

基于主成分分析的协同克里格插值模型 对土壤铜含量的空间分布预测

章 清¹ 张海涛¹ 郭 龙² 杜佩颖¹ 李林蔚¹ 李锐娟¹ 唐晓霏³

1. 华中农业大学资源与环境学院, 武汉 430070; 2. 武汉大学资源与环境学院, 武汉 430072;
3. 湖北省测绘工程院, 武汉 430074

摘要 以湖北省宜都市红花套镇的柑橘产区为例, 选取土壤重金属全铜含量为研究对象, 利用 Pearson 相关系数方法选择与土壤铜含量具有较高相关性的土壤因子(全 K、全 Cr、CEC、全 Al、全 N($P < 0.05$))作为辅助变量, 利用主成分分析(PCA)对辅助变量进行降维求总得分处理后, 与协同克里格插值相结合构建土壤铜含量空间模型(COK_{PCA}), 同时构建土壤铜含量的普通克里格插值模型(OK)和以全 K 作为辅助变量构建协同克里格插值模型(COK_K), 对不同的空间模型进行模拟精度的对比和分析之后提出最优空间模型方法, 进一步分析土壤铜含量在研究区域的空间分布特征。结果表明: 普通克里格插值(OK)、协同克里格插值(COK_K)和与主成分相结合的协同克里格插值(COK_{PCA}) 3 种模型的 RMSE 分别为 7.044、6.907 和 6.309, R^2 分别为 0.716、0.743 和 0.852, 赤池信息量准则(AIC)分别为 101.591、96.908 和 87.203。综合比较, COK_{PCA} 具有最优的模拟插值结果, 其次为 COK_K, 而 OK 则相对较差。

关键词 主成分分析; 协同克里格插值; Pearson 相关系数; 铜; 微量元素; 空间分布; 估测

中图分类号 S 151.9⁺5 **文献标识码** A **文章编号** 1000-2421(2016)01-0060-09

随着计算机技术、空间技术和信息技术的逐步发展, 地理信息技术和各种数学模型算法作为空间分析的工具, 被广泛应用于土壤领域。利用地统计学和 GIS (geographic information system, GIS) 制图技术相结合的方法探索土壤空间分布情况和变异规律为目前最有效方法之一^[1]。土壤重金属元素在空间分布上具有较强的空间异质性和依赖性, 为此构建精确的空间模型是研究土壤重金属空间分布特征必不可少的手段和基础条件^[2-4]。普通克里格插值方法综合考虑了土壤属性的空间异质性和依赖性, 其最优无偏特性在模型的模拟和预测中得到较好的体现。因此, 普通克里格插值方法在土壤制图领域得到广泛的应用^[5]。在实际应用中, 土壤属性元素之间和土壤属性与环境影响因子之间存在着一定的相关性, 利用一些资料丰富并且有价值的辅助因素对于精准模拟和预测土壤属性具有重要的意义^[6]。普通克里格插值方法只能单纯地进行土壤属性的空间预测和模拟, 并不能利用多变量的优势, 而

协同克里格插值方法则可以进行多变量的估值, 不仅考虑变量的空间连续性, 同时还可以利用变量之间的相关关系。研究表明, 利用协同克里格方法能缩小极值误差范围、减小均方根误差, 提高实测值与预测值的拟合精度^[6]。因此, 采样点数量相同时协同克里格方法的结果精度明显高于普通克里格插值法^[7-8]。利用协同克里格法进行插值的关键在于辅助变量的选取, 以往研究方法在选取协助变量时单一地选取与目标变量相关性较好的变量, 不能全面地反映辅助变量的信息^[9-10]。在较多的变量中去除数据的冗余性和共线性, 提取有价值的信息构建新的变量, 可以更好地辅助协同克里格插值方法提高模型的预测精度。主成分分析方法可以利用降维的思想将多变量转换成几个综合指标, 简化数据集, 提取辅助变量中有价值的信息以用于模型的进一步研究^[11]。

