归算法。同时使用反向传播算法得到的结果的均方根误差为 28.64, 优于后者的 29.23, 说明使用反向传播算法得到的结果分布更加集中。此外, 通过分析关联矩阵中各数值的大小与各元素谱线的位置的对应关系, 表明使用反向传播算法反演出来的关联矩阵 F 与之所代表的物理含义的相关性更高。说明反向传播运算不仅与传统的多元线性回归算法性能相当, 而且在预测数据的一致性上表现得更好。此外使用反向传播算法可以直接对激光诱导击穿光谱得到的橄榄石全谱数据进行数据反演, 而不需要经过光谱寻峰这一步骤, 简化了数据分析流程, 弥补了多元线性回归算法难以分析全谱数据的不足。

关键词 激光诱导击穿光谱; 定量分析; 反向传播; 橄榄石

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)12-3861-07

引言

激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)技术是一种利用高能脉冲激光激发的等离子体光谱来检测材料所含物质信息的技术^[1]。产生等离子体光谱的原理是高能脉冲激光轰击靶材表面, 靶材表面产生的高达 10 000 K 的局部高温使物质样本被解离成等离子体状态, 随着处于高能状态的等离子体逐渐冷却, 离子中的电子会向低能级跃迁, 在这一过程中会发射出与元素种类相关的特定波长的光子。通过观察等离子体光谱中谱峰的波长位置和强度就可以分析靶材中所含物质的成分及含量信息^[1-2]。

由于 LIBS 具有高效、无损、低检测限和可远距离探测等优势, 在成分分析方面具备很好的实用性与应用前景^[3-5]。具体应用场所包括冶金、矿产资源勘测、水体检测和地质分

析等领域。目前 LIBS 技术在利用传统的多元线性回归(multiple linear regression, MLR)算法^[6-7]对物质成分进行定量分析时需要人工参与手动查找谱峰位置等准备过程。另外当样品条件、激光能量变化时, 与实际物理含义无关的 MLR 方法对目标的成分预测值会比实际值有较大的误差。本工作利用反向传播(back-propagation, BP)算法^[8]预测待检样品中的橄榄石成分比传统的多元线性回归算法的精度稍有改善, 改善了物理含义匹配度不足的缺点。此外还可以对全谱数据进行直接分析, 免除人工寻找谱峰位置的过程。因此结合反向传播算法的方式是一种可以改善激光诱导击穿光谱对物质成分进行准确分析的有效方法。

橄榄石是一种镁与铁的硅酸盐, 化学式为 $(Mg, Fe)_2SiO_4$ 。在地质成分分类中, 橄榄石属于火成岩中的一个重要的组成部分。橄榄石的化学组成通常由两部分组成, 一部分是镁橄榄石 (Mg_2SiO_4) , 简称 Fo, 另一部分是铁橄榄石

收稿日期: 2018-10-19, 修订日期: 2019-02-08

基金项目: 民用航天技术预先研究项目(D030207), 载人航天领域预先研究项目(0209), 上海技术物理研究所创新专项(CX-110)资助

作者简介: 袁汝俊, 1993 年生, 中国科学院上海技术物理研究所博士研究生 e-mail: yrujun@163.com

* 通讯联系人 e-mail: wanxiong@mail.sitp.ac.cn

(Fe_2SiO_4 , 简称 Fa)。地质中形成的橄榄石就是这两种端元组分形成的固溶体。橄榄石的成分可由 Fo 和 Fa 这两种端元组分的摩尔百分比表示, 例如 Fo70 Fa30 表示镁橄榄石端元组分占 70% 的橄榄石。不同的橄榄石成分反映了地质形成过程中不同的物理条件, 例如: 不同的结晶温度下形成的橄榄石中 Fo 和 Fa 分子数的比例关系如图 1 所示。通过分析岩石中橄榄石的成分信息, 结合岩石采集地点的地理位置和地质状况, 可以大致分析出地质运动时的成矿条件以及成矿之后的岩石演化和侵蚀作用。因此使用 LIBS 定量分析橄榄石成分可以实现研究地质变化及地球演变的目的。由于 LIBS 可在极端条件下进行远程探测, 因此在航天领域中对行星表面矿物的物质成分分析有十分积极的作用。本工作就包含对火成岩中橄榄石成分的定量分析就是一个实例。

