

doi:10.6041/j.issn.1000-1298.2013.03.016

基于 MC - UVE 的土壤碱解氮和速效钾近红外光谱检测*

刘雪梅^{1,2} 柳建设¹

(1. 东华大学环境科学与工程学院, 上海 201620; 2. 华东交通大学土木建筑学院, 南昌 330013)

摘要: 应用可见/短波近红外光谱分析测量土壤碱解氮和速效钾含量。为了提高该分析方法的预测精度, 消除无信息建模变量对模型稳定性的影响, 原始光谱平滑后采用蒙特卡罗无信息变量消除方法 (MC - UVE) 对土壤碱解氮和速效钾的建模变量进行筛选, 应用偏最小二乘方法 (PLS) 建立校正模型。对于碱解氮模型, 采用 MC - UVE PLS 方法, 建模变量减少为 210, 相关系数和预测均方差分别为 0.84 和 17.1 mg/kg。对于速效钾的预测模型, 采用 MC - UVE 方法后, 建模变量减少为 150, 模型的预测相关系数为 0.76, 预测均方根误差为 15.4 mg/kg。

关键词: 土壤 碱解氮 速效钾 近红外光谱 蒙特卡罗无信息变量消除

中图分类号: O657.33; S158.2 文献标识码: A 文章编号: 1000-1298(2013)03-0088-04

Near-infrared Spectroscopy Determination of Soil Available N and Available K Based on MC - UVE Method

Liu Xuemei^{1,2} Liu Jianshe¹

(1. College of Environmental Science and Engineering, Donghua University, Shanghai 201620, China

2. School of Civil Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: Visible/near-infrared spectroscopy (Vis/NIRS) was investigated for determination of soil properties, namely, available nitrogen (N) and available potassium (K). In order to improve the predictive precision and eliminate the influence of uninformative variables for model robustness, Monte Carlo uninformative variables elimination (MC - UVE) methods were proposed for variable selection in available N and available K NIR spectral modeling. Partial least squares (PLS) models analysis was implemented for calibration models. The modeling variable number was reduced to 210 from 751 for available N calibration model and 150 for available K calibration model. The performance of the model was evaluated by the correlation coefficient (R), RMSEP. The optimal MC - UVE PLS models were achieved, and R , RMSEP were 0.84, 17.1 mg/kg for N and 0.76, 15.4 mg/kg for K, respectively.

Key words: Soil Available N Available K Near-infrared spectroscopy Monte Carlo uninformative variables elimination

引言

探测土壤的碱解氮和速效钾含量是了解土壤肥力的重要途径, 土壤碱解氮和速效钾含量信息的获取对于农业和环境等方面的研究很有意义^[1]。近红外光谱 (NIRS) 分析方法具有快速、高效、无损和

适合在线分析等优点, 近年来, 采用近红外光谱技术获取土壤养分信息, 成为国内外学者研究的重点^[2-8]。

近红外光谱分析过程中, 变量筛选是重要的数据前处理方法, 目前已有许多相关报道^[9-12], 其中无信息变量消除 (Uninformative variables elimination,

收稿日期: 2012-06-08 修回日期: 2012-07-07

* 国家自然科学基金资助项目 (41073060) 和江西省科技支撑计划资助项目 (2010EHB02000、2009AE01603)

作者简介: 刘雪梅, 博士生, 华东交通大学副教授, 主要从事环境工程和农产品无损检测研究, E-mail: lyumu@163.com

通讯作者: 柳建设, 教授, 博士生导师, 主要从事生物及生化工程研究, E-mail: liujianshe@dhu.edu.cn

UVE)偏最小二乘(Partial least squares, PLS)方法是基于PLS回归系数提出的一种波长选择方法。该方法通过一定的变量筛选标准,例如,引入稳定性值来评价模型中每个变量的可靠性,从而来决定每个变量的取舍。该方法已被广泛应用于光谱变量的选择,与其他波长选择方法相比,蒙特卡罗无信息变量消除(Monte Carlo uninformative variables elimination, MC-UVE)PLS取得了满意的效果^[13]。

本文采用近红外光谱分析方法对土壤碱解氮和速效钾含量进行定量分析,探讨建模变量选择方法对土壤碱解氮和速效钾含量指标预测模型的精度和稳定性的影响,以实现土壤碱解氮和速效钾含量的快速无损分析。

