

应用基于 PLSR 的土壤-环境模型预测土壤属性*

王昌昆^{1,2} 潘贤章^{1†} 周睿^{1,2} 刘娅^{1,2} 李燕丽^{1,2} 解宪丽¹

(1 中国科学院南京土壤研究所, 南京 210008)

(2 中国科学院研究生院, 北京 100049)

摘要 土壤-环境模型对于正确理解土壤属性与环境因子间的关系, 以及进行土壤属性预测与制图均具有重要的意义。研究区位于陕西省长武县内多年退耕还林还草沟壑区域, 采集 72 个土壤表层样本, 选择 3/4 的样本作为建模集, 其余 1/4 的样本作为验证集; 环境因子选择容易获取的地形因子和由遥感影像提取的植被因子和湿度因子, 建立基于偏最小二乘回归 (PLSR) 的土壤-环境模型。结果表明: 全氮、速效钾、全钾、有机质与环境因子间均有显著相关性; 建立的 PLSR 模型可解释土壤属性的空间变异从 23% (全氮) 到 27% (全钾); 与逐步回归方法构建的模型相比, 利用 PLSR 构建的土壤-环境模型可以更好地表征土壤属性与环境变量间的关系, 拟合精度和预测精度也相对较高, 说明 PLSR 建立的模型可以更好地应用于相似区域的土壤属性预测。

关键词 土壤-环境模型; 土壤属性预测; 偏最小二乘回归 (PLSR)

中图分类号 P934 **文献标识码** A

分析与预测土壤属性的空间分布与动态变化是管理土壤资源的重要内容^[1]。1941 年 Jenny 从土壤形成的角度提出了土壤与成土因素间的关系方程: $S = f(\text{cl}, \text{o}, \text{r}, \text{p}, \text{t}, \dots)$ ^[2], 试图表达土壤与成土因素气候 (cl)、生物 (o)、地形 (f)、母质 (p)、时间 (t) 等之间的定量关系。该方程提供了土壤形成的概念模型, 为土壤属性预测提供了理论基础^[3]。

随着现代分析测量技术、3S 技术、统计技术的发展, 原来无法定量表达的成土因素信息可以实现量化表达, 使之可以作为模型输入参数的环境因子, 通过建立土壤属性与环境因子间的定量关系, 进而实现土壤属性的定量预测。由于量化的环境因子越来越容易获取, 近 30 年来, 世界各国土壤学家在利用环境因子定量预测土壤属性方面进行了大量的研究工作。McBratney^[4]总结了前人的研究成果, 概括了土壤属性定量预测模型主要包括: 地统计、分类树、模糊聚类、多元线性回归、人工神经网络等模型; 所使用的环境因子主要反映: 地形、植被、气候、母质成土因素信息。其中基于 DEM 数据源建立的土壤-景观模型^[1,5-10], 以及基于 DEM 和遥感数据源建立的土壤-环境相关模型^[11-17]更是得

到了广泛、成功的应用。

由于多元线性回归模型具有模型简单, 变量间的关系易于解释说明等优点, 因此被广泛地应用于构建土壤属性预测模型, 但由于各土壤成土因素相互作用, 相互影响共同决定了土壤的形成, 因此成土因素量化后的环境因子间不可避免地会存在多重共线性, 导致建立的土壤属性预测模型具有稳定性不强, 预测能力差的缺点。伍德和阿巴诺于 1983 年提出了一种新型的多元统计分析方法: 偏最小二乘回归 (Partial Least Squares Regression, PLSR), PLSR 集成了多元线性回归分析、主成分分析和典型相关分析的特点, 因而可以有效地解决环境因子间存在的多重线性问题^[18], 使所建的土壤属性预测模型具有更强的稳定性和更好的预测能力, 但国内外基于 PLSR 建立土壤-环境模型的例子较为少见。

本文研究区选在黄土高原沟壑退耕还林还草区域, 利用数字地形技术和遥感技术分别提取对土壤形成、发育具有较大影响的环境因子, 结合野外土壤样品采集, 分析土壤属性与环境因子间的关系, 并建立基于 PLSR 的土壤环境预测模型, 为

* 中国科学院战略性先导科技专项 (XDA05050509) 和国家自然科学基金项目 (41071140, 40801081) 资助

† 通讯作者, 潘贤章, 男, 研究员, 从事土壤景观及定量遥感研究。E-mail: panxz@issas.ac.cn

作者简介: 王昌昆 (1984—), 男, 博士研究生, 主要研究方向: GIS 与 RS 应用。E-mail: chrisgardenr@126.com

收稿日期: 2011-08-28; 收到修改稿日期: 2011-12-02

建立土壤-环境模型提供方法上的探索;并为该区域的土地利用,水土保持工作提供决策上的理论依据。

1 材料与方 法

1.1 研究区简介

研究区位于陕西省长武县北部的沟壑区。该县位于渭北与陇东高原结合部的过渡地带,具有沟

壑纵横,塬高沟深的地貌特征,土壤母质为黄土母质,植被主要为草类、农作物、林木。年均降雨量 578.5 mm,年均气温 9.1 °C,属暖温带半湿润大陆型季风气候,是典型的雨养农业区。研究区选择一个矩形区域(8.5 km × 3.9 km),面积约 33 km²(图 1),为多年退耕还林还草区域。研究区高程范围 902 ~ 1 211 m,坡度介于 0 ~ 78°,土地利用方式包括草地、林地,土壤为黄盖黏黑垆土(系统分类为简育干润均腐土)。

