

# 基于协同克里格插值和地理加权回归模型的 土壤属性空间预测比较\*

郭龙 张海涛<sup>†</sup> 陈家赢 李锐娟 秦聪

(华中农业大学资源与环境学院, 武汉 430070)

## COMPARISON BETWEEN CO-KRIGING MODEL AND GEOGRAPHICALLY WEIGHTED REGRESSION MODEL IN SPATIAL PREDICTION OF SOIL ATTRIBUTES

Guo Long Zhang Haitao<sup>†</sup> Chen Jiaying Li Ruijuan Qin Cong

(College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China)

**摘要** 选取宜昌市红花套镇作为研究区域,研究土壤 pH、有机质、有效磷、速效钾、碱解氮与土壤属性指标变量之间的关系,选择与预测变量之间具有较高相关性的变量作为辅助变量用以提高预测精度,本文试图将地理加权回归模型应用于土壤属性空间模拟中,以此与协同克里格插值的预测结果进行对照,从而比较它们的预测精度以提出更适合土壤属性预测的模型。结果表明:协同克里格插值和地理加权回归模型对土壤属性的空间模拟均有较高的预测精度,在辅助变量较多的情况下地理加权回归模型具有比协同克里格插值更为简单的算法,并且比较预测值相对误差的范围跨度和标准差以及均方根误差等方面,地理加权回归模型在土壤属性指标预测方面具有更高的预测精度,也具有更大的优势。

**关键词** 协同克里格插值;地理加权回归模型;权函数;土壤属性;空间预测

**中图分类号** S151.9<sup>+</sup>5 **文献标识码** A

土壤是一种形态和演化过程均十分复杂的自然综合体,土壤属性具有复杂的空间变异性和非平稳性,对土壤形态和性质作定量化描述,尤其是对土壤过程的变异以及空间相关性和依赖性作定量描述的难度大<sup>[1]</sup>。协同克里格插值可以同时分析多个土壤属性之间的相互依赖性和地域性,不仅能够利用目标变量预测变量的相关性,同时也可以利用预测变量于目标变量间空间变异结构的相似性来提供预测精度,因此在土壤属性空间差异性分析中应用较为广泛<sup>[2-9]</sup>,但由于其算法复杂,并且在采样点密度的不均一性以及土壤属性权重强弱的取舍方面均存在一定的缺陷,为此其应用性有一定的局限性。地理加权回归模型由于方法简单易行,估计结果有明确的解析表示,还可以

对得到的参数估计进行统计检验,因此目前已被广泛的应用于社会经济学、城市地理学、气象学、森林学等诸多多学科领域<sup>[10-13]</sup>,并且借助于其权函数的灵活性可以很方便地弥补协同克里格插值所反映出的缺陷,从而为土壤属性的空间模拟提供更合适的模型算法。

本研究选取宜都市红花套地区作为研究区域,以土壤 pH、有机质(OM)、有效磷、速效钾、碱解氮在空间位置上与土壤属性指标等变量之间的关系为出发点,利用协同克里格插值和 GWR 模型来对 pH、OM、有效磷、速效钾、碱解氮分别建立回归拟合模型,并比较它们之间的估值精度,从而为土壤空间模拟提供更好的数学模型算法,为土壤属性空间变异性的研究提供新的思想。

\* 国家自然科学基金项目(40601073,41101192)、中央高校基本科研业务费专项资金项目(2011PY112,2011QC041)和华中农业大学自主科技创新基金(2011SC21)资助

<sup>†</sup> 通讯作者,张海涛(1974—),男,湖北京山人,博士,副教授,主要从事 3S 技术及其应用、土壤与土地资源调查与评价、水土资源数字化管理与数字国土。E-mail:gis@mail.hzau.edu.cn

