

基于高分辨率反射光谱的土壤营养元素估算模型*

徐永明¹ 蔺启忠² 王璐³ 黄秀华²

(1 南京信息工程大学中美合作遥感中心, 南京 210044)

(2 中国科学院遥感应用研究所, 北京 100101)

(3 南京大学地球科学系, 南京 210093)

摘要 研究了土壤中营养元素含量(N、P、K)与土壤可见光/近红外光谱之间的关系。在对原始光谱进行预处理分析后,计算出了4种光谱指标:反射率 Reflectance、一阶导数 FDR、倒数之对数 $\log(1/R)$ 和波段深度 Depth。通过偏最小二乘回归分析建立了营养元素与4种光谱指标的经验模型,并且利用验证样本集对回归模型进行了验证。结果表明,可见光/近红外反射光谱具有快速估算土壤中营养元素含量的潜力。

关键词 土壤;营养元素;反射光谱;偏最小二乘回归

中图分类号 TP79; S151.9 **文献标识码** A

土壤营养元素的监测,在农业环境综合评价中具有重要的意义。随着现代农业的迅速发展,将成像技术与光谱技术结合起来的高光谱遥感凭借其极高的光谱分辨率,具备了定量获取土壤化学组分的研究潜力。利用土壤反射光谱估算土壤中营养元素含量,将为高光谱遥感在土壤遥感的实用化进行前期理论研究,具有深远的意义。

近20年来,国内外相继开展了土壤地球化学定量遥感的相关研究。可见-近红外土壤光谱可以对土壤性质,如有机质、黏土含量、水分含量和总氮含量等进行估算^[1~3]。此外,一些学者还研究了土壤中氧化铁以及一些重金属元素与土壤光谱之间的关系^[4,5]。很多学者分析了原始光谱、标准化光谱、一阶导数等不同的光谱指标对估算精度的影响^[1,4,6]。目前,运用土壤光谱对土壤各种性质的估算多是利用统计的方法,如:多元逐步回归、偏最小二乘回归等等^[1~4,6]。本研究将利用偏最小二乘回归法建立土壤反射率 Reflectance、一阶导数 FDR、倒数之对数 $\log(1/R)$ 、波段深度 Depth 估算土壤营养元素(N、P、K)的经验模型。

1 材料与方 法

1.1 样品及其分析测定

依据天津市国土农业环境地球化学地面调查结

果,在天津南部城区周围选择了采样点共70个。2004年5月对所选研究区进行了实地考察,并采集了土壤表层(0~20 cm)样本,自然风干,过200目筛,4分法筛选。样本土壤类型主要为盐化湿潮土和淋溶褐土,覆盖了菜地、棉花田、玉米田、西瓜田、葡萄园、桃园、芦苇地等各类农业土地利用类型。

土壤化学分析工作由中国科学院资源与地理研究所完成,P、K元素用ICP法测定,N元素用开氏法测定。

土壤的实验室光谱测量是在南京大学成矿作用重点实验室完成。具体测量过程如下:将过200目筛的土壤约0.3 g置于载玻片上,滴入纯净水,搅拌使之均匀分布,将载玻片放在烘盘上,加热烘干,烘干之后的载玻片上的样本不出现龟裂或气泡则说明样本分布均匀,可以用于反射率的测量。取标准白板置于样品仓中,对光谱仪进行校零;将制好的载玻片放入光谱仪的样品仓中,测量样品的反射率。所用仪器为Perkin Elmer公司生产的Lambda 900紫外/可见/近红外分光光度计,采样间隔2 nm,波长范围从320 nm到2480 nm,共1081个波段。获得的土壤光谱曲线需要进行光谱分区断点的抛物线拟合校正以及去噪声平滑处理。

1.2 数据处理

1.2.1 光谱曲线校正 由于Lambda 900光谱仪在紫外/可见光和近红外/短波红外波长部分采用了

* 国家自然科学基金项目“农业环境土壤地球化学的遥感机理研究”(40371085)资助

作者简介:徐永明(1980~),男,助理工程师,硕士研究生,研究方向为资源环境遥感。E-mail: xym30@263.net

收稿日期:2005-05-19;收到修改稿日期:2005-09-28

不同的探测元件(紫外/可见光部分为光电倍增管,近红外/短波红外部分为恒温硫化铅),因此在两类探测元件分区结合处会出现光谱的偏差,如图1所示,在858 nm和860 nm处的光谱曲线有一个陡变。为了消除不同探测元件所造成的误差,需要对原始光谱曲线进行校正。