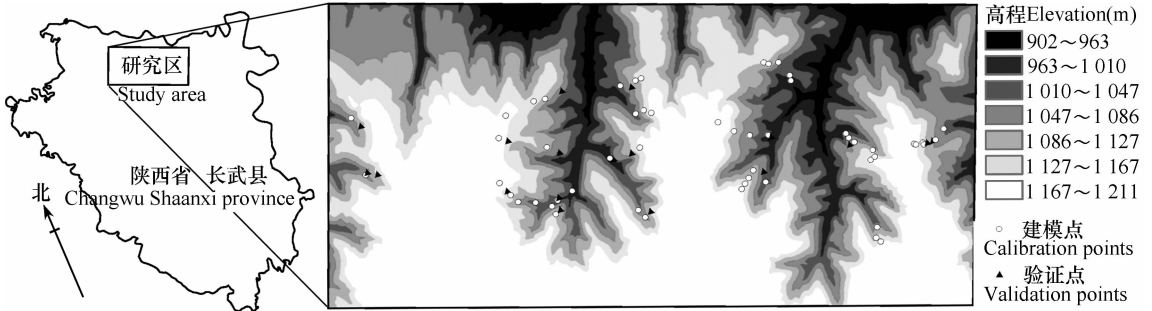


图 1 研究区位置、DEM 与采样点分布图

Fig. 1 Location and DEM of the study area and the distribution of sampling points

1.2 土壤样品采集

为了更好地研究土壤属性与环境因子间的关系,土壤样品采集仅在沟壑区内进行,在塬面平坡区不进行样品采集。在室内采样点布设时,主要考虑地形信息和植被信息,按照不同高程、坡度、坡向、坡位和土地利用方式布设采样点;在野外根据实际地形进行调整,采用 5 点采样法进行表层(0 ~ 20 cm)样品采集,等量均匀(四分法)混合后为一个样品,采样量为 1 kg,共采集 72 个土壤样品;并利用 GEO XT(2008) 高精度 GPS 设备记录采样点的三维坐标信息,定位精度可达亚米级。采样时间为 2009 年 11 月份,采样点分布情况见图 1。

1.3 土壤属性测定

测量的土壤属性包括:有机质、全氮、全钾、速效钾;各土壤属性具体测量方法参照国标^[19],即:有机质采用重铬酸钾容量法;全氮采用半微量开氏法;全钾采用氢氧化钠熔融-火焰光度法;速效钾采用乙酸铵浸提-火焰光度法。

1.4 环境因子提取

土壤-环境模型首先获取容易获得的环境因子,以及较少的野外土壤采集样品,然后建立土壤-环境模型,利用建立的预测模型进行土壤属性的预测^[12]。本文选取的环境因子包括有关的地形因子、反映植被状况的植被因子以及反映水分状况的湿度因子。

1.4.1 地形因子提取 地形是影响土壤与环境间进行物质、能量交换的一个重要场所。地形主要通过影响降水和辐射的再分配来影响土壤性质空间变异^[5]。地形因子通过 DEM 来提取,本研究采用 10 m 分辨率的 DEM 进行地形因子的提取,通过 ArcGis9.2 实现。地形因子分为基本地形因子和复合地形因子,基本地形因子是直接从 DEM 提取的,复合地形因子是通过基本地形因子计算得到的^[1]。本研究提取的部分地形因子如表 1 所示,已有的研究表明,这些地形因子与土壤属性间有着较好的关系^[5-6, 12, 16]。在定量分析中,对坡度进行正弦变换;由于坡向为环形变量(自北顺时针方向范围为 0 ~ 359°),同时由于研究区沟壑的走向为东南方向,因此对坡向做 $\sin(A/2 - \pi/4)$ 变换,表示朝南的程度。

1.4.2 植被因子和湿度因子提取 随着遥感技术的不断发展,利用遥感手段定量探测地表信息得到越来越广泛的应用。植被指数能很好地表征地表植被覆盖状况,由于比值植被指数(RVI)能增强植被与土壤背景间的辐射差异^[20],本文采用 RVI 作为植被因子。本文选择经穗帽(TC)变换提取的湿度特征(WI)作为湿度因子。TC 变换是由 Kauth 和 Thomas 两位学者提出的一种经验性的多波段图像线性正交变换,经 TC 变换得到的第三分量可以反映出土壤和植被的湿度特征^[20]。

表 1 地形因子及定义

Table 1 Topographic factors and their respective formulae^[1]

