

文章编号: 1004-8227(2009)05-0489-06

ArcGIS 支持下的样本稀疏山区空间插值模拟探讨

秦建成

(重庆理工大学工商管理学院土地资源管理系, 重庆 400054)

摘要: 样本稀疏地区空间插值法对区域化变量的精准管理具有重要意义。基于 ArcGIS 9.0, 在分析土壤属性空间分布特征的基础上, 提出并构建了基于不同土壤类型的土壤特性空间预测模拟模型, 对比了传统方法与改进方法空间插值精度, 实现了数值插值在复杂地理环境区域的应用, 得到以下结论: (1) 基于经度、纬度、海拔高度及坡度等地理因子的土壤基础环境因子的空间预测模拟模型, 突破以往只能描述土壤属性在水平方向变化的局限, 较客观、合理地反映土壤属性随地理位置及海拔高度的立体变化特征; (2) 基于不同土壤类型回归模型来增加样本点以推断评价指标在无取样地区的分布状况的处理方式具有一定的数学理论支撑, 有效降低了插值误差, 提高了评价精度, 使评价结果更加接近现实。

关键词: 地统计学; 空间预测模型; 空间插值; 稀疏样本; 山区

文献标识码: A

土地资源自然组成要素的空间分布信息对于区域土地资源科学管理及可持续利用具有重要的意义。目前, 多数研究中所采用的基础数据一般都来自于有限的采样点数据, 进而在地理信息系统的空间插值功能的支持下, 生成连续分布的空间信息^[1~3]。研究表明, 只有在采样点密度比较大的地区, 插值方法的精度才比较可靠, 而在现有的研究理论和实际资料的基础上, 针对地理环境相对复杂的山区, 在采样点稀疏、组成要素空间变异强度大的条件下, 应该如何实现组成要素的区域化精准管理, 这是摆在我们面前亟需解决的问题。针对这一问题, 研究者们也提出了大量的插值模型和改进算法: 一类是数据上的改进, 即增加与组成要素相关的辅助地理信息^[4~6]; 另一类则是数学模型上的改进, 即在同样的数据源上, 改进新的数学算法, 以获取更优的插值模型^[7,8]。然而, 在采样点十分稀疏且分布不合理的山区, 数据和数学上的多元化对插值精度的影响还需作进一步探讨。

因此, 本文以土壤有机质为例, 试图提出一种改进的适宜山区土地组成要素精准管理的方法, 在不排除采样点原始数据空间变异特征的基础上, 系统分析自然地理因子与插值数据之间的强相关性, 在样本稀疏区域模拟采样点数据, 以增加原始采样数

据的空间变异信息, 再进行地统计学分析及空间插值。

1 材料与方法

1.1 研究区域与取样方法

研究区域在重庆市彭水县(107°48' ~ 108°36' E, 28°57' ~ 29°51' N), 属三峡低山丘陵地区, 地质地貌复杂, 相对高差 1 600 m, 温暖多雨, 光热资源偏低, 气候立体差异大; 成土母质种类多样, 土壤以黄壤、黄棕壤、潮土、紫色土、石灰岩土、水稻土为主, 主要分布在海拔 600~ 2 000 m 的低山丘陵区。根据研究区土壤类型, 兼顾不同生态条件及不同肥力水平的地块, 采用 GPS 定位, 采集耕层(0~ 20 cm)土壤, 最小样本间距为 400 m, 每个样点设置 4 次重复, 将 4 个重复的样品充分混合后, 用四分法取大约 1 kg 的土样带回实验室进行室内系统分析, 共采集 235 个样品(见图 1)。

1.2 数据测定方法

土壤有机质采用重铬酸钾容量法测定^[9]。

1.3 数据分析方法

采用 ArcGIS 9.0 软件的地统计学模块完成相关参数的计算。

收稿日期: 2008-04-17; 修回日期: 2008-07-04

基金项目: 重庆工学院博士科研启动基金(2007Q51); 重庆市烟叶公司专项基金项目(CQT006010)

作者简介: 秦建成(1976~), 男, 甘肃省兰州人, 副教授, 主要从事土地资源信息管理研究。E-mail: qinjiancheng1976@126.com