1 材料和方法

1.1 收集土壤样本

土壤样本分别采集于江西省南昌县昌东镇三联村砖红土、遂川县枚江乡张塘村黄土和东江村水稻土各60个土样,为了消除土壤粒径对预测碱解氮和速效钾含量的影响,使用研钵将所有样本磨碎,并分别通过孔径0.5 mm的筛子,得到粒径大致相同的样本。土壤碱解氮含量采用1 mol/L NaOH碱解扩散法测定,速效钾含量采用火焰光度计测定^[14]。土壤样品参数如表1所示。

表1 土壤样品理化值统计

参数	最小值	最大值	均值	标准偏差
碱解氮质量比	60	212	168	31.20
速效钾质量比	25	89	66	15.86

1.2 检测仪器及参数设置

光谱测定使用ASD FieldSpec便携式光谱仪(Analytical Spectral Devices, Inc, USA),测定范围325~1 075 nm,光谱采样间隔:1 nm,扫描时间:100 ms/次,检测距离3 cm,扫描次数:10次,探头视场角为5°。

1.3 光谱采集

土壤样品放入直径为80 mm、厚度为15 mm的玻璃培养皿中,采用漫反射方式进行样品光谱采样。光源入射角与垂直方向成45°,光源、土壤样品和光谱仪位于同一平面且光谱仪探头与垂直方向成45°,光源至土壤样品直线距离为15 cm。测试中仪器噪声、样本粒径大小引起的散射会影响有效光谱信息的分析和提取,为保证光谱数据具有代表性,采集一个土壤样本光谱保存3次,将保存的3次数据求平均值,将平均值作为土壤样本最终的光谱数据。

土壤样品的近红外漫反射光谱如图1所示。

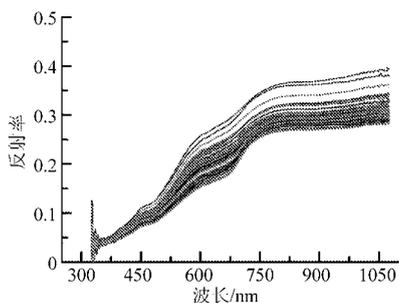


图1 土壤样品的近红外漫反射光谱

Fig.1 Diffuse reflectance NIRS of soil samples

1.4 偏最小二乘回归模型的建立

PLS方法用于回归模型的建立,模型的因子数采用留一法交叉验证进行选取。模型分别采用完全交叉验证和外部验证对其性能进行评价,由相关系数(R)、验证均方差(RMSECV)和预测均方差(RMSEP)进行评价。在建模分析中, R 偏高为好,RMSECV和RMSEP偏小为好,建模方法的选取一定要适当,以避免出现过拟合现象,又保证预测具有较高的精度,RMSEP要小。对原始光谱进行SG平滑预处理后应用PLS建立校正模型,模型结果如表2所示。通过比较分析可知,原始光谱进行平滑预处理后建立的PLS回归模型碱解氮质量比的 R 和RMSECV分别为0.78和18.39 mg/kg,速效钾质量比的 R 和RMSECV分别为0.69和16.81 mg/kg。

表2 应用PLS模型预测建模集和验证集结果

Tab.2 Prediction results of soil properties in calibration and validation set by PLS models

参数	预处理	建模集		验证集	
		R	RMSECV/ mg·kg ⁻¹	R	RMSEP/ mg·kg ⁻¹
碱解氮质量比	原始光谱	0.76	17.69	0.75	18.39
	SG平滑	0.78	18.39	0.76	18.46
速效钾质量比	原始光谱	0.65	15.61	0.64	16.06
	SG平滑	0.69	16.81	0.65	16.97

2 结果和讨论

2.1 主成分因子数的选择

图2表示土壤碱解氮和速效钾模型中PLS和MC-UVE PLS两种方法的RMSECV和RMSEP随主成分数的变化图。从图中可以清楚地看出,随着主成分数的增加,均方差随之减小,当主成分数增加到14以后,均方差值均趋于稳定,对于土壤碱解氮和速效钾的模型,因子数选14时即满足计算要求。

