

小样本量复合肥中氮磷钾高精度预测光谱方法

马骁¹, 李安¹, 王宪双¹, 孔德男¹, 邱苏玲¹, 何雅格¹, 殷允嵩¹, 柳宇飞², 刘瑞斌^{1*}

¹北京理工大学物理学院, 北京 100081

²宝瑞激光科技(常州)有限公司, 江苏 常州 213000

摘要 采用激光诱导击穿光谱(LIBS)技术对复合肥样品中氮(N)、磷(P)和钾(K)等营养元素进行定量分析。实验中一共选取 20 个样品, 由于样品量较少, 为了提高预测精度, 采取一种新的数据提取方式来建立训练集和预测集。利用 PLS 结合主成分分析(PCA)为训练集的光谱数据建立定标模型, 定标过程中选取 12 个主成分。N、P、K 三种元素定标模型的决定系数分别为 0.99、0.98、0.99; 20 个样品中 N、P、K 元素含量(质量分数, 下同)预测的平均相对误差分别为 2.33%、0.70%、3.00%。之后对定标模型的鲁棒性进行检验, 其中 N、P 的平均相对误差大多维持在 12% 以下。采用基于统计学原理的数据抽取方式扩充样本光谱数据后, 与未扩充时相比, 被测元素含量的平均相对误差降幅均在 10% 以上。实验结果表明, 当样本数量较少时, 利用所提的数据提取方法结合 PLS 定量分析可以提高检测的准确度, 实现复合肥样品中 N、P、K 等营养元素的快速检测。

关键词 光谱学; 激光诱导击穿光谱; 复合肥; 偏最小二乘法; 小样本量

中图分类号 O433.4

文献标志码 A

doi: 10.3788/CJL202148.2311003

1 引言

氮(N)、磷(P)和钾(K)是复合肥产品的三种主要营养元素, 是植物生长所需要的必要营养元素。而随着我国农业的快速发展, 复合肥的消耗需求越来越大, 因此复合肥的质量检测也就变得非常重要。目前复合肥的检测过程主要包括人工采样、制样、送实验室检测分析, 这种检测分析周期一般比较长。实验室检测的方法主要有火焰原子吸收光谱法^[1]、电感耦合等离子体发射光谱(ICP-AES)法^[2]、近红外反射光谱(NIRS)法^[3]等。这些方法缩短了检测时间, 但是需要对样品进行预处理, 不仅无法实现实时监测, 还有可能导致测量结果的准确性降低。

与上述检测技术相比, 激光诱导击穿光谱(LIBS)技术具有无需对样品进行预处理、检测时间短、可实现全元素实时检测、检测过程绿色安全等优点。在采用 LIBS 技术的检测中, 通过将一束高能量、短脉冲的激光束聚焦到待检测的样品表面, 待激光烧蚀样品后产生等离子体, 采集等离子体的辐射

光谱并进行分析, 从而实现对样品的定性、定量分析。目前 LIBS 技术已经广泛应用于很多领域, 如生物医学^[4-5]、工业应用^[6]、环境检测^[7-10]等领域, 包括复合肥检测领域。Farooq 等^[11]利用 LIBS 技术对 DAP 肥料中的元素进行分析, 实验结果表明该肥料中不仅含有营养元素, 还存在一些重金属元素, 可能会对环境和人类健康造成危害。廖索引等^[12]利用 LIBS 技术对复合肥中的 P 元素进行多元非线性定量分析, 测量相对误差仅为 0.38%~1.70%, 大大提高了测量的准确性。陈凯等^[13]利用 LIBS 技术对复合肥中 K 元素的含量进行分析, 在考虑自吸收之后选取 404.40 nm 的谱线作为 K 元素的特征谱线, 得到定标曲线的线性拟合度为 0.989, 检验样品的绝对误差小于 0.3%。卢伟业等^[14]应用 LIBS 技术同时检测出 5 个样品的 N、P、K 元素的含量, 检测结果的平均相对误差(质量分数)小于 8%, 平均相对标准偏差(质量分数)小于 7%, 检测极限值分别为 0.16%、0.21% 和 0.50%(质量分数)。Yao 等^[15]采用 LIBS 技术结合偏最小二乘法对肥料样品

收稿日期: 2021-03-22; 修回日期: 2021-04-23; 录用日期: 2021-05-14

通信作者: *1299854386@qq.com

进行了快速、准确的多元素分析,对于 P 和 K,PLS 回归模型的相关系数高于 0.97。PLS 回归模型准确性的验证结果显示 P_2O_5 和 K_2O 的质量分数绝对误差在 0.16%~0.63% 范围内。Zhang 等^[16] 利用 LIBS 技术对复合肥料中的 P 元素进行单变量和多元分析,多元分析时采用偏最小二乘回归(PLSR)法对 P 元素进行建模。多元分析得到的校准集和预测集的相关系数分别为 0.977 和 0.976,表明 PLSR 法可以提高 LIBS 测量的准确性。

