

支持向量机在土壤镁含量高光谱估算中的应用^①

田 焯¹, 沈润平^{1*}, 丁国香²

(1 南京信息工程大学遥感学院, 南京 210044; 2 安徽省气象局, 合肥 230001)

摘要: 研究利用土壤样本实验反射光谱, 分析了土壤镁(Mg)含量与土壤反射光谱的关系, 比较了主成分回归分析(PCR)、偏最小二乘回归分析(PLSR)和支持向量机回归分析(SVMR)等方法, 以及土壤反射光谱及其变换光谱与土壤 Mg 含量之间的估算模型, 为土壤 Mg 含量高光谱估算提供依据。结果表明: PCR、PLSR、SVMR 3 种建模方法在 Mg 含量的估算中, SVMR 的估算精度相对较高, 估算精度平均达到 80.96%, 分别比 PCR 和 PLSR 提高了 6.16%、4.20%; 对于不同的数学变换处理方法, 一阶微分变换相对较好, 估算精度平均为 80.76%, 分别比反射率、倒数对数变换提高了 4.95%、4.61%。因此, 运用土壤反射光谱一阶微分变换的 SVMR 进行建模, 可以相对较好地估算全 Mg 含量, 精度达 84.04%。

关键词: 高光谱; 土壤镁含量; 支持向量机回归

中图分类号: S151.9

土壤不仅是珍贵的自然资源, 更是人类农业生产的基地和陆地生态系统的基础。镁(Mg)是自然界广泛存在的化学元素, 土壤中的 Mg 大多数以矿物态存在于土壤黏土矿物中, 由于风化和淋溶作用, 使得土壤中 Mg 的平均含量只有 5 g/kg^[1]。Mg 是植物生长所必需的营养元素, 是叶绿素的重要组成部分, 同时也是植物光合作用和呼吸作用过程中许多酶的活化剂。缺 Mg 会造成植物叶片失绿变黄, 体内代谢活动受到影响, 光合作用受阻, 使植物不能正常地生长发育, 严重时叶片会枯萎、掉落, 这将大大降低产量和质量。本研究区为我国中亚热带典型红壤丘陵区, 土壤呈酸性, Mg 含量相对较少, 可能存在缺 Mg 现象, 因此对土壤 Mg 含量进行系统研究十分必要。

高光谱具备定量获取土壤化学组分的研究潜力, 并在土壤水分、土壤有机质、土壤 Fe 和土壤营养元素含量估算方面进行了较多的研究, 一些研究学者通过利用实验室和野外测量的土壤高光谱数据, 分析可见光-近红外光谱范围内土壤含水量、有机质、Fe、氮磷钾等营养元素的特征波段, 采用多元线性回归、主成分回归、偏最小二乘回归和人工神经网络等方法, 建立了土壤各组分含量估算模型, 并进行定量估

算, 得到了较好的预测精度^[2-10]。对于土壤 Mg 元素的研究, 则相对较少, Chang 等^[11]运用主成分回归分析方法, 分析并预测了 N、Ca、K、Mg 等多种化学成分的含量, 其中 Mg 元素的定标模型决定系数为 0.70。Malley 等^[12]利用近红外光谱分析技术(NIRS)对潮土和风干土壤进行了分析, 预测的潮土和风干土壤中 Mg 的定标系数分别为 0.85 和 0.92, 结果显示经过风干后的土壤更有利于土壤元素的定标分析。Chodak 等^[13]运用线性回归分析方法研究了森林覆盖地区土壤物理、生化性质(C、N、P、S、Na、K、Ca、Mg、Mn、Fe 等)与反射光谱之间的关系, 得到 Mg 的相关系数为 0.68。Thomasson 等^[14]研究了密西西比河流域土壤样本, 发现其中一个研究区内土壤 Mg、Ca 含量与反射光谱相关性很高, 得出土壤特性与反射光谱之间存在一定的相关性, 但是这种相关性并不稳定。以上研究说明土壤 Mg 含量的高光谱遥感具有一定可行性, 但是对于红壤及其发育的土壤中 Mg 含量的高光谱遥感尚缺乏深入研究。红壤是我国长江以南重要的地带性土壤, 养分贫瘠化制约南方地区农业发展, Mg 是作物生长发育所需的中量元素, 部分红泥质红壤和硅质红壤土壤有效 Mg 含量低于

