

水体重金属激光诱导击穿光谱定量分析方法对比研究*

王春龙 刘建国 赵南京[†] 马明俊 王寅 胡丽 张大海 余洋
孟德硕 章炜 刘晶 张玉钧 刘文清

(中国科学院安徽光学精密机械研究所, 中国科学院环境光学与技术重点实验室, 合肥 230031)

(2012年11月1日收到; 2013年1月6日收到修改稿)

建立了适用于激光诱导击穿光谱探测的多元线性回归、神经网络回归和支持向量机回归三种定量反演算法模型, 以水体重金属 Ni 为例进行了回归实验测试和对比分析. 多元线性回归、神经网络回归和支持向量机回归的平均相对标准偏差分别为 7.60%, 4.86%, 2.35%; 最大相对标准偏差分别为 23.35%, 15.20%, 8.29%; 平均相对误差分别为 25.98%, 10.58%, 2.72%, 最大相对误差分别为 116.47%, 47.38%, 9.89%. 研究为进一步实现水中痕量金属元素的快速定量分析提供了方法和数据参考.

关键词: 光谱学, 激光诱导击穿光谱, 支持向量机回归, 重金属

PACS: 52.50.Jm, 32.30.-r, 42.62.Fi

DOI: 10.7498/aps.62.125201

1 引言

激光诱导击穿光谱技术 (laser-induced breakdown spectroscopy, LIBS) 是一种通过探测激光诱导等离子体, 由分析线的位置及信号强度获取物质成分和浓度的快速物质成分分析技术^[1], 已被广泛应用于化学、医学和生物学等诸多研究领域^[2,3]. 由于固态样品不需要预处理就可以获得很好的信号灵敏度和重复性, LIBS 定量分析方法研究主要集中在煤、合金、土壤等测量领域. 董美蓉等^[4] 基于多元定标法对煤粉碳元素进行了 LIBS 定量分析, 绝对偏差和相对偏差均在 5% 之内; Wang 等^[5,6] 应用基于主导因素的偏最小二乘法 (partial least squares, PLS) 模型和基于 PLS 的非线性多变量主导因素模型对黄铜合金成分进行了定量测量, 相关系数达到 0.999; 周卫东等^[7] 基于遗传神经网络对土壤中的 Cr 和 Ba 元素进行了定量检测, 相关系数分别为 0.958 和 0.992, 相对误差分别低

于 4.85% 和 8.63%, 相对标准偏差在 9.48% 以内. 相对而言, 水中元素定量检测却困难重重. 虽然采用水面、水滴、水柱、喷雾、冰冻样品^[8]、化学固化^[9]、滤纸富集^[10]、木片富集^[11]、电化学沉积^[12]、超声波雾化^[13] 等采样方式可以减小水体压力、波动、强吸收、溅射污染光学元件、基体效应、富集不均匀等因素的影响^[14], 但检测限和稳定性仍不理想, 定量分析基本仍停留在单曲线定标拟合阶段^[15]. LIBS 技术的快速发展迫切需要稳定可靠的定量分析方法.

2 定量分析模型

激光诱导击穿光谱谱线强度可表示为^[5]

$$I_{\lambda}^{ki} = FC_S \frac{A_{ki} g_k}{U_S(T)} e^{-E_k/k_B T}, \quad (1)$$

其中, I_{λ}^{ki} 为实验中测得的谱线强度, C_S 为重金属元素 S 的质量浓度, $U_S(T)$ 为其配分函数, 谱线跃迁的高能级为 k , 低能级为 i , g_k 为高能级的简并度, A_{ki}

* 国家自然科学基金 (批准号: 60908018) 和国家重大科技专项 (批准号: 2009ZX07527-007, 2009ZX07420-008, 2011ZX05051) 资助的课题.

[†] 通讯作者. E-mail: njzhao@aiofm.ac.cn

为特征谱线的跃迁概率, E_k 为跃迁能级的高能级能量, T 为等离子体温度, F 为与光收集装置的效率相关的实验参数, k_B 为玻尔兹曼常数. 理论上, C_S 和 I_λ^{ki} 成线性关系, 但由于受自吸收、多组分互扰、激光能量抖动、透镜与样品距离、延时、门宽、重复频率等因素的影响, 实际测量结果和理论值有很大偏差, 如图 1^[5] 所示.

