

doi:10.6041/j.issn.1000-1298.2014.S0.045

生菜叶中磷含量的光谱定量分析*

高洪燕 毛罕平 张晓东

(江苏大学现代农业装备与技术教育部重点实验室, 镇江 212013)

摘要:为快速、准确检测生菜叶内的磷含量,提出了应用光谱技术结合化学计量法无损检测生菜叶内磷含量的方法。通过获取不同施磷量下生菜叶片于波长350~2500 nm处的反射光谱,对光谱数据进行5点平滑和一阶导数变换后,利用联合区间偏最小二乘算法(siPLS)提取了与生菜叶磷元素相关的4个特征波段,即950~1070 nm, 1430~1549 nm, 1906~2025 nm和2144~2263 nm。进一步利用连续投影算法(SPA)对全光谱波段和4个特征波段进行特征波长提取,分别筛选出变量63个和25个。分别对4个特征波段、63个和25个特征波长进行主成分降维,当主成分数分别为7、5和4时,隐含层神经元数分别为7、5和3时,建立了siPLS+BPANN, SPA+BPANN, siPLS+SPA+BPANN生菜叶磷含量检测模型。研究结果表明:siPLS+SPA+BPANN模型的预测结果优于其他模型,验证集相关系数为0.911,验证均方根误差为479 mg/kg。

关键词:生菜叶 磷 光谱分析 联合区间偏最小二乘 连续投影算法 BP神经网络

中图分类号: O657.33; S636.2 **文献标识码:** A **文章编号:** 1000-1298(2014)S0-0276-05

引言

生菜(*Lactuca sativa*)是叶菜类作物的典型代表,我国是生菜的主要生产国,总产量约占世界总量的1/2。磷(P)是维持作物生长所需营养元素之一,是核酸、核蛋白、磷脂、ATP和辅酶等化合物的重要组成部分^[1]。磷在作物体内以多种形式参与光合作用、呼吸作用、酶活性调节和信号转导等多种重要生理生化过程。例如:磷能促进作物氮素代谢过程,磷是氨基转移酶的组成成分之一,这种酶的活性基为磷酸吡哆醛素,它能促进氨基化、脱氨基和氨基转移作用的进行。磷还能加强有氧呼吸作用中糖类的转化,有利于各种有机酸(如 α -酮酸)和ATP的形成,其中 α -酮酸可作为氨的受体形成氨基酸,ATP可为氨基酸和蛋白质合成提供能量^[2]。

传统的作物磷含量的检测方法主要是化学测定法,此方法虽然检测度较高,但分析测定过程存在高耗性、繁冗复杂性和时滞性等缺点,同时需要破坏作物样本,不适用于大范围、全面和快速的营养诊断。与传统化学分析方法相比,光谱技术具有快速、低成本、测量方便和非破坏性等优点^[3]。目前国内外对基于光谱技术的作物磷元素无论是定性还是定量的

研究均较少^[4],其中一个原因是磷在作物体内含量较少,仅为氮素含量的几分之一^[5-6]。

本文通过获取不同施磷量下生菜叶的反射光谱,在对光谱数据进行平滑去噪及求导变换后,利用联合区间偏最小二乘(siPLS)算法提取与生菜磷元素相关的特征波段;再利用连续投影算法(SPA)提取特征波长;然后对所提取的波段和波长进行主成分降维,将提取的主成分因子作为BP神经网络(BPANN)的输入,建立多种形式的生菜磷含量定量检测模型,比较得到最佳检测模型。

1 材料与方 法

1.1 样本培育

供试材料为意大利全年耐抽苔半结球生菜。在生长到“五叶一心”时,将长势状况相似的穴盘苗移植到花盆中定植,花盆内径为20 cm。试验中采用营养液加珍珠岩的栽培模式,根据山崎营养液配方^[7],将样本分为5个处理,分别为标准配方中P浓度的25%、50%、75%、100%和125%(均为质量分数)。样本培育过程中早晚各供液一次,以保证生菜始终在固定的磷浓度环境中生长,每个处理水平15株,共75个样本。