作者简介:郭龙(1987—),男,山东临沂市人,硕士研究生,主要从事地理信息系统与国土资源研究。E-mail:Guolong027@gmail.com

收稿日期:2011-07-21;收到修改稿日期:2011-12-22

# 1 材料与方法

## 1.1 研究区域概况

本研究选择宜都市红花套地区为研究样区。宜都市位于长江流域中游,地处东经  $111^{\circ}05' \sim 111^{\circ}36'$ , 北纬  $30^{\circ}05' \sim 30^{\circ}36'$ , 全区面积  $2\,357\text{ km}^2$ 。红花套镇位于宜都市北部,总面积  $149\text{ km}^2$ , 现有耕地面积  $831.5\text{ hm}^2$ 。该区属于中亚热带向北亚热带过渡的气候、亚热带季风性区,年降雨量丰富。境内地形地貌以丘陵为主,土壤类型主要由黄棕壤土类、紫色土类、石灰岩性土、潮土类与水稻土类组成。

## 1.2 样品采集分析

将研究区 1:10 000 比例尺地形图与实地调查

相结合,并查阅该地区第二次土壤普查相关资料,选取一个地貌类型多样、地形起伏变化有较大差异、土壤类型具有典型性特点等具有代表性的研究区域,采用网状均匀布点模式,采样间隔为  $200\text{ m} \times 200\text{ m}$ ,利用差分式全球定位系统(DGPS)进行实地样点空间定位,最后在平原与耕地地区采集 129 个样点,在丘陵及地势不平坦地区则采集 200 个样点,采样深度为耕地种植区域表层采集  $0 \sim 20\text{ cm}$ ,果树林种植区域表层采集  $0 \sim 30\text{ cm}$ ,共采集得到 329 个土壤样本。样品分析采用文献[14]的方法进行测试,得到土壤 OM、pH、CEC、交换性镁、碱解氮、全氮、全钾、全铬、全铅、有效磷、有效铁、交换性钙、有效锌、速效钾、全磷以及全铜数据指标<sup>[14]</sup>。

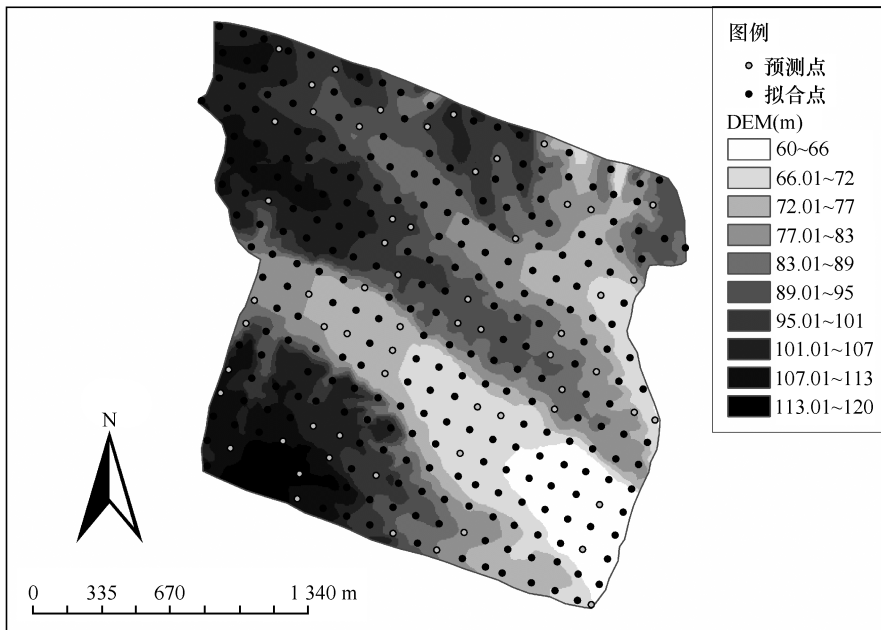


图1 湖北省红花套镇土壤采样点分布图

## 1.3 研究方法

影响土壤 pH、OM、有效磷、速效钾、碱解氮的因素很多,为得到较好的土壤属性空间分析结果,首先将这几个土壤属性指标分别与土壤全氮、全钾、全铬、全铅、有效铁、交换性钙、有效锌、全磷、全铜等进行相关性分析。然后选用协同克里格插值模型和地理加权回归模型分别对土壤属性建立模型,最后结合相对误差和均方根误差对模型的精度进行比较。