地形因子 Topography factors	公式 Formula
坡度 G (°) Slope gradient	$G = \arctan(\sqrt{p^2 + q^2})$
坡向 A (°) Slope aspect	$A = \arctan(q/p)$
平面曲率 K_h (1/100m) Horizontal curvature	$K_h = \frac{p^2 r + 2pqs + q^2 t}{(p^2 + q^2) \sqrt{1 + p^2 + q^2}}$
剖面曲率 K_v (1/100m) Vertical curvature	$K_v = \frac{p^2 r + 2pqs + q^2 t}{(p^2 + q^2) \sqrt{(1 + p^2 + q^2)^3}}$
地形湿度指数 TWI Topographic wetness index	$TWI = \ln(CA/\tan(G))$
径流强度指数 SPI Stream power index	$SPI = \ln(CA * \tan(G))$

注: 表格公式中, r, t, s, p 和 q 为函数 $z = f(x, y)$ 的偏导数。 $r = d^2 z/dx^2$, $t = d^2 z/dy^2$, $s = d^2 z/dxdy$, $p = dz/dx$, $q = dz/dy$ 。 Note: in the formulae, r, t, s, p and q are partial derivatives of the function $z = f(x, y)$

本文选择 TM 影像作为遥感数据源(数据来源于中国科学院计算机网络信息中心国际科学数据服务平台(<http://datamirror.csdb.cn>)), 影像为 Level 1T 标准地形校正产品; 影像日期选择在植被的生长季: 2009 年 7 月 23 日; 平均云量为 4.9%, 研究区位置处云量低于平均云量。遥感影像受到传感器、大气的影 响, 因此在定量分析时应当进行传感器定标和大气校正处理。随着时间的推移, TM 传感器出现了老化问题, 本文采用最新的传感器定标参数^[21]; 大气校正采用 MODTRAN4 辐射传输模型^[22]。

1.5 数据建模方法

本文采用 PLSR 方法建立土壤属性与环境因子间的定量模型。为了评价 PLSR 模型在构建土壤-环境模型的适用性与合理性, 同时利用逐步回归方法建立土壤-环境模型, 对两种方法建立的模型进行对比分析。

1.5.1 建模集和验证集的划分 本研究将所有样本划分为建模集和验证集两部分, 建模集用于构建土壤-环境模型, 验证集用于对构建的模型进行验证、评价。每个采样点有 4 个土壤属性, 各土壤属性量纲不同, 首先对土壤属性数据进行 Z 标准化(均值为 0, 方差为 1)处理, 这样每个采样点就组成一个土壤属性向量; 然后求取各个土壤属性向量的模, 按照土壤属性模值的大小进行排序, 由小到大每隔 3 个点(模最小值点作为验证点)选择一个作为验证点, 共选择 18 个验证点作为验证数据集, 即 1/4 的

采样点作为验证数据集^[4], 剩余 54 个点组成建模集。建模集和验证集的分布情况见图 1。

1.5.2 PLSR 模型建立 在建立 PLSR 模型之前, 对建模集环境因子进行多重共线性的诊断, 本文采用特征根判断法进行多重共线性的诊断, 通过分析环境因子特征根的值和条件数来判定多重共线性。

在构建 PLSR 模型时, 异常点对模型的稳定性及预测精度有较大的影响, 本文在构建模型时去除残差 (Residual) 大于 3、马氏距离 (leverage) 大于其 3 倍平均值的异常样本^[23]。在 PLSR 过程中, 提取成分的个数至关重要, 采用单一交叉验证的方法来确定提取成分的个数。即每个观测值既参与模型的建立, 又参与模型的评价, 以此来求得残差平方和 (Prediction Residual Error Sum of Squares, PRESS, 其体现了观测点的变动所引起的扰动误差), 然后求得所有残差平方和之和作为总的残差平方和, PRESS 值越大, 模型越不稳定, 最后按照总的残差平方和最小的原则确定提取成分个数^[18]。PLSR 模型的建立使用 SAS8.0 实现。

1.5.3 模型对比与评价 利用验证集数据对建立的模型进行评价。首先利用模型计算验证集数据的预测值, 然后与实测值比较, 以评价模型的预测效果。评价的指标有: 平均误差 (mean error, ME), 均方根误差 (root mean of squared error, RMSE)^[10]。

平均误差:

$$ME = \frac{1}{N} \sum_{j=1}^N (P_j - O_j)^2 \quad (1)$$

均方根误差:

$$RMSE = \frac{1}{N} \left[\sum_{j=1}^N (P_j - O_j)^2 \right]^{0.5} \quad (2)$$

式中, P, O, N, j 分别为预测值、实测值、观测数、观测样点。平均误差 (ME) 是对模型的总体准确度的度量, 其值越接近 0, 说明模型的总体偏差越小, 准确度越高; 均方根误差 (RMSE) 是对模型精度及稳定性的度量, 值越小说明模型的精度与稳定性越高。

2 结果分析

2.1 土壤属性描述统计

土壤属性描述统计分析结果见表 2。通常认为样本的变异系数 $CV \leq 10\%$ 时为弱变异, $10\% < CV \leq 100\%$ 为中等变异, $CV > 100\%$ 为强变异^[24]。