收稿日期:2014-07-31 修回日期:2014-08-20

*国家自然科学基金资助项目(61233006)、“十二五”国家科技支撑计划资助项目(2014BAD08B03)、江苏省普通高校研究生科研创新计划资助项目(CXZZ13_0690)和江苏高校优势学科建设工程资助项目(苏政办发(2011)6号)

作者简介:高洪燕,博士生,主要从事作物生长信息智能监测研究,E-mail:ghy06@126.com

通讯作者:毛罕平,教授,博士生导师,主要从事现代农业装备和设施农业环境控制技术研究,E-mail:maohp@ujs.edu.cn

1.2 光谱数据获取

光谱测量设备采用美国 ASD 公司生产的 FieldSpec® 3 型手持便携式光谱分析仪。该仪器光谱测量波长范围 350 ~ 2 500 nm; 在波长 350 ~ 1 000 nm 光谱区采样间隔为 1.4 nm, 分辨率为 3 nm; 在波长 1 000 ~ 2 500 nm 光谱区采样间隔为 2 nm, 分辨率为 10 nm。将采集于同一级层的新鲜叶片置于黑布上, 以减少底面对光谱测量的影响。光谱仪的视场角为 25°, 光纤垂直固定于叶片上方 50 mm 处, 光源与叶片水平方向呈 45° 角。试验前先进行标准白板标定, 以消除环境因素引起的误差, 每个样本采集 5 个光谱, 取平均值作为采样光谱值。

1.3 磷含量测定

样本磷含量测定采用比色法, 叶片在浓硫酸和双氧水的作用下, 将不同形态的磷转化为磷酸盐, 并转入溶液中。采用英国 SEAL 公司生产的 AA3 型连续流动分析仪测定消化液中的磷含量^[8], 单位为 mg/kg。

2 结果与分析

2.1 生菜叶磷含量分析

试验共采集样本 75 个, 随机分为 2 组, 分别为校正集 (50 个) 和验证集 (25 个), 磷元素在各组样本中的含量统计如表 1 所示, 可以看出, 2 个样本集无明显差异, 分配方式合理。

表 1 校正集和验证集样本中磷含量

Tab. 1 Phosphorus content in calibration set and

样本集	样本数	prediction set			mg/kg
		最小值	最大值	均值	
校正集	50	3 568	7 694	5 707	1 068
验证集	25	3 726	7 546	5 736	1 049

2.2 光谱数据预处理

光谱数据中除包含与被测物有关的信息外, 同时还包含仪器以及外界因素所带来的干扰信号 (图 1)。光谱数据的预处理采用 5 点平滑法, 以消除光谱曲线存在的微弱噪声^[9]。然后对平滑后的反射光谱进行一阶导数变换, 消除光谱的基线偏移、漂移和背景干扰, 并增强光谱信号, 提高分辨率^[10-12] (图 2)。

2.3 提取特征波段

联合区间偏最小二乘算法 (siPLS) 是间隔偏最小二乘算法 (iPLS) 的一种改进算法, 该方法将整个光谱范围划分为 n 个等宽的子区间, 联合其中 m 个区间进行 PLS 建模, 共有 C_n^m 种组合方式, 当 PLS 模型的交互验证均方根误差 (RMSECV) 最小时, 所对应的子区间为入选的最优联合子区间。此方法的计

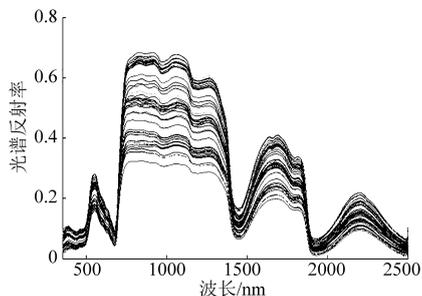


图 1 生菜叶片反射光谱

Fig. 1 Reflectance spectra of lettuce leaves

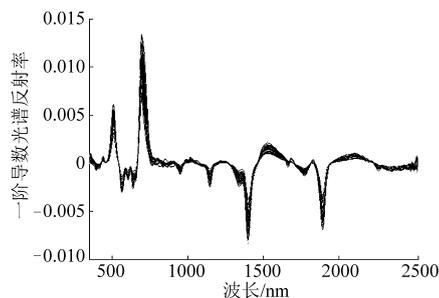


图 2 5 点平滑后生菜叶片一阶导数反射光谱

Fig. 2 First-order derivative spectra of lettuce leaves after five points smoothing