以上研究表明 LIBS 技术可用于复合肥的成分检测和种类识别,但用于小样本容量的研究较少且精度不高。因此,当样本容量较少时,预测精度的提高显得尤为重要。为了提高小样本容量的预测精度,本文在定标数据的选取上进行了创新,实验中一共选取 20 个样品,每个样品采集 175 组光谱数据,其中 100 组光谱数据用于模型训练,剩余 75 组取全谱平均值后作为模型检验集。在模型训练集中,每个样品随机抽取 80% 的光谱数据进行组合,并且每个样品随机抽取数据 1000 次以扩充样本光谱数据,抽取得到的组合光谱再取平均值作为该样品的最终光谱。经过这种数据抽取方式扩充样本光谱数据后,与未扩充时相比,小样本量复合肥中营养元素的预测精度明显提高。

2 实验过程

2.1 实验装置

实验装置如图 1 所示,其中:激光器为 1064 nm 主动调 Q 的 Nd:YAG 脉冲激光器(Newwave, USA),脉宽为 3~5 ns,可调重复频率为 1~10 Hz,激光器的最大输出能量为 130 mJ,实验中所用能量为 50 mJ;光谱仪为四通道光纤光谱仪(AvaSpec-ULS2048CL-4-EVO, Netherlands;190~790 nm),分辨率为 0.1 nm。利用数字脉冲发生器(DG535, Stabford Research Systems, USA)为激光器和光谱仪提供精准的外触发信号。脉冲激光经过反射镜反射后由透镜聚焦到样品表面,样品表面经激光烧蚀产生等离子体,等离子体的光信号由光学收集系统进行收集。光收集系统由两个透镜组成,在与激光作用方向成 45° 的方向进行收集,并将收集到的光通过光纤传输到四通道光纤光谱仪。实验过程中,光谱仪采集光谱的延迟时间由数字脉冲发生器精确控制,以便得到具有较小连续背景辐射的光谱图。光谱仪的延迟时间为 1 μ s,积分时间为 1 ms。

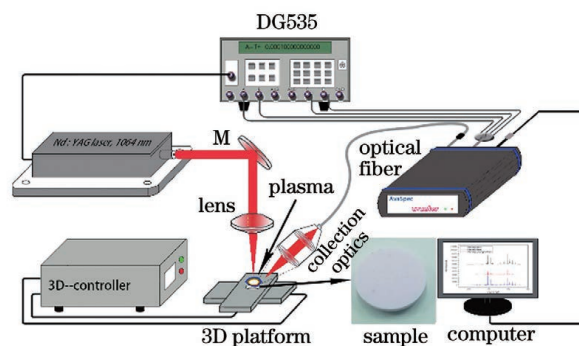


图 1 实验装置

Fig. 1 Experimental setup

2.2 样品制备

将 20 种营养成分类别相同,但各成分含量不同的复合肥料作为样品,该样品的主要营养成分为 N、 P_2O_5 、 K_2O ,各成分的含量(下文的“含量”均指质量分数)见表 1。实验中每种样品均取 0.5 g 进行 3~5 min 的研磨,之后使用机械粉末压片机将其压制厚度为 3 mm、直径为 13 mm 的圆饼,机械粉末压片机的压力为 20 MPa,压制时间为 200 s。

表 1 复合肥料样品编号及主要营养成分的含量

Table 1 Sample number and content of main nutrients of compound fertilizer

Sample number	Content of N / %	Content of P_2O_5 / %	Content of K_2O / %
2009-3	15.86	17.62	15.11
2009-4	15.26	17.22	15.71
2009-6	15.48	17.22	15.34
2009-7	15.99	16.75	15.12
2009-9	23.22	17.00	6.18
2009-11	23.03	16.68	6.35
2009-16	23.48	16.40	6.04
2009-17	23.55	17.54	6.32
2009-18	23.16	17.46	6.32
2009-22	23.73	16.23	6.11
2009-24	23.66	16.32	6.32
2010-21	15.74	17.46	15.47
2010-16	15.08	17.02	15.92
2010-40	13.82	16.42	15.40
2011-5	15.57	16.00	16.25
2011-10	16.22	16.70	15.45
2011-15	13.96	16.17	14.58
2011-20	14.65	15.46	15.02
2011-25	14.29	15.93	14.65
2011-30	14.76	15.72	14.08