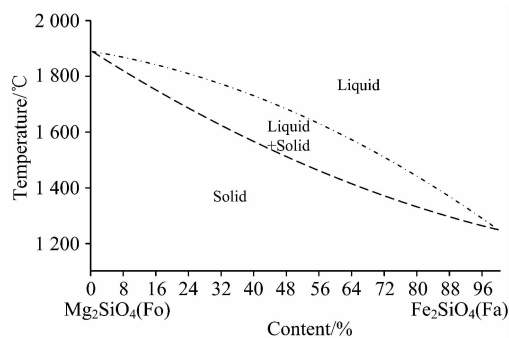


图 1 橄榄石成分与结晶温度的关系

Fig. 1 Relationship between olivine compositions and crystallization temperature

1 实验部分

1.1 靶材制备

为分析出自然界中的各种橄榄石, 以组分比 Fo : Fa 从 0 : 10 至 10 : 0 制作标准橄榄石样品。测试样品选用另一组自制样品及自然界中存在的橄榄石样品。

由于自然界中的火成岩成分都是由物质的氧化物反应形成的, 因此可以按照氧化物混合的配料方式来模拟镁橄榄石和铁橄榄石的化学组成。方案中采用 MgO 和 SiO_2 的 2 : 1 混合的方式模拟橄榄石中的镁橄榄石端元矿物 (Mg_2SiO_4 , Fo) 的化学组成, 采用 Fe_2O_3 和 SiO_2 的 1 : 1 混合的方式模拟橄榄石中的铁橄榄石端元矿物 (Fe_2SiO_4 , Fa) 的化学组成。最终的镁橄榄石端元矿物的化学式和配比方案中的组成成分 ($2(\text{MgO}) + \text{SiO}_2$) 的氧分子数完全相同。最终的铁橄榄石端元矿物的化学式和配比方案中的组成成分 ($\text{Fe}_2\text{O}_3 + \text{SiO}_2$) 的氧分子数稍有不同, 但由于空气中有氧元素, 因此氧元素不用作成分分析, 不会影响实验的结果, 这种配比方法是可行的。最终制定的物质配比方案如表 1、表 2 所示。实验中所用的氧化物粉末均购自阿拉丁化学试剂公司, 所有物质的纯净度等于或优于色谱级 (SP)。

以上配制好的样品粉末经充分混合均匀后, 再经 20 MPa 压力压制成约 10g 的靶材。用第一组样品的测试数据作为标准数据, 第二组样品的数据则用作最终的测试数据。图

2 为混合完成的粉末以及最终压制成的实验样品。

表 1 第一组训练样品

Table 1 First group of samples for training

Numbering	Content /%		Proportional coefficient of moles content		
	Mg_2SiO_4 (Fo)	Fe_2SiO_4 (Fa)	MgO	Fe_2O_3	SiO_2
1_1	0	100	0	100	100
1_2	10	90	20	90	100
1_3	20	80	40	80	100
1_4	30	70	60	70	100
1_5	40	60	80	60	100
1_6	50	50	100	50	100
1_7	60	40	120	40	100
1_8	70	30	140	30	100
1_9	80	20	160	20	100
1_10	90	10	180	10	100
1_11	100	0	200	0	100

表 2 第二组测试样品

Table 2 Second group of samples for testing

Numbering	Content /%		Proportional coefficient of moles content		
	Mg_2SiO_4 (Fo)	Fe_2SiO_4 (Fa)	MgO	Fe_2O_3	SiO_2
2_1	25	75	50	75	100
2_2	55	45	110	45	100
2_3	75	25	150	25	100
2_4 (Natural sample)	90.4	9.6	180.8	19.2	100