算量非常大, m 一般为 2、3 或 4^[13-14]。RMSECV 计算式为

$$D_{\text{RMSECV}} = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{y}_{ci} - y_{ci})^2} \quad (1)$$

式中 y_{ci} ——校正集中第 i 个样本的实测值

\hat{y}_{ci} ——校正集中剔除第 i 个样本后, 用余下样本建立的检测模型中第 i 个样本的预测值

n_c ——校正集的样本数

由于尚不能从理论上确定子区间的划分数和参与建模的联合子区间数, 故本研究将整个光谱划分为 10 ~ 25 个子区间, 并建立了联合 2、3 和 4 个子区间的 PLS 模型, 由表 2 可知, 当全光谱区域被划分为 18 个子区间时, 联合 [6 10 14 16] 4 个子区间后, 所建模型最优, 此时得到的 RMSECV 最小。4 个子区间分别对应波长 950 ~ 1 070 nm、1 430 ~ 1 549 nm、1 906 ~ 2 025 nm 和 2 144 ~ 2 263 nm (图 3)。

由图 3 可知, 与磷相关的特征波段集中在近红外区域, 这主要是由于磷在作物体内参与淀粉和糖类的代谢过程^[5,15]。

2.4 连续投影算法提取特征波长

连续投影算法 (SPA) 是通过计算光谱矩阵中某一波长对其他波长的投影, 以投影信息量来选择特征波长。该方法能够有效地解决波长间的共线性问题, 降低了模型的复杂程度, 提高了建模效率^[16]。

表2 不同子区间下 siPLS 模型结果

Tab.2 Results of siPLS model with different number of subintervals

区间数	联合区间	计算模型数	主因子数	联合子区间	RMSECV
10	2	34	7	[5 8]	449
11	3	165	7	[6 8 9]	434
12	4	495	7	[4 6 9 10]	446
13	3	286	6	[6 9 11]	445
14	3	364	8	[4 7 12]	406
15	4	1 365	7	[5 7 12 13]	416
16	4	1 820	9	[5 9 11 13]	410
17	4	2 380	6	[6 8 14 15]	421
18	4	3 060	7	[6 10 14 16]	375
19	2	171	7	[9 15]	416
20	3	1 140	7	[11 14 16]	424
21	4	5 985	10	[7 9 10 16]	420
22	4	7 315	6	[11 15 17 19]	405
23	4	8 855	8	[1 6 19 20]	411
24	4	10 626	9	[10 12 16 19]	505
25	3	2 300	6	[12 17 20]	497

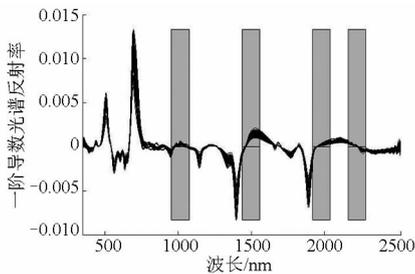


图3 siPLS 优选的最佳联合子区间

Fig.3 Optimal spectra regions by siPLS

本文利用 SPA 分别对全光谱波段和 siPLS 所提取的 4 个特征波段进行特征波长提取。根据校正集 RMSECV 选出最优波长组合,运行结果如图 4 所示,分别筛选出特征波长 63 和 25 个。