将全部样本点数 329 个分为两组,一组 268 个作为拟合数据,另外一组 61 个作为检验数据,选择

检验数据对两种拟合方法进行检验,主要针对相对误差的范围、平均值、标准差、方差、均方根误差等进行比较。

## 2 结果分析

### 2.1 土壤养分含量基本统计

使用协同克里格插值和地理加权回归模型的前提是实验数据要符合正态分布,因此首先对实验所测得的 16 个指标进行基本描述和 K-S 检验,如表 1 所示。

表 1 土壤养分含量描述性统计结果

	样本数	总和	均值	标准差	方差	变异系数(%)	渐进显著性
OM	329	5 583.22	16.97	4.22	17.82	24.87	0.99
pH	329	1 750.16	5.32	0.67	0.45	12.61	0.80
CEC	329	4 029.51	12.25	1.86	3.45	15.16	0.52
交换性钙	329	2 263.41	6.88	3.10	9.63	45.11	0.21
交换性镁	329	472.47	1.44	1.29	1.66	89.60	0.20
有效锌	329	2 388.23	7.26	3.16	9.97	43.50	0.19
有效铁	329	152 634.90	463.94	317.34	100 703.18	68.40	0.16
碱解氮	329	27 631.48	83.99	23.59	556.52	28.09	0.06
有效磷	329	6 449.24	19.60	24.02	577.00	122.54	0.00
速效钾	329	23 849.70	72.49	67.93	4 615.14	93.71	0.00
全氮	329	248.12	0.75	0.22	0.05	29.81	0.00
全磷	329	232.51	0.71	0.16	0.03	22.71	0.00
全钾	329	1 374.43	4.18	1.25	1.56	29.92	0.00
全铬	329	12 939.40	39.33	11.32	128.18	28.79	0.00
全铜	329	9 174.53	27.89	10.53	110.88	37.76	0.00
全铅	329	5 208.57	15.83	5.27	27.78	33.29	0.00

注:土壤属性元素总和与均值的单位为:OM( $\text{g kg}^{-1}$ ),全氮( $\text{g kg}^{-1}$ ),全磷( $\text{g kg}^{-1}$ ),全钾( $\text{g kg}^{-1}$ ),碱解氮( $\text{mg kg}^{-1}$ ),有效磷( $\text{mg kg}^{-1}$ ),速效钾( $\text{mg kg}^{-1}$ ),有效锌( $\text{mg kg}^{-1}$ ),有效铁( $\text{mg kg}^{-1}$ ),全铬( $\text{mg kg}^{-1}$ ),全铜( $\text{mg kg}^{-1}$ ),全铅( $\text{mg kg}^{-1}$ ),CEC( $\text{cmol kg}^{-1}$ ),交换性钙( $\text{cmol kg}^{-1}$ ),交换性镁( $\text{cmol kg}^{-1}$ )

变异系数的大小表示土壤属性空间变异性的 大小,变异系数  $< 10\%$  时为弱变异性,变异系数  $> 100\%$  时为强变异性,位于二者之间是中等变异性。从表 1 可知,研究区域土壤 pH、OM、有效磷、速效钾、碱解氮的变异系数均较大,说明这些元素的分布存在一定的空间变异性,且除了有效磷属于强变异类型之外,其他均属于中等变异类型。

当渐进显著性的值大于 0.05 时一般可以认为实验数据服从正态分布,因此对于不满足这一条件的实验数据先进行对数变化或者是 Box-Cox 变换,以期使数据服从正态分布,经过验证,有效磷、速效钾、交换性钙、交换性镁、有效铁、全铜、有效锌、全磷等经过对数变换之后所得到的数据组服从正态分布,而交换性镁无论进行何种变换均不符合要求,为此舍弃该组数据。

## 2.2 土壤养分含量之间的相关分析

土壤属性指标在土壤形成的过程中,经过长时间的相互作用和影响,其中的某些属性指标对另一些指标表现出不同程度的相关性,充分利用这一联系不仅有利于分析土壤属性内部的关系还能提高土壤属性指标空间预测精度,因此选用 Person 相关系数对需要进行估值的五个指标与土壤其他属性