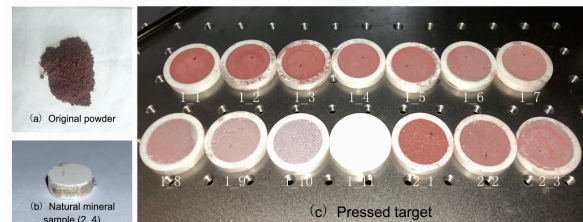


图 2 混合完成的粉末以及最终制成的实验样品

Fig. 2 Mixed powder and final experimental sample

1.2 系统装置

激光诱导击穿光谱实验系统如图 3 所示。

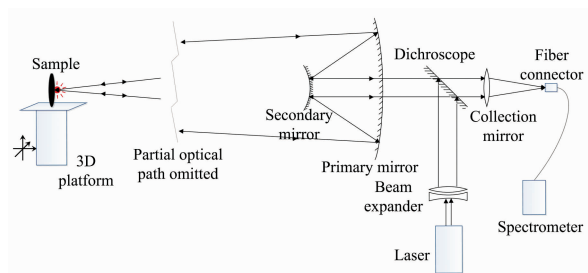


图 3 激光诱导击穿光谱实验系统结构

Fig. 3 Structure of LIBS experimental system

1 064 nm 波长的输出激光，到达样品表面的激光脉冲能量为 $18.2 \text{ mJ} \cdot \text{pulse}^{-1}$ ，脉冲频率使用 DG-645 脉冲延时发生器调节为 5 Hz。激光从激光器中输出后，首先经过 4 倍的扩束镜进行扩束以减小激光的能量密度，然后经由一面镀有 1 064 nm 反射膜的 45° 放置的二向色镜将激光引入一组由主次镜组成的施密特镜。激光经由施密特镜反射后聚焦至 1.2 m 处靶材的位置。在靶材表面产生等离子体 LIBS 的光谱经过同光路返回至二向色镜，300~1 000 nm 的有效信号透过

二向色镜进入光纤输出的收集镜中。最终的光谱信号由光纤引入 Avantes 2048USB 型光纤光谱仪中。该型光谱仪的信噪比为 200 : 1，波长范围为 200~1 100 nm，分辨率为 0.04~20 nm。

由 DG546 控制曝光的曝光时间为 1.05 ms 使整个等离子体发光的过程中产生的光子都能有效的收集。最终的数据由 PC 机读出。

试验参数如表 3 所示。

表 3 激光诱导击穿光谱试验参数
Table 3 Parameters of LIBS experiment

Experiment parameters	Laser Power / ($\text{mJ} \cdot \text{pulse}^{-1}$)	Exposure time/ms	Delay time/ μs	Samples number/times	Detecting distance/m	Spot diameter (on target)/mm
	18.2	1.05	0	50	1.2	0.5

在上述条件下一共测试了 15 个样品，每个样品包含 50 条光谱信息，共 750 条原始光谱数据。将 1_1 组至 1_11 组的 550 条光谱作为训练数据，2_1 组至 2_4 组的 200 条光谱作为测试数据。

1.3 数据预处理及分析

使用 LIBS 数据分析橄榄石中的 Fo 与 Fa 的含量的方法可分为数据准备、回归分析两个过程，详细流程见图 4。为保证回归分析中数据的统一性和有效性，首先需要进行数据预处理^[9]。预处理之后，以往需要利用寻峰算法找出原始数据中的谱峰的数据。对谱峰数据可以采用多元线性回归算法和反向传播算法，分别预测测试样品中的橄榄石成分信息，具体的分析方法包括单变量分析^[10]和多变量^[10-12]。本工作不利用寻峰算法，创新性地提出采用反向传播算法直接对全谱段数据进行迭代运算以得到最终橄榄石中 Fo 含量的预测结果。

析算法的准确性，原始的光谱数据需要经过预处理才可进行下一步的数据分析过程。预处理的流程示意图如图 5 所示。

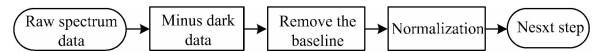


图 5 数据预处理流程图

Fig. 5 Flow chart of data preprocessing

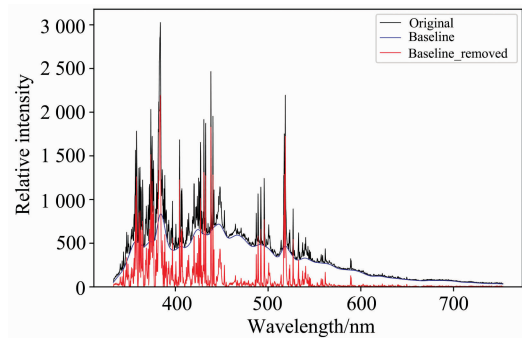


图 6 去基线前后光谱对比图 (第 1_6 组)

Fig. 6 Comparison between baseline removed and original specura

根据等离子体光谱的光谱发射特性可知：在等离子体发射光谱的过程中，存在两个主要过程^[12]。因为实验中采用的 1.05 ms 曝光时间远大于连续光谱背景和 LIBS 光谱的持续时间。因此可以使用非对称最小二乘算法去除光谱信号中的连续光谱背景基线(Baseline)，保留有效的 LIBS 光谱信号。经去除连续光谱背景基线后的 LIBS 光谱见图 6。此外对去除连续光谱背景基线后的数据进行规范化(normalization)处理，以确保信号的一致性、有效性，有利于提高最终分析精度。