土壤铜是作物生长发育必需的微量营养元素, 在作物的生理代谢过程中具有重要作用, 适量的铜

收稿日期: 2014-12-22

基金项目: 国家自然科学基金项目(41371227, 41101192); 中央高校基本科研业务费专项资金项目(2013JC016)

章 清, 硕士研究生, 研究方向: 地理信息系统与国土资源. E-mail: 841481270@qq.com

通信作者: 张海涛, 博士, 副教授, 研究方向: 3S 技术及其应用、土壤与土地资源调查与评价、土地资源数字化管理与数字国土. E-mail: zht@mail.hzau.edu.cn

能促进作物的正常生长和发育。研究土壤重金属的空间分布特征,并探讨影响其空间分布特征的主要影响因素对于土壤污染风险性评价和土壤环境质量保护具有重要意义^[1]。笔者以湖北省宜都市红花套镇(位于长江流域中游,地处东经 111° 05′ ~ 111° 36′,北纬 30° 05′ ~ 30° 36′,全区面积 2 357 km²,具有多年的柑橘种植历史)为研究区域,通过普通克里格插值法、协同克里格插值法及与主成分相结合的协同克里格插值法对土壤重金属铜含量进行插值分析,找出较优的内插方法,为小尺度下土壤全铜含量的空间分布研究提供参考。

1 材料与方法

1.1 研究区域概况

将红花套镇 1 : 10 000 地形图与实地调查相结

合,并查阅该地区第 2 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、土壤类型典型的研究区域(图 1)。该研究区域位于红花套镇东南部,总面积 4.25 km²,其中果园面积 3.02 km²。采用网状均匀布点模式,利用差分式全球定位系统(DGPS)进行实地样点空间定位,采样间隔为 200 m × 200 m。共采集土壤样本 329 个,其中平原与耕地地区采集样点 129 个,丘陵及地势不平坦地区采集样点 200 个。耕地种植区域和果树林种植区域均为表层采集,采样深度分别为 0~20、0~30 cm。测量土壤全 Cu、SOM、pH、CEC、碱解氮、全 N、全 K、全 Cr、全 Al、速效 P、速效 K 以及全 P 等^[12]指标。