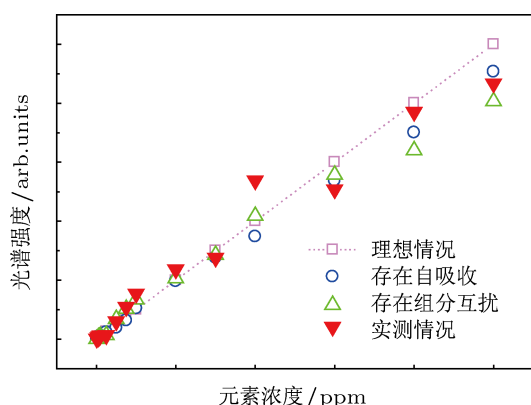


图 1 LIBS 测量中典型光谱强度随浓度的变化

为克服上述因素的影响, 提高水体重金属激光诱导击穿光谱定量精度, 实验中建立了多元线性回归, 神经网络回归和支持向量机回归三种多变量定量回归分析模型.

1) 多元线性回归模型

多元线性回归是回归分析中应用最为广泛、理论上已经成熟的方法, 很多非线性的回归问题也可以通过变换转化为线性回归求解. 简写 I_λ^{ki} 为 I , C_S 为 C , 带入标准方程, 得到多元线性回归模型^[4]

$$C_{MLR} = b_0 + b_1 I_1 + b_2 I_2 + \dots + b_m I_m + \xi, \quad (2)$$

式中 C_{MLR} 为多元线性回归得到的重金属浓度, b_0, b_1, \dots, b_m 为常数, I_1, I_2, \dots, I_m 为某元素多个特征峰的积分谱线强度, ξ 为服从正态分布的误差.

2) 神经网络回归模型

误差反向传播 (back propagation, BP) 神经网络是基于误差反向传播算法的多层前向神经网络, 具有很强的非线性映射能力和泛化能力. 其输入的正向传播方程^[16] 为

$$y_h = f_1 \left(\sum_i w_{ih} I_i \right), \quad (3)$$

$$C_{NNR} = f_2 \left(\sum_h w_{hj} f_1(w_{ih} I_i) \right), \quad (4)$$

式中 C_{NNR} 为输出层输出, 即神经网络预测得到的重金属浓度, I_i 为某元素多个特征峰的积分谱线强

度; y_h 是隐含层输出; f_1 是输入层和隐含层之间的 S 型传递函数; f_2 是隐含层与输出层之间的线性传递函数; w_{ih} 是输入层至隐含层连接权值; w_{hj} 是隐含层至输出层连接权值. 如输出层得不到目标值, 将目标值与输出值之差反向传播, 逐层修改各层神经元连接权值, 使输出误差减小至允许范围内. 其权值修正公式^[16] 为

$$w_{hj}(n+1) = w_{hj}(n) + \eta \delta_j y_h, \quad (5)$$

其中: $\delta_j = f_2' \left(\sum_h (w_{hj}(n) y_h) \right) \cdot (T_j - y_j) \cdot w_{ih}(n+1)$,

$$w_{ih}(n+1) = w_{ih}(n) + \eta \delta_h I_i, \quad (6)$$

其中: $\delta_h = f_1' \left(\sum_i (w_{ih}(n) I_i) \cdot \sum_j (w_{hj}(n) \delta_j) \right)$, 式中 $w_{hj}(n+1)$, $w_{ih}(n+1)$ 分别表示第 $(n+1)$ 次权值系数, $w_{hj}(n)$, $w_{ih}(n)$ 分别表示第 n 次权值系数; T_j 为输出层第 j 个结点目标值; η 为训练速度系数, 取值 0—1.