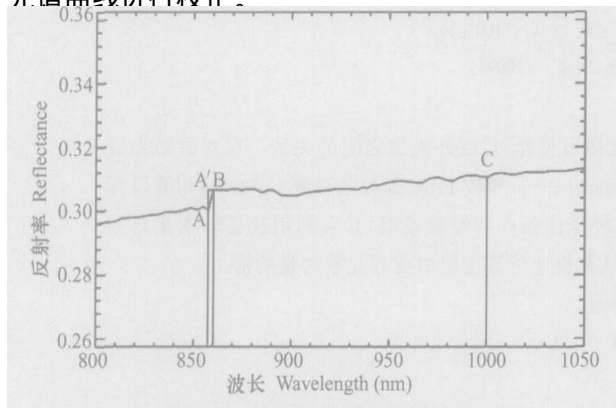


图1 光谱曲线分区断点的抛物线校正
Fig. 1 Parabola correction of spectra breaking

以可见光部分的反射率为准,对靠近分区断点的近红外部分(860~1 000 nm)反射率采用抛物线方

法进行校正,消除陡变。

校正的抛物线方程为:

$$R(\lambda) / R(\lambda_0) = a(\lambda - b)^2 + c \quad (1)$$

其中 λ 为波长, $R(\lambda)$ 为原始反射率, $R(\lambda_0)$ 为校正后的反射率, a, b, c 为拟合抛物线的参数。

已知条件包括:两个已知点:(858, $R(860)/R(858)$)、(1 000, 1),此外,第一个已知点为顶点。根据这些已知条件,可以求出抛物线方程的参数 a, b, c 。将 a, b, c 代入抛物线方程,根据此方程可以求得校正后的反射率:

$$R(\lambda) = \frac{R(\lambda_0)}{a(\lambda - b)^2 + c} \quad (2)$$

由于光谱仪波段之间对能量响应上的差别,光谱曲线上存在许多“毛刺”噪声,反射率变化剧烈,需要对光谱曲线进行平滑优化处理。本研究中,选用了长度为9的海明窗(Hamming)作为滤波器模板。对光谱曲线进行低通滤波后,基本消除了高频成分(毛刺),使曲线变得平滑。