1.2 数据预处理

进行克里格插值前,首先对土壤属性数据进行

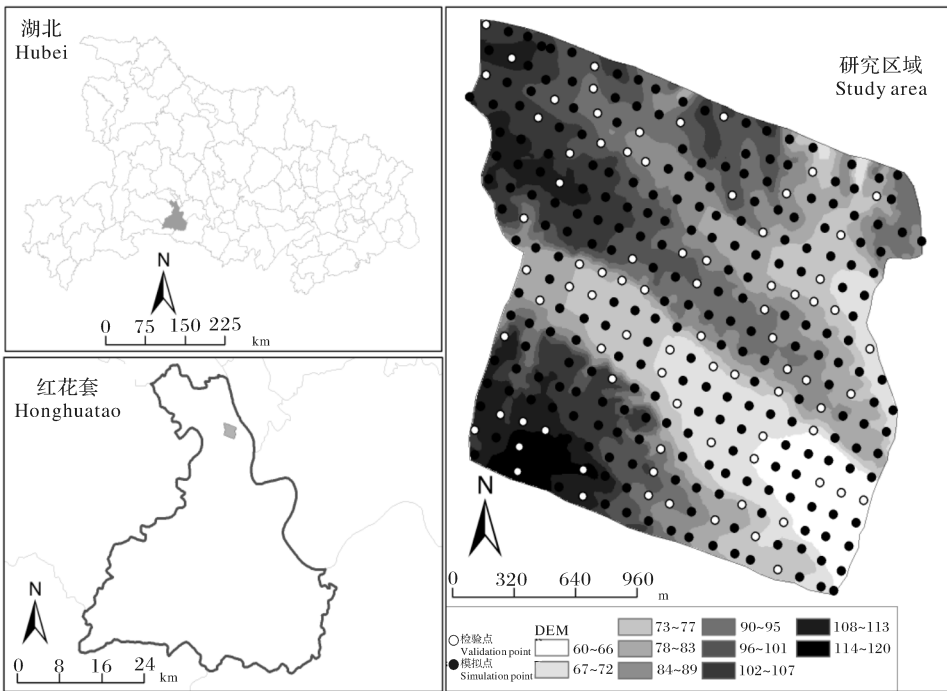


图 1 研究区域样点分布图

Fig.1 Map of sample sites in the study area marked with sampling cites

初步检查,包括数据的常规统计特征以及是否符合正态分布等,再对预处理后的数据进行相应插值,然后利用验证方法检验数据内插精度^[13-14]。从 329 个土壤样点中随机选取 248 个数据点作为模拟数据集构建土壤铜含量空间分布模型,预留 81 个数据点为验证数据集以作内插精度的比较。基于数据的统计特征分析、正态分布检验 SPSS 10.0 及 Microsoft Excel 软件实现,地统计学分析采用 ArcGIS 10.2 软

件中的 Geostatistical Analyst 模块进行半变异函数计算、理论模型的选择及交叉验证,绘制目标变量的空间分布图^[12-13]。

1.3 模型方法

在 1 组变量中找出其方差和协方差矩阵的特征量,将多个变量通过降维转化为少数几个综合变量,再转化为 1 个综合数值。土壤综合得分计算公式如下:

$$Z = \sum_{i=1, j=1}^n (\alpha_i \beta_j Z_{ij} / \sqrt{\lambda_i})$$

$$(i=1, 2, 3; j=1, 2, 3 \dots 329) \quad (1)$$

式(1)中, i 表示主成分编号, j 表示样点编号, α_i 表示该成分得分系数, Z_{ij} 表示第 i 主成分的第 j 样点主成分得分, λ_i 表示第 i 主成分的特征值, β_j 表示第 i 主成分的贡献率。

1.4 模型验证指标

选用评价指标 R^2 、赤池信息量准则(AIC)和均方根预测误差(RMSE), 其中 $T_{R_i^2}$ 、 T_{AIC} 和 T_{RMSE} 3 个指标分别表示以普通克里格插值模型(OK)的 R^2 、AIC 和 RMSE 指标作为最低标准, 来求取其他模型相对于 OK 模型的提高精度, 可以验证模型的模拟预测能力。其计算公式为:

$$A_{IC} = 2k - 2\ln(L) \quad (2)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{oi} - V_{pi})^2} \quad (3)$$

式(2)中, A_{IC} 表示赤池信息量准则(AIC), k 是参数的数量, L 是似然函数。式(3)中, σ 表示 RMSE, V_{oi} 为验证点土壤重金属铜的实测数据, V_{pi} 为预测数据, n 为验证点数量。

$$T_{R_i^2} = |(R_i^2 - R_{OK}^2) / R_{OK}^2| \times 100\% \quad (4)$$

式(4)中, $i = COK_K$ 、 COK_{PCA} , 而 R_{OK}^2 、 $R_{COK_K}^2$ 和 $R_{COK_{PCA}}^2$ 分别表示在对应的内插方法下的 R^2 , 其值越大表明对应的模型预测效果越好。 $T_{R_i^2}$ 是本研究中其他 2 种模型对比普通克里格插值其 R^2 提高程度。 T_{AIC} 和 T_{RMSE} 分别表示 COK_{PCA} 和 COK_K 模型

相比 OK 模型 AIC 和 RMSE 的精度提高率, 其计算公式与 $T_{R_i^2}$ 一致。

1.5 主要步骤

主要步骤为:(1)将理化分析后的原数据进行预处理, 包括正态检验、变异函数计算等。(2)数据经预处理后, 遵循随机选取的原则将 248 个数据点作模拟数据集, 另外 81 个数据点作验证数据集。(3)在模拟数据集中选取目标变量与其他土壤因子作相关性分析, 再进行相关性强弱排列($P < 0.05$), 其中相关性最高的作为辅助变量 W_1 。(4)将步骤(3)中与目标变量符合相关性条件的变量进行主成分分析^[14], 计算土壤综合得分作为辅助变量 W_2 。(5)通过普通克里格和协同克里格(协助变量分别为 W_1 和 W_2) 等 3 种方法对目标变量的空间分布进行预测。(6)用预留的验证数据集对 3 种预测方法的结果进行验证, 其中利用 6 种指标比较不同模型之间精度的大小。(7)依据土壤环境质量标准对土壤进行 3 个级别的划分, 直观地呈现出采样区土壤铜含量空间分布区域图。