3) 支持向量机回归模型

支持向量回归 (support vector regression, SVR) 是以统计学习理论 (statistical learning theory, SLT) 为基础, 建立在 VC (vapnik-chervonenkis dimension) 维、结构风险最小和核函数原理基础上的新型机器学习方法^[17], 适用于小样本^[18], 可以融合先验知识^[19], 通过核函数能和小波等先进算法相结合^[20], 已被广泛应用于各种预测领域^[21-24]. 其回归模型^[23] 为

$$C_{SVR} = \sum_{i \in SV} \alpha_i \cdot k_{libS}(I_i, I) + b^*, \quad (7)$$

式中 C_{SVR} 为支持向量机回归得到的重金属浓度, SV 为支持向量集, α_i 和 I_i 分别为相应支持向量的拉格朗日乘子和谱线强度, b^* 为常数, k_{libS} 为自适应混合核函数, 表达式为

$$k_{libS} = mx \cdot x_i + (1 - m) \exp \left(\frac{-\|x - x_i\|^2}{2p^2} \right) \quad (0 \leq m \leq 1), \quad (8)$$

式中第一项为线性核函数 k_{liner} , 第二项为径向基核函数 k_{rbf} , 两种核函数的贡献比重通过参数 m 自适应调整.

3 实验样本

3.1 实验装置

水中重金属等离子体寿命短且不稳定, 光谱信

号微弱, 直接采样比较困难. 实验中采用石墨富集的方式进行预处理, 获得更强、更丰富的水中重金属激光诱导击穿光谱信号. 实验原理框图如图 2 所示.

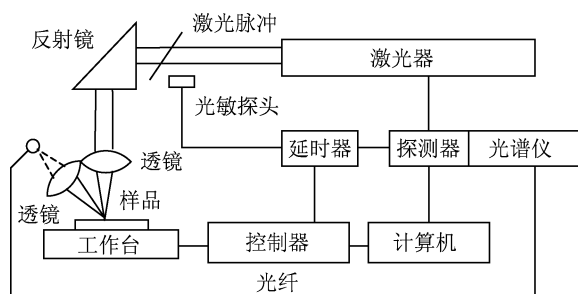


图2 实验系统原理框图

采用 Nd:YAG 激光器作为激发光源 (波长 1064 nm, 单脉冲能量 320 mJ, 脉冲宽度 6 ns), 激光束经焦距为 100 mm 的透镜后聚焦在富集有 1.5 mL 待测样品的带凹槽圆饼型高纯石墨基体表面, 产生的激光等离子体光谱信号经焦距为 50 mm 的石英透镜耦合至光纤; 传输至光谱仪完成光谱的分光与探测, 其中光谱仪 (Mechell 5000 型, 英国 Andor 公司) 测量波长范围为 200—975 nm, 分辨率 0.1 nm, 探测器为 ICCD (iStar 型, 英国 Andor 公司), 像素 1024×1024 . 样品置于程控二维旋转工作台上, 按预定轨迹实现无重复采样打点, 以保证样品测量的均匀性和稳定性.

3.2 样本获取及优化

实验中, 选择镍作为分析元素, 其积分强度衰减特性曲线如图 3 所示.

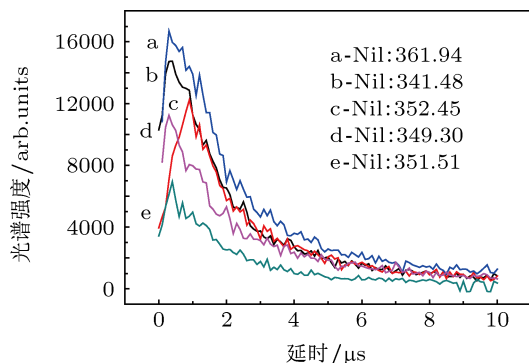


图3 镍元素的衰减特性

综合考虑五条谱线衰减特性, 确定最佳延时为 $0.5 \mu\text{s}$, 最佳门宽为 $5 \mu\text{s}$, 此时各特征峰信噪

比较好. 采用 2.1 中实验装置对镍含量为 100, 80, 60, 40, 30, 20, 10, 7.5, 5, 2.5, 1, 0.75, 0.5, 0.1 ppm ($1 \text{ ppm} = 10^{-6}$) 的 14 种浓度混合样本分别进行 30 组测试, 随机抽取每种浓度中的 20 组数据作为训练集, 剩余数据作为测试集. 采用下述四步提取样本信息.