经过上述步骤处理后,最终得到了70个土壤样本的反射率曲线,如图2所示。

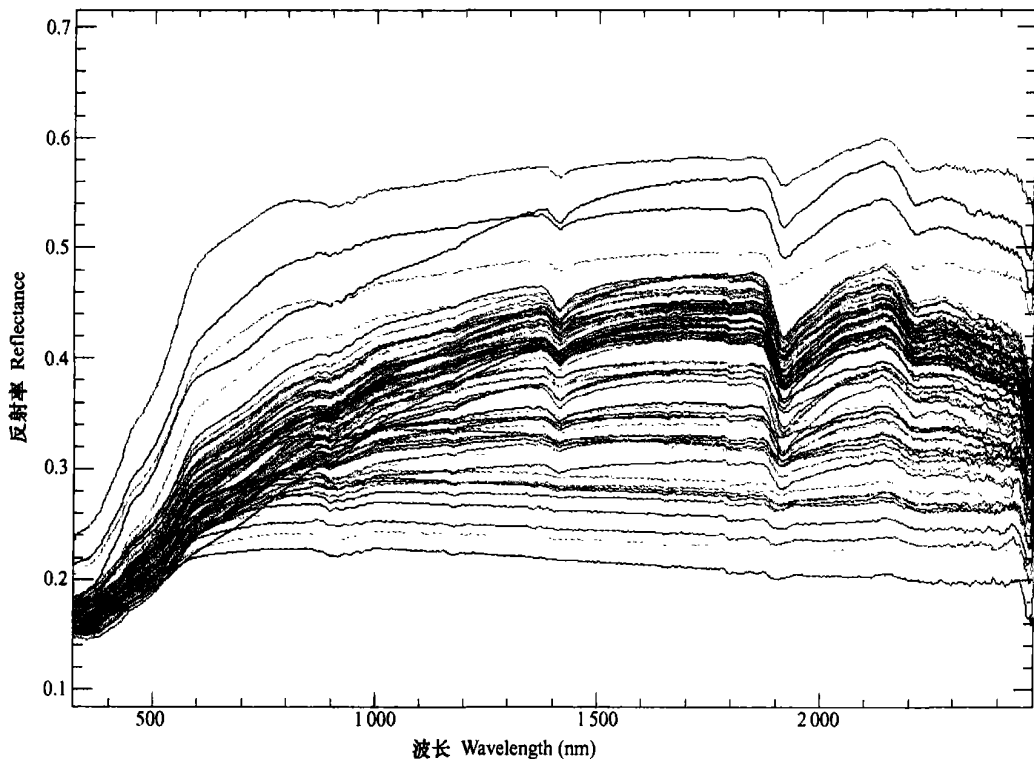


图2 经过预处理后的70个样本的光谱曲线
Fig. 2 Pre-processed spectra of 70 soil samples

1.2.2 偏最小二乘回归方法 单因变量偏最小二乘回归方法的思路如下:设已知因变量 y 和 k 个

自变量 x_1, x_2, \dots, x_k , 样本数为 n , 构成数据表 $X = [x_1, x_2, \dots, x_k]_{n \times k}$ 和 $y = [y]_{n \times 1}$ 。在 X 中提取成分

t_1, t_1 是 x_1, x_2, \dots, x_k 的线性组合,要求 t_1 尽可能大地携带 X 矩阵中的变异信息,且与 y 的相关程度最大。这样, t_1 既能很好的代表 X 的信息,同时对 y 又具有最强的解释能力。提取第一个主成分 t_1 后,实施 y 对 t_1 的回归,如果此时回归方程已经达到满意的精度,则算法停止;否则,将利用 X 被 t_1 解释后的残余信息以及 y 被 t_1 解释后的残余信息再进行回归分析。如此反复,直到能达到一个较满意的精度为止。

2 结果与分析

2.1 营养元素含量分析

对 70 个土壤样本营养元素含量进行了简单的统计分析,结果如表 1、表 2 所示。从相关系数矩阵看,N 和 P 元素的相关系数很高,而 K 与 N、P 两种元素的相关性不强。

表 1 土壤样本营养元素数据集统计结果

Table 1 Statistics of nutrient elements in soil samples

	最大值 Max (mg g ⁻¹)	最小值 Min (mg g ⁻¹)	均值 Mean (mg g ⁻¹)	标准差 Std (mg g ⁻¹)	变异系数 CV (%)
氮 N	3.17	0.128	1.52	0.690	45.33
磷 P	3.23	0.088 2	1.09	0.708	64.91
钾 K	21.6	13.9	17.5	2.85	16.27

表 2 三种营养元素的相关系数矩阵

Table 2 Correlation matrix of the three nutrient elements

	氮 N	磷 P	钾 K
氮 N	1	0.761 1	- 0.080 8
磷 P	0.761 1	1	- 0.249 5
钾 K	- 0.080 8	- 0.249 5	1

2.2 建模分析

在前人对叶片生化组分及土壤化学组分的研究中,除了直接利用光谱反射率 R 进行多元统计分析外,还提出了一阶导数(FDS)、倒数(1/R)、倒数的对数 $\log(1/R)$ 、波段深度(Band Depth)等光谱指标^[1,4,7]。在本研究中对这几种光谱指标进行了分析,比较其对精度的影响。

将 70 个样本数据分为两组,第一组 50 个样本,用于偏最小二乘回归分析,建立回归模型,第二组 20 个样本用于回归模型的检验。在偏最小二乘回归分析中,一般并不需要采用全部的成分进行回归建模,而是可以像在主成分分析时一样,采用截尾的方式选择前 h 个成分就可以得到一个预测性能很好的模型。本研究中,采用了下述方法选取最佳成分个数:设入选的成分为 h ,则回归分析的判定系数 R_h^2 随着 h 的增大而增大,当 $R_{h+1}^2 - R_h^2 < 1\%$ 时,认为增加新的成分对减少方程的预测误差无明显的改善作用,因而选取前 h 个成分进行偏最小二乘回归分析。

回归建模和检验结果如表 3、表 4 所示, R^2 为回归方程的判定系数, R 为实测值与估算值之间的相关系数, RMSE 为均方根误差。

表 3 N、P、K 元素建模样本偏最小二乘回归分析的结果

Table 3 PLSR analysis of calibration dataset for N, P and K

光谱指标 Spectral parameter		氮 N	磷 P	钾 K
反射率 Reflectance	判定系数 R^2	0.978 8	0.982 9	0.984 0
	均方根误差 RMSE	0.090 9	0.078 4	0.254 5
	成分个数 Number of factors	13	14	14
一阶导数 FDR	判定系数 R^2	0.993 4	0.983 9	0.985 5
	均方根误差 RMSE	0.050 9	0.076 1	0.242 5
	成分个数 Number of factors	6	5	6
倒数之对数 $\log(1/R)$	判定系数 R^2	0.984 9	0.982 6	0.982 2
	均方根误差 RMSE	0.076 7	0.079 1	0.268 3
	成分个数 Number of factors	14	14	14
波段深度 Depth	判定系数 R^2	0.988 7	0.986 2	0.976 7
	均方根误差 RMSE	0.066 3	0.070 5	0.306 7
	成分个数 Number of factors	10	9	9