2 结果与分析

2.1 数据分析

1) 数据统计特征分析。由表 1 可知, 研究区域土壤铬含量为 16.91~77.07 mg/kg, 平均值为 39.33 mg/kg; 铅含量为 2.09~40.97 mg/kg, 平均值为

表 1 土壤理化指标基本描述及正态分布检验

Table 1 Basic description of soil physical and chemical indicators and normal inspection

变量 Variable	N	极小值 Min	极大值 Max	均值 Mean	标准差 Standard deviation	正态 W 检验 W-test of normality	变换方式 Transformation method	变换后 W 检验 W-test of normality after transformation	变异系数/% Coefficient of variation
TCu/(mg/kg)	329	10.54	77.45	27.89	10.53	0.13	无 No		37.76
SOM/(g/kg)	329	6.07	29.18	16.97	4.22	0.77	无 No		24.87
pH	329	4.02	7.90	5.32	0.67	0.00	倒数 Reciprocal	0.77	12.61
CEC/(cmol/kg)	329	7.11	18.21	12.25	1.86	0.09	无 No		15.16
AN/(mg/kg)	329	7.42	202.22	83.99	23.59	0.10	无 No		28.09
AP/(mg/kg)	329	0.59	153.45	19.60	24.02	0.00	ln	0.19	122.54
AK/(mg/kg)	329	5.57	473.16	72.49	67.93	0.00	ln	0.19	93.71
TN/(g/kg)	329	0.32	1.72	0.75	0.22	0.00	平方根 Squareroot	0.16	29.81
TP/(g/kg)	329	0.33	1.32	0.71	0.16	0.00	ln	0.43	22.71
TK/(g/kg)	329	1.51	8.37	4.18	1.25	0.00	平方根 Squareroot	0.96	29.92
TCr/(mg/kg)	329	16.91	77.07	39.33	11.32	0.00	平方根 Squareroot	0.08	28.79
TPb/(mg/kg)	329	2.09	40.97	15.83	5.27	0.15	无 No		33.29

注 Note: SOM: 有机质 Organic matter; AN: 碱解氮 Available nitrogen; AP: 速效磷 Available phosphorus; AK: 速效钾 Available potassium; TN: 全氮 Total nitrogen; TP: 全磷 Total phosphorus; TK: 全钾 Total potassium; TCu: 全铜 Total Cu; TCr: 全铬 Total Cr; TPb: 全铅 Total Pb. 下同 The same as below.

15.83 mg/kg;铜含量为 10.54~77.45 mg/kg,平均值为 27.89 mg/kg。根据土壤环境质量标准(GB15618-1995)的规定,铬和铅属于一级土壤标准范围内,而铜含量有一定的超标($Cu \leq 35$ mg/kg,一级土壤; 35 mg/kg $< Cu \leq 50$ mg/kg,二级土壤; $Cu > 50$ mg/kg,三级土壤)。

对模拟的变量进行正态检验,并对不符合正态分布的数据集进行转换,以使其满足数据建模的基本条件。土壤全 Cu、有机质、CEC、碱解 N 和全 Al 都呈正态分布,而对其他不符合正态 W 检验的土壤因子进行相应的变换,最终变换后的所有数据都通过了正态 W 检验($P < 0.05$)。12 种土壤因子最大值和最小值之间差异显著,表明其在研究区内的值域分布广泛。样点土壤因子的变异系数为 12.61%~122.54%,属于中等强度的变异。速效 P 和速效 K 的变异系数分别为 122.54%和 93.71%,