基金项目: 国家“973”计划项目(2010CB950701, G20000779)资助。

* 通讯作者(rpshen@nuist.edu.cn)

作者简介: 田焯(1989—), 女, 江苏苏州人, 硕士研究生, 主要从事土壤光谱遥感研究。E-mail: 13770852866@163.com

20 mg/kg, 供 Mg 能力低, Mg 元素的供需不平衡日益加重, 严重影响作物产量和品质^[15]。本文综合光谱数学变换因素, 并运用支持向量机回归分析, 以克服处理复杂的非线性问题时常规建模方法的不足, 为土壤 Mg 含量高光谱估算提供依据。

1 材料与方 法

1.1 土壤样本采集与光谱测定

土壤样本采集于亚热带典型红壤丘陵区的江西余江县和泰和县, 共 34 个土壤样本, 土层深度为 0~20 cm, 其中余江县样本 22 个, 泰和县样本 12 个。土壤样本经风干、碾磨, 过 1 mm 筛。土壤化学分析工作由中国科学院南京土壤研究所红壤生态实验站完成, 土壤全 Mg 含量测定采用 HF-HClO₄ 消煮, 原子吸收光谱法测定^[16]。

土壤反射光谱测量在室内进行, 使用 ASD Pro FieldSpec3 便携式光谱仪, 土壤样品为 1 mm 粒径风干土, 光源为一盏 1 000 W 的卤素灯, 光照角度为 15°, 距离土样 30 cm, 采用 5° 视场角探头 ($B = 2.5^\circ$) 垂直于土样观测, 探头距离土样 15 cm, 测试之前先以白板进行定标, 获取绝对反射率^[17]。测定得到 350~2 500 nm 波长范围的反射光谱, 重采样间隔为 1 nm, 输出波段数为 2 151。将每个样本测定 10 次, 经算术平均后作为该样本的实际反射光谱数据。为了减少数据量, 并去除冗余信息, 对原始反射率光谱进行 10 nm 间隔的重采样, 经过重采样后的光谱分辨率有所降低, 但数据在一定程度上得到了平滑, 仍然较好地保持了原光谱的主要特征^[18]。

1.2 光谱变换与估算方法

光谱变换有利于消除测定背景及环境因子对光谱的影响, 提高高光谱遥感估算精度, 为了研究不同变换处理方法对土壤 Mg 估算的影响, 研究选用了常见的一阶微分变换和倒数对数变换方法^[19-21], 对原始光谱进行处理, 并与原始光谱比较, 以确定最优的光谱变换方法。

主成分回归分析方法 (PCR) 和偏最小二乘回归分析方法 (PLSR) 是使用较多的土壤物质含量估算方法^[22-24]。近年来, 支持向量机回归分析方法 (support vector machine regression, SVMR) 在数据挖掘、遥感图像分类、目标识别等方面得到应用, 取得了较好的结果^[25], 但对于土壤 Mg 的光谱估算尚鲜见报道, 本文对其进行了研究, 并选用上述 PCR 和 PLSR 方法进行比较分析。其中, PCR 和 PLSR 模型建立在 SPSS Statistics 20.0 中完成。

支持向量机是 Vapnik^[26]于 1995 年基于针对小样本的统计学习理论 (statistical learning theory, SLT) 提出的, 有别于现有的统计方法, 该方法不涉及概率测度和大数定理理论, 并且采用了结构风险最小化原理, 较好地解决了人工神经网络方法的网络结构难以确定、过学习和欠学习、局部极小的问题, 针对小样本分类和回归问题, 此方法具有明显的优势。

本文选取 22 个样本建模, 10 个样本检验, 根据反射率、一阶微分、倒数对数光谱与 Mg 含量的相关性, 选择显著性水平大于 0.1 的波段。为了降低相邻特征波段间共线性的影响, 将光谱信息进行归一化处理作主成分分析, 选择适当的主成分数 (累积贡献率超过 95%), 运用 SVMR 方法建立估算模型。

支持向量机的结构与神经网络的结构相似, 输出是中间节点的线性组合, 每个中间节点对应一个支持向量。基本思想是寻找一个最优分类面使得所有训练样本离该最优分类面的误差最小。设含有 l 个建模样本的建模样本集为 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中, $x_i (x_i \in R^d)$ 是第 i 个建模样本的输入列向量, 即经预处理后的各主成分, $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$, $y_i \in R$ 为对应的土壤 Mg 含量实测值。