2.2 基于MC-UVE PLS方法模型的变量选择

图3给出了波长325~1 075 nm范围内,碱解氮和速效钾通过MC-UVE方法得到的每个变量的稳

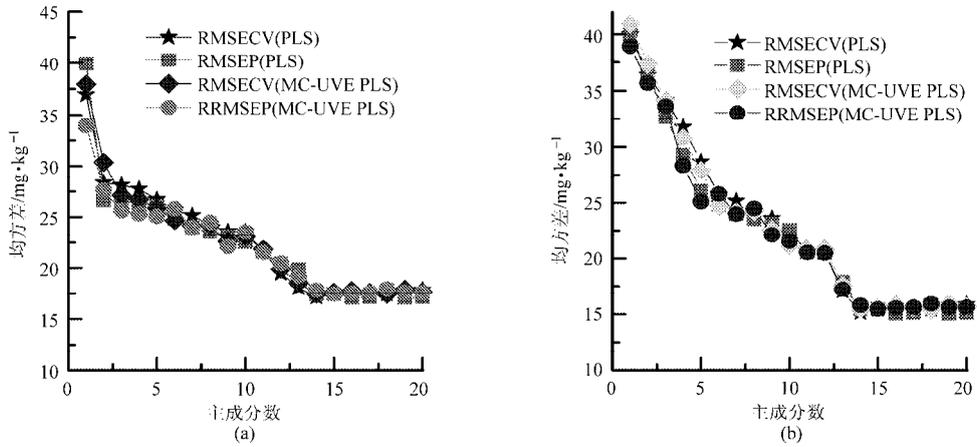


图2 模型的均方差随因子数的变化图

Fig. 2 Variation of RMSE with number of factors by PLS and MC-UVE PLS methods

(a) 碱解氮 (b) 速效钾

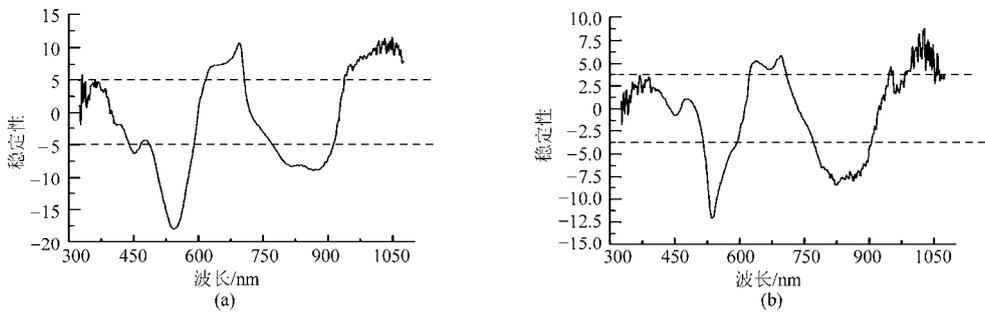


图3 建模变量的稳定性分布图

Fig. 3 Stability distribution of variables for prediction by MC-UVE method

(a) 碱解氮 (b) 速效钾

定性值。图中,虚线表示变量筛选的阈值。位于两个虚线之间的稳定性值所对应的变量将被舍弃,保留虚线之外的变量用于 PLS 模型的构建。比较图 3a 和图 3b 可知,虽然都是采用同一批土壤样品的近红外光谱建模,然而对于不同的定量指标,选择的建模变量是不同的。

在 MC-UVE 方法中,保留建模变量数目的确定非常重要,因为保留变量的数目决定着模型的预测稳定性和精确度。如果保留的变量个数过少,可能会造成有用信息变量的丢失;相反,如果保留的变量个数过多,多余的无用信息变量会使模型质量变差。为此,考察了测试集预测均方差随保留变量数目变化的情况。图 4 给出了变量数目从 1 到 751 之间每隔 10 个数目所得的预测均方差。采用每一组保留的变量建立一个 PLS 模型,用来计算测试集的预测均方差。从图中可以看出,最初两者的数值都很大。随着保留变量数目的增加,两者的数值均急剧减小。对于碱解氮的模型,当保留变量数目为 210 时,预测均方差最小为 17.1 mg/kg。当保留变量数目继续增多时,预测均方差变化不大;对于速效钾模型,当保留变量数目为 150 时,预测均方差最小为 15.4 mg/kg。当保留变量数目继续增加时,预测