复合肥样品的等离子体光谱如图 2 所示,根据美国国家标准与技术研究院(NIST)的标准原子光谱数据库收录的光谱数据,对复合肥样品的元素组

成进行鉴定和标记。从图 2 可以看到,该复合肥样品中主要包含 N、P、K 等特征元素,以及 Fe、Na、Zn 等微量元素。

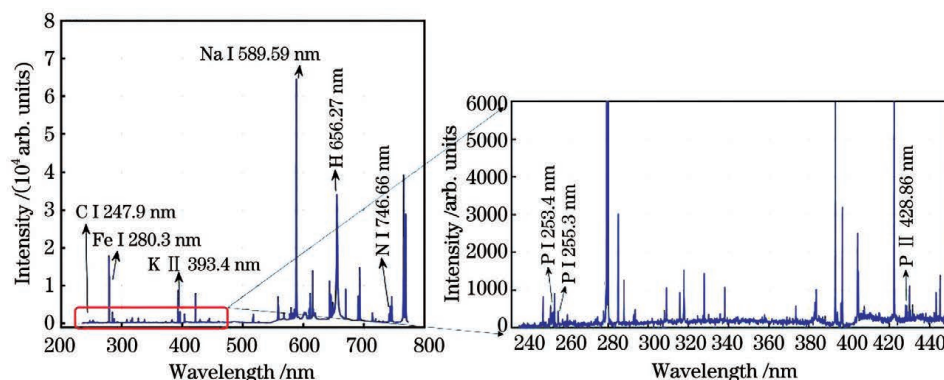


图 2 复合肥样品的等离子体发射光谱

Fig. 2 Plasma emission spectra of compound fertilizer samples

3 结果与讨论

3.1 光谱数据预处理

实验过程中在每个样品上取 5 行 7 列共 35 个点作为脉冲激光的作用点,每个点作用 5 次,相邻作用点的间隔为 1 mm。每个样品共采集 175 组光谱数据,20 个样品共计采集 $20 \times 175 = 3500$ 组光谱数据。对光谱数据进行预处理的步骤如下:筛选与剔除异常光谱,对光谱进行去背景,并将分通道背景面积归一化。

异常光谱的筛选与剔除遵循拉依达准则^[17];光谱去背景常用的方法为插值法和窗口平移平滑法^[18],本实验中使用窗口平移平滑法去除光谱背景基线;对预处理之后的光谱数据进行分通道背景面积归一化^[19]。对光谱数据进行预处理后的光谱如图 3 所示。预处理后的光谱与原始光谱相比,主要的变化出现在长波波段,背景被消除。由于采

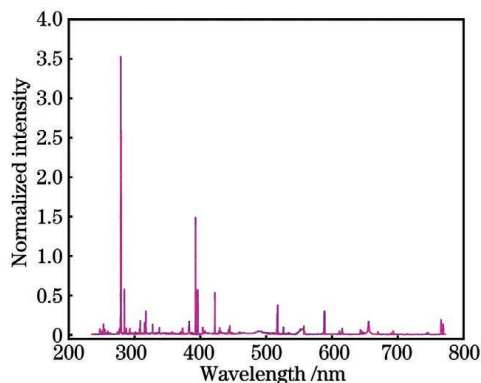


图 3 预处理之后复合肥样品的光谱

Fig. 3 Spectrum of a compound fertilizer sample after pretreatment

用的是分通道背景面积归一化,通道与通道之间的相对强度出现变化,但通道之间各条谱线的相对强度不变。

3.2 PLS 定量分析模型的建立

在使用 PLS 定量分析方法建模时,通常是将一部分样品作为训练集,另一部分样品作为预测集,然后利用根据训练集建立的模型对预测集进行预测。对于样本容量较多的数据,建立模型可以对每个样品的多组光谱取平均,但对于样本量较少的数据,直接对全部光谱数据取平均值时,训练模型的精度和鲁棒性都会出现较大的波动^[14],本实验中对样本量较少的光谱数据的抽取方法做了如下改变:一共有 20 个样品,对每个样品采集 175 组光谱数据,其中前 100 组光谱数据用于模型训练,标记为 T1 类光谱,后 75 组取全谱平均值后作为模型检验集,标记为 T2 类光谱。在 T1 类光谱中,对每个样品随机抽取 80% 的光谱数据进行组合。图 4 为所测复合肥样品 T1 类光谱数据的存储方式和数据抽取示意图,将每个样品所抽取的光谱数据存储为 Sample 1, Sample 2, ..., Sample N, 其中每一次数据采集记为 Pulse K ($K \leq 100$),粗体的光谱数据为每个样品随机抽取作为训练集的数据,即每个样品抽取的光谱属于不同的数据采集点并且每个样品随机抽取数据 1000 次以扩充样本光谱数据,抽取得到的组合光谱再取平均值作为该样品的最终光谱,20 个样品在 1000 次抽取后得到的样品光谱矩阵大小为 $20 \times 8192 \times 1000$,将其中每一个 20×8192 的光谱矩阵作为模型的训练基元,选取其中 75% 的数据即 15×8192 的光谱矩阵进行 PCA-PLS 建模,剩余