说明人为因素对速效磷和速效钾空间变异性的影响较其他土壤因子强。

2) Pearson 相关系数分析。采用 Pearson(entered)相关系数 P_c 计算土壤其他因子与重金属铜之间的对应相关程度,并遴选出其中相关性较强的土壤因子进行分析。 P_c 主要是用来衡量 2 组分析对象线性关系的强弱,其取值范围为 $[0, 1]$,表达式为:

$$P_c = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (5)$$

其中, x 和 y 分别对应 2 组分析对象的数据序列, N 为样本个数, P_c 绝对值越大,相关性越强。

由表 2 可知,与全铜具有显著相关性的为 5 种土壤因子,其显著性程度为全 K > 全 Cr > 全 Pb > CEC > 全 N ($P < 0.05$),达到极显著水平,说明它们在研究区内受同样的区域化现象或空间过程的影响,

表 2 土壤因子与全铜的相关性

Table 2 Correlation of soil attributes and Cu

项目 Item	SOM	pH	CEC	AN	AP	AK	TN	TP	TK	TCr	TPb
相关性 Pearson correlation	-0.006	0.096	0.181**	-0.018	-0.051	0.058	-0.124*	0.044	0.309**	0.255**	0.147**
显著性 Significant	0.916	0.083	0.001	0.750	0.356	0.294	0.025	0.422	0.000	0.000	0.008

注 Note: ** , $P < 0.01$; * , $P < 0.05$.

响,属于协同区域化变量。再对这 5 种土壤因子进行主成分分析,得出其综合得分作协助变量来进行目标变量的协同克里格插值。

3) 土壤因子主成分分析。对全 K、全 Cr、全 Pb、CEC 和全 N 等变量进行主成分分析,转化为几个重要的主成分变量(表 3)。前 3 个主成分的累积贡献率已经达到 73%,而第一主成分的贡献率为 32.658%,即前 3 个主成分已经对 5 个检测指标所

涵盖的大部分土壤属性信息进行了概括。主成分的初始特征值分别为 1.633、1.197、0.844、0.672 和 0.653。表 4 表示的是成分得分系数矩阵,其中第一主成分主要反映的是 CEC(0.740)、全 K(0.635)和全 Pb(0.565);第二主成分主要反映的是全 Cr(0.640);第三主成分主要反映的是全 N(0.613)。主成分分析所观察到的关系表明,整个研究区域中的各种土壤因子间存在着难以直接得知的复杂关系。

表 3 解释的总方差

Table 3 Total variance explained

成分 Component	初始特征值 Initial eigenvalue			提取平方和载入 Extraction of square and load		
	合计 Total	方差比例/% Variance	累积/% Cumulation	合计 Total	方差比例/% Variance	累积/% Cumulation
1	1.633	32.658	32.658	1.633	32.658	32.658
2	1.197	23.942	56.599	1.197	23.942	56.599
3	0.844	16.889	73.489	0.844	16.889	73.489
4	0.672	13.442	86.931			
5	0.653	13.069	100.000			

表 4 成分得分系数矩阵

Table 4 Component score coefficient matrix

变量 Variable	成分 Component		
	1	2	3
CEC	0.740	-0.048	0.275
TN	-0.457	0.541	0.613
TK	0.635	-0.439	0.243
TCr	0.393	0.640	-0.529
TPb	0.565	0.547	0.232

在本研究中,选择的主成分变量用来解释多个土壤因子,进行全铜的空间插值模型的构建。主成分变量的分数代表从被选择的 5 种土壤因子中提取的信息,这些分数不仅实现了原始数据的标准化,也包括大部分的原始信息。得到计算各主成分变量的形式为:

$$Z_1 = 0.740 \times w_{\text{CEC}} - 0.457 \times w_{\text{全N}} + 0.635 \times w_{\text{全K}} + 0.393 \times w_{\text{全Cr}} + 0.565 \times w_{\text{全Al}} \quad (5)$$

$$Z_2 = -0.048 \times w_{\text{CEC}} + 0.541 \times w_{\text{全N}} - 0.439 \times w_{\text{全K}} + 0.640 \times w_{\text{全Cr}} + 0.547 \times w_{\text{全Al}} \quad (6)$$

$$Z_3 = 0.275 \times w_{\text{CEC}} + 0.613 \times w_{\text{全N}} + 0.243 \times w_{\text{全K}} - 0.529 \times w_{\text{全Cr}} + 0.232 \times w_{\text{全Pb}} \quad (7)$$