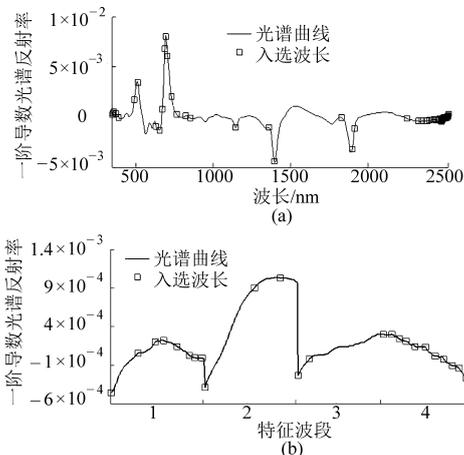


图4 连续投影算法筛选最优组合波长结果

Fig.4 Results of optimal wavelengths select by SPA

(a) 全光谱波段 (b) 4 个特征波段

2.5 BPANN 模型建立

BPANN 是利用非线性可微函数进行权值训练

的反向传播网络^[17],该网络拓扑具有 3 层结构,即输入层、隐含层和输出层。研究中分别对 siPLS 提取的 4 个特征波段,SPA 提取的 63 个特征波长和 siPLS + SPA 提取的 25 个特征波长进行建模分析。在利用 BPANN 建模之前,首先对光谱数据进行主成分分析 (PCA) 降维,用压缩的主成分作为 BPANN 的输入。PCA 通过线性变换保留方差大、含信息多的分量,丢掉信息量少的方向,从而降低数据的维数^[18-19]。在文中,对特征波段和特征波长进行主成分分析,将提取的主成分作为网络的输入,最佳的主成分因子数通过校正模型的交互验证进行优化,由所建模型的 RMSECV 值确定。其中隐含层的传递函数为 tansig,输出层的传递函数为 purelin,学习速率为 0.01,目标函数误差为 0.001,最大训练 1 000 次。最终当主成分数为 7、5 和 4 时,隐含层神经元数为 7、5 和 3 时,RMSECV 最小。利用训练好的模型对验证集样本进行预测,结果如表 3 所示。

表3 不同模型的结果

Tab.3 Results of different models

模型	RMSECV/ ($\text{mg}\cdot\text{kg}^{-1}$)	R_c	RMSEP/ ($\text{mg}\cdot\text{kg}^{-1}$)	R_p
siPLS	375	0.905	624	0.866
siPLS + BPANN	338	0.927	556	0.891
SPA + BPANN	573	0.842	821	0.803
siPLS + SPA + BPANN	314	0.935	479	0.911

注: R_c 为校正集相关系数, RMSEP 为验证均方根误差, R_p 为验证集相关系数。

2.6 模型对比

由表 3 可知, siPLS + SPA + BPANN 模型预测结果优于其他模型 (R_p 为 0.911, RMSEP 为 479 mg/kg)。这主要是由于:从特征波长提取方面考虑,由于 siPLS 提取的特征波段中仍然存在一定的共线性变量,而 SPA 具有很强的消除共线性能力,虽然 siPLS + SPA + BPANN 模型预测精度上相较于 siPLS + BPANN 模型提高较小,但模型复杂程度大大降低。从建模算法上考虑,施肥量的变化会导致作物体内发生复杂的动态的化学变化,故光谱特征与磷含量之间存在着一定的非线性关系,PLS 算法是一种多元线性回归的建模方法,而 BPANN 则是一种典型的非线性算法,且具有较强的自适应能力和信息综合能力,所以 BPANN 模型的预测结果优于 PLS 模型。

3 结束语

快速、精确地检测作物体内磷含量是磷肥精确调控的重要基础。本文获取了不同施磷量下生菜叶

的反射光谱,对反射光谱进行5点平滑处理和一阶导数变换,用以消除光谱的微弱噪声、基线偏移、漂移和背景干扰,并增强光谱信号。利用 siPLS 算法将 2 151 个波长点划分为 18 个子区间,当联合 [6 10 14 16] 4 个子区间时,所建的 PLS 模型最优,4 个子区间分别对应波长 950 ~ 1 070 nm、1 430 ~ 1 549 nm、1 906 ~ 2 025 nm 和 2 144 ~ 2 263 nm。利用 SPA 分别对全光谱波段和 siPLS 提取的 4 个联合区间进行