1) 冗余信息剔除 单次测量产生 24000 组光谱数据, 其中绝大多数信息与被分析元素无关, 实验中分别选取镍的五个特征峰光谱段 (NiI: 361.94, 341.48, 352.45, 349.30, 351.51 nm) 作为基本输入信息, 这将比传统单特征峰分析方法获得更多有用信息;

2) 基体背景剔除 石墨富集方式测量水体重金属的好处之一就是基体背景可以通过多次测量空白石墨得到, 用实测光谱减去校正后的基体背景就可得到只含有待测水样成分的纯净光谱;

3) 环境噪声剔除 由于仪器自身特点和测量过程中不可避免的干扰, 获得的特征峰会发生变形、毛刺和漂移, 实验中通过洛伦兹拟合获得特征峰谱线的积分强度和拟合峰强度, 这比传统未拟合方法获得的数据更准确;

4) 能量抖动剔除 普通实验室用激光器的能量抖动范围在 5% 左右, 高端激光器也只能控制在 2%, 这对定量测量影响很大, 实验中以石墨基体中的碳 (CI: 247.86 nm) 作为内标, 对五个特征峰光谱的积分强度和峰值强度进行归一化处理, 减小能量抖动产生的测量误差. 按上述方法获得 Ni 的五个特征峰光谱的积分强度和峰值强度共 10 个特征向量作为定量分析的输入样本.

4 定量结果及对比分析

4.1 评价指标

实验中用相对标准偏差和相对误差指标评定定量结果.

相对标准偏差 (RSD) 用于衡量回归预测的精密程度, 即独立预测结果之间的一致程度, 计算公式如下:

$$RSD = \left(\frac{1}{\bar{x}} \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \right) \times 100\%, \quad (9)$$

式中: RSD 为相对标准偏差, x_i 为预测值, \bar{x} 为预测均值, n 为预测次数.

相对误差用于衡量回归预测的准确度,即预测结果与真实值之间的一致程度,计算公式如下:

$$RE = \left| \frac{\bar{x} - \mu}{\mu} \right| \times 100\%, \quad (10)$$

式中: RE 为相对误差, \bar{x} 为预测均值, μ 为真实浓度.

4.2 定量结果

应用第 2 部分中定量分析模型和第 3 部分中获取的样本,在 matlab 中对水中重金属 Ni 含量

进行定量回归分析.多元线性回归、神经网络回归和回归支持向量机回归分别基于 regress 函数、nntool 工具箱和台湾大学林智仁的 libsvm 3.1 软件包编程实现,三种定量反演算法的相对标准偏差和相对误差对比如图 4 所示.

多元线性回归, BP 神经网络回归和支持向量机回归的平均相对标准偏差分别为 7.60%, 4.86%, 2.35%, 最大相对标准偏差分别为 23.35%, 15.20%, 8.29%; 平均相对误差分别为 25.98%, 10.58%, 2.72%, 最大相对误差分别为 116.47%, 47.38%, 9.89%. 详细实验数据如表 1 所示.