设在高维特征空间中建立的线性回归函数为:

$$f(x) = w\Phi(x) + b \quad (1)$$

式中: $\Phi(x)$ 为非线性映射函数。

定义 ε 线性不敏感函数损失函数:

$$L(f(x), y, \varepsilon) = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon \end{cases} \quad (2)$$

式中: $f(x)$ 为回归函数返回的 Mg 含量预测值; y 为对应的实测值。若 $f(x)$ 与 y 之间的差别 ε , 则损失等于 0。

引入松弛变量 ζ_i, ζ_i^* , 并将上述寻找 w, b 的问题用数学语言描述出来, 即:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \\ s.t. \begin{cases} y_i - w\Phi(x_i) - b \leq \varepsilon + \zeta_i, i = 1, 2, \dots, l \\ -y_i + w\Phi(x_i) + b \leq \varepsilon + \zeta_i^* \\ \zeta_i \geq 0, \zeta_i^* \geq 0 \end{cases} \end{cases} \quad (3)$$

式中: C 为惩罚因子, C 越大表示对训练误差大于 ε 的样本惩罚越大, ε 规定了回归函数的误差要求, ε 越小表示回归函数的误差越小。

求解公式(3)时, 本项研究引入 Lagrange 函数, 并转换为对偶形式:

$$\left\{ \begin{array}{l} \max_{\alpha, \alpha^*} \left[\begin{array}{l} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) \\ -\sum_{i=1}^l (a_i + a_i^*)\varepsilon + \sum_{i=1}^l (a_i - a_i^*)y_i \end{array} \right] \\ s.t. \left\{ \begin{array}{l} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i \leq C \\ 0 \leq a_i^* \leq C \end{array} \right. \end{array} \right. \quad (4)$$

式中： $K(x_i, x_j) = \Phi(x_i)\Phi(x_j)$ 为核函数。

设求解公式(4)得到的最优解为 $\alpha = [a_1, a_2, \dots, a_l]$ ，

$\alpha^* = [a_1^*, a_2^*, \dots, a_l^*]$ ，则有：

$$w^* = \sum_{i=1}^l (a_i - a_i^*)\Phi(x_i) \quad (5)$$

$$b^* = \frac{1}{N_{nsv}} \left\{ \sum_{0 < a_i < C} \left[y_i - \sum_{x_j \in SV} (a_i - a_i^*)K(x_i, x_j) - \varepsilon \right] + \sum_{0 < a_i < C} \left[y_i - \sum_{x_j \in SV} (a_j - a_j^*)K(x_i, x_j) + \varepsilon \right] \right\} \quad (6)$$

式中： N_{nsv} 为支持向量个数。

于是，回归函数为：

$$\begin{aligned} f(x) &= w^* \Phi(x) + b^* \\ &= \sum_{i=1}^l (a_i - a_i^*)\Phi(x_i)\Phi(x) + b^* \\ &= \sum_{i=1}^l (a_i - a_i^*)K(x_i, x) + b^* \end{aligned} \quad (7)$$

式中：只有部分参数 $(a_i - a_i^*)$ 不为零，其对应的样本 x_i 即为问题中的支持向量。

研究运用 MATLAB2012a 软件中的 LIBSVM-MAT-2.89 工具箱完成，选择自带的 epsilon—SVR 模型类型和径向基函数作为核函数，利用交叉验证 (Cross-validation) 选取惩罚参数 c 和核函数参数 g 进行建模。将估算值与检验样本实测值进行对比分析，估算精度通过相关系数 r 、均方根误差 (RMSE)、预测偏差比 (RPD)^[27]、平均相对误差、显著性检验等参数进行评价。

2 结果与分析

2.1 土壤化学成分的统计分析

由表 1 可以看出，研究采集的样本 Mg 含量相对较低，Fe、Al 和 Mn 含量则相对较高，这与研究区土壤为红壤及其发育的土壤有关，在土壤形成过程中存在强烈风化淋溶作用、脱硅富铝过程和生物富集作用，盐基离子流失。Mg 是土壤胶体重要的组成成分，对 K^+ 有较强的吸附能力，从而使得土壤中 Mg 与 K 的含量存在相关性；另外，pH 增加会提高土壤对 Mg^{2+} 的电性吸附和专性吸附，从而使土壤对 Mg^{2+} 的吸收能力增强^[28]，相关分析表明，土壤 Mg 与 K、pH 呈极显著相关，相关系数分别为 0.85、0.60。另外，Mg 与 Cu、Fe 含量之间也呈现显著相关性，相关系数分别为 0.33 和 0.32，Fe 的含水氧化物 $Fe_2O_3 \cdot nH_2O$ 对土壤 Mg^{2+} 的吸附有一定影响作用^[29]，说明研究区 K、Cu、Fe 等元素对 Mg 有一定的赋存作用。