1.3.1 寻找特征峰

经过预处理之后的光谱为了方便数据反演，通常会通过与标准数据库对比选取各元素对应特征峰进行手动的特征峰提取，然后再利用回归算法进行分析。

在此采用计算机自动选取光谱中的公共谱峰来做回归分析。寻峰条件是：谱峰间距大于 1 nm，光谱强度大于光谱最

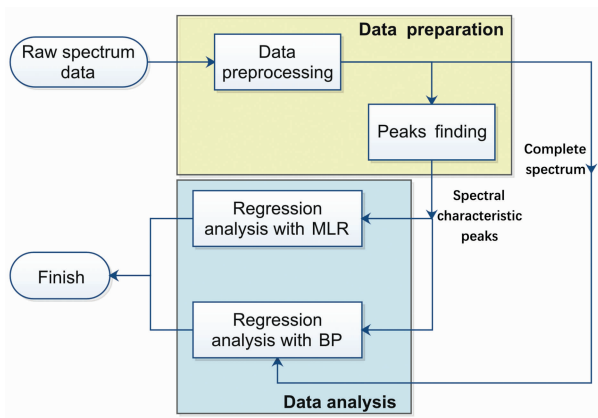


图 4 数据分析流程图

Fig. 4 Flow chart of data analysis

同时也对比分析了多元线性回归算法和反向传播算法的优缺点；对比分析了使用两种不同数据即基于特征峰的数据和使用全谱段的数据进行反演分析的结果差异。

实验直接获取的 LIBS 信号会受到 CCD 暗电流、连续光谱背景和激光脉冲能量变化等因素的干扰。为了提高定量分

高强度值的 1/10。最终每条光谱信息选取出 28 条公共谱峰，的谱峰的所述元素的信息见表 4。利用这组特征峰数据就可以进行下一步的分析工作。寻找到

表 4 峰位及元素信息对照表
Table 4 Contrast table between spectral peaks and elements

Elements	Mg	Fe	Si	O	Others
Correlation with c(Fo)	Positive	Negative	None	None	Unknown
wavelength/nm	517.412, 518.505	361.84, 372.7, 376.41, 406.5, 425.11, 426.23, 432.69, 438.3, 489.16, 492.19, 526.96	375.84, 407.06, 430.73, 440.54, 495.77	385.80, 374.98, 385.80	356.98, 360.7, 373.56, 379.83, 380.68, 388.64, 396.87, 404.52

1.3.2 裁剪全谱数据

由于反射、透射镜片的衰减作用，仪器最终响应的波长范围在 333~600 nm 之间，因此全谱段的信息由最原始的每条光谱 2 048 点剪切至 964 点。最终全谱数据的数据格式统一为每条光谱 964 点。

1.4 数据分析

\mathbf{X} 矩阵为由测试集中各组物质的光谱经数据预处理得到的光谱信息矩阵， \mathbf{Y} 为对应的物质含量信息矩阵， \mathbf{F} 为它们 \mathbf{X} 与 \mathbf{Y} 之间的关系矩阵。关联矩阵 \mathbf{F} 的物理含义：光谱图中波长位置的强度与物质含量之间的相关关系。利用式 (1) 就可以描述它们之间的基本关系

$$\mathbf{XF} = \mathbf{Y} \quad (1)$$

(1) 只是分析光谱的特征峰时，由于选用光谱中的 28 个特征峰作为 \mathbf{X} 矩阵(形状为 550×28)，对应的物质信息只有 $c(\text{Fo})\%$ 这一个未知量，所以 \mathbf{Y} 矩阵的形状为 550×1 。因为需要求解的矩阵是关联矩阵，因此需要包含 28 个系数项和 1 个常数项，所以最终需要求解的关联矩阵就是一个形状为 $(28+1) \times 1$ 的向量。

(2) 只是分析全谱数据时，将完整光谱作为 \mathbf{X} 矩阵(形状为 550×964)，对应的物质信息只有 $c(\text{Mg})\% : c(\text{Fe})\%$ 这一个未知量，所以 \mathbf{Y} 矩阵的形状依旧是 550×1 。所以我们需要求解的关联矩阵就是一个形状为 $(964+1) \times 1$ 的向量。