进行相关性分析,得出与土壤 pH、OM、有效磷、速效钾、碱解氮相关程度最高的元素。与土壤 pH 的为:交换性钙、全铬、全铅;OM:碱解氮、有效铁、全氮;与有效磷的为:全磷、速效钾、交换性钙;与速效钾的为:CEC、有效磷、交换性钙;与碱解氮的为:有效铁、全氮、OM。从结论中可以发现,不同的土壤属性指标受到不同的土壤属性的影响,这主要是由于土壤在形成过程中受到当地地形、气候、环境以及母质等因素共同表现出来的结果,使土壤内部的理化性质表现出一定的地域性,借助于这种关系选取合适的土壤属性指标做为辅助因素用来提高土壤属性指标的空间预测精度,并结合相应的局部空间异质性和依赖性对不同的土壤属性指标进行预测。

## 2.3 协同克里格插值

利用在相关性分析中得出的与土壤 pH、OM、有效磷、速效钾和碱解氮有较高相关系数的土壤属性指标进行协同克里格插值,得出土壤属性值的区域分布图(图 2)。从图中可以发现不同的土壤属性指标在空间上表现出不一致性,碱解氮、速效钾、pH、有机质以及有效磷在土壤中含量的大小呈现出交互错落的特性,没有一定的空间规律,在局部特征上表现的比较明显。

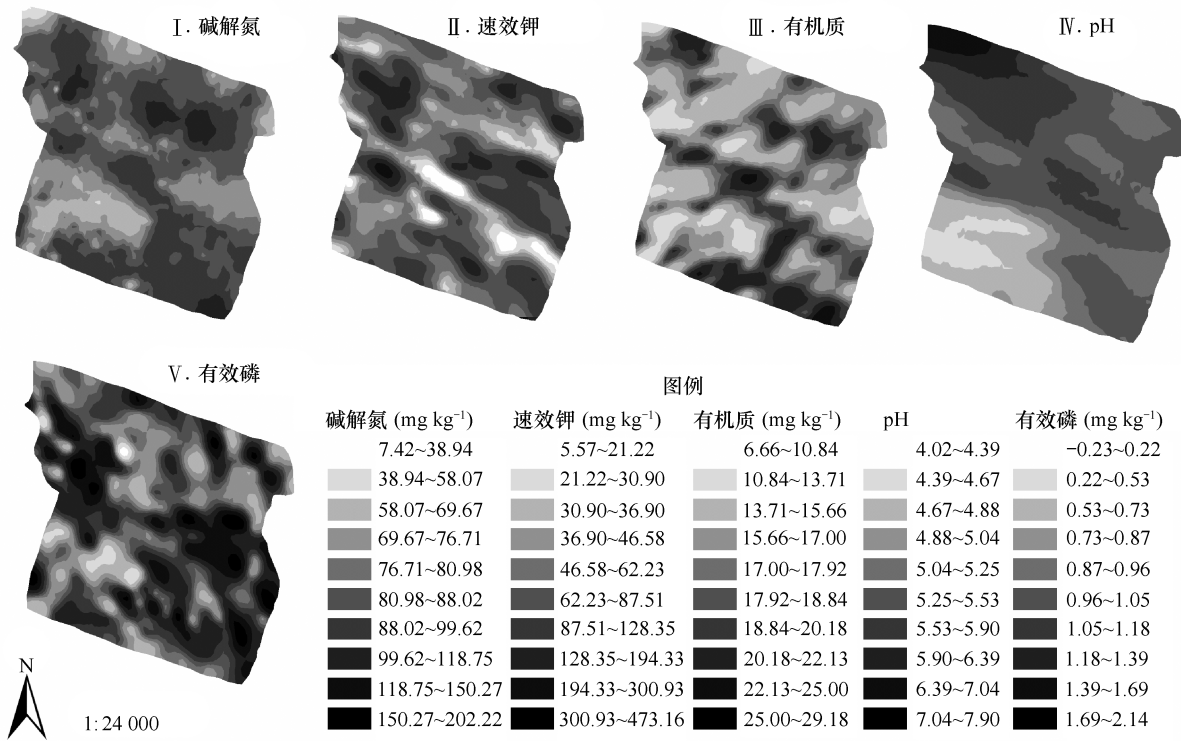


图2 利用协同克里格插值预测土壤属性空间分布图

表2 反应了每种土壤属性指标通过所构建的协同克里格插值模型进行交叉验证之后得到的实际变量与预测变量之间预测误差的统计量,表中的平均值基本接近于0 表明了所构建模型的预测具有无偏性,标准均方根误差的范围为 0.889 ~ 1.107 在 1