特征波长提取,入选波长为 63 和 25 个。对 siPLS 提取的 4 个特征波段,SPA 提取的 63 个特征波长和 siPLS + SPA 模型提取的 25 个特征波长进行主成分分析,建立基于主成分的 BPANN 模型,根据建立好的模型对验证集样本进行预测,研究结果表明 siPLS + SPA + BPANN 模型可以很好地预测生菜叶的磷含量,为合理调控磷肥施用量提供了依据。

参 考 文 献

- 1 张喜杰,李民赞. 基于反射光谱的温室黄瓜叶片磷素含量分析与预测[J]. 光谱学与光谱分析,2008,28(10): 2404 - 2408. Zhang Xijie, Li Minzan. Analysis and estimation of the phosphorus content in cucumber leaf in greenhouse by spectroscopy [J]. Spectroscopy and Spectral Analysis, 2008, 28(10): 2404 - 2408. (in Chinese)
- 2 马磊,梅凤娴,郑少玲,等. 不同氮磷钾水平对生菜产量及体内养分的影响[J]. 仲恺农业技术学院学报,2006,19(4): 13 - 16. Ma Lei, Mei Fengxian, Zheng Shaoling, et al. Effects of different N, P and K levels on yield and nutrient content in Lettuce [J]. Journal of Zhongkai University of Agriculture and Technology, 2006, 19(4): 13 - 16. (in Chinese)
- 3 郝勇,陈斌. 茶叶中低含量氨基酸近红外光谱定量分析模型研究[J]. 农业机械学报,2014,45(6): 216 - 220. Hao Yong, Chen Bin. Quantitative determination of low amino acid contents in tea by using near-infrared spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(6): 216 - 220. (in Chinese)
- 4 黄双萍,洪添胜,岳学军,等. 基于高光谱的柑橘叶片磷含量估算模型实验[J]. 农业机械学报,2013,44(4): 202 - 207, 195. Huang Shuangping, Hong Tiansheng, Yue Xuejun, et al. Hyperspectral estimation model of total phosphorus content for citrus leaves [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(4): 202 - 207, 195. (in Chinese)
- 5 Ramoelo A, Skidmore A K, Schlerf M, et al. Savanna grass nitrogen to phosphorous ratio estimation using field spectroscopy and the potential for estimation with imaging spectroscopy [J]. Journal of Applied Earth Observation and Geoinformation, 2013, 23: 334 - 343.
- 6 Kokaly R F, Asner G P, Ollinger S V, et al. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies [J]. Remote Sensing of Environment, 2009, 113 (Supp. 1): S78 - S91.
- 7 王正银. 蔬菜营养与品质[M]. 北京: 科学出版社, 2009: 36 - 37.
- 8 贝美容,罗雪华,杨红竹. AA3 型连续流动分析仪(CFA)同时测定橡胶叶全氮、全磷、全钾的方法研究[J]. 热带作物学报, 2011, 32(7): 1258 - 1264. Bei Meirong, Luo Xuehua, Yang Hongzhu. Simultaneous determination of nitrogen, phosphorus and potassium in rubber leaf samples by AA3 continuous flow analyzer (CFA) [J]. Chinese Journal of Tropical Crops, 2011, 32(7): 1258 - 1264. (in Chinese)
- 9 蒋金豹,李一凡,郭海强,等. 不同病害胁迫下大豆的光谱特征及识别研究[J]. 光谱学与光谱分析,2012,32(10): 2775 - 2779. Jiang Jinbao, Li Yifan, Guo Haiqiang, et al. Spectral characteristics and identification research of soybean under different disease stressed [J]. Spectroscopy and Spectral Analysis, 2012, 32(10): 2775 - 2779. (in Chinese)
- 10 Zhao D, Reddy K R, Kakani V G, et al. Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum [J]. European Journal of Agronomy, 2005, 22(4): 391 - 403.
- 11 Abdel-Rahman E M, Ahmed F B, van den Berg M. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy [J]. International Journal of Applied Earth Observation and Geoinformation, 2010, 12(Supp. 1): S52 - S57.
- 12 Wang Fumin, Huang Jingfeng, Wang Yuan, et al. Estimating nitrogen concentration in rape from hyperspectral data at canopy level using support vector machines [J]. Precision Agriculture, 2013, 14(2): 172 - 183.
- 13 McGovern C M, Weeranatanaphan J, Downey G, et al. The application of near infrared spectroscopy to the measurement of bioactive compounds in food commodities [J]. Journal of Near Infrared Spectroscopy, 2010, 18(2): 87 - 111.
- 14 Muller A L H, Picolo R S, Mello P D, et al. Total sulfur determination in residues of crude oil distillation using FT-IR/ATR and variable selection methods [J]. Spectrochimica Acta Part A-Molecular and Biomolecular, 2012, 89: 82 - 87.
- 15 Knox N M, Skidmore A K, Prins H H T, et al. Remote sensing of forage nutrients: combining ecological and spectral absorption feature data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 72: 27 - 35.
- 16 欧阳爱国,谢小强,刘燕德. 苹果可溶性固形物近红外在线光谱变量优选[J]. 农业机械学报,2014,45(4): 220 - 225. Ouyang Aiguo, Xie Xiaoqiang, Liu Yande. Selection of NIR variables for online detecting soluble solids content of apple [J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(4): 220 - 225. (in Chinese)
- 17 郑立华,李民赞,潘雯,等. 基于近红外光谱技术的土壤参数 BP 神经网络预测[J]. 光谱学与光谱分析,2008,28(5): 1160 - 1164.