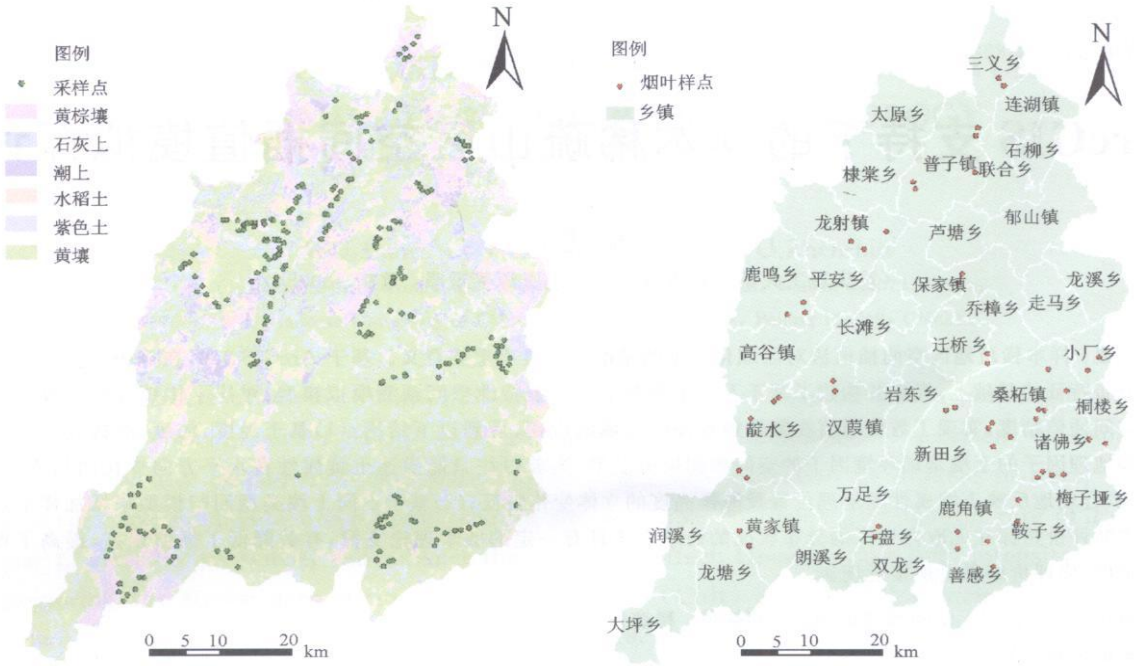


图 1 土壤样点和烟叶检验点分布图

Fig. 1 Distribution of Soil Samples and Tobacco Samples in Pengshui County

2 土壤有机质空间预测模型构建

土壤养分的空间分布依据不同土壤母质发育的土壤类型的不同而相应有所变化,具有相对独立的连续性^[10]。在传统地统计学插值原理中,假想研究区域是一个相对较为平坦的平面、区域化变量分布较为连续且有一定规律的范围,仅仅是通过周围点指标的样本数据来近似估算未测区域的指标值。但是作为土壤这个特殊的连续体,其区域化特征变量分布受母质、气候、人为等多种因素的影响。从土壤自身形成发育的过程来看,母质对土壤特性起到决定性影响。不同的土壤母质发育的土壤类型不同,从而导致多项土壤特性及养分状况存在显著的差异。此时如果不充分考虑土壤类型间的差异,应用插值法进行处理,将会在不同土壤类型边际区域产生较大的插值误差,特别是当取样点离边际区域较远时,插值误差将会更大。运用高等数学中函数理论可以将土壤类型间的空间差异解释为:同一种土壤,其理化性质及养分状况的分布函数具有连续性的特征;而在不同土壤类型间,当把土壤特性等作为一个整体进行研究时,其分布函数从理论上应该是一个分段函数,即 2 种土壤类型间存在跳跃间断点。而在目前多数研究中都没有将其进行考虑,而是作