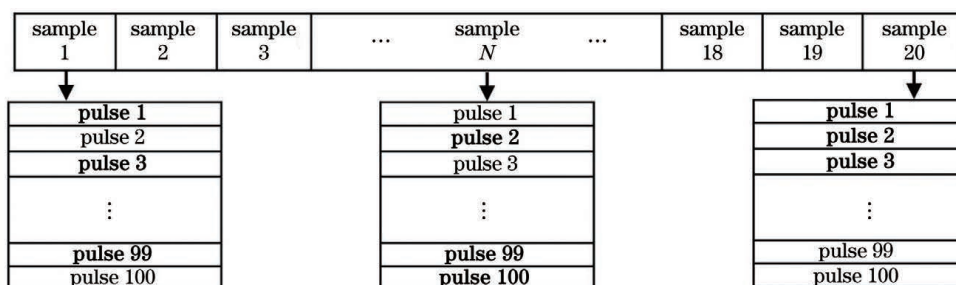


图 4 训练集与预测集光谱数据抽取原理图

Fig. 4 Schematic of spectral data extraction of training set and prediction set

25%数据即 5×8192 的光谱矩阵进行预测。

利用以上的数据抽取方法扩充样本光谱数据,通过 PCA-PLS 化学计量方法分别建立复合肥料样品中 N、P、K 元素的定量分析模型。目前在定量分析领域较普遍的做法是在分析时采用基于主成分分析^[20]的偏最小二乘法^[21](PCA-PLS)。主成分分析是将高维光谱数据中的变量进行矩阵变换和线性组合,在主成分空间提取方差最大的几个维度对应的变量信息,用所选取的这些新的综合变量可以最大限度地反映原始变量数据所包含的信息。模型训练时提取训练集的前 12 个主成分,用这 12 个主成分代表全谱的原始数据变量,建立预测模型,分别用决定系数(R^2)、训练集的均方根误差(RMSEC)、预测集的均方根误差(RMSEP)、相对误差(RE)来综合衡量模型的回归效果、定标精度、预测精度、预测误差,并将 RMSEP 最小值对应的回归系数作为模型系数。根据 T1 类光谱数据中的建模集和预测集,图 5 给出了 PLS 定量分析模型分别对训练集和预测集中 N 元素含量预测值与实际值的拟合曲线,得到训练集模型的决定系数 $R^2 = 0.99$, RMSEC 为 0.4309,这表明训练集中 N 元素的实际含量和预测含量具有非常良好的对应关系,所建立的定量分析

模型非常可靠。利用上述 PLS 所建立的定量分析模型,计算 T1 类光谱预测集中 N 元素的含量,最终得到预测样品中 N 元素含量预测值与实际值的拟合曲线,如图 5 所示,其决定系数 $R^2 = 0.99$, RMSEP 为 0.4787。为了进一步得出每个样品中 N 元素含量真值与预测值之间的相对误差,在 T2 类光谱中计算了 20 个样品光谱数据中 N 元素含量真值与预测值的分布曲线和相对误差,如图 6 所示。

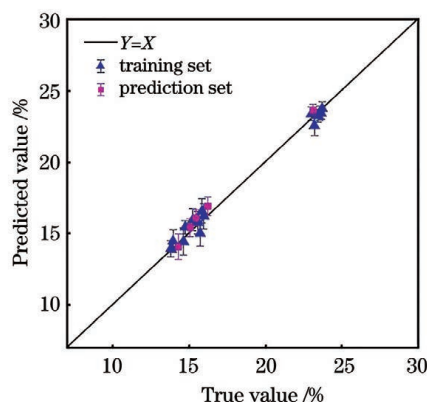


图 5 训练集和预测集中 N 元素含量预测值与真值的关系
Fig. 5 Relationship between the predicted value and the true value of N element content in training set and prediction set

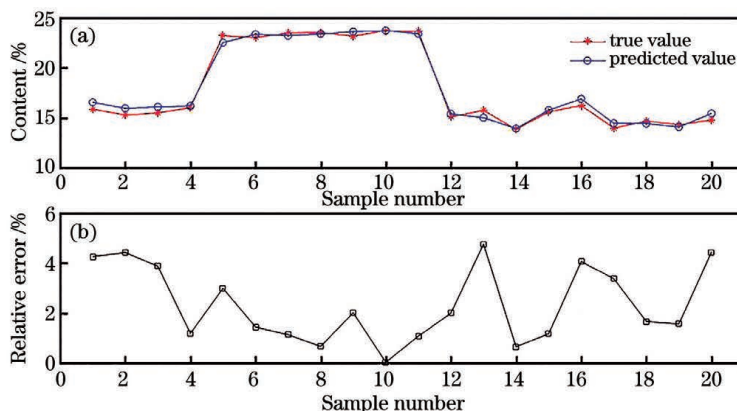


图 6 20 个样品中 N 元素含量真值与预测值的分布及预测的相对误差。(a)真值与预测值的分布曲线;(b)预测的相对误差
Fig. 6 Distribution of true and predicted values of N element content and relative error of prediction for 20 samples.

(a) Distribution of true and predicted values; (b) predicted relative error

其中图 6(a)给出了 N 元素含量真值与预测值的分布曲线,图 6(b)所示为 N 元素含量预测的相对误差,20 个样品的平均相对误差(ARE)为 2.33%。

重复上述过程,得到复合肥中 P 元素的建模效果和预测结果如图 7 所示,其训练集模型的决定系数 $R^2=0.98$,RMSEC 为 0.0979;预测集的 $R^2=0.98$,RMSEP 为 0.0706。图 8 所示为此 20 个样品中 P 元素含量真值与预测值的分布曲线和相对误差。图 8(a)给出了 20 个样品中 P 元素含量真值与预测值的分布曲线,图 8(b)则给出了 20 个样品 P 元素含量预测的相对误差。由图 8(b)可知,20 个样品的 ARE 为 0.70%。对 P 元素的建模-预测结果显示,P 元素的预测结果优于 N 元素的预测结果,原因可能是虽然 P 元素光谱的峰值较弱,但不

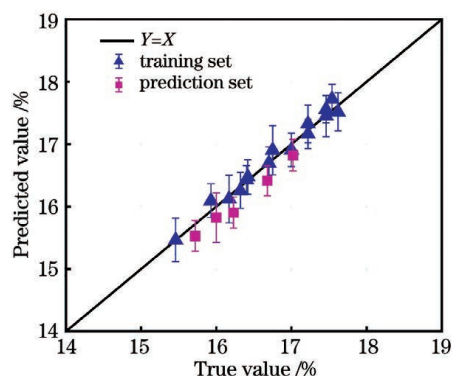


图 7 训练集和预测集中 P 元素含量预测值与真值的关系
Fig. 7 Relationship between the predicted value and the true value of P element content in training set and prediction set