表 1 土壤样本化学成分含量统计值
Table 1 Statistical values for measured chemical element contents of soil samples

统计值	K	Ca	Mg	Fe	Al	Mn	Cu	Zn
	(g/kg)			(mg/kg)				
平均值	10.37	0.50	2.80	26.84	67.04	134.90	11.32	66.01
最小值	3.63	0.03	1.21	10.66	35.81	47.16	0.00	25.14
最大值	19.15	0.84	6.46	50.46	96.42	249.69	24.66	98.47
标准差	3.687 6	0.20	1.13	8.95	14.28	49.23	7.05	17.15
变异系数(%)	35.44	39.96	40.41	33.36	21.30	36.49	62.31	25.98

2.2 土壤镁含量与土壤光谱的相关分析

从土壤 Mg 含量与土壤反射率及其变换光谱的相关分析 (图 1) 可以看出，在 1 480 ~ 1 860 nm 区域内，Mg 含量与反射率之间呈负相关关系，两者相关系数绝对值在 0.30 以上，达到 0.1 显著性水平，在其他波段范围相关系数绝对值较小，在 -0.39 ~ 0.26 范

围内波动。经过一阶微分变换后，相关性显著提高，在 1 010 ~ 1 270、1 410 ~ 1 560、2 240 ~ 2 300 nm 波长范围内皆达到 0.01 显著性水平，其中 2 240 ~ 2 300 nm 与王璐等^[30]估算土壤中 N、P、K 含量模型入选波段相似。在 1 100 nm 附近，两者负相关性最好，相关系数达到 -0.82，该波段范围是 Fe-OH 的光谱吸收带，

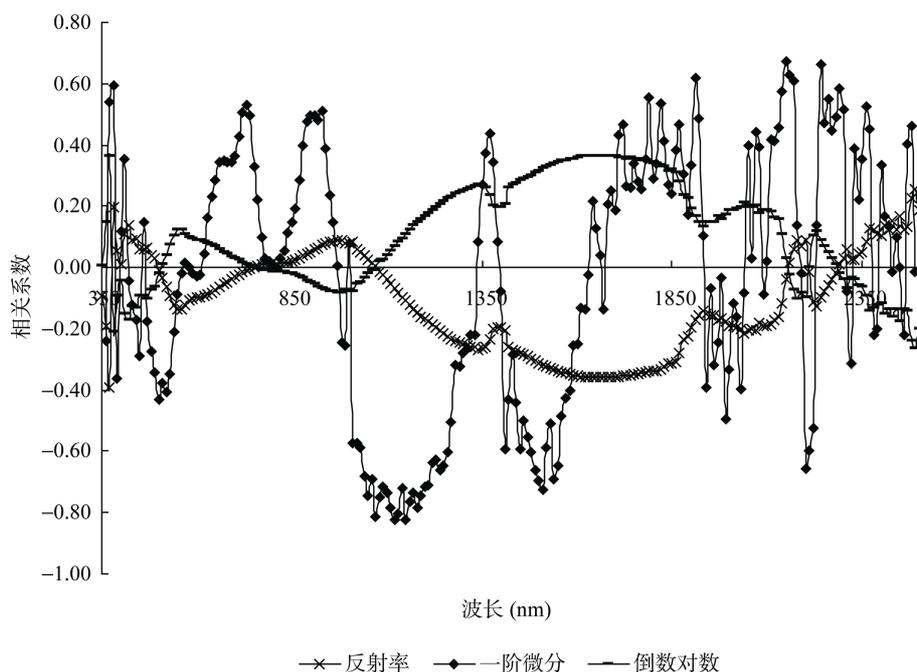


图 1 土壤 Mg 含量与反射率及其变换光谱的相关性

Fig. 1 Correlation between soil magnesium content and reflectance spectrum and its transformations