为均质单元进行处理,随着研究尺度的不同,相应产生了不同程度的误差。

土地的自然环境评价因子与其地学因子,如经度、纬度、海拔高度及坡度等之间有着较为密切的关系。利用研究区 235 个实地取样得到的土壤有机质数据及相对应取样点的经度、纬度、海拔高度、坡度等地理信息数据,应用数理统计中的多元线性回归分析方法,将经度、纬度、海拔高度、坡度等地理因子作为自变量,把其他土壤适宜性评价的基础环境要素指标分别作为因变量,依据土壤类型,应用 SPSS 11.0 软件中多元线性回归分析方法分别构建土壤适宜性评价指标的空间预测推算模型。首先通过土壤图与采样点图的叠加,统计各种土壤类型的采样点数据;其次通过生成土壤有机质与各地学因子之间的数据散点图,确定两者之间的线性特点,应用 SPSS 11.0 软件程序查找并剔除各指标数据标准化残差的绝对值大于 3 的奇异值以及影响点;应用多元回归分析中默认的强行进入法使各地理因子全部进入回归模型,并计算土壤有机质在不同土壤类型中回归模型的复相关系数及 F 值。

从表 1 可以看出,土壤有机质与地学因子在不同土壤类型上构建的回归模型方程多数回归效果良好,可利用这些回归模型方程进行土壤适宜性评价各指标因子的小网格推算及推算未采样区域在空间

补点后的数据。研究依据上述不同土壤类型有机质的模型方程,采用在未采样区域随机、相对均匀布点的方法在研究区进行布点并依据不同土壤对应的指

标空间预测推算模型计算所布每个点的各指标变量特征值,共 535 组模拟点。

表 1 土壤有机质空间预测推算模型
Tab. 1 Spatial Estimating Models of Soil Organic Matter(SOM)

土壤类型	样本数	方程表达式	复相关系数	F 值
黄壤	150	$y = (22.100 - 0.246a + 0.185b + 0.00172v + 0.00378) \times 10$	0.828	305.115
黄棕壤	21	$y = (-168.785 - 1.843a - 1.038b + 0.0023v + 0.22206) \times 10$	0.741	164.433
石灰土	55	$y = (34.483 - 0.247a - 0.241b + 0.00231v - 0.00598) \times 10$	0.697	113.191
紫色土	9	$y = (67.476 - 0.428a - 0.663b + 0.00088v - 0.00428) \times 10$	0.679	99.165

3 结果与分析

为便于区域模拟精度比较,基于地统计学的土壤有机质空间变异特征研究采用传统方法与改进方法分别进行插值模拟。传统方式是以 235 个样本实测值作为基础数据进行插值;改进方式是以样本点实测值及用指标空间预测推算模型模拟的模拟值共 770 个数据值作为基础数据进行插值模拟。

3.1 土壤有机质 Fisher 统计特征

采用域法处理数据特异值,即样本平均值加减 3 倍标准差,在此区间以外的数据均定为特异值,然

后分别用正常最大和最小值代替特异值,最后根据 Fisher 理论对土壤有机质样本数据进行传统统计分析,具体采用样点的均值、标准差、峰度、偏斜度、最大值、最小值、变异系数等统计特征来描述土壤适宜性评价指标分布的统计特征值及其分布状态。

从表 2 可以看出:按传统数据方式,研究区域内土壤有机质的变异系数为 31.44%,属于中度变异;按改进数据方式,研究区域内土壤有机质的变异系数为 25.21%,也为中度变异,2 种数据处理方式对土壤有机质变异程度的强弱变化差异不显著,说明改进数据处理方式对土壤有机质自身的空间异质性影响小。

表 2 土壤有机质统计特征值
Tab. 2 Statistical Values of SOM

处理	均值	最大值	最小值	峰度	偏斜度	变异系数	标准差	分布
传统	26.59	53.02	5.75	7.07	1.33	31.44	8.36	非正态
改进	25.15	45.32	5.75	3.64	0.73	25.21	6.34	正态

3.2 土壤有机质地统计学特征

3.2.1 土壤有机质空间分布状态及正态转换

为了消除比较效应、提高预测精度,采用直方图法和正态 QQ 分布方法对不同数据处理方式下土壤有机质样本数据分布状况进行检验。从表 2 中可以看出,传统方式数据分布不是正态分布,需要转换成正态,而改进方式不需要转正态。研究采用对数变换使数据趋于正态^①。

3.2.2 土壤有机质空间趋势效应及方向效应

由于受成土因素、自然地理条件及人类种植活动的影响,区域土壤属性的空间格局常呈显著的趋势分布和异向性特征。在考虑各向异性的情况下,比较不同数据方式趋势效应参数为 0 阶、一阶、二阶和三阶所造成的插值误差(见表 3),依据趋势判断标准^[11],二阶趋势预测 ME 和 MSE 的绝对值最接近于 0,ASE 与 RMSE 最接近,且 RMSSE 最接近