均方差逐渐增大。这表明当保留较少变量时,有用信息变量不能全部被模型所包含;而当选用过多变量时,无用的变量信息也会影响预测结果。

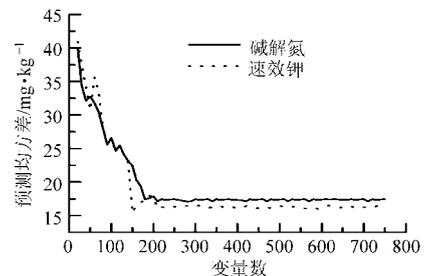


图4 测试集的预测均方差随保留变量数目的变化曲线

Fig. 4 Variation of RMSEP with number of selected wavelengths

2.3 PLS 和 MC-UVE PLS 方法预测结果比较

为了进一步比较 PLS 和 MC-UVE PLS 两种方法对土壤样品碱解氮和速效钾含量的预测结果,分别采用上述步骤优化后的参数对测试集样品进行预测,分别以校正集的 RMSECV、测试集的 RMSEP 和预测相关系数 R 对上述方法的模型质量进行评价,结果如表 3 所示。

由表 3 可知,对于碱解氮的模型,采用 MC-UVE PLS 得到了比原始光谱平滑后更好的预测结

表3 PLS和MC-UVE PLS两种方法建模结果比较
Tab.3 Comparison of results obtained by PLS and MC-UVE PLS

处理方法	碱解氮				速效钾			
	变量数目	RMSECV/mg·kg ⁻¹	R	RMSEP/mg·kg ⁻¹	变量数目	RMSECV/mg·kg ⁻¹	R	RMSEP/mg·kg ⁻¹
PLS	751	18.39	0.76	18.46	751	16.81	0.65	16.97
MC-UVE PLS	210	17.58	0.84	17.10	150	15.82	0.76	15.40

果,模型的预测相关系数从0.76提高到0.84, RMSEP从18.46 mg/kg减小为17.1 mg/kg,建模变量从751减少到210;对于速效钾的模型,模型的预测相关系数从0.65提高到0.76, RMSEP从16.97 mg/kg减小为15.4 mg/kg,建模变量从751减少到150。通过对上述结果的综合分析可知,为了减少模型的复杂性,对于土壤碱解氮和速效钾的模型,统一选用MC-UVE PLS方法即可保证模型的预测精度,同时也极大地减少了建模变量数目。

对碱解氮和速效钾的预测结果进行综合对比分析后可知,碱解氮和速效钾模型拟合效果都较好,说明碱解氮和速效钾的模型不存在过拟合和欠拟合现象,碱解氮校正模型的稳定性和适应性在实际应用中更加稳定与可靠。主要是因为土壤碱解氮相对速效钾来说,碱解氮含有C—H + C—H, C—H + C—C和N—H键组合,其对近红外光谱反应更灵敏,而速效钾之所以有较好的预测效果,或许是因为近红外

通过活性成分和矿物质间接地捕捉到所致,这个结论与Mouazen等^[15]的研究结论相同。

3 结束语

将PLS与MC-UVE方法相结合,应用于土壤的碱解氮和速效钾的定量分析。通过采用Monte Carlo技术,产生多个训练子集,并基于这些训练子集建立大量回归模型,然后根据模型的回归系数计算变量的稳定性,最后根据这些稳定性值的大小对变量进行筛选。结果表明,与使用全谱的PLS方法相比,MC-UVE PLS方法得到了更好的预测结果。模型获得的最好预测精度是碱解氮,它是近红外区域内唯一与光谱预测有直接相关的特性。土壤样品经过处理后,速效钾含量可以通过可见-近红外光谱法检测出来,部分原因是土壤活性成分与速效钾间接相关。