表 2 土壤属性描述统计分析

Table 2 Descriptive statistics of soil properties

土壤属性 Soil properties	极小值 Min	极大值 Max	均值 Mean	标准差 SD	变异系数 CV (%)
速效钾 AK(mg kg^{-1})	108.0	324.0	181.6	54.39	29.95
全钾 TK(g kg^{-1})	20.58	24.59	22.65	0.80	3.54
全氮 TN(g kg^{-1})	0.34	1.65	0.80	0.25	31.63
有机质 OM(g kg^{-1})	5.10	30.45	13.99	5.23	37.38

从表 2 可以看出速效钾、全氮、有机质具有中等程度变异;全钾具有较弱的变异。从均值和标准差来看,与黄土高原地区有关研究结果^[25]基本一致。

2.2 土壤属性与环境因子相关分析

土壤属性与环境因子间相关分析结果见表 3。

由表 3 可知,土壤属性与植被因子、湿度因子间均存在显著相关性,除 $\text{Sin}(G)$ 外,其他地形因子与土壤属性间不存在显著的相关性,说明在该沟壑区域,植被因子、湿度因子更适合于土壤-景观模型的构建。

表 3 土壤属性与环境因子相关矩阵

Table 3 Correlation matrix of soil properties and environmental factors

环境因子 Environment factors	速效钾 AK	全钾 TK	全氮 TN	有机质 OM
高程 Height(H)	-0.21	-0.2	-0.19	-0.22
坡度正弦 $\text{sin}(\text{Slope gradient})(\text{sin}(G))$	0.08	0.09	0.28*	0.31**
坡向正弦 $\text{sin}(A/2 - \pi/4)$	-0.07	-0.21	-0.12	-0.04
平面曲率 Horizontal curvature(K_h)	0.02	0.05	0.04	0.03
剖面曲率 Vertical curvature(K_v)	0.01	0.04	0.09	0.1
比汇水面积 Specific catchment area(SCA)	-0.05	-0.08	0.04	0.01
地形湿度指数 Topographic wetness index(TWI)	0.09	0.03	0.12	0.12
径流强度指数 Stream power index(SPI)	0.09	0.01	-0.11	-0.14
比值植被指数 Ratio vegetation index(RVI)	0.43**	0.44**	0.28*	0.29*
湿度指数 Wetness index(WI)	0.34**	0.40**	0.31**	0.27*

注: * 显著性水平为 0.05, ** 显著性水平为 0.01 Note: * Correlation is significant at the 0.05 level, ** Correlation is significant at the 0.01 level

速效钾、全钾与 RVI、WI 有显著的正相关,这主要是由于钾含量的增加能促进该区域作物的生长,因此钾与 RVI 呈现显著的正相关。全氮、有机质与 $\text{Sin}(G)$ 、RVI、WI 有着显著的正相关,在本研究区内,原来地形较为平缓地区多为耕地,而较陡区域则为自然次生草地或林地,在退耕还林还草后,较平缓区域有机质积累过程较短,因此有机质、全氮与 $\text{Sin}(G)$ 有着显著的正相关;另一方面,全氮、有机质含量的增加有助于该区域作物的生长,因此全氮与 RVI 间有着显著的正相关。

由于土壤形成受到多种因素的影响,土壤属性与某个地形因子的作用关系可能会因其他地形因子的作用而变小或变弱^[1,5],因而在相关分析中,H、 K_h 、 K_v 、SCA、TWI 与土壤属性之间的相关性并不显著。由表 3 可以看出,整体而言,土壤属性与地形因子间的相关性较弱,而与植被因子和湿

度因子相关性较好,这主要是由于研究区域长期土壤流失对地形造成较大影响,改变了原有的土壤属性与地形因子间的关系;另一方面,经过多年的退耕还林还草措施,研究区内的植被对土壤属性产生了更为重要的影响,因此土壤属性与植被因子间的关系明显好于地形因子与土壤属性间的关系。

2.3 土壤-环境模型构建

尽管 K_h 、 K_v 、SCA 与土壤属性之间的相关性不显著,但由于 K_h 、 K_v 、SCA 均具有明确的物理意义,前人的研究已经表明 K_h 、 K_v 、SCA 与土壤属性间具有较好的关系^[1,6,10,12],因此 K_h 、 K_v 、SCA 仍然参与 PLSR 模型的构建;由于复合地形因子 SPI、CTI 是由 SCA、G 计算得到的,因此在建模型时将 SPI、CTI 从环境因子中剔除,不参与 PLSR 模型的建立^[1]。

2.3.1 环境因子多重共线性诊断 从环境因子

多重共线性诊断结果(表 4)可以看出: 特征根的最小值为 0.001, 条件索引值最大可达到 85.84, 因此, 环境因子间必存在较强的多重共线性。通常认为, 如果环境因子的方差比例值在同一行同时较大(接近 1), 则二者之间存在多重共线性, 从表 4 可知, K_h 、 K_v 在同一行的方差比例值为 0.65、0.64, RVI、WI 在同一行的方差比例值为 0.47、0.69, Con(常数项)、H 在同一行的方差比例值为 0.99、0.94, 因此可以确定 K_h 、 K_v 间存在多重共线性, RVI、WI 间存在多重共线性, Constant 与 H 间存在多重共线性。由于环境变量间存在着多重共线性, 利用简单多元线性回归方法构建的模型稳