于 1,故趋势参数宜选择二阶。

表 3 土壤有机质趋势效应分析
Tab. 3 Semi-variogram Parameters with Different Trends of SOM

处理	趋势效应	预测误差				
		ME	RMSE	ASE	MSE	RMSSE
传统	无 Non	-0.03474	7.511	6.755	-0.00757	1.1140
	一阶 1-order	-0.03947	7.524	6.769	-0.00821	1.1130
	二阶 2-order	0.01290	7.803	7.297	-0.00425	1.0680
	三阶 3-order	-0.01883	7.858	7.224	-0.00476	1.0870
改进	无 Non	-0.02866	5.255	5.624	-0.00554	0.9368
	一阶 1-order	-0.02813	5.249	5.603	-0.00548	0.9393
	二阶 2-order	-0.01089	5.232	5.545	-0.00256	0.9473
	三阶 3-order	-0.01211	5.378	5.700	-0.00252	0.9439

3.2.3 土壤有机质空间变异特征

从表 4 可以看出,在各向同性条件下,2 种处理

① 曾 杉, ArcGIS 地统计分析实用指南[M]. 北京: ArcInfo 中国技术咨询与培训中心, 2002: 110~ 116.
©1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

土壤有机质的块金值与基台值比值分别为 68.18% 和 72.04%, 说明在各向同性条件下土壤有机质在对应变程范围内均为中等程度空间相关; 在各向异性条件下, 2 种处理在不同方向上指标的拟合模型和基台值有所差异。土壤有机质在不同方向基台值相对较大, 说明在相应方向上变异程度较大, 但在总体上各方向相应变程内均呈现中等程度空间相关。从整体来看, 2 种处理有机质在 4 个方向上的变异程度十分接近, 但具有不同的变程, 因此可近似认为几何异向性。

通过对 2 种不同数据处理方式得到的指标半方差函数模型类型及其参数值分析结果表明: 有机质所对应的自相关阈值均大于最小样本尺度 0.4 km, 具有较强空间相关性, 其分布特性由区域因素和非区域因素共同决定^[12, 13]。一方面, 中、低山地形严重影响了土壤基础环境的生态演变, 使得有机质的空间随机变异、结构变异差异显著; 另一方面, 施肥、栽培管理制度等非区域因素也强烈影响土壤有机质的空间相关性。因此, 在 2 种因素共同作用下, 研究区土壤有机质空间变异呈现出 2 种处理方式分析所得的空间相关性。