- Zheng Lihua, Li Minzan, Pan Luan, et al. Estimation of soil organic matter and soil total nitrogen based on NIR spectroscopy and BP neural network [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(5): 1160 - 1164. (in Chinese)
- 18 van der Maaten L J, Postma E O, van den Herik H J. Dimensionality reduction: a comparative review [R]. Ticc, Tilburg University, Ticc TR2009 - 005.
- 19 Cozzolino D, Cynkar W U, Shah N, et al. Multivariate data analysis applied to spectroscopy: potential application to juice and fruit quality [J]. *Food Research International*, 2011, 44(7): 1888 - 1896.

Quantitative Determination of Phosphorus in Lettuce Leaf Using Spectroscopy

Gao Hongyan Mao Hanping Zhang Xiaodong

(Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China)

Abstract: In order to evaluate comprehensively lettuce leaves phosphorus content and precisely control phosphorus fertilizer quantity, the specific aim of this study was to attempt a strategy for measurement of phosphorus content integrating spectroscopy together with synergy interval partial least square (siPLS) and back propagation artificial neural network (BPANN). Leaves reflectance was acquired with a Fieldspec 3 spectroradiometer that provides measurements in the 350 ~ 2 500 nm spectral range, and then five points smoothed and first-order derivative transform were used to eliminate noise effects. siPLS was used to search for the optimal spectral intervals, which corresponded to 950 ~ 1 070 nm, 1 430 ~ 1 549 nm, 1 906 ~ 2 025 nm and 2 144 ~ 2 263 nm. 63 wavelengths were selected from 350 ~ 2 500 nm and 25 wavelengths were selected from four optimal intervals by successive projections algorithm (SPA). Principal component analysis (PCA) was implemented on the spectra intervals or variables, finally 7, 4 and 5 PCs were obtained. The siPLS + BPANN, SPA + BPANN, siPLS + SPA + BPANN models were achieved when the number of neurons in the hidden layer was 7, 5 and 3. It was conclude that spectroscopy combined with siPLS + SPA + BPANN were feasible to measure phosphorus content in lettuce, which had better performance than others model, correlation coefficient for the prediction set was 0.911, root mean square error of the prediction set was 479 mg/kg.

Key words: Lettuce leaf Phosphorus Spectroscopic analysis Synergy interval partial least square Successive projections algorithm Back propagation artificial neural network