利用上述矩阵关系以及用于训练的 550 条光谱数据，结合适当的优化算法就可以求解出关系矩阵 \mathbf{F} 。最后就可以利用上述公式及关系矩阵对测试集的 200 条光谱数据进行定量反演。

1.4.1 多元线性回归算法

多元线性回归是利用矩阵运算求解最小二乘解的一种方法^[6]。使用多元线性回归算法是利用最小二乘算法最小化误差寻找最优解的方法，可以用来求解两个矩阵(\mathbf{X} 和 \mathbf{Y})的关系矩阵 \mathbf{F} 的近似解。在使用 MLR 进行光谱分析时，就是在求解光谱强度矩阵 \mathbf{X} 与物质含量矩阵 \mathbf{Y} 之间的关联矩阵 \mathbf{F} ，即

$$\mathbf{F} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2)$$

式(2)中， \mathbf{X}^T 为 \mathbf{X} 的转置矩阵， $(\mathbf{X}^T \mathbf{X})^{-1}$ 为 $\mathbf{X}^T \mathbf{X}$ 的逆矩阵。

利用求解到的关联矩阵 \mathbf{F} 对测试集进行数据分析：结合已知测试样品的光谱信息矩阵 \mathbf{X} ，根据式(1)就可以计算出

该光谱对应的物质含量信息矩阵 \mathbf{Y} 。

只是分析光谱的特征峰时，求解的关联矩阵 \mathbf{F}_{mlr} 是一个形状为 $(28+1) \times 1$ 的向量。但是在分析全谱数据时，因为此时的未知数(965 个)大于建立方程的个数(550)，因此利用多元线性回归的方法是无法求解出全谱数据的关系矩阵 \mathbf{F}_{mlr} 。

1.4.2 反向传播算法

反向传播算法是一种利用迭代方式逐渐优化参数，以获得近似最优解的优化算法。在上式中求解关系矩阵 \mathbf{F}_{bp} 的过程中就可以使用反向传播算法来优化最终的测试结果。

使用反向传播算法计算关联矩阵 \mathbf{F} 的方法包含以下三个基本过程：

(1) 随机选取所有数据中的一部分，计算出在当前参数 \mathbf{F}_{tmp} 下使用前向传播的方式得出的预测值 \mathbf{Y}_{pre} 。

(2) 利用 L2 损失函数，计算预测值 \mathbf{Y}_{pre} 与实际值 \mathbf{Y}_{rel} 的误差值 \mathbf{I}_{tmp} 。

(3) 使用反向传播算法在所得的损失值 \mathbf{I}_{tmp} 的条件下重新调整参数矩阵 \mathbf{F}_{tmp} 中各元素的值，并回到步骤(1)。

不断地重复以上过程就可以实现优化矩阵 \mathbf{F}_{tmp} 的目的。在此条件下得到的预测值 \mathbf{Y}_{pre} 趋近于实际值 \mathbf{Y}_{rel} ，此时得到的关联矩阵 \mathbf{F}_{tmp} 即为关联矩阵的最优解 \mathbf{F}_{bp} 矩阵。

1.5 预测待测样品

通过以上两种方法分别求解出的关联矩阵 \mathbf{F}_{mlr} 和 \mathbf{F}_{bp} ，再结合方程就可以计算出测试集数据对应的橄榄石成分。此时测试集是 4 种不同的含量的橄榄石样品，每个样品测试出 50 条 LIBS 光谱，通过使用矩阵运算的方法就可以得出最终的含量信息。由于每条光谱都会有一定的差异，因此将最终的 50 个预测值进行平均作为预测值。

2 结果与讨论

按照上述方式获得的相应橄榄石含量的预测结果见表 5。

表 5 中 R^2 为决定系数(coefficient of determination)，表明预测结果与真实值之间的相关关系，以此来判断模型的解释力。RMSE 为均方根误差(root-mean-square error)，表明数据的偏差程度，以此来判断数据的集中程度。

2.1 使用 MLR 和 BP 算法分别对谱峰数据进行分析

2.1.1 使用单一谱峰数据进行分析

最终的预测结果可以说明在使用单一 Mg 元素的谱峰作为橄榄石成分分析的条件下, 采用 MLR 和 BP 算法反演求

得的橄榄石成分分析结果是十分接近, R^2 值分别为 0.821 和 0.820, RMSE 分别为 28.76 和 28.73。说明二者在性能上没有表现出明显差别。具体预测结果见图 7 和表 5。