易受环境气体影响。

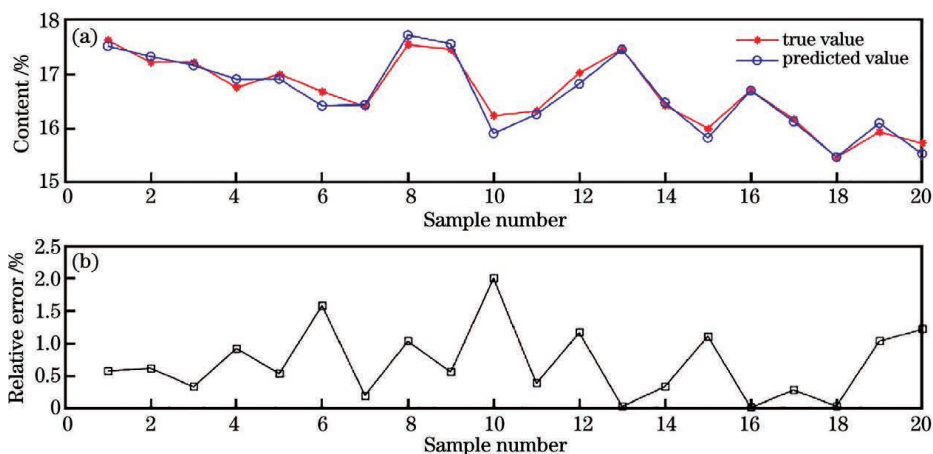


图 8 20 个样品中 P 元素含量真值与预测值的分布及预测的相对误差。(a)真值与预测值的分布曲线;(b)预测的相对误差
Fig. 8 Distribution of true and predicted values of P element content and relative error of prediction for 20 samples.

(a) Distribution curve of true and predicted values; (b) predicted relative error

图 9 给出了训练集和预测集中 K 元素含量预测值与实际值的拟合曲线,得到训练集模型的决定系数 $R^2=0.99$,RMSEC 为 0.3385;预测集中 $R^2=0.99$,RMSEP 为 0.0195。图 10(a)给出了 20 个样品中 K 元素含量真值与预测值的分布曲线,图 10(b)给出了 K 元素含量预测的相对误差,20 个样品的 ARE 为 3.00%。

根据以上过程,可以得到样本数据扩充后和未扩充时预测集中被测元素含量的 R^2 、RMSEP 和 ARE,如表 2 所示。从表 2 可以看出,样本数据扩充后,对 20 个复合肥样本中 N、P、K 元素的 ARE 均小于 3%,其中 P 的预测平均相对误差小于 1%。如果没有采取数据抽取方式进行样本数据扩充,在样本容量较少的情况下直接对全谱数据进行平均训练,那么模型的精度和鲁棒性都容易出现较大的波

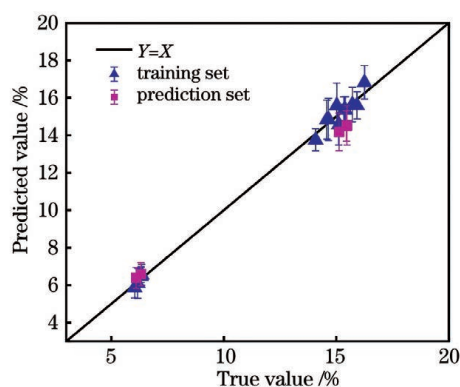


图 9 训练集和预测集中 K 元素含量预测值与真值的关系

Fig. 9 Relationship between the predicted value and the true value of K element content in training set and prediction set

动。当对每个样品随机抽取 1000 次数据来扩充样

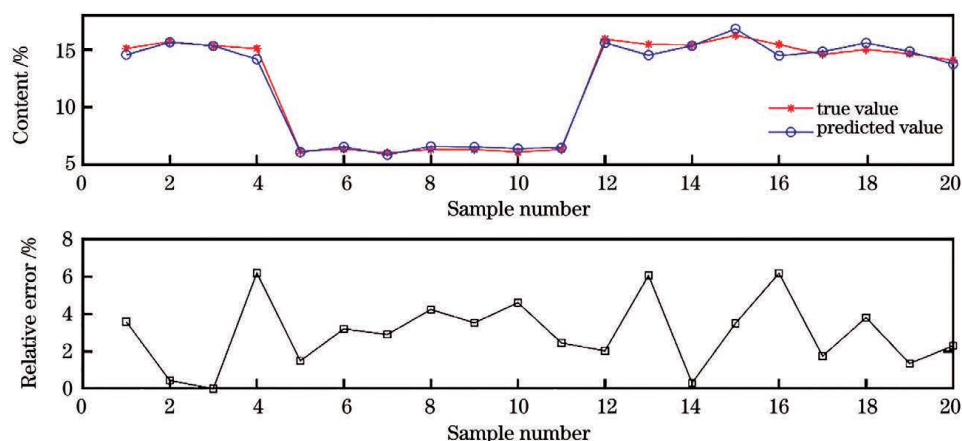


图 10 20 个样品 K 元素含量真值与预测值的分布及预测的相对误差。(a)真值与预测值的分布曲线;(b)预测的相对误差
Fig. 10 Distribution of true and predicted values of K element content and relative error of prediction for 20 samples.