表 4 回归方程的检验结果
Table 4 Verification of the regression equations

光谱指标 Spectral parameter	氮 N		磷 P		钾 K	
	相关系数	均方根误差	相关系数	均方根误差	相关系数	均方根误差
	r	RMSE	r	RMSE	r	RMSE
反射率 Reflectance	0.761 4	0.542 1	0.693 8	0.700 8	0.699 8	1.483 6
一阶导数 FDR	0.839 8	0.485 1	0.675 5	0.694 8	0.648 4	1.520 3
倒数之对数 $\log(1/R)$	0.755 6	0.543 8	0.717 1	0.696 1	0.650 9	1.495 1
波段深度 Depth	0.897 0	0.434 5	0.728 6	0.657 2	0.458 4	1.948 3

从表 3、表 4 可以看出:对于 N 元素, FDR 和 Depth 的偏最小二乘回归收敛速度最快,分别选用了 6 个和 10 个成分,而 Reflectance 和 $\log(1/R)$ 的收敛速度要慢很多,分别选用了 13 个和 14 个成分。各个光谱指标的偏最小二乘回归建模精度都非常高,但是验证精度却低很多。利用回归模型对检验样本进行验证时发现 Depth 的精度最高, RMSE 为 0.434 5。FDR 的检验精度也比较高, RMSE 为 0.485 1。Reflectance 与 $\log(1/R)$ 的验证精度相近, RMSE 为 0.542 1 和 0.543 8。此外,对 PLSR 模型的系数进行研究, P 元素的分析结果与 N 元素比较相似,但精度要低一些。FDR 和 Depth 的偏最小二乘回归收敛速度最快,分别选用了 5 个和 9 个成分,而 Reflectance 和 $\log(1/R)$ 都选用了 14 个成分。利用回归模型对检验样本进行验证时发现 Depth 的精度最高, RMSE 为 0.657 2,而其余三个光谱指标 Reflectance、FDR 和 $\log(1/R)$ 的验证精度比较接近, RMSE 分别为 0.700 8、0.694 8 和 0.696 1。对于 K 元素, FDR 和 Depth 的偏最小二乘回归收敛速度仍然最快,分别选用了 6 个和 9 个成分,而 Reflectance 和 $\log(1/R)$ 同样选用了 14 个成分。但是利用回归模型对检验样本进行验证时发现 Reflectance、FDR 和 $\log(1/R)$ 的精度相差不多, RMSE 分别为 1.483 6、1.520 3 和 1.495,而 Depth 的精度最低, RMSE 高达 1.948 3。K 元素的反演精度最低,原因可能有多种,如土壤中 K 元素含量比较稳定,变异系数较小,值域较窄会影响模型的精度,也有可能是 K 元素本身与土壤反射率之间的相关性本来就比较小。

通过上面的分析可以看出: FDR 和 Depth 的偏最小二乘回归收敛速度非常快,一般成分数量都小

于 10,而 Reflectance 和 $\log(1/R)$ 的收敛速度比较慢,需要的成分数较多,回归的验证精度也很接近。对于 N 元素和 P 元素, Depth 为最合适进行偏最小二乘回归分析的光谱指标,而对于 K 元素来说,虽然 Reflectance、FDR 和 Depth 的回归验证精度差不多,但 FDR 的收敛速度最快,需要的成分最少,因此,我们认为 FDR 是最适用来估算 K 元素含量的光谱指标。

图 3 给出了 FDR 光谱指标建立的 PLSR 模型的系数,限于篇幅, Reflectance、 $\log(1/R)$ 和 Depth 的模型系数图没有列出,但与 FDR 都比较接近。对图 3 进行观察可以发现,以 1750 nm 为中心的短波红外谱段对于 N、P、K 元素的 PLSR 回归模型的贡献度最大,以 1200 nm 为中心的近红外谱段对 PLSR 回归模型也有较大的贡献,其余谱段的回归模型系数普遍偏小。