参 考 文 献

- 岑益郎,宋韬,何勇,等. 基于可见/近红外漫反射光谱的土壤有机质含量估算方法研究[J]. 浙江大学学报:农业与生命科学版,2011,37(3):300~306.
Cen Yilang, Song Tao, He Yong, et al. Rapid detection method of soil organic matter contents using visible/ near infrared diffuse reflectance spectral data [J]. Journal of Zhejiang University: Agriculture and Life Sciences, 2011, 37(3): 300~306. (in Chinese)
- 陈鹏飞,刘良云,王纪华,等. 近红外光谱技术实时测定土壤中总氮及磷含量的初步研究[J]. 光谱学与光谱分析,2008,28(2):295~298.
Chen Pengfei, Liu Liangyun, Wang Jihua, et al. Real-time analysis of soil N and P with near infrared diffuse reflectance spectroscopy [J]. Spectroscopy and Spectral Analysis, 2008, 28(2): 295~298. (in Chinese)
- 李颖,张小超,苑严伟,等. 北京典型耕作土壤养分的近红外光谱分析[J]. 农业工程学报,2012,28(2):176~179.
Li Jie, Zhang Xiaochao, Yuan Yanwei, et al. Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region [J]. Transactions of the CSAE, 2012, 28(2): 176~179. (in Chinese)
- Kuang B, Mouazen A M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms [J]. European Journal of Soil Science, 2011, 62(4): 629~636.
- 卢艳丽,白由路,王磊,等. 黑土土壤中全氮含量的高光谱预测分析[J]. 农业工程学报,2010,26(1):256~260.
Lu Yanli, Bai Youlu, Wang Lei, et al. Determination for total nitrogen content in black soil using hyperspectral data [J]. Transactions of the CSAE, 2010, 26(1): 256~260. (in Chinese)
- 宋海燕,何勇. 基于OSC和PLS的土壤有机质近红外光谱测定[J]. 农业机械学报,2007,38(12):113~116.
Song Haiyan, He Yong. Nearinfrared determination of organic matter content in soil based on OSC and PLS [J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(12): 113~116. (in Chinese)
- 朱登胜,吴迪,宋海燕,等. 应用近红外光谱法测定土壤的有机质和pH值[J]. 农业工程学报,2008,24(6):196~199.
Zhu Dengsheng, Wu Di, Song Haiyan, et al. Determination of organic matter contents and pH values of soil using near infrared spectroscopy [J]. Transactions of the CSAE, 2008, 24(6): 196~199. (in Chinese)
- 郭志新,梁亮,何见. 一种林地土壤氮磷钾含量快速测定的新方法[J]. 中国农学通报,2011,27(2):61~65.