Fe 和 Mg 显著相关, Fe^{3+} 对光谱的吸收提高了此波段与 Mg 的负相关性; 在 1 910、2 250 nm 附近正相关性最好, 相关系数达到 0.62 以上, 该波段范围是土壤中结合水 H_2O 和金属 -OH 振动合频产生的吸收带, 其中 Mg-OH 的弯折振动基波波长在 2 300 nm 附近^[31-32]。Mg 含量与倒数对数在 1 460 ~ 1 860 nm 波长内呈现显著的正相关。由此可见, 土壤反射光谱存在与 Mg 含量连续的相关性, 且波段范围相对较大, 为 Mg 含量的估算提供可能。

2.3 不同光谱变换方法对估算精度的影响

估算结果表明(表 2), 反射率模型及由一阶微分、倒数对数变换后, 建立的土壤 Mg 含量回归模型检验相关系数均在 0.85 以上, 达到显著或极显著水平。从各项精度检验指标分析, 对于同一估算模型, 一阶微分变换均优于反射率和倒数对数变换, 其中均方根误差最小为 0.47 g/kg, 最高为 0.55 g/kg, 明显小于反射率和倒数对数变换; 预测偏差比相对较高, 最高可达 1.72; 平均相对误差相对较小, 其算数平均值为 19.24%, 平均估算精度达到 80% 以上; 倒数对数变换模型精度相近。这与微分变换可以减弱或消除背景影响和提高不同吸收特征的对比度有关, 倒数对数变换虽然增强了可见光区域的光谱差异, 减少了因光照条件变化引起的乘性因素影响, 但是变换后的光谱与 Mg 含量的相关性不高, 进而影响了模型精度^[18-20]。因此, 在对土壤全 Mg 的定量估算中, 一阶微分变换较反射率和倒数对数变换具有更好的应用潜力。

2.4 土壤镁含量的 SVMR 模型估算精度分析

与主成分回归分析方法(PCR) 和偏最小二乘回归分析方法(PLSR) 相比, 支持向量机回归分析(SVMR) 模型在土壤 Mg 含量的估算中, 均取得了相对较好的效果(表 2), 无论是原始反射光谱, 还是经过一阶微分或倒数对数变换, 估算值与检验样本值的相关系数最高, 并且一阶微分变换和倒数对数变换后的支持向量机回归分析模型均达到极显著水平。由于土壤组分复杂, 功能团多样, PCR 和 PLSR 方法仅将各主成分与 Mg 含量进行线性回归, 忽略了两之间可能存在的非线性关系, 存在一定的缺陷; 而 SVMR 方法最终决策函数可由少数的支持向量所决定, 可以抓住关键样本, 能够去除冗余样本, 且计算简单, 在处理非线性问题时优势明显。从检验精度指标看, 以一阶导数变换后支持向量机回归分析模型更好, 平均相对误差为 15.96%, 平均估算精度为 84.04%, 均方根误差只有 0.47 g/kg。

3 小结与讨论

本文将反射率、一阶微分、倒数对数光谱与土壤 Mg 含量进行相关分析, 得出相关性较高的波段, 运用主成分回归分析方法、偏最小二乘回归分析方法和支持向量机方法, 建立了反射率及其变换光谱的土壤 Mg 含量估算模型。结果显示, PCR、PLSR、SVMR 3 种建模方法对于 Mg 含量的估算中, SVMR 的预测精度相对较高。其中 PCR、PLSR 的平均估算精度分

表 2 基于不同模型和光谱变换的土壤 Mg 估算精度比较
Table 2 Precision comparison of magnesium content estimation based on different models and spectral transformations

模型	光谱变换	引入的主成分数	相关系数	RMSE(g/kg)	RPD	平均相对误差(%)	系数检验
PCR	反射率	1	0.60	0.68	1.18	24.46	显著
	一阶微分	11	0.86	0.55	1.47	22.50	极显著
	倒数对数	1	0.60	0.68	1.18	24.27	显著
PLSR	反射率	2	0.69	0.65	1.24	24.44	显著
	一阶微分	2	0.85	0.50	1.62	19.26	极显著
	倒数对数	2	0.69	0.64	1.26	23.22	显著
SVMR	反射率	1	0.73	0.53	1.52	20.25	显著
	一阶微分	11	0.86	0.47	1.72	15.96	极显著
	倒数对数	1	0.78	0.61	1.32	20.92	极显著