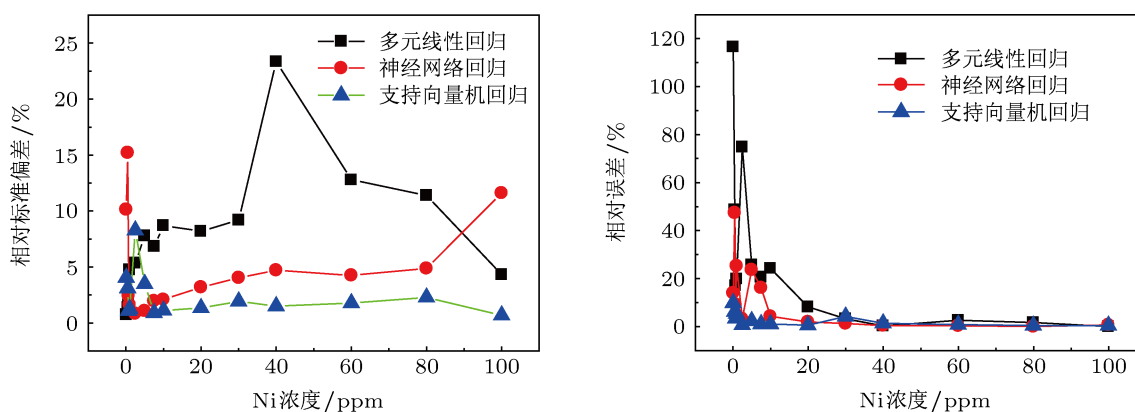


图 4 相对标准偏差和相对误差对比

表 1 回归预测结果

样本组号	真实浓度 /ppm	多元线性回归/ppm	相对标准偏差/%	相对误差/%	神经网络回归/ppm	相对标准偏差/%	相对误差/%	支持向量回归/ppm	相对标准偏差/%	相对误差/%
1	100.00	99.91	4.32	0.09	100.68	11.61	0.68	99.53	0.68	0.47
2	80.00	78.61	11.40	1.73	80.06	4.88	0.08	80.34	2.29	0.43
3	60.00	61.61	12.79	2.69	59.76	4.26	0.41	59.48	1.78	0.87
4	50.00	40.10	23.35	0.26	39.83	4.71	0.42	40.57	1.51	1.43
5	40.00	29.02	9.19	3.25	29.61	4.03	1.29	28.74	1.93	4.19
6	20.00	18.36	8.19	8.21	19.62	3.19	1.92	19.88	1.35	0.59
7	10.00	12.42	8.69	24.23	10.43	2.11	4.28	10.11	1.12	1.12
8	7.50	9.04	6.84	20.58	6.29	1.96	16.14	7.58	0.90	1.11
9	5.00	6.28	7.77	25.67	3.82	1.09	23.64	4.88	3.50	2.37
10	2.50	0.63	5.35	74.83	2.42	0.83	3.28	2.52	8.29	0.68
11	1.00	1.20	4.76	19.88	1.25	1.62	25.33	0.95	1.28	5.38
12	0.75	0.62	1.69	17.11	0.68	2.44	9.27	0.78	1.10	3.36
13	0.50	0.26	1.38	48.70	0.26	15.20	47.38	0.53	3.10	6.23
14	0.10	0.22	0.73	116.47	0.11	10.13	14.07	0.09	4.05	9.89
平均值	—	—	7.60	25.98	—	4.86	10.58	—	2.35	2.72
最大值	—	—	23.35	116.47	—	15.20	47.38	—	8.29	9.89

4.3 分析讨论

从实验结果来看,三种方法都可以用于定量分析,具体如下.

1) 支持向量机回归的效果最好,平均相对标准偏差和平均相对误差都低于 3%,明显优于其他两种方法.

实验中采用的自适应混合核函数 k_{libs} 由线性核函数 k_{liner} 和径向基核函数 k_{rbf} 线性组合而成. k_{liner} 善于提取样本的全局特性,具有良好外推能力. k_{rbf} 善于提取样本的局部性质,学习能力强. 通过改变两种核函数的贡献比重,可以实现水体重金属浓度宽范围变化时模型的自适应调整,获得良好的回归效果. 当水中重金属浓度较高,谱线自吸收明显时, k_{rbf} 发挥主要作用,反之, k_{liner} 发挥主要作用.

2) 神经网络回归训练效果优于支持向量机回归,但在测试集中表现一般,平均相对误差高于 10%.

BP 神经网络有着很强的非线性拟合能力,能以任意精度逼近任意非线性连续函数,由于实验中的 LIBS 数据属于小样本,造成网络训练不充分,影响其预测精度. 相对而言,支持向量机回归基于结构风险最小化,而不是经验风险最小化,可以根据有限样本信息在模型的复杂性和学习能力之间寻求最佳折中,在保证泛化能力的前提下,达到最优学习效果,使其泛化能力明显优于 BP 神经网络.

3) 多元线性回归的平均相对标准偏差为 7.60%,明显劣于支持向量机回归和神经网络回归.