- 4 Rapagnà S, Provendier H, Petit C, et al. Development of catalysts suitable for hydrogen or syngas production from biomass gasification [J]. *Biomass and Bioenergy*, 2002, 22(3): 377 ~ 388.
- 5 Yohan Richardson, Joël Blin, Ghislaine Volle, et al. In situ generation of Ni metal nanoparticles as catalyst for H₂-rich syngas production from biomass gasification [J]. *Applied Catalysis A: General*, 2010, 382(2): 220 ~ 230.
- 6 Keiichi Tomishige, Mohammad Asadullah, Kimio Kunimori. Syngas production by biomass gasification using Rh/CeO₂/SiO₂ catalysts and fluidized bed reactor [J]. *Catalysis Today*, 2004, 89(4): 389 ~ 403.
- 7 苏德仁,周肇秋,谢建军,等. 生物质流化床富氧-水蒸气气化制备合成气研究[J]. *农业机械学报*, 2011, 42(3): 100 ~ 104.
Su Deren, Zhou Zhaoqiu, Xie Jianjun, et al. Biomass oxygen enriched-steam gasification in an atmospheric fluidized bed for syngas production [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2011, 42(3): 100 ~ 104. (in Chinese)
- 8 José Corella, José M Toledo, Gregorio Molina. A review on dual fluidized-bed biomass gasifiers [J]. *Industrial & Engineering Chemistry Research*, 2007, 46(21): 6 831 ~ 6 839.
- 9 Christoph Pfeifer, Reinhard Rauch, Hermann Hofbauer. In-bed catalytic tar reduction in a dual fluidized bed biomass steam gasifier [J]. *Industrial & Engineering Chemistry Research*, 2004, 43(7): 1 634 ~ 1 640.
- 10 吴家桦,沈来宏,肖军,等. 串行流化床生物质气化制取合成气试验研究 [J]. *中国电机工程学报*, 2009, 29(11): 111 ~ 118.
Wu Jiahua, Shen Laihong, Xiao Jun, et al. Experimental study on syngas production from biomass gasification in interconnected fluidized beds [J]. *Proceedings of the CSEE*, 2009, 29(11): 111 ~ 118. (in Chinese)
- 11 Hugo de Lasa, Enrique Salaices, Jahirul Mazumder, et al. Catalytic steam gasification of biomass: catalysts, thermodynamics and kinetics [J]. *Chemical Reviews*, 2011, 111(9): 5 404 ~ 5 433.
- 12 高杨,肖军,沈来宏. 串行流化床生物质气化制取富氢气体模拟研究[J]. *太阳能学报*, 2008, 29(7): 894 ~ 899.
Gao Yang, Xiao Jun, Shen Laihong. Hydrogen production from biomass gasification in interconnected fluidized beds [J]. *Acta Energetica Solaris Sinica*, 2008, 29(7): 894 ~ 899. (in Chinese)
- 13 陈汉平,赵向富,米铁,等. 基于 ASPEN PLUS 平台的生物质气化模拟[J]. *华中科技大学学报:自然科学版*, 2007, 35(9): 49 ~ 52.
Chen Hanping, Zhao Xiangfu, Mi Tie, et al. Simulation of biomass gasification by ASPEN PLUS [J]. *Journal of Huazhong University of Sci. & Tech.: Nature Science Edition*, 2007, 35(9): 49 ~ 52. (in Chinese)
- 14 屈一新. 化工过程数值模拟及软件[M]. 北京: 化学工业出版社, 2011.
- 15 张亚男,肖军,沈来宏. 串行流化床稻秸气化合成甲醇的模拟[J]. *中国电机工程学报*, 2009, 29(32): 103 ~ 111.
Zhang Yanan, Xiao Jun, Shen Laihong. Simulation of methanol production via rice straw gasification in interconnected fluidized beds [J]. *Proceedings of the CSEE*, 2009, 29(32): 103 ~ 111. (in Chinese)
- 16 Shen Laihong, Gao Yang, Xiao Jun. Simulation of hydrogen production from biomass gasification in interconnected fluidized beds [J]. *Biomass and Bioenergy*, 2008, 32(2): 120 ~ 127.

(上接第 91 页)

- 9 Centner V, Massart D L. Elimination of uninformative variables for multivariate calibration [J]. *Analytical Chemistry*, 1996, 68(21): 3 851 ~ 3 858.
- 10 Chen D, Shao X G, Hu B, et al. A background and noise elimination method for quantitative calibration of near infrared spectra [J]. *Analytica Chimica Acta*, 2004, 511(1): 37 ~ 45.
- 11 Put R, Daszykowski M, Baczek T, et al. Retention prediction of peptides based on uninformative variable elimination by partial least squares [J]. *Journal of Proteome Research*, 2006, 5(7): 1 618 ~ 1 625.
- 12 Polanski J, Gieleciak R. The comparative molecular surface analysis (CoMSA) with modified uninformative variable elimination PLS (UVE-PLS) method: application to the steroids binding the aromatase enzyme [J]. *Journal of Chemical Information and Computer Sciences*, 2003, 43(2): 656 ~ 666.
- 13 Cai W S, Li Y K, Shao X G. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 90(2): 188 ~ 194.
- 14 鲍士旦. 土壤农化分析[M]. 北京: 中国农业出版社, 1999: 30 ~ 34.
- 15 Mouazen A M, Kuang B, Baerdemaeker J D, et al. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy [J]. *Geoderma*, 2010, 158(1 ~ 2): 23 ~ 31.
- 16 宋海燕,秦刚,韩小平,等. 基于可见光谱的不同质地土壤有机质快速测定[J]. *农业机械学报*, 2012, 43(7): 69 ~ 72.
Song Haiyan, Qin Gang, Han Xiaoping, et al. Rapid prediction of soil organic matter by using visible infrared spectral technology [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2012, 43(7): 69 ~ 72. (in Chinese)
- 17 张淼,盛明雅,张丽楠,等. 基于电极阵列的土壤速效养分快速检测系统[J]. *农业机械学报*, 2012, 43(增刊): 277 ~ 282.
Zhang Miao, Sheng Mingya, Zhang Li'nan, et al. Development of soil available macronutrients detection system based on ion-selective electrodes array [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2012, 43(Supp.): 277 ~ 282. (in Chinese)