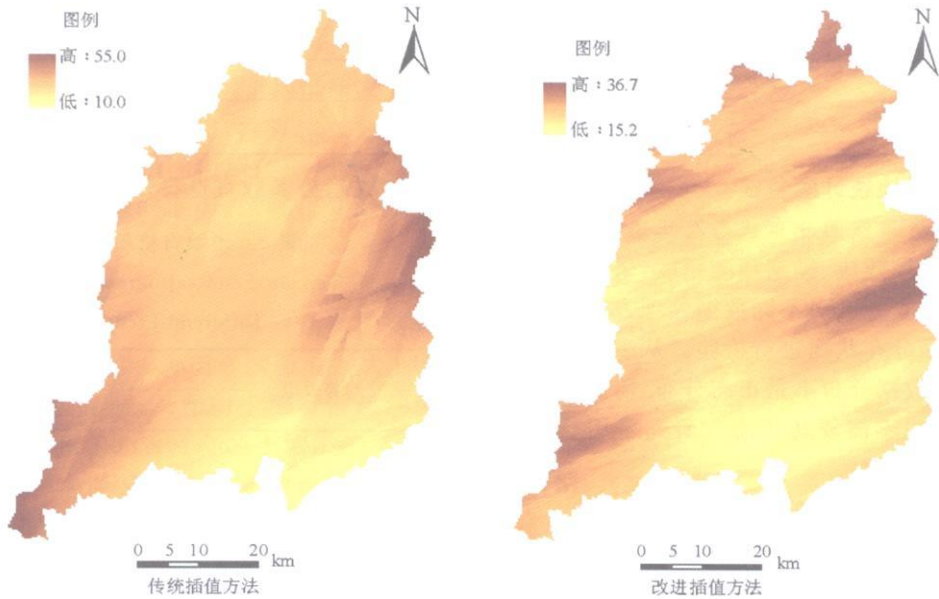


图 2 2 种方法所得有机质空间预测图
Fig. 2 Spatial Prediction Maps of SOM Based on Two Methods

3.3 土壤有机质空间插值模拟结果检验

土壤有机质空间插值模拟结果的正确性必须通过实际验证才能得到认可, 传统检验方法就是通过研究区针对某种作物选取一些经济性状指标如产

表 4 半方差函数模型类型及其参数值

Tab. 4 Semivariogram Model of SOM and Its Parameters

处理	主轴方向	模型	块金值	基台值	Nugget / Sill	变程 (km)
传统	—	E	0.060	0.088	68.18	26.9
	0	S	0.066	0.088	75.00	27.6
	45	E	0.059	0.090	65.56	67.0
	90	S	0.065	0.088	73.86	41.4
	135	E	0.059	0.090	65.56	23.2
改进	—	E	0.268	0.372	72.04	14.8
	0	S	0.304	0.374	81.28	14.8
	45	S	0.326	0.374	87.17	79.2
	90	S	0.304	0.374	81.28	45.4
	135	S	0.326	0.374	87.17	14.8

注: S 表示半球模型; E 表示指数模型。

3.2.4 土壤有机质空间模拟预测

在趋势分析、方向效应和半方差模型分析的基础上, 考虑评价指标的各向异性特征, 采用普通 Kriging 插值法对 2 种数据处理方式下土壤有机质进行区域空间插值和预测, 分别得到预测图(图 2)。依据烟叶种植不同适宜度的划分标准^[14], 应用 ArcGIS 中的 Spatial Analyst 模块生成土壤适宜性评价指标空间格局的 Grid 等值线图并进行分类和计算(表 5), 进行有机质空间预测分析。

量、品质等对其进行一致性检验。研究采用 2005 年研究区涉及的 17 个乡镇的 54 个烟叶样品的等级指标与 2 种数据处理所得模拟结果进行检验。土壤有机质模拟结果的检验标准为: 高度适宜对应上等烟

或中等烟,中度适宜对应中等烟,勉强适宜和不适宜对应下等烟。将 54 个烟叶样点分别叠加到不同数据处理评价图上,检验土壤有机质评价结果与烟叶样本等级之间的对应关系。检验结果表明:传统数据处理所得结果与烟叶品质等级的对应数目达到 37 个,符合率为 68.51%;改进数据处理所得结果与烟叶品质等级的对应数目达到 40 个,符合率为 74.07%。检验结果表明,基于改进数据的插值处理可以有效提高评价精度,使评价结果更加接近现实。

表 5 有机质适宜性评价

Tab.5 Evaluation on Suitability of SOM Based on Two Methods				
数据处理方式	不适宜	勉强适宜	中度适宜	高度适宜
传统 比例(%)	1.54	27.55	46.66	24.25
面积(hm ²)	6 010.62	107 527.65	182 113.98	94 647.75
改进 比例(%)	3.09	12.53	36.28	48.10
面积(hm ²)	12 060.27	48 904.59	141 600.84	187 734.30

4 讨论

地统计学理论和方法在土壤养分空间变异结构分析、空间局部估算及误差估计中的优越性已被证明^[15],而稀疏样点区域的插值方法发展至今,仍无法摆脱数据缺失所带来的插值难题。解决途径只能是增加辅助信息,然而以地统计学为基础的插值方法即使增加辅助变量,也无法从根本上解决稀疏样点数据的变异函数模型模拟问题,从而造成辅助变量的作用并不能充分发挥;另一方面以多元回归模型为基础的插值方法则完全依赖于回归变量的空间变异特征,这对于土壤这种受多种因素影响的自然要素来说,显然也不尽合理。本文试图从中另辟捷径,结合 2 种方法的优点,应用 GIS 强大的数据综合、地理模拟和空间分析能力,基于不同类型土壤的养分空间分布差异实际,以回归模型为基础,增设模拟站点反演其数据,增加实测数据的空间变异信息,突破以往只能描述土壤属性在水平方向变化的局限,较直观地反映土壤属性随地理位置及海拔高度的立体变化特征,再以地统计学模型进行分析插值,插值结果精度与传统插值结果相比具有较显著的提高。可见,在样本数据稀疏区域,利用原始数据通过评价指标的不同土壤类型回归模拟来增加样本点的

处理方法具有一定的数学理论支撑,不失为一种细化土壤属性空间变异过程的便利方法。此外,本方法也在理论上存在一些不足之处,如在模拟样点的布局上,仍然有较多的主观性和不确定性,这在进一步的研究工作中仍需要继续探讨和研究。

参考文献:

[1] 刘冬碧,余常兵,熊桂云,等.大比例尺度下土壤的养分特征及其空间变异性研究[J].华中农业大学学报,2004,23(5):524~527.