定性较差。

2.3.2 PLSR 模型 利用 PLSR 构建的土壤-环境模型见表 5, 提取的成分个数均为一个, 这说明通过提取较少的成分即可获得较多的环境变量信息, 从而有效地去除了环境变量间的冗余信息, 同时尽管提取的成分个数较少, 但建立的模型中仍然保有所有变量的信息, 进而使构建的模型具有较好的稳定性、移植性。

从拟合模型 R^2 可以看出, 可解释土壤属性空间变异的变化范围从全氮的 23% 到全钾的 27%, 不同的土壤属性的预测效果总体差别不大, 这与已有的研究成果^[26]类似。

表 4 环境因子多重共线性诊断结果

Table 4 Collinearity diagnosis of environmental factors

维数 Dim	特征根值 Eigenvalue	条件索引 Condition index	方差比例 Variance proportions								
			Con	H	Sin(G)	Sin(A/2 - $\pi/4$)	K_h	K_v	SCA	RVI	WI
1	5.819	1.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
2	1.499	1.97	0.00	0.00	0.00	0.00	0.22	0.22	0.02	0.00	0.00
3	0.605	3.10	0.00	0.00	0.00	0.01	0.02	0.18	0.66	0.00	0.00
4	0.528	3.32	0.00	0.00	0.00	0.01	0.65	0.44	0.11	0.00	0.00
5	0.260	4.73	0.00	0.00	0.03	0.87	0.03	0.00	0.03	0.01	0.00
6	0.174	5.78	0.00	0.00	0.50	0.02	0.01	0.00	0.07	0.01	0.04
7	0.101	7.59	0.00	0.00	0.36	0.01	0.03	0.14	0.07	0.12	0.06
8	0.012	21.68	0.01	0.06	0.07	0.01	0.03	0.01	0.00	0.47	0.69
9	0.001	85.84	0.99	0.94	0.04	0.06	0.01	0.00	0.05	0.39	0.20

注: Con 为常数项, H 为高程, Sin(G) 为坡度正弦值, Sin(A/2 - $\pi/4$) 为坡向正弦值, K_h 为平面曲率, K_v 为剖面曲率, SCA 为比汇水面积, RVI 为比值植被指数, WI 为湿度指数 Note: Con stands for constant, H for height, Sin(G) for sin(slope gradient), Sin(A/2 - $\pi/4$) for sin(A/2 - $\pi/4$), K_h for Horizontal curvature, K_v for Vertical curvature, SCA for Specific catchment area, RVI for Ratio vegetation index, and WI for wetness index

表 5 土壤属性与环境因子的 PLSR 和逐步回归结果

Table 5 PLSR and multiple stepwise regression between soil properties and environmental factors

土壤属性 Soil properties	偏最小二乘回归 PLSR			逐步回归 Stepwise regression		
	样本数(N) Number of samples	成分数 Number of factors	决定系数(R^2) Coefficient of determination	样本数(N) Number of samples	入选变量数 Number of selected variables	决定系数(R^2) Coefficient of determination
速效钾 AK	54	1	0.23	54	1	0.18
全钾 TK	51	1	0.27	54	2	0.23
全氮 TN	53	1	0.23	54	1	0.16
有机质 OM	53	1	0.23	54	2	0.23

注: 成分数为 PLSR 模型中提取的成分个数, 入选变量数为逐步回归模型中入选模型的变量个数 Note: "Number of factors" is the number of selected factors in the PLSR models, and "Number of selected variables", the number of selected variables in the stepwise regression

2.3.3 PLSR 与逐步回归模型对比与验证 利用逐步回归构建土壤-环境模型时,引入自变量的显著性水平为 0.05,剔除自变量的显著性水平为 0.1。从表 5 可以看出,最多只有两个环境因子被引入到模型中。从拟合模型的 R^2 来看,利用 PLSR 构建的模型拟合精度略高于利用逐步回归方法构建的模型。

对构建的土壤-环境模型,利用验证数据集进行土壤属性预测,与实测值的比较结果见图 2、表 6。从表 6 可以看出,利用 PLSR 和逐步回归方法构建的模型进行土壤属性预测,各土壤属性的平均误差均较小,说明模型的预测结果在整体上具有较高的准确性;各土壤属性的均方根误差较小,说明预测误差的变异较小,预测模型具有较高的可信度,可以利用构建的模型进行土壤属性的预测。对于速效钾而言,PLSR 模型预测的平均误差和均方根误差均小于逐步回归模型预测的平均误差和均方根误差,说明 PLSR 预测模型具有更高的预测精度。