图 4、图 5、图 6 分别给出了 20 个验证样本的实测 N、P、K 元素含量与不同光谱指标建立的 PLSR 经验方程估算值的比较, a、b、c、d 分别对应于 Reflectance、FDR、 $\log(1/R)$ 、Depth 这 4 种光谱指标。从图可以看出,利用 Depth 得到的回归方程对于 N 元素和 P 元素的反演结果比较好,而 K 元素的估算值与实测值大多偏离了 1:1 线, Reflectance 和 $\log(1/R)$ 得到的估算值也比较接近。此外,对于 N 元素和 P 元素,数值较小的样本其估算值与实测值更为相近,高含量样本同样存在低估现象,说明 PLSR 回归方程对中高含量的 N 元素、P 元素具有更好的解释能力;而对于 K 元素,正好相反,数值较大的样本其估算值与实测值更为相近,说明 PLSR 回归方程对中高含量的 K 元素具有比较好的解释能力。

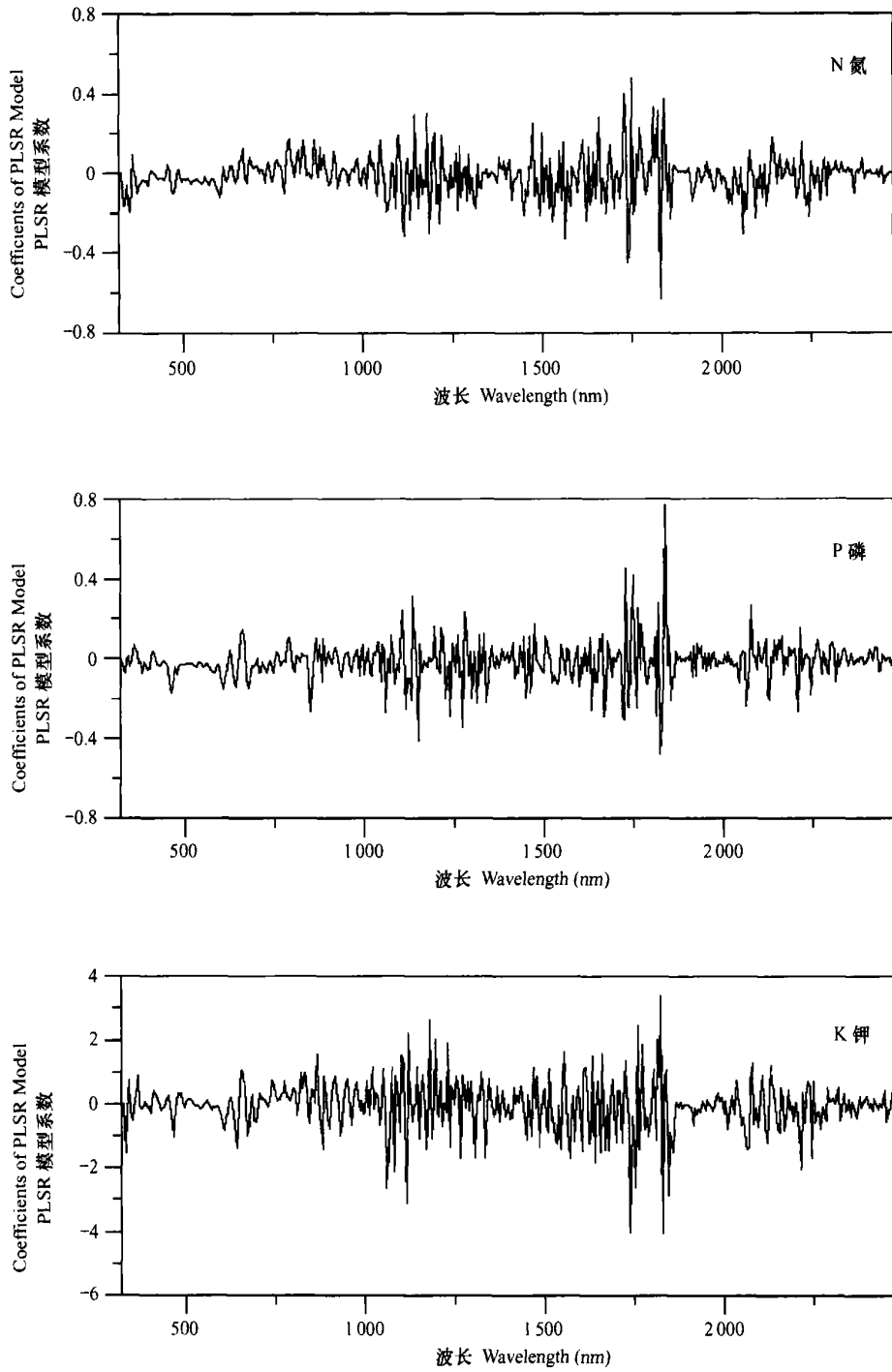


图3 基于 FDR 光谱指标建立的 PLSR 模型的系数

Fig.3 Coefficients of PLSR models based on FDR

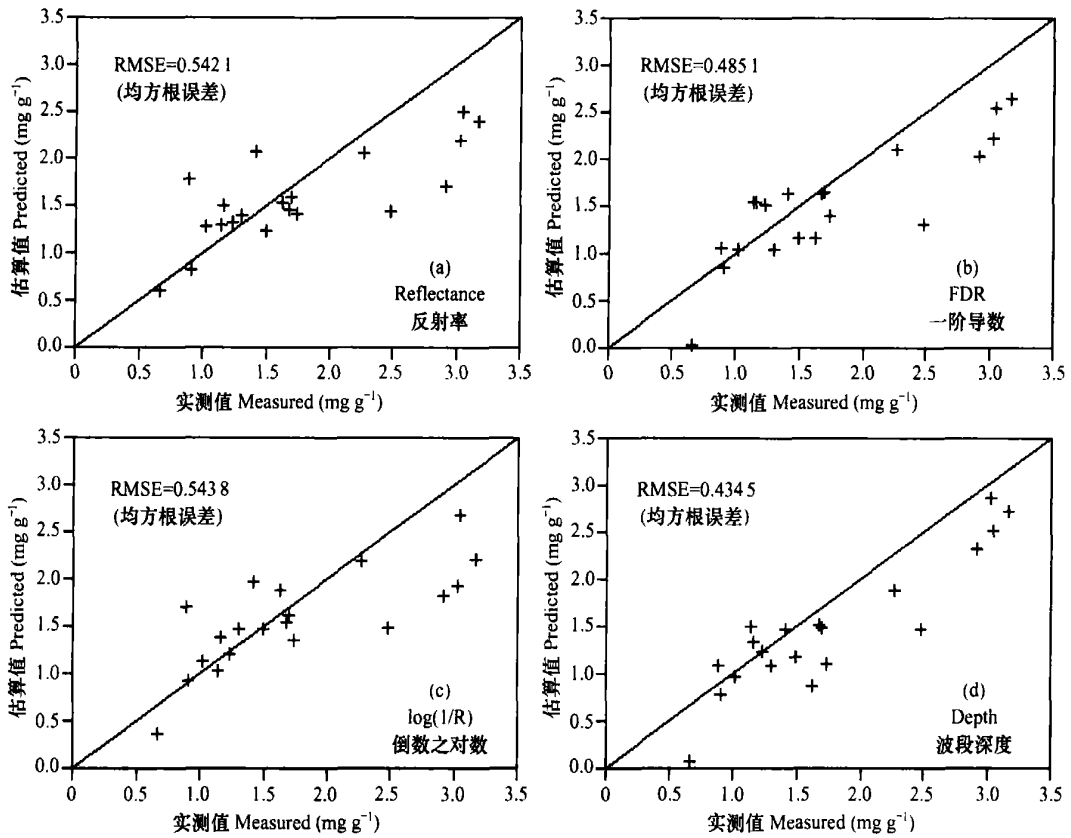


图 4 N 元素含量实测值与 PLSR 模型的估算值之间的比较

Fig. 4 Comparison between predicted N and measured N in the validation dataset

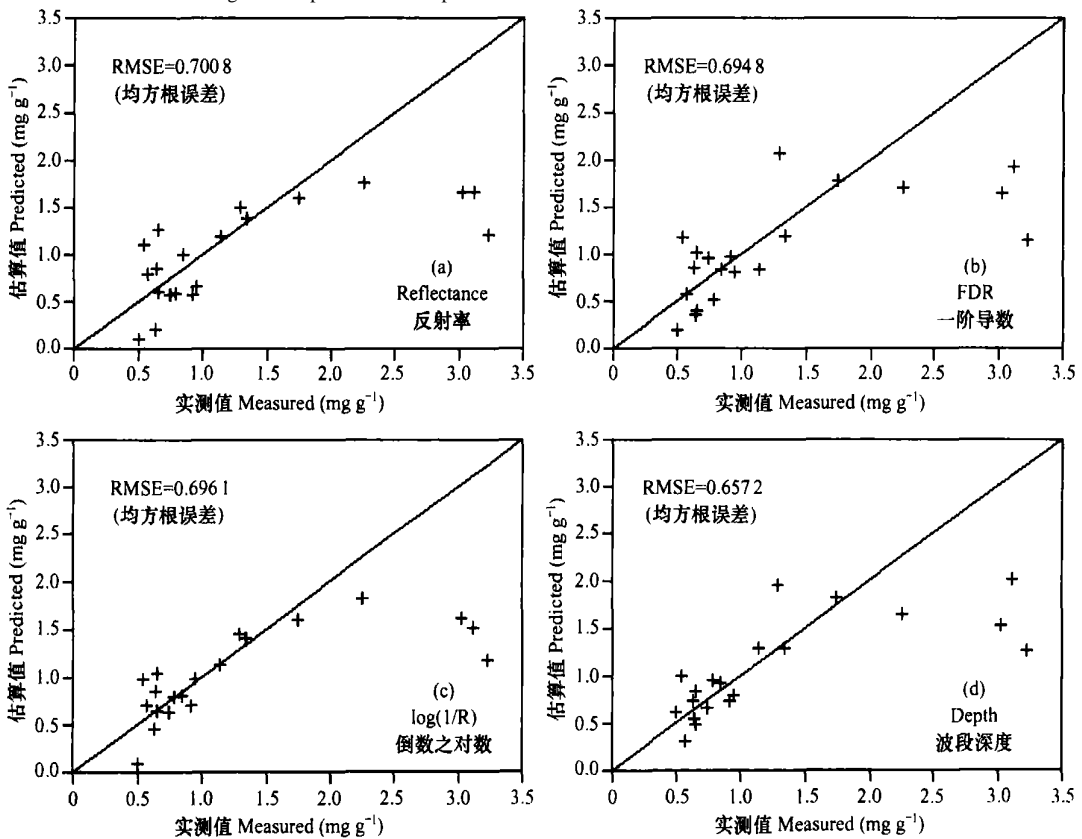