表 5 橄榄石成分信息的预测结果

Table 5 Prediction results of olivine composition information

Numbering	Analysis data	Method	2_1	2_2	2_3	2_4	R^2	RMSE
Real Fo content/%			25	55	75	90.4		
	517.41 nm peak	MLR	25.72	65.8	66.34	99.5	0.821	28.76
		BP	28.19	65.96	65.08	97.4	0.820	28.73
Predicted Fo content/%	28 spectral characteristic peaks	MLR	25.15	53.83	63.25	92.63	0.911	29.23
		BP	22.05	53.12	61.58	98.72	0.887	29.53
	Complete spectrum	MLR	Failed(The predicted value is far beyond the normal range)					
		BP	24.03	52.10	62.27	92.65	0.901	28.64

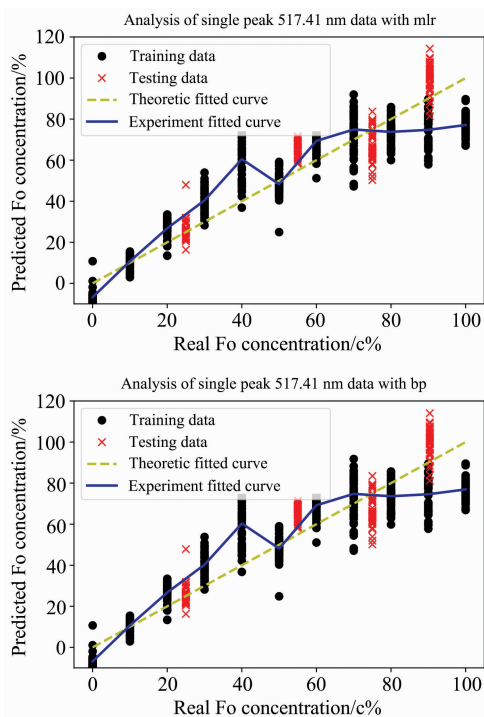


图 7 使用 517.41 nm 谱峰对 Fo 含量进行数据反演的结果
Fig. 7 Calibration and concentration prediction of Fo with the peak at 517.41 nm

2.1.2 使用多条谱峰数据进行分析

图 8 和表 5 可知使用橄榄石中能查找到的 28 条谱峰分析橄榄石成分, 采用 MLR 和 BP 算法反演求得的橄榄石成分分析结果为: R^2 值分别为 0.911 和 0.888, RMSE 分别为 29.23 和 29.53。说明 MLR 算法的精度稍好, 但是数据的集中性与 BP 算法相当。

此外观察使用 BP 算法得到的谱峰的关联矩阵(决定系数)(图 9)与使用 MLR 方法得到的结果相比, 可以明显发现关联矩阵 F 在实线框内谱峰的绝对值后者大于前者。且该谱峰与 Mg 元素(Fo 含量)呈正相关趋势; 同理, 在虚线框中呈现相同的特性。这说明使用 BP 算法相比于 MLR 方法得到的结果更加符合关联矩阵的物理含义(光谱中 Mg 和 Fe 元素

谱峰的强度与橄榄石中 Fo 含量之间的正、负相关关系)。

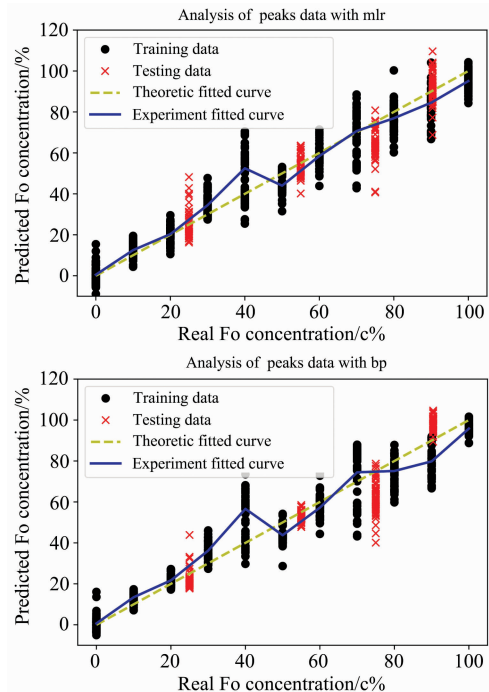


图 8 使用所有 28 条谱峰对 Fo 含量进行数据反演的结果
Fig. 8 Calibration and concentration prediction of Fo with all 28 peaks