因此,从拟合模型的 R^2 以及预测结果的平均误差和均方根误差来看,可以认为在该研究区 PLSR 模型略好于逐步回归模型,但总体差异不大。

表 6 土壤-环境模型验证结果

Table 6 Validation of established soil-environment models

土壤属性 Soil properties	偏最小二乘回归 PLSR		逐步回归 Stepwise Regression	
	平均误差 ME	均方根误差 RMSE	平均误差 ME	均方根误差 RMSE
	速效钾 AK (mg kg^{-1})	-0.34	8.64	0.67
全钾 TK (g kg^{-1})	0.44	0.32	-0.07	0.15
全氮 TN (g kg^{-1})	0.24	1.12	0.35	1.16
有机质 OM (g kg^{-1})	0.01	0.05	-0.04	0.06

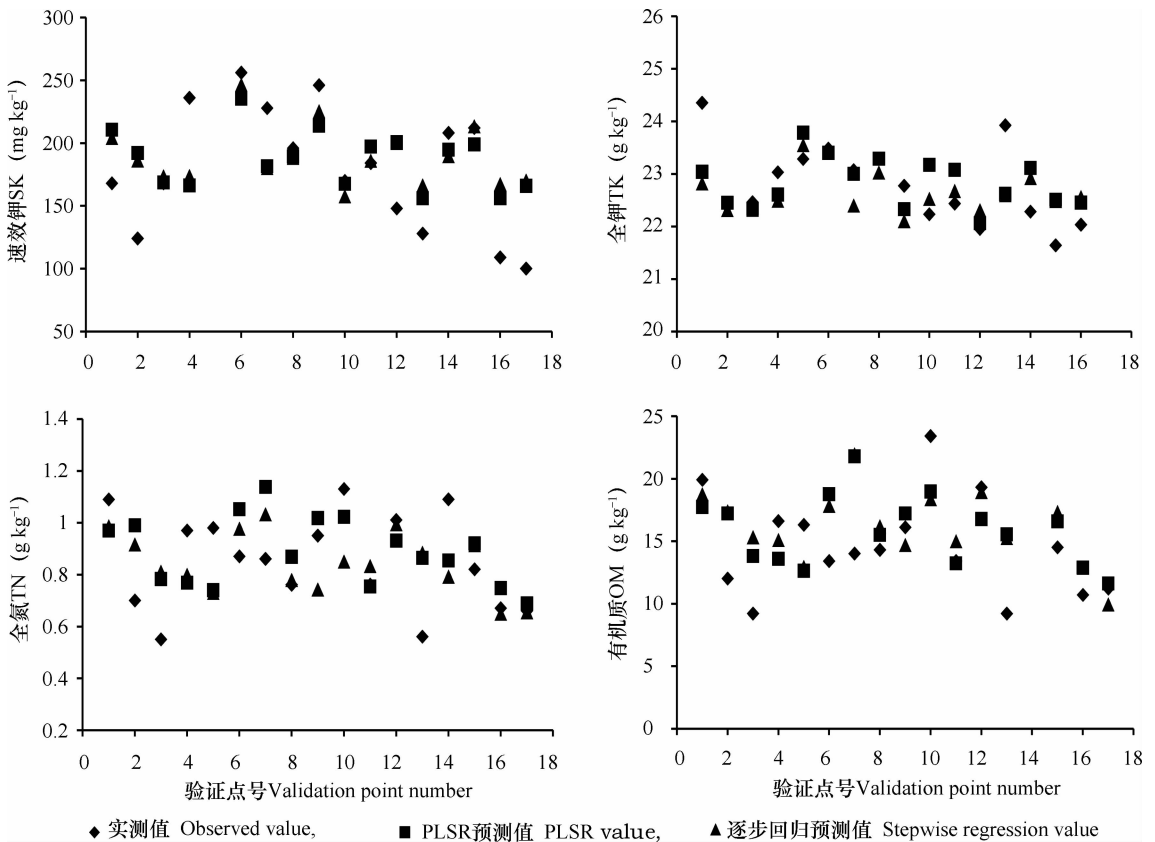


图 2 实测值与预测值比较

Fig. 2 Comparison between observed and predicted value of soil properties

3 结论与讨论

本文利用容易获取的环境因子,构建了基于 PLSR 方法的土壤-环境模型,研究结果表明:

1) 土壤属性与环境因子的相关分析表明,由遥感影像提取的植被因子、湿度因子与土壤属性间的关系较好;地形因子与土壤属性间的关系较差,主要的原因是在沟深坡陡区域,土壤侵蚀较为严重,改变了原有土壤属性与地形因子间的关系,可以选择其他的环境因子来定量地反映土壤属性信息。

2) 构建的 PLSR 模型,综合了地形因子与植被因子、湿度因子与土壤属性的作用关系,能够较好的表达土壤属性的空间变异,可解释土壤属性空间变异的变化范围从全氮的 23% 到全钾的 27%。

3) 与逐步回归方法构建的模型相比,利用 PLSR 构建的土壤-景观模型可以综合反映环境变量与土壤属性间的关系,同时拟合模型的精度也相对较高。

4) 应用构建的 PLSR 模型,可以进行土壤属性的预测,预测值与实测值的对比结果表明:各土壤属性的平均误差和均方根误差均较小,说明模型的预测结果在整体上具有较高的准确性,预测误差的变异较小,预测模型具有较高的可信度。