图 5 P 元素含量实测值与 PLSR 模型的估算值之间的比较

Fig. 5 Comparison between predicted P and measured P in the validation dataset

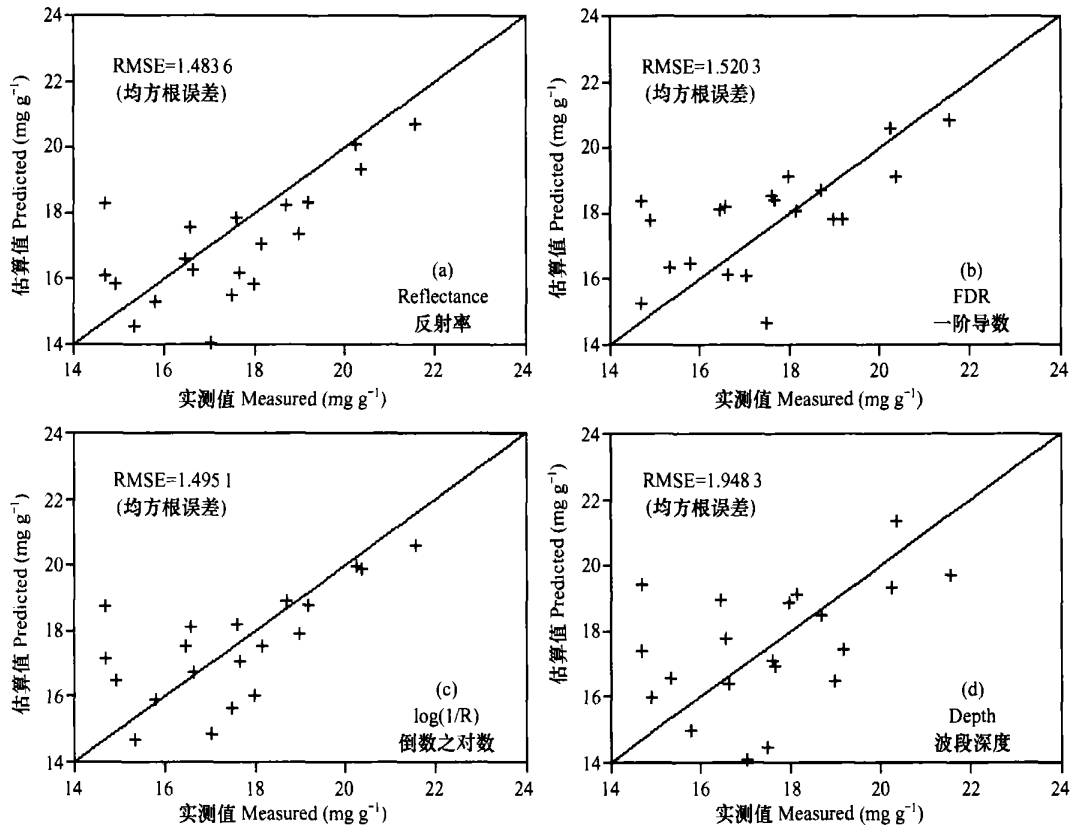


图 6 K 元素含量实测值与 HLSR 模型的估算值之间的比较

Fig. 6 Comparison between predicted K and measured K in the validation dataset

3 结 论

1) 图表分析表明,土壤反射率光谱与营养元素含量之间存在良好的相关性,尤其是 N 元素,预测值与实测值之间的相关系数 R 可达到 0.89。说明利用土壤可见-近红外光谱对土壤营养元素含量进行测定是有可能的。