注：PCR：主成分回归；PLSR：偏最小二乘回归；SVMR：支持向量机回归。RMSE：均方根误差，RPD：预测偏差比。

别为 76.26%、77.70%，经 SVMR 处理后，平均估算精度为 80.96%，分别提高了 6.16% 和 4.20%。运用反射率、一阶微分、倒数对数光谱数据参与回归分析发现，一阶微分光谱优于反射率和倒数对数光谱，其平均估算精度为 80.76%，比反射率和倒数对数光谱分别提高了 4.95% 和 4.61%。研究区土壤样本经一阶微分变换，运用支持向量机回归进行建模，可以相对较好地估算土壤中全 Mg 含量，精度达 84.04%，确定为三者中最优模型。本研究建立的支持向量机估算模型对于土壤 Mg 含量的预测具有一定适用性，较传统的主成分回归和偏最小二乘回归预测精度显著提高。土壤反射光谱是土壤内部理化性质共同作用的结果，实现某一组分的准确估算，必须考虑其他性质对光谱的影响，这些性质会影响该组分的光谱特性，造成光谱信息提取困难，从而降低模型的估算精度。目前，土壤理化特性遥感估算主要采用统计方法进行某一特性与土壤光谱信息的相关性分析，进而建立预测模型，因此不具备普适性，精度也难以得到显著提高，今后需要进一步研究土壤的光谱机理，建立光谱机理模型，以提高模型的精度和适用性。

致谢：中国科学院遥感应用研究所博士后路鹏、研究生黄晶晶，北京师范大学研究生孙知文和南京信息工程大学遥感学院杨安民、朱万强同学参与了野外土壤样品采集工作，特此致谢！

参考文献：

- [1] 陆景陵. 植物营养学[M]. 北京: 北京农业大学出版社, 2003: 66-71
- [2] 姚艳敏, 魏娜, 唐鹏钦, 李志斌, 余强毅, 许新国, 陈佑启, 何英彬. 黑土土壤水分高光谱特征及反演模型[J]. 农业工程学报, 2011, 27 (8): 95-100
- [3] 彭杰, 向红英, 王家强, 柳维扬, 迟春明, 牛建龙. 基于

- 野外实测高光谱数据的干旱区耕作土壤含水量反演研究[J]. 干旱地区农业研究, 2013, 31 (2): 241-246
- [4] 陈红艳, 赵庚星, 李希灿, 朱西存, 隋龙, 王银娟. 基于小波变换的土壤有机质含量高光谱估测[J]. 应用生态学报, 2011, 22 (11): 2 935-2 942
- [5] 栾福明, 张小雷, 熊黑钢, 张芳, 王芳. 基于不同模型的土壤有机质含量高光谱反演比较分析[J]. 光谱学与光谱分析, 2013, 33 (1): 196-200
- [6] 王淼, 潘贤章, 解宪丽, 王昌昆, 刘娅, 李燕丽, 潘剑君. 土壤含水量对反射光谱法预测红壤土壤有机质的影响研究[J]. 土壤, 2012, 44 (4): 645-651
- [7] 汪善勤, 舒宁, 蔡崇法. 基于 PCA 和 SRRI 的潮土土壤属性与田间光谱关系研究[J]. 土壤, 2008, 40 (6): 966-970
- [8] 彭杰, 周清, 张杨珠, 向红英. 有机质对土壤光谱特性的影响研究[J]. 土壤学报, 2013, 50 (3): 517-524
- [9] 徐永明. 基于实验室光谱的土壤营养元素反演研究[D]. 北京: 中国科学院遥感应用研究所, 2005: 1-49
- [10] 陈红艳. 土壤主要养分含量的高光谱估测研究[D]. 泰安: 山东农业大学, 2012: 1-96
- [11] Cheng WC, Laird DA. Near-infrared reflectance spectroscopic analysis of soil C and N[J]. Soil Science, 2002, 167(2): 110-116
- [12] Malley DF, Yesmin L, Eilers RG. Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy[J]. Soil Science Society of America Journal, 2002, 66 (5): 1 677-1 686
- [13] Chodak M, Ludwig B, Khanna P, Beese F. Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands[J]. Journal of Plant Nutrition and Soil Science, 2002, 165 (1): 27-33
- [14] Thomasson JA, Sui R, Cox MS, Al-Rajehy A. Soil reflectance sensing for determining soil properties in precision agriculture[J]. Transactions of the ASAE, 2001, 44 (6): 1 445-1 453
- [15] 黄鸿翔, 陈福兴, 徐明岗, 秦道株, 高菊生, 朱永兴. 红壤地区土壤镁素状况及镁肥施用技术的研究[J]. 土壤肥料, 2000(5): 19-23
- [16] 刘光崧. 土壤理化分析与剖面描述[M]. 北京: 中国标准出版社, 1996: 52