左右浮动,表明所构建的模型的误差有效,在可以接收的范围之内。均方根误差的范围为 0.29 ~ 58.79,平均标准误差的范围为 0.26 ~ 57.24,表明模型对于有效磷有较高的预测精度,而相对于速效钾而言则预测精度较差。

表2 土壤属性指标协同克里格插值预测精度统计量

	平均值	均方根误差	标准平均误差	标准均方根误差	平均标准误差
有效磷	0.003	0.290	0.010	1.107	0.262
pH	-0.004	0.556	-0.007	1.045	0.532
有机质	0.053	3.187	0.013	0.889	3.578
速效钾	0.099	58.797	0.003	1.026	57.240
碱解氮	0.157	15.474	0.009	1.003	15.463

## 2.4 地理加权回归模型

利用与土壤 pH、OM、有效磷、速效钾和碱解氮有较高相关系数的土壤属性指标作为自变量,利用 ArcGIS 提供的地理加权回归模型工具,带宽单位选择自适应,以满足采样点密度不均一的情况,同时结合校正赤池信息量准则 AICc 选择最优宽度,得出这五个土壤属性指标与对应的土壤属性指标之间的地理加权回归模型,AICc 代表的是模型的一种度

量,用于比较同种元素不同的回归模型。有效数量的取值和带宽的取值有关系,根据地理加权回归模型的自适应带宽选择,能够选择局部方差较小,而偏差也较小的中间最优值,根据表3 的结果可以发现有机质的邻域点个数为 9.10(平均值),更能表现出其局部空间性,而有效磷的局部特征相对较小一点,反应在其邻域个数上为 31.95 个(平均值)。均方差误差和  $R^2$  作为衡量拟合回归之后结果的好坏



来进行评定,均方差误差是正规化剩余平方和(剩余平方和除以残差的有效自由度)的平方根,它是残差的估计标准差,此统计值越小越好。分析表 5 可以得到除了碱解氮有较大的误差之外,其他土壤属性指标的结果比较满意,由于碱解氮直接选用的实验原始数据,考虑到其均值为  $83.99 \text{ mg kg}^{-1}$ ,因此会造成较大的均方差误差。 $R^2$  是拟合度的一种度量,其值在 0.0 到 1.0 范围内变化,值越大越好。此值可解释为回归模型所涵盖的因变量方差的比例。比较表 5 之后可以得到  $R^2$  值分布范围 0.57 ~ 0.66,考虑到样本的数量较大,这一范围满足数据预测的要求,为此应用地理加权回归模型对土壤 pH、OM、有效磷、速效钾和碱解氮拟合的数学模型可以用来预测未知区域的土壤元素值。

限于篇幅原因在此处仅以碱解氮为例来说明 GWR 模型对于土壤属性预测自变量随空间位置的

不同而表现出不同的权重,图 3 揭示了利用 GWR 模型对于碱解氮属性预测时模型中土壤有机质、全氮和有效铁的权重,从图中可以看出有机质的权重由北向南呈现出递减的趋势,在试验样区的北部有机质对于碱解氮的影响相对于南部来说贡献较大,与碱解氮呈现正相关;全氮的权重由西北向东南方向逐渐减小,与碱解氮表现出负相关;有效铁的权重范围为  $-2.4 \sim 3.68$ ,在不同位置表现出不同的相关性,并且在该研究区内北部主要表现出负相关,而西部则表现出正相关。从中可以得到每种元素所对应的土壤属性元素的权重空间分布值会因地理位置的不同而不同,表现出了土壤属性指标的局部空间依赖性和异质性,GWR 模型因自身的特性可以很好地结合土壤属性指标的空间性对于主变量进行预测,从而提高土壤属性的空间预测精度。