[2] 刘英华,张世熔,张素兰,等.成都平原地下水硝酸盐含量空间变异研究[J].长江流域资源与环境,2005,14(1):114~118.

[3] 陈防,刘冬碧,熊桂云,等.中亚热带两种水稻土壤养分空间变异的对比研究[J].土壤学报,2006,43(4):169~173.

[4] DROUGUE G, HUMBERT J, DERAISME J, et al. A statistical topographic model using an omnidirectional parameterization of the relief for mapping orographic rainfall[J]. International Journal of Climatology, 2002, 22(5): 599~613.

[5] MARQUEZ J, LASTRA J, GARCA P. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis[J]. Journal of Hydrology, 2003, 270(1-2): 1~11.

[6] 孙鹏森,刘世荣,李崇巍.基于地形和主风向效应模拟山区降水空间分布[J].生态学报,2004,24(9):1910~1915.

[7] DALY C, HELMER E H, QUINONES M. Mapping the climate of Puerto Rico, Vieques and Culebra[J]. International Journal of Climatology, 2003, 23(11): 1359~1381.

[8] WONG K W, WONG P M, GEDEON T D, et al. Rainfall prediction model using soft computing technique[J]. Soft Computing A Fusion of Foundations, Methodologies and Applications, 2003, 7(6): 434~438.

[9] 鲁如坤.土壤农业化学分析方法[M].北京:中国农业科学出版社,1999:205~206.

[10] HUGGETT R J. Soil chronosequences, soil development, and soil evolution: a critical review[J]. Catena, 1998, 32(3-4): 155~172.

[11] 黄元仿,周志宇,苑小勇,等.干旱荒漠区土壤有机质空间变异特征[J].生态学报,2004,24(12):2777~2781.

[12] 陈文辉,谢高地,卓庆卿.农田基础环境信息空间变异性分析[J].生态学报,2004,24(2):347~351.

[13] 侯文广,江聪世,熊庆文,等.基于GIS的土壤质量评价研究[J].武汉大学学报(信息科学版),2003,28(1):60~65.

[14] 刘国顺.烟草栽培学[M].北京:中国农业出版社,2003.

[15] 邱炳文,池天河,王钦敏,等.GIS在土地适宜性评价中的应用与展望[J].地理与地理信息科学,2004,20(5):20~23.

**SPATIAL INTERPOLATION RESEARCH IN HILLY REGIONS
WITH SPARSE SAMPLES BASED ON ARCGIS**

Q IN Jian-cheng

(Department of Business Administration, Chongqing University of Technology, Chongqing 400054, China)

Abstract: A study on distribution of soil nutrient on the field scale is important for improving agricultural management, and for assessing effects of agriculture on environmental quality and degrees of the influence of some random factors. However, soil nutrients are highly heterogeneous whether on a large scale or a small scale, and their heterogeneity results from many processes acting and interacting across a continuum of spatial and temporal scales. In this paper, a new method setting suppositional samples over regions with no measured data is presented. According to the high relationship between soil property and other physical factors, the data of suppositional samples could be simulated. Then a geo-statistical analysis and spatial interpolation method could be used with the enough samples data. This method doesn't completely rely on the relationship of soil property and other physical factors, and could be an easy way for soil property interpolation over regions with sparse measured samples. A case study in Pengshui County using geo-statistics of ArcGIS shows that (1) Tridimensional characters of soil property is represented reasonably by the spatial prediction model based on geographical factors; (2) the improved method integrating geographical environmental factors and spatial distributions of regional soil properties could get a better result with the least prediction error and more abundant spatial information than the traditional methods.

Key words: geo-statistics; spatial prediction model; spatial interpolation; sparse samples; hilly region