多元线性回归的相对标准偏差偏大主要有两点原因:一是实验中的输入矩阵由 Ni 的各个特征峰拟合得到,具有一定的共线性,可能形成病态矩

阵无法求逆;二是输入矩阵存在噪声,导致了过度拟合,某种程度上降低了模型的预测能力. 而支持向量机回归和神经网络回归等机器学习方法具有一定的容错能力,可以克服噪声的影响,减小预测结果的波动性.

4) 三种方法的相对误差均在高、中浓度时小,低浓度时大.

多元线性回归在检测限附近时相对误差达到 116.47%,即使是预测效果较好的支持向量机回归也达到了 9.89%. 原因主要是探测浓度在检测限附近时,激光诱导击穿光谱的信背比变小,预测结果与真实值之间的偏差变大. 这也说明定量分析方法只能在一定量程上提高准确度和精密度,无法改变实验方法固有的检测限.

5 结论

本文建立了适用于激光诱导击穿光谱探测的多元线性回归, BP 神经网络回归和支持向量机回归三种定量反演算法模型,以水体重金属 Ni 为例进行了回归实验测试和对比分析. 研究表明: 1) 支持向量机回归的效果最好,平均相对标准偏差和平均相对误差都低于 3%,明显优于其他两种方法; 2) 神经网络回归训练效果优于支持向量机回归,但在测试集中表现一般,平均相对误差高于 10%; 3) 多元线性回归的平均相对标准偏差为 7.60%,明显劣于支持向量机回归和神经网络回归; 4) 三种方法的相对误差均在高、中浓度时小,低浓度时大,说明定量分析方法只能在一定量程上提高准确度和精密度,无法改变实验方法固有的检测限.

- [1] Sadegh Cheri M, Tavassoli S H 2011 *Appl. Opt.* **50** 1227
- [2] Sun D X, Su M G, Dong C Z 2010 *Acta Phys. Sin.* **59** 4571 (in Chinese) [孙对兄, 苏茂根, 董晨钟 2010 物理学报 **59** 4571]
- [3] Lu C P, Liu W Q, Zhao N J 2011 *Acta Phys. Sin.* **60** 045206 (in Chinese) [鲁翠平, 刘文清, 赵南京 2011 物理学报 **60** 045206]
- [4] Dong M R, Lu J D, Yao S C 2012 *J. Engineer. Thermophys.* **33** 175 (in Chinese) [董美蓉, 陆继东, 姚顺春 2012 工程热物理学报 **33** 175]
- [5] Wang Z, Feng J, Li L Z 2011 *J. Anal. At. Spectrom.* **26** 2289
- [6] Wang Z, Feng J, Li L Z 2011 *J. Anal. At. Spectrom.* **26** 2175
- [7] Shen Q M, Zhou W D, Li K X 2010 *Acta Photon. Sin.* **39** 2134 (in Chinese) [沈沁梅, 周卫东, 李科学 2010 光子学报 **39** 2134]
- [8] Alsulmi A S 2008 *Ph. D. Dissertation* (Saudi Arabia: Kingdom of Saudi Arabia Ministry of Higher Education King Saud University)
- [9] Díaz Pace D M 2006 *Spectrochim. Acta B* **61** 929
- [10] Wu J Q, Chang L, Liu L M 2010 *Appl. Laser.* **30** 340 (in Chinese) [吴金泉, 常亮, 刘林美 2010 应用激光 **30** 340]
- [11] Chen Z J, Li H K 2008 *Spectrochim. Acta B* **63** 64
- [12] Zhao F, Zhang Q, Xiong W 2010 *Environ. Sci. Technol.* **33** 137 (in Chinese) [赵芳, 张谦, 熊威 2010 环境科学与技术 **33** 137]
- [13] Zhong S L, Lu Y, Cheng K 2011 *Spectrosc. Spectral Anal.* **31** 1458 (in Chinese) [钟石磊, 卢渊, 程凯 2011 光谱学与光谱分析 **31** 1458]
- [14] Wu J L, Fu Y X 2008 *Spectrosc. Spectral Anal.* **28** 1979 (in Chinese) [吴江来, 傅院霞 2008 光谱学与光谱分析 **28** 1979]
- [15] Xiu J S, Hou H M, Zheng R E 2011 *Chin. J. Lasers* **38** 0815003 (in Chinese) [修俊山, 侯华明, 郑荣儿 2011 中国激光 **38** 0815003]
- [16] Zhu B F, Xu G Y, Li J M 2007 *Instrum. Metrol. Technol.* **1** 39 (in Chinese) [朱滨峰, 徐桂云, 李俊敏 2007 仪器仪表标准化与计量 **1** 39]