- [17] 沈润平, 丁国香, 魏国柱, 孙波. 基于人工神经网络的土壤有机质含量高光谱反演[J]. 土壤学报, 2009, 46(3): 391-397
- [18] Kemper T, Sommer S. Estimate of heavy metal contamination in soil after a mining accident using reflectance spectroscopy[J]. Environmental Science & Technology, 2002, 36(12): 2 742-2 747
- [19] Tsai F, Philpot W. Derivative analysis of hyper-spectral data[J]. Remote Sensing of Environment, 1998, 66: 41-51
- [20] Yoder BJ, Pettigrew-Crosby RE. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales[J]. Remote Sensing of Environment, 1995, 53 (2): 199-211
- [21] Serrano L, Penuelas J, Ustin SL. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals[J]. Remote Sensing of Environment, 2002, 81(2-3): 337-354
- [22] Jolliffe IT. Principal Component Analysis[M]. New York: Springer, 1986: 167-198
- [23] 王慧文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999: 200-222
- [24] 黄长平, 刘波, 张霞, 童庆禧. 土壤重金属 Cu 含量遥感反演的波段选择与最佳光谱分辨率研究[J]. 遥感技术与应用, 2010, 25 (3): 353-357
- [25] 叶磊, 骆兴国. 支持向量机应用概述[J]. 电脑知识与技术, 2010, 6(34): 9 815-9 816
- [26] Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer-Verlag, 1999: 123-170
- [27] 郑光辉, 周生路, 吴绍华. 土壤砷含量高光谱估算模型研究[J]. 光谱学与光谱分析, 2011, 31(1): 173-176
- [28] 廖丽霞, 胡红青, 贺纪正. 中南地区几种地带性土壤表面电荷特性与 pH 的关系[J]. 华中农业大学学报, 2005, 24(1): 29-32
- [29] 于天仁, 季国亮, 丁昌璞. 可变电荷土壤的电化学[M]. 北京: 科学出版社, 1996: 294-309
- [30] 王璐, 葡启忠, 贾东, 石火生, 黄秀华. 多光谱数据定量反演土壤营养元素含量可行性分析[J]. 环境科学, 2007, 28 (8): 1 822-1 828
- [31] Viscarra RRA, Walvoort DJJ, McBratney AB. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties[J]. Geoderma, 2006, 131 (1-2): 59-75
- [32] Viscarra RRA, Chen C. Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils[J]. Remote Sensing of Environment, 2011, 115 (6): 1 443-1 455

Application of Support Vector Machine on Soil Magnesium Content Estimation Based on Hyper-Spectra

TIAN Ye¹, SHEN Run-ping^{1*}, DING Guo-xiang²

(1 School of Remote Sensing, Nanjing University of Information Science & Technology, Nanjing 210044, China;

2 Anhui Meteorological Bureau, Hefei 230001, China)

Abstract: Based on the measured reflectance spectra of soil samples in the laboratory, the relationship between soil magnesium content and reflectance spectra was analyzed, and the methods of principal component regression (PCR) and partial least squares regression (PLSR) were compared with support vector machine regression (SVMR) analysis for soil magnesium content estimation, then the estimation models of magnesium content between the reflectance spectra and its transforms were established to provide the basis for soil magnesium content estimation by hyper-spectra. The results showed that the average prediction accuracy of SVMR model reached 80.96%, higher than PCR and PLSR models by 6.16% and 4.20%, respectively. The first derivative of the reflectance spectra obtained the best outcome in the different mathematical transforms with an average estimation accuracy of 80.76%, higher by 4.95% and 4.61% than reflectivity and reciprocal logarithmic transforms, respectively. Therefore, SVMR model of first derivative transform was optimal to estimate the total magnesium content with an accuracy of 84.04%.

Key words: Hyper-spectra; Soil magnesium content; Support vector machine regression