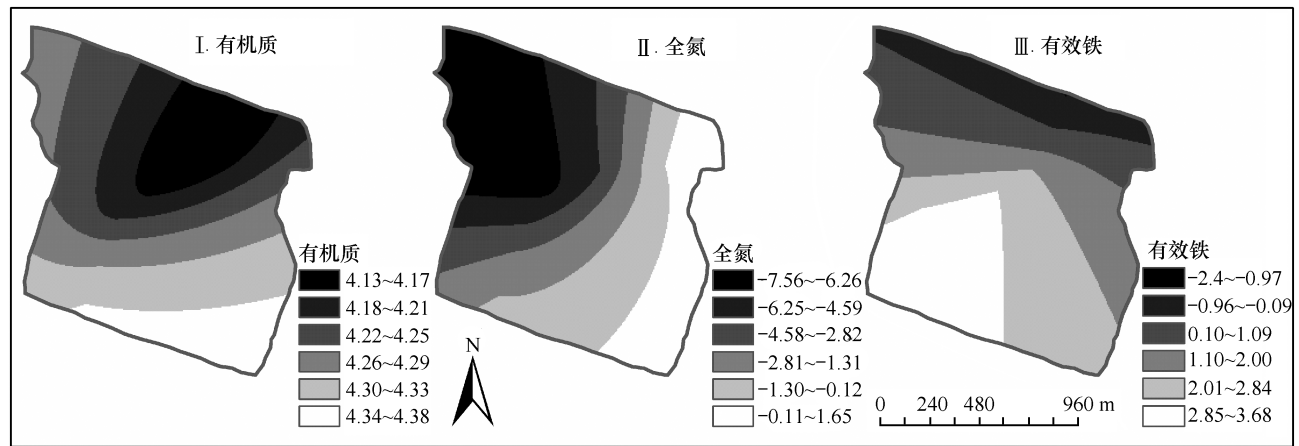


图 3 GWR 模拟土壤碱解氮属性指标的权重空间分布图

表 3 地理加权回归模型拟合的诊断变量

	OM	速效钾	碱解氮	pH	有效磷
邻域样点个数	9.10	14.25	9.00	27.71	31.95
均方根误差	2.46	0.26	15.13	0.47	0.28
校正赤池信息量准则(AICc)	1 250.09	53.37	2 224.30	372.28	99.47
模型拟合度( $R^2$ )	0.66	0.61	0.60	0.58	0.57
校正模型拟合度(Adjusted $R^2$ )	0.65	0.60	0.59	0.53	0.51

## 2.5 预测精准度检验

利用事先预留出来的 61 个样点来检验协同克里格插值和地理加权回归模型对土壤属性空间值模拟的精确度。利用相对误差(预测值与实际值之间的插值绝对值和实际值之间的百分比)和均方根

误差来对其预测精度进行说明。

$$y = \frac{x_{\text{预测}} - x_{\text{实际}}}{x_{\text{实际}}} \times 100\%$$

统计两种模型分别对五个土壤属性指标的预测值的结果,见表 4。

表4 协同克里格插值和地理加权回归模型的预测相对误差比较

预测个数	极小值		极大值		均值		标准差		均方根误差		
	CK	GWR	CK	GWR	CK	GWR	CK	GWR	CK	GWR	
有机质	61	0.16%	1.47%	172.23%	58.72%	22.30%	20.09%	28.23%	11.36%	4.47	3.92
有效磷	61	1.07%	0.06%	211.76%	168.46%	29.25%	22.10%	31.68%	29.48%	0.37	0.26
速效钾	61	0.16%	0.04%	42.86%	43.62%	13.35%	14.02%	10.62%	10.89%	0.56	0.42
碱解氮	61	0.17%	0.08%	223.10%	86.02%	4.24%	2.06%	33.08%	16.06%	0.32	0.30
pH	61	0.01%	0.16%	24.50%	23.84%	0.85%	0.65%	6.61%	5.08%	23.72	13.59