- [17] Xu H M, Wang H Y, Liang J 2010 *J. Beijing Institute Petro-Chem. Technol.* **18** 62 (in Chinese) [徐红敏, 王海英, 梁瑾 2010 北京石油化工学院学报 **18** 62]
- [18] Chen G, Zhou J 2008 *Acta Metrol. Sin.* **29** 92 (in Chinese) [陈果, 周伽 2008 计量学报 **29** 92]
- [19] Liu X F, Wang J D 2011 *Informat. Res.* **37** 46 (in Chinese) [刘小飞, 王建东 2011 信息化研究 **37** 46]
- [20] Zhang X S, Wang L, Pan F 2012 *Comput. Engineer.* **38** 175 (in Chinese) [张相胜, 王蕾, 潘丰 2012 计算机工程 **38** 175]
- [21] Hou Z Y, Yao S W, Gu Y Q 2006 *J. Henan Normal Univ.* **34** 75 (in Chinese) [侯振雨, 姚树文, 谷永庆 2006 河南师范大学学报 **34** 75]
- [22] Zhang X, Liu X M, Xiao X L 2005 *Chin. J. Eng. Geophys.* **2** 191 (in Chinese) [张翔, 刘晓敏, 肖小玲 2005 工程地球物理学报 **2** 191]
- [23] Guo S X, Wang Y F, Chen A 2007 *Comput. Engineer. Appl.* **43** 12 (in Chinese) [郭水霞, 王一夫, 陈安 2007 计算机工程与应用 **43** 12]
- [24] Zheng Y, Cheng W M, Cheng Y 2011 *Mech. Sci. Technol.* **30** 52 (in Chinese) [郑严, 程文明, 程跃 2011 机械科学与技术 **30** 52]

Comparative analysis of quantitative method on heavy metal detection in water with laser-induced breakdown spectroscopy*

Wang Chun-Long Liu Jian-Guo Zhao Nan-Jing[†] Ma Ming-Jun Wang Yin
Hu Li Zhang Da-Hai Yu Yang Meng De-Shuo Zhang Wei Liu Jing
Zhang Yu-Jun Liu Wen-Qing

(Key Laboratory of Environment Optics and Technology, Institute of Anhui Optics Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China)

(Received 1 November 2012; revised manuscript received 6 January 2013)

Abstract

The quantitative analysis models of multiple linear regression, neural network regression and support vector machine regression with laser-induced breakdown spectroscopy are established in this paper. Heavy metal Ni in water selected as research object is tested and comparatively analyzed. The average relative standard deviations of multiple linear regression, neural network regression and support vector machine regression are 7.60%, 4.86% and 2.35%, and the maximum standard deviations are 23.35%, 15.20% and 8.29% respectively, the average relative errors are 25.98%, 10.58% and 2.72%, and the maximum relative errors are 116.47%, 47.38% and 9.89% respectively. Methods and reference data are provided for the further study of fast measurement of tracing heavy metals in water by laser induced breakdown spectroscopy technique.

Keywords: spectroscopy, laser-induced breakdown spectroscopy, support vector machine regression, heavy metals

PACS: 52.50.Jm, 32.30.-r, 42.62.Fi

DOI: 10.7498/aps.62.125201

* Project supported by the National Natural Science Foundation of China (Grant No. 60908018) and the Major Project of the Ministry of Science and Technology of China (Grant Nos. 2009ZX07527-007, 2009ZX07420-008, 2011ZX05051).

[†] Corresponding author. E-mail: njzhao@aiofm.ac.cn