本研究利用 PLSR 方法构建的模型与已有研究^[1,5-6,13,26]采用的逐步回归方法构建的模型有着显著的不同,PLSR 构建的模型是利用土壤属性与提取的成分间建立回归方程,该成分包含所有环境因子的信息,而且有效去除了环境变量间的多重共线性,提高了模型的稳定性,而逐步回归方法是利用土壤属性与入选的变量间建立回归方程,该回归方程中仅包含入选变量的信息,从而忽视了其他环境变量的信息,由于逐步回归方程通常入选的变量过少,降低了模型的可移植性。因此,利用 PLSR 构建的土壤-环境模型可以更好地表征土壤属性与环境变量间的关系,使得构建的模型具有更好的稳定性和可移植性。利用 PLSR 方法构建的模型可以进行土壤属性的预测,并为土壤属性制图提供方法上的参考。本文研究区域为黄土沟壑退耕还林还草区,地形因子与土壤属性间的关系相对较差,通过结合地形因子和遥感因子,

进行土壤养分的预测取得了较好的效果,可以为类似区域的研究提供借鉴。

土壤-环境模型近年来得到广泛应用,构建模型主要基于地形因子及遥感因子,受到环境因子数据的空间分辨率和时相等其他因素的影响,预测精度普遍不高。已有的研究^[1,5-6,13,26]表明利用环境因子构建的土壤属性预测模型的决定系数通常在 0.15~0.82 的范围内,预测模型精度不高的原因除了土壤属性本身受环境因素影响的差异外,环境变量的选择也会对预测结果产生重要的影响。在研究土壤属性与地形因子间的关系时,不同的土壤属性对 DEM 分辨率的要求不同,采用合适分辨率的 DEM 能够提高模型的拟合精度^[27-28];在提取遥感因子时,遥感影像的时相选择和空间分辨率大小亦会影响土壤属性与遥感因子间的关系;McKenzie 等^[12]认为在利用土壤-环境模型预测土壤属性时,由于土壤属性空间变异的复杂性,有可能选择的环境因子不能有效地解释土壤属性的空间变异。Florinsky 等^[1]认为在研究土壤-环境模型时,应当考虑四种类型即区域、尺度、土层深度和时间之间的差异,因为这些因素最终都有可能影响模型的精度。

因此,后续研究应当充分考虑影响土壤属性空间变异的多个因素,避免漏掉对土壤属性有重要影响的环境变量;在构建土壤-环境模型时,应当选择合适分辨率的 DEM,遥感影像本身的质量亦会对预测结果产生重要的影响,因此还应当选择合适空间和时间分辨率的遥感影像,同时应考虑不同区域、不同土层深度、不同空间尺度和不同时间尺度对土壤-环境模型拟合精度和预测结果的影响。

参考文献

- [1] Florinsky I V, Eilers R G, Manning G R, et al. Prediction of soil properties by digital terrain modelling. *Environmental Modelling & Software*, 2002, 17(3): 295—311
- [2] Jenny H. *Factors of soil formation-A system of quantitative pedology*. New York: McGraw-Hill, 1941
- [3] Huggett R J. *Soillandscape systems-Model of soil genesis*. *Geoderma*, 1975, 13(1): 1—22
- [4] McBratney A B, Santos M L M, Minasny B. On digital soil mapping. *Geoderma*, 2003, 117(1/2): 3—52
- [5] Moore I D, Gessler P E, Nielsen G A, et al. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal*, 1993, 57(2): 443—452
- [6] Gessler P E, Moore I D, Mckenzie N J, et al. Soil-landscape

- modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Science*, 1995, 9 (4): 421—432
- [7] Chaplot V, Walter C, Curmi P. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 2000, 97(3/4): 405—422
- [8] Gessler P E, Chadwick O A, Chamran F, et al. Modeling soil-landscape and ecosystem properties using terrain attributes. *Soil Science Society of America Journal*, 2000, 64 (6): 2 046—2 056
- [9] Gobin A, Campling P, Feyen J. Soil-landscape modelling to quantify spatial variability of soil texture. *Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere*, 2001, 26(1): 41—45
- [10] Thompson J A, Pena-Yewtukhiw E M, Grove J H. Soil-landscape modeling across a physiographic region: Topographic patterns and model transportability. *Geoderma*, 2006, 133(1/2): 57—70
- [11] McKenzie N J, Austin M P. A quantitative Australian approach to medium and small-scale surveys based on soil stratigraphy and environmental correlation. *Geoderma*, 1993, 57(4): 329—355
- [12] McKenzie N J, Ryan P J. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 1999, 89 (1/2): 67—94
- [13] Dobos E, Micheli E, Baumgardner M F, et al. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*, 2000, 97(3/4): 367—391
- [14] Ryan P J, McKenzie N J, O'Connell D, et al. Integrating forest soils information across scales: spatial prediction of soil properties under Australian forests. *Forest Ecology and Management*, 2000, 138(1/3): 139—157
- [15] Park S J, Vlek P L G. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. *Geoderma*, 2002, 109(1/2): 117—140
- [16] Laffan S W, Lees B G. Predicting regolith properties using environmental correlation: a comparison of spatially global and spatially local approaches. *Geoderma*, 2004, 120(3/4): 241—258
- [17] Hughes M W, Schmidt J, Almond P C. Automatic landform stratification and environmental correlation for modelling loess landscapes in North Otago, South Island, New Zealand. *Geoderma*, 2009, 149(1/2): 92—100
- [18] 王惠文, 吴载斌, 孟洁. 偏最小二乘回归的线性与非线性方法. 北京: 国防工业出版社, 2006. Wang H W, Wu Z B, Meng J. Partial least-squares regression-linear and nonlinear methods (In Chinese). Beijing: National Defence Industry Press, 2006
- [19] 鲍士旦. 土壤农化分析. 北京: 中国农业出版社, 2000. Bao S D. Soil analysis in agricultural chemistry (In Chinese). Beijing: China Agriculture Press, 2000
- [20] 赵英时. 遥感应用分析原理与方法. 北京: 科学出版社, 2003. Zhao Y S. The principle and method of analysis of remote sensing application (In Chinese). Beijing: Science Press, 2003
- [21] Chander G, Markham B L, Helder D L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM +, and EO-1 ALI sensors. *Remote Sensing of Environment*, 2009, 113(5): 893—903
- [22] Guanter L, Richter R, Kaufmann H. On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing. *International Journal of Remote Sensing*, 2009, 30(6): 1 407—1 424
- [23] 何晓群. 现代统计分析方法与应用. 北京: 中国人民大学出版社, 2007. He X Q. Modern statistical analysis method and application (In Chinese). Beijing: China Renmin University Press, 2007
- [24] 马斌, 马琨, 何宪平, 等. 宁夏南部黄土高原流域土壤速效养分空间变异研究. *农业科学研究*, 2006, 27(3): 16—20. Ma B, Ma K, He X P, et al. Spatial heterogeneity of available soil nutrients in the loess plateau of southern Ningxia (In Chinese). *Journal of Agricultural Sciences*, 2006, 27(3): 16—20
- [25] 宋娟丽. 黄土高原草地土壤质量特征及评价研究. 西北农林科技大学, 2010. Song J L. The characteristics and evaluation on soil quality of grassland in loess plateau (In Chinese). Northwest A & F University, 2010
- [26] 连纲, 郭旭东, 傅伯杰, 等. 黄土高原县域土壤养分空间变异特征及预测——以陕西省横山县为例. *土壤学报*, 2008, 45(4): 577—584. Lian G, Guo X D, Fu B J, et al. Spatial variability and prediction of soil nutrients on a county scale on the Loess Plateau—a case study of Hengshan county, Shanxi province (In Chinese). *Acta Pedologica Sinica*, 2008, 45 (4): 577—584
- [27] Wu W, Fan Y, Wang Z Y, et al. Assessing effects of digital elevation model resolutions on soil-landscape correlations in a hilly area. *Agriculture Ecosystems & Environment*, 2008, 126(3/4): 209—216
- [28] 孙孝林, 赵玉国, 赵量, 等. 应用土壤-景观定量模型预测土壤属性空间分布及制图. *土壤*, 2008, 40 (5): 837—842. Sun X L, Zhao Y G, Zhao L, et al. Prediction and mapping of spatial distribution of soil attributes by using soil-landscape models (In Chinese). *Soils*, 2008, 40(5): 837—842

PREDICTION OF SOIL PROPERTIES USING PLSR-BASED SOIL-ENVIRONMENT MODELS

Wang Changkun^{1,2} Pan Xianzhang^{1†} Zhou Rui^{1,2} Liu Ya^{1,2} Li Yanli^{1,2} Xie Xianli¹

(1 *Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*)

(2 *Graduate University of Chinese Academy of Science, Beijing 100049, China*)

Abstract Soil-environment models can be of great importance to proper understanding of relationships between soil properties and environmental factors, and to predicting and mapping of soil properties, as well. A gully area where the “Grain for Green” policy had been implemented for years was selected in Changwu County, Shaanxi Province, a Loess Plateau region in China. A total of 72 surface soil samples were collected, and 3 fourths of the samples were used as a calibration set of samples and the rest as a validation set. Several easily acquired environmental factors, such as topographic factor, vegetation index and wetness index were used in a PLSR (partial least squares regression)-based soil-environment model established for the study. Quantitative analysis of relationships between environmental factors and soil properties of the samples was done. Results show that soil properties, including available potassium, total potassium, organic matter and total nitrogen, were significantly correlated with environmental factors. The PLSR-based model could well explain 23% to 27% the spatial variability of soil properties. Compared with the stepwise regression model used, the PLSR model was much better at characterizing soil-environment relationships with better fitting and prediction accuracy, suggesting that the PLSR-based model is applicable to prediction of soil properties of similar regions.

Key words Soil-environment model; Prediction of soil properties; PLSR