注:CK表示协同克里格插值模型,GWR表示地理加权回归模型

由表4可知,两种模型对五种土壤属性指标预测的相对误差的均值范围为0.85%~29.25%,说明两种模型的预测准确率均较高,表现出了较高的预测精度。利用协同克里格模型预测pH的相对误差的极小值能达0.01%。但是利用协同克里格插值预测碱解氮的极大值的误差为223.10%。

利用协同克里格插值预测土壤属性指标相对误差的标准差范围为6.61%~33.08%,地理加权回归模型相对误差的标准差范围为5.08%~29.48%,比较同一土壤属性指标的相对标准误差和均方根误差之后,发现地理加权回归模型所产生的误差相对于协同克里格插值而言均要小一些,这说明在选取的五种土壤属性指标的预测中,地理加权回归模型具有较高的空间模拟预测精度。

### 3 结论

1) 通过进行相关分析得出了与土壤pH、OM、速效钾、碱解氮、有效磷相关系数较高的土壤属性指标。与土壤pH相关性较高的为:交换性钙、全铬、全铅;OM的为:碱解氮、有效铁、全氮;有效磷的为:全磷、速效钾、交换性钙;速效钾的为:CEC、有效磷、交换性钙;碱解氮的为:有效铁、全氮、OM。

2) 利用协同克里格插值和地理加权回归模型对土壤pH、OM、速效钾、碱解氮、有效磷进行预测均可以得到较好的估值精度,其中土壤pH的预测精度最高。

3) 在本实验条件下,通过相对误差的范围以及标准差和均方根误差的比较,得出地理加权回归模型的预测精度相对协同克里格插值来说误差较小,误差范围跨度较小,更适合于对土壤属性空间变异性的分析和预测。

4) 本文考虑到研究土壤属性指标之间的内在关联特性,同时由于受到所选区域面积的限制,在辅助元素的选择上没有考虑地形、环境以及气候

对土壤属性指标的影响,这些将有待于进一步的研究。

### 参考文献

- [1] Campell J B. Spatial variation of sand content and pH within single contiguous delineation of two soil mapping units. *Soil Sci Soc Am J*, 1978, 42:460—464
- [2] Webster B, Burgess T M. Optimal interpolation and isarithmic mapping of soil properties III: Changing drift and universal kriging. *Soil Sci*, 1981, 31:505—524
- [3] Hurvich C M, Simonoff J S, Tsai C L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc*, 1998, 60:271—293
- [4] Leung Y, Mei C L, Zhang W X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ Plan*, 2000, 32:9—32
- [5] Mariusz S, Maciej K. Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wrocław (SW Poland). *Mariusz Szymanowski and Maciej Kryza/Procedia Environmental Sciences*, 2011, 3:87—92
- [6] Jansen B, Tonneijck F H, Verstraten J M. Selective extraction methods for aluminium, iron and organic carbon from montane volcanic ash soils. *Pedosphere*, 2011, 21(5): 549—565
- [7] 宋莎,李廷轩,王永东,等. 县域农田土壤有机质空间变异及其影响因素分析. *土壤*, 2011, 43(1): 44—49
- [8] 赵彦锋,化全县,陈杰. Kriging插值和序贯高斯条件模拟的原理比较及在土壤空间变异研究中的案例分析. *土壤学报*, 2011, 48(4): 856—862
- [9] 石淑芹,陈佑启,李正国,等. 基于土壤类型和微量元素辅助信息的土壤属性空间模拟. *农业工程学报*, 2010, 26(12): 199—206
- [10] Pan Z Q, Liu G H. The research progress of areal interpolation. *Progress in Human Geography*, 2002, 21(2): 146—152
- [11] 苏方林. 基于地理加权回归模型的县域经济发展的空间因素分析——以辽宁省县域为例. *学术论坛*, 2005, 5(5): 81—84
- [12] 覃文忠,王建梅,刘妙龙. 地理加权回归分析模型空间数据的空间非平稳性. *辽宁师范大学学报:自然科学版*, 2005, 28(4): 476—479
- [13] 覃文忠,王建梅,刘妙龙. 混合地理加权回归模型算法研究. *武汉大学学报:信息科学版*, 2007, 32(2): 115—119
- [14] 鲍士旦. 土壤农化分析. 北京:中国农业出版社, 2005