(a) Distribution curve of true and predicted values; (b) predicted relative error

表 2 样本数据扩充后和未扩充时利用 PLS 模型预测被测元素含量的 R^2 、RMSEP 和 ARE 值

Table 2 Comparison of R^2 , RMSEP, and ARE values of the predicted content of measured elements using PLS model when the sample data are expanded and not expanded

Element	Origin			Data expansion		
	R^2	RMSEP	ARE / %	R^2	RMSEP	ARE / %
N	0.83	2.8461	18.40	0.99	0.4787	2.33
P	0.91	1.4048	13.40	0.98	0.0706	0.70
K	0.75	2.5092	14.43	0.99	0.0195	3.00

本光谱数据后,选取回归效果最好的一次进行建模,这样不仅提高了建模的精度,还提高了测量的准确性,与未扩充时相比,被测元素含量的平均相对误差均下降了 10% 以上。

3.3 模型的鲁棒性检验

为了验证模型的鲁棒性,随机去除 5 个样品数据,利用剩余的 15 个样品进行建模,并利用 PLS 定量分析模型对去除的 5 个样品进行 N、P、K 元素含量的真值预测。N、P、K 元素含量的预测结果及相对误差如表 3~5 所示。结果表明:N、P 含量的预

表 3 利用 PLS 模型测得的 N 元素含量预测值与真值的比较
Table 3 Comparison of predicted value and true value of N element content determined by PLS model

Removed sample number	True value / %	Predicted value / %	Relative error / %
2009-7	15.99	13.51	15.51
2009-16	23.48	26.28	23.36
2009-18	23.16	23.36	0.43
2010-40	13.28	12.89	6.73
2011-20	14.65	12.94	11.67

表 4 利用 PLS 模型测得的 P 元素含量预测值与真值的比较
Table 4 Comparison of predicted value and true value of P element content determined by PLS model

Removed sample number	True value / %	Predicted value / %	Relative error / %
2009-7	16.75	17.28	3.22
2009-16	16.40	16.56	0.99
2009-18	17.46	16.82	3.65
2010-40	16.42	17.78	8.33
2011-20	15.46	17.14	10.93

表 5 利用 PLS 模型测得的 K 元素含量预测值与真值的比较
Table 5 Comparison of predicted value and true value of K element content determined by PLS model

Removed sample number	True value / %	Predicted value / %	Relative error / %
2009-7	15.12	18.16	20.12
2009-16	6.04	3.24	46.34
2009-18	6.32	6.48	2.59
2010-40	15.40	17.97	16.74
2011-20	15.02	18.41	22.61

测值与实际值吻合较好,相对误差均在 12% 以下;K 含量的预测值与实际值吻合较差,相对误差大都在 20% 以上。从模型鲁棒性检验结果来看,PLS 结合上述数据抽取方式的方法也可以对同类化肥的未知样品进行高精度预测。

4 结 论

使用 PLS 定量分析方法建立回归模型,采用基于统计学原理的数据抽取方式扩充小样本量的复合肥光谱数据并建立预测模型,对复合肥样品中 N、P、K 元素的含量进行建模和预测。20 个样品中 N、P、K 元素含量预测的平均相对误差分别为 2.33%、0.70%、3.00%。P 元素的预测结果最优,N、K 元素的预测误差稍大,其原因可能是环境气体的干扰和元素谱线的波动。最后,采用随机去除 5 个样品数据的方法对所建模型进行鲁棒性检验,结果表明,N、P 元素含量的预测值与实际值吻合得较好,相对误差大多维持在 12% 以下。总之,采用基于统计学原理的数据抽取方式扩充样本光谱数据后,与未扩充时相比,被测元素含量的平均相对误差的降幅均在 10% 以上。综上所述,当样本容量较少时,可以采用所提的数据抽取方式建立训练集,并结合 PLS 定量分析模型进行回归建模,以有效提高测量的准确度。