2) 在偏最小二乘回归分析中,Depth(波段深度)和 FDR(光谱一阶导数)这两种光谱指标的反演精度较高,且收敛速度很快;Reflectance(反射率)和 $\log(1/R)$ (光谱倒数之对数)这两种光谱指标的建模精度与验证精度都很相近,收敛速度也很相近。从本论文的研究来看,对原始光谱进行倒数之对数计算对于提高估算精度作用不大。

3) 本研究采用了统计方法对土壤的实验室光谱进行了分析,经验方法固有的缺陷使得该方法对数据集的变化极为敏感,其应用的空间与时间范围受到限制。以后的研究中我们希望能运用尽可能多的样本来进行回归分析,不同含量的、宽域值的样本

能够大大提高模型的精度与适用性。

4) 实验室光谱有效去除了很多因素的影响,可以得到比较好的结果。但在实际田间情况下,土壤反射率还受到植被覆盖、大气状况等多种因素的干扰,需要进行大量的实验室与田间的光谱测量与分析才能建立准确的土壤营养元素的光谱反演模型。将反演模型推广到遥感影像则需要进一步分析不同光谱分辨率的影响、混合像元的影响、波段是否位于大气窗口以及对图像进行大气校正与定标等等。

参 考 文 献

- [1] Krishnan P, Alexander J D, Butler B J, *et al.* Reflectance technique for predicting soil organic matter. *Soil Science Society of America Journal*, 1980, 44: 1 282 ~ 1 285
- [2] Ben-Dor E, Banin A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, 1995, 59: 364 ~ 372
- [3] 彭玉魁,张建新,何绪生,等. 土壤水分、有机质和总氮含量的近红外光谱分析研究. *土壤学报*, 1998, 35(4): 553 ~ 559. Peng Y K, Zhang J X, He X S, *et al.* Analysis of soil moisture, organic matter and total nitrogen content in loess in China with near infrared spectroscopy (In Chinese). *Acta Pedologica Sinica*, 1998, 35(4):

553 ~ 559

- [4] Kooistra L , Wehrens L , Leuven R S E W , *et al.* Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. *Analytica Chimica Acta* , 2001 , 446 : 97 ~ 105
- [5] Chodak M , Ludwig B , Khanna P , *et al.* Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands. *Journal of Plant Nutrition and Soil Science* , 2002 , 165 : 27 ~ 33
- [6] Coops N C , Ryan P , Bishop , A W. Investigating CASI responses to soil properties and disturbance across an Australian eucalypt forest. *Canadian Journal of Remote Sensing* , 1998 , 24 (2) : 153 ~ 168
- [7] Kokaly R F , Clark R N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment* , 1999 , 76 : 267 ~ 287

MODEL FOR ESTIMATING SOIL NUTRIENT ELEMENTS BASED ON HIGH RESOLUTION REFLECTANCE SPECTRA

Xu Yongming¹ Lin Qizhong² Wang Lu³ Huang Xiuhua²

(1 *SinoAmerica Cooperative Remote Sensing Center, Nanjing University of Information Science and Technology, Nanjing 210044, China*)

(2 *Institute of Remote Sensing Application, Chinese Academy of Sciences, Beijing 100101, China*)

(3 *Department of Earth Science, Nanjing University, Nanjing 210093, China*)

Abstract Relationship between contents of soil nutrient elements (N, P and K) and laboratory VNIR (visible-near infrared) spectra of the soil is analyzed in this article. Some spectra are pre-processed before calculation of four spectral indices, including inverse spectra ($1/R$), first derivative reflectance spectra (FDR), inverse-log spectra ($\log(1/R)$) and Depth. Based on the partial least square regression method, statistical relationships between these indices and contents of soil nutrient elements are established by calibration set. Then the regression model is verified by test set. The results showed that VNIR (visible-near infrared) spectra of soil are potentially useful for rapid estimation of content of the nutrient elements in the soil.

Key words Soil; Nutrient elements; Reflectance spectra; PLSR