2.2 使用 BP 算法对全谱数据进行分析

因为全谱数据中的未知量太多, 使用 MLR 方法无法求解出全谱信息对应的橄榄石中 Fo 含量的信息。但是采用反向传播算法则可以得出较为理想的反演结果。故对于全谱数据只采用 BP 方法进行对关联矩阵 F 的求解。

由图 10 和表 5 可知使用全谱数据分析橄榄石成分, 采用 BP 算法反演求得的橄榄石成分分析结果与图 8 中采用 MLR 算法对 28 个谱峰数据进行分析得到的结果是十分接近, R^2 值分别为 0.911 和 0.901, RMSE 分别为 29.23 和 28.64。这说明二者预测精度相当, 但是对全谱数据分析的结果更加集中, 可信度更高。

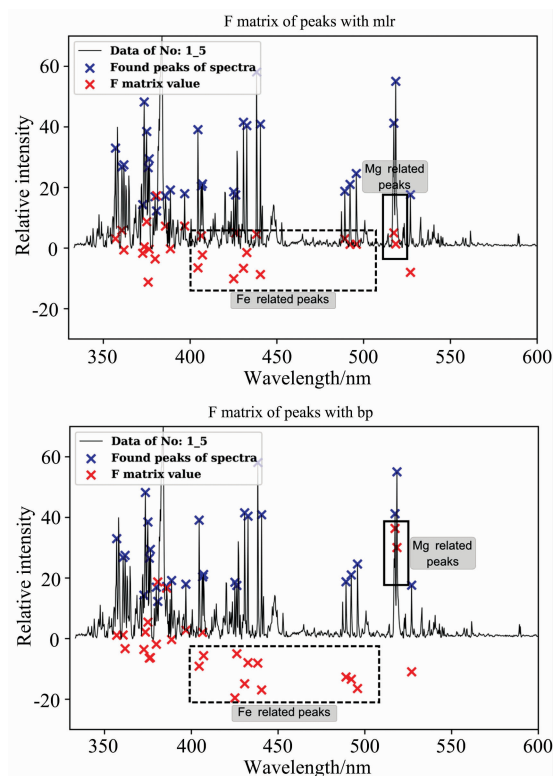


图 9 使用所有 28 条谱峰对 Fo 含量进行数据反演得到的关联矩阵 F

Fig. 9 Correlation matrix F obtained by data inversion of Fo with all 28 peaks

此外观察使用 BP 算法得到谱峰的关联矩阵(决定系数)(图 9、图 10)与使用 MLR 方法得到的结果相比,可以明显发现关联矩阵 F 在实线框内谱峰的绝对值后者大于前者。且该谱峰与 Mg 元素(Fo 含量)呈正相关趋势;同理,在虚线框中呈现相同的特性。

这说明使用 BP 算法得到的结果更加符合关联矩阵的物理含义(即光谱中 Mg 和 Fe 元素谱峰的强度与橄榄石中 Fo 含量之间的正、负相关关系)。相比之下,使用 MLR 方法得到的关联矩阵 F 并没有与物理含义相对应的表现。从这方面

讨论,使用 BP 算法得到的 Fo 含量的预测结果更具可信度。

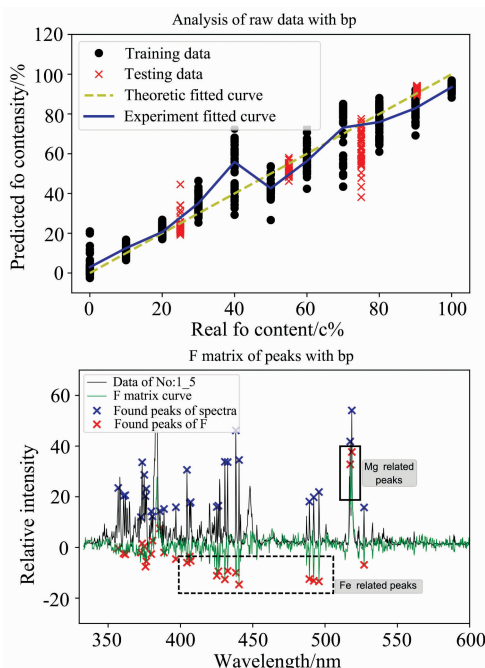


图 10 使用全谱数据对 Fo 含量进行数据反演得到的预测结果和关联矩阵 F

Fig. 10 Prediction results and correlation matrix F obtained by data inversion of Fo with full-spectrum data