参 考 文 献

- [1] Yang Q Q, Zhou C H. The test of potassium oxide on mixed and compound fertilizer by flame atomic absorption spectrometry[J]. Contemporary Chemical Industry, 2007, 36(2): 209-211.
杨清群, 周成洪. 火焰原子吸收光谱法测定复(混)合肥料中的钾[J]. 当代化工, 2007, 36(2): 209-211.
- [2] Huang Y F, Deng W L, Lin W Y. Determination of potassium and available phosphorus content in compound fertilizers by ICP-AES [J]. Journal of Guangxi Academy of Sciences, 2006, 22(S1): 435-436.
黄一帆, 邓卫利, 林文业. ICP 光谱法测定复混肥料中钾和有效磷的含量[J]. 广西科学院学报, 2006, 22(S1): 435-436.
- [3] Song L, Zhang H, Ni X Y, et al. Quantitative analysis of contents in compound fertilizer and application research using near infrared reflectance spectroscopy[J]. Spectroscopy and Spectral Analysis, 2014, 34(1): 73-77.
宋乐, 张红, 倪晓宇, 等. 基于 NIRS 技术的复合化肥成分定量分析及应用研究[J]. 光谱学与光谱分析, 2014, 34(1): 73-77.
- [4] Kumar A, Yueh F Y, Singh J P, et al. Characterization of malignant tissue cells by laser-induced breakdown spectroscopy[J]. Applied Optics, 2004, 43(28): 5399-5403.
- [5] Dong P K, Zhao S Y, Zheng K X, et al. Rapid identification of ginseng origin by laser induced breakdown spectroscopy combined with neural network and support vector machine algorithm [J]. Acta Physica Sinica, 2021, 70(4): 040201.
董鹏凯, 赵上勇, 郑柯鑫, 等. 激光诱导击穿光谱技术结合神经网络和支持向量机算法的人参产地快速识别研究[J]. 物理学报, 2021, 70(4): 040201.
- [6] Blevins L G, Shaddix C R, Sickafoose S M, et al. Laser-induced breakdown spectroscopy at high temperatures in industrial boilers and furnaces [J]. Applied Optics, 2003, 42(30): 6107-6118.
- [7] López-Moreno C, Palanco S, Laserna J J. Remote laser-induced plasma spectrometry for elemental analysis of samples of environmental interest [J]. Journal of Analytical Atomic Spectrometry, 2004, 19(11): 1479-1484.
- [8] Yang L, Yang H B, Zhang J K, et al. Review on determination of nutrients and heavy metals in soils and plants by laser-induced breakdown spectroscopy [J]. Laser & Optoelectronics Progress, 2020, 57(23): 230005.
杨柳, 杨海波, 张加康, 等. 激光诱导击穿光谱技术测定土壤和植物中养分及重金属的研究进展[J]. 激光与光电子学进展, 2020, 57(23): 230005.
- [9] Sha W, Li J T, Lu C P. Quantitative analysis of Mn in soil based on laser-induced breakdown spectroscopy optimization [J]. Chinese Journal of Lasers, 2020, 47(5): 0511001.
沙文, 李江涛, 鲁翠萍. 基于激光诱导击穿光谱技术寻优定量分析土壤中 Mn 元素[J]. 中国激光, 2020, 47(5): 0511001.
- [10] Song H S, Ma L Z, Zhu E G, et al. Plastic classification and recognition by laser-induced breakdown spectroscopy and GA-BP neural network [J]. Laser & Optoelectronics Progress, 2020, 57(15): 153002.
宋海声, 麻林召, 朱恩功, 等. 基于激光诱导击穿光谱与 GA-BP 神经网络的塑料分类识别[J]. 激光与光电子学进展, 2020, 57(15): 153002.
- [11] Farooq W A, Al-Mutairi F N, Khater A E M, et al. Elemental analysis of fertilizer using laser induced breakdown spectroscopy [J]. Optics and Spectroscopy, 2012, 112(6): 874-880.
- [12] Liao S Y, Wu X L, Li G H, et al. Quantitative analysis of P in fertilizer by laser-induced breakdown

- spectroscopy with multivariate nonlinear method[J]. Spectroscopy and Spectral Analysis, 2018, 38(1): 271-275.
- 廖索引, 吴先良, 李桂华, 等. 复合肥中磷元素的激光诱导击穿光谱多元非线性定量分析[J]. 光谱学与光谱分析, 2018, 38(1): 271-275.
- [13] Chen K, Lu J D. Determination of potassium concentrations in compound fertilizer with laser induced breakdown spectroscopy [J]. High Power Laser and Particle Beams, 2011, 23(2): 293-297. 陈凯, 陆继东. 复混肥中钾含量的激光诱导击穿光谱分析[J]. 强激光与粒子束, 2011, 23(2): 293-297.
- [14] Lu W Y, Lu J D, Yao S C, et al. Synchronization detection of nitrogen phosphorus potassium in compound fertilizer with laser induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2011, 38(10): 1008003. 卢伟业, 陆继东, 姚顺春, 等. 复合肥氮磷钾元素含量的激光诱导击穿光谱同步测量[J]. 中国激光, 2011, 38(10): 1008003.
- [15] Yao S C, Lu J D, Li J Y, et al. Multi-elemental analysis of fertilizer using laser-induced breakdown spectroscopy coupled with partial least squares regression [J]. Journal of Analytical Atomic Spectrometry, 2010, 25(11): 1733-1738.
- [16] Zhang B H, Ling P P, Sha W, et al. Univariate and multivariate analysis of phosphorus element in fertilizers using laser-induced breakdown spectroscopy [J]. Sensors, 2019, 19(7): 1727.
- [17] Zhang M, Yuan H. The PauTa criterion and rejecting the abnormal value[J]. Journal of Zhengzhou University of Technology, 1997, 18(1): 87-91. 张敏, 袁辉. 拉依达(PauTa)准则与异常值剔除[J]. 郑州工业大学学报, 1997, 18(1): 87-91.
- [18] Yang C R. Study of the spectral data processing in laser induced breakdown spectroscopy technology [D]. Beijing: Beijing Jiaotong University, 2014: 20-21. 杨崇瑞. 激光诱导击穿光谱数据处理方法研究[D]. 北京: 北京交通大学, 2014: 20-21.
- [19] Body D, Chadwick B L. Optimization of the spectral data processing in a LIBS simultaneous elemental analysis system[J]. Spectrochimica Acta B, 2001, 56(6): 725-736.
- [20] Wang Q Q, Huang Z W, Liu K, et al. Classification of plastics with laser-induced breakdown spectroscopy based on principal component analysis and artificial neural network model[J]. Spectroscopy and Spectral Analysis, 2012, 32(12): 3179-3182. 王茜茜, 黄志文, 刘凯, 等. 基于主成分分析和人工神经网络的激光诱导击穿光谱塑料分类识别方法研究[J]. 光谱学与光谱分析, 2012, 32(12): 3179-3182.
- [21] Yang Y W, Li C, Liu S, et al. Classification and identification of brands of iron ores using laser-induced breakdown spectroscopy combined with principal component analysis and artificial neural networks[J]. Analytical Methods, 2020, 12(10): 1316-1323.

Spectral Method for Predicting Nitrogen, Phosphorus, and Potassium in Small Amount of Compound Fertilizer

Ma Xiao¹, Li An¹, Wang Xianshuang¹, Kong Denan¹, Qiu Suling¹, He Yage¹,
Yin Yunsong¹, Liu Yufei², Liu Ruibin¹

¹ School of Physics, Beijing Institute of Technology, Beijing 100081, China;

² Bright-Ray Laser Technology (Changzhou) Co., Ltd., Changzhou, Jiangsu 213000, China

Abstract

Objective N, P, and K, the three main nutrient elements of compound fertilizer products, are essential nutrient elements for crop growth; therefore, the quality inspection of compound fertilizers is very important. Currently, the main methods for detection of compound fertilizer include flame atomic absorption spectrometry, inductively coupled plasma emission spectroscopy (ICP-AES), and near infrared reflectance spectroscopy (NIRS). The detection time of these methods is short. However, samples must be pretreated, thereby preventing real-time monitoring of the samples, and possibly reducing the accuracy of the measurements. Compared with the above technologies, laser-induced breakdown spectroscopy (LIBS) technology requires no sample pretreatment and is characterized by a green and safe detection process, short detection time, and real-time detection of all elements. This technology has been

used in various fields. When LIBS technology is combined with partial least squares to model and predict the nutrient elements of compound fertilizers, numerous samples are usually needed to improve the prediction accuracy of the model. The accuracy of prediction is generally low when the sample size is small. Therefore, improving the prediction accuracy for small sample sizes is important. In this article, we propose a data extraction method based on statistical principles to expand sample spectral data for small sample sizes, thereby improving the measurement accuracy.

Methods Twenty types of compound fertilizers, with N, P_2O_5 , and K_2O as the main components, were investigated in this work. A LIBS detection system was set up for collecting and analyzing the radiation spectrum of the plasma; then, we used a new proposed method of data extraction. Afterward, the N, P, and K elements in the compound fertilizer samples were modeled and predicted using a partial least square method combined with principal component analysis. The last 75 sets of spectra were used as T2 spectra to obtain the relative error between the true N, P, and K element content of each sample and the predicted content. To determine the robustness of the calibration model, we randomly removed five samples for modeling and predicting the N, P, and K content of these samples.

Results and Discussions The spectral data were preprocessed. Compared with the original spectrum, the background of the preprocessed spectrum is eliminated, and the relative intensity between channels changes, but the relative intensity of each spectral line between the channels remains unchanged (Fig. 3). The N, P, and K elements are modeled and predicted after the preprocessing steps. The coefficients of determination for N, P, and K element content modeling in the training set are 0.99, 0.98, and 0.99, respectively, and the root mean square errors are 0.4309, 0.0979, and 0.3385, respectively; moreover, the coefficients of determination obtained for the fitting curves of the predicted and true values of the N, P, and K element contents in the prediction set are 0.99, 0.98, and 0.99, respectively, and the root mean square errors are 0.4787, 0.0706, and 0.0195, respectively (Fig. 5, Fig. 7, and Fig. 9). The average relative errors between the true and predicted values of the N, P, and K element contents of 20 samples obtained from T2 spectra are 2.33%, 0.70%, and 3%, respectively (Fig. 6, Fig. 8, and Fig. 10). After the sample data are expanded, the average predicted relative error (ARE) values of N, P, and K elements in the 20 compound fertilizer samples are all $<3\%$, and ARE of P is $<1\%$. Compared with the unexpanded condition, the average relative error of the measured element content dropped by more than 10% (Table 2). The robustness of the model is determined. The predicted values of N and P content agree well with the true values, the relative errors are all below 12%, and the relative errors of K element are mostly above 20% (Table 3, Table 4, and Table 5).

Conclusions In this work, the partial least square quantitative analysis method is used to establish a regression model, and a data extraction method based on statistical principles is used to expand the small sample size of the compound fertilizer spectral data. The N, P, and K element content of a compound fertilizer sample is modeled and predicted. The average relative errors of the content prediction in 20 samples are 2.33%, 0.70%, and 3.00%, respectively. Further, the robustness of the model is determined by randomly removing the data of five samples. The results reveal that the predicted values of N and P element content concur with the actual values, and the relative errors are mainly below 12%. Thus, after using the data extraction method based on statistical principles to expand the sample spectrum data, the average relative error of the measured element content is reduced by more than 10% compared with the unexpanded time. The experimental results show that when the sample size is small, the accuracy of the measurement can be improved using this new data extraction method combined with the partial least square quantitative analysis model for regression modeling.

Key words spectroscopy; laser-induced breakdown spectroscopy; compound fertilizer; partial least squares; small sample size

OCIS codes 300.2530; 300.6210; 300.6365