3 结 论

在使用 MLR 和 BP 算法对橄榄石中 Fo 的含量进行数据反演分析时, BP 算法通过全谱数据对物质成分预测出的结果更加集中。此外使用 BP 算法对谱峰数据进行数据反演结果比 MLR 算法更符合实际的各元素谱峰的强度与橄榄石中 Fo 含量之间的相关关系,因此预测结果在理论上分析更加完备。这说明使用 BP 算法可以免除在特定条件下对 LIBS 寻峰的要求,简化了数据分析的过程,分析结果精度与传统的 MLR 方法相当。因此在特定条件下,可以采用 BP 算法替代传统的 MLR 方法进行数据反演。

References

- [1] Musazzi S, Perini U. Laser-Induced Breakdown Spectroscopy. Springer Berlin Heidelberg, 2014.
- [2] Hahn D W, Omenetto N. Applied Spectroscopy, 2012, 66(4): 347.
- [3] Fortes F J, Moros J, Lucena P, et al. Analytical Chemistry, 2012, 85(2): 640.
- [4] Luna A S, Gonzaga F B, da Rocha W F C, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 139: 20.
- [5] Li X, Zhou W, Li K, et al. Optics Communications, 2012, 285(1): 54.
- [6] Su Y, Gao X, Li X, et al. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(6): 1560.
- [7] Dyar M D, Carmosino M L, Breves E A, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2012, 70: 51.
- [8] Zhu G, Zhang P, Tshukudu T, et al. Atmospheric Pollution Research, 2015, 6(6): 1087.
- [9] Bowles M. Machine Learning in Python: Essential Techniques for Predictive Analysis. John Wiley & Sons, 2015.
- [10] Motto-Ros V, Syvilay D, Bassel L, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 140: 54.
- [11] Ammari F, Bassel L, Ferrier C, et al. Talanta, 2016, 159: 287.
- [12] Radziemski, Leon J. Handbook of Laser-Induced Breakdown Spectroscopy. 2013, 44: 223.

Research on Olivine Component Analysis Using LIBS Combined with Back-Propagation Algorithm

YUAN Ru-jun^{1, 2, 3}, WAN Xiong^{1, 2*}, HE Qiang^{1, 2, 3}, WANG Hong-peng^{1, 2}

1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China

2. Key Laboratory of Space Active Opto-Electronics Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China

3. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Laser-induced breakdown spectroscopy analysis technology is a powerful method for analyzing material composition, but the use of it for quantitative analysis has the disadvantages of inaccurate analysis results and low reproducibility. In order to predict the composition information of olivine in nature accurately, this paper made 15 samples according to the composition information of olivine in nature, and 11 of them were used as standard samples, and the other 4 samples were used as test samples for LIBS quantitative analysis. Lastly, the laser-induced breakdown spectroscopy database of olivine was established with 50 spectra per sample. Then, the multiple linear regression algorithm and the back-propagation algorithm were used to analyze the 50 sets of data of this series of samples, which effectively reduced the inaccuracy of the test results caused by random errors. In the study of the content of forsterite and fayalite in olivine using back-propagation algorithm with data of laser-induced breakdown spectroscopy, the final results showed that the coefficient of determination of the prediction result was 0.901, close to the 0.911 which was yielded with conventional multiple linear regression algorithm. This indicated that the backward propagation algorithm's prediction accuracy for olivine content was close to the multiple linear regression algorithm. Furthermore, the root mean square error of the result obtained with back-propagation algorithm was 28.64, which was better than the latter's 29.23, which indicated that the result distribution obtained by the back-propagation algorithm was more concentrated. In addition, by analyzing the correspondence between the size of each numerical value in the correlation matrix and the position of each element's spectral line, the results showed that the correlation matrix F inverted with back-propagation algorithm had a higher correlation with the physical meaning represented by it. This showed that the performance of the back-propagation operation was comparable to that of the traditional multiple linear regression algorithm, and it performed better in predicting the consistency of the data. In addition, the back-propagation algorithm could directly analyze the data of the olivine full spectrum data obtained by laser induced breakdown spectroscopy without the step of spectral peak finding, simplifying the process of data analysis and making up the shortcomings of the multiple linear regression algorithm in analyzing full-spectrum data.

Keywords Laser-induced breakdown spectroscopy; Quantitative analysis; Back-propagation; Olivine

(Received Oct. 19, 2018; accepted Feb. 8, 2019)

* Corresponding author