将上步骤得出的各主成分得分对应其贡献率计算出综合得分,如下:

$$W_2 = 0.3266Z_1 + 0.2394Z_2 + 0.1689Z_3 \quad (8)$$

由式(5)、(6)和(7)得出的各主成分得分 Z_1 、 Z_2 和 Z_3 代入到式(8),计算综合得分(W_2),其符合正态性检验($P < 0.05$),得分为 4.01~51.97,以 W_2 为

辅助变量进行全铜的协同克里格插值。

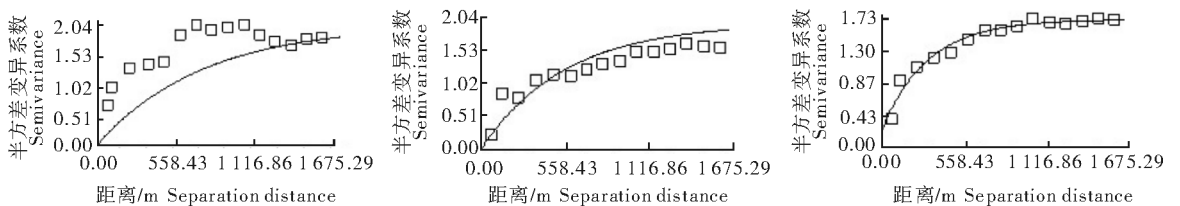
2.2 模拟空间分析

1)半方差函数分析。将铜作为目标变量依次进行普通克里格插值和协同克里格插值,其中协同克里格插值的辅助变量分别是与目标变量相关性最高的土壤养分全钾(W_1)及主成分计算出来的综合得分(W_2)。在不同的插值方法下对该研究区的土壤铜含量进行半方差函数分析,得到相关的参数(表 5)。表 5 中,块金值与基台值的比值($C_0/(C_0 + C)$)表明区域化变量的空间相关性程度,一般情况下其标准为:该比值 $> 75\%$ 说明该研究区土壤铜的空间相关性较弱,比值在 $25\% \sim 75\%$ 之间说明该研究区土壤铜的空间相关性中等,该比值 $< 25\%$,说明该研究区土壤铜具有较强的空间相关性。由表 5 可知,Cu 的块金值与基台值之比 $< 25\%$,说明具有较强的空间相关性。从变程的大小可知,总体上研究区土壤铜含量的空间相关范围为 1 100~1 800 m。其中,残差是指实际观测值与预测值之间的差,残差值越大,表明半方差函数的拟合度也就较差,对比各模型的残差值大小可以看出,COK_{PCA} 模型对土壤铜空间分布的模拟效果相对较好,而 COK_K 模型次之,OK 模型相对较差。根据表 5 分析结果,分别绘制 3 种不同插值下的半方差函数图,直观的反映 OK、COK_K 和 COK_{PCA} 模型的半方差函数拟合精度依次增加的趋势(图 2)。

表 5 土壤铜元素地统计学参数

Table 5 Geo-statistical parameters of Cu elements

内插方法 Interpolation method	模型 Model	步长/m Lag	变量 Variable	块金值 Nugget	偏基台值 Partial sill	变程/m Range	(块金值/基台值)/% Nugget/Sill	残差 Residuals error
OK	指数函数 Index function	111.41	Cu	0.001 0	2.037 0	1 782.37	0.05	0.994
COK _K	高斯函数 Gaussian function	96.67	Cov(Cu, W_1)	0	0.245 6	1 753.26	0	0.981
COK _{PCA}	指数函数 Index function	72.05	Cov(Cu, W_2)	0	0.154 4	1 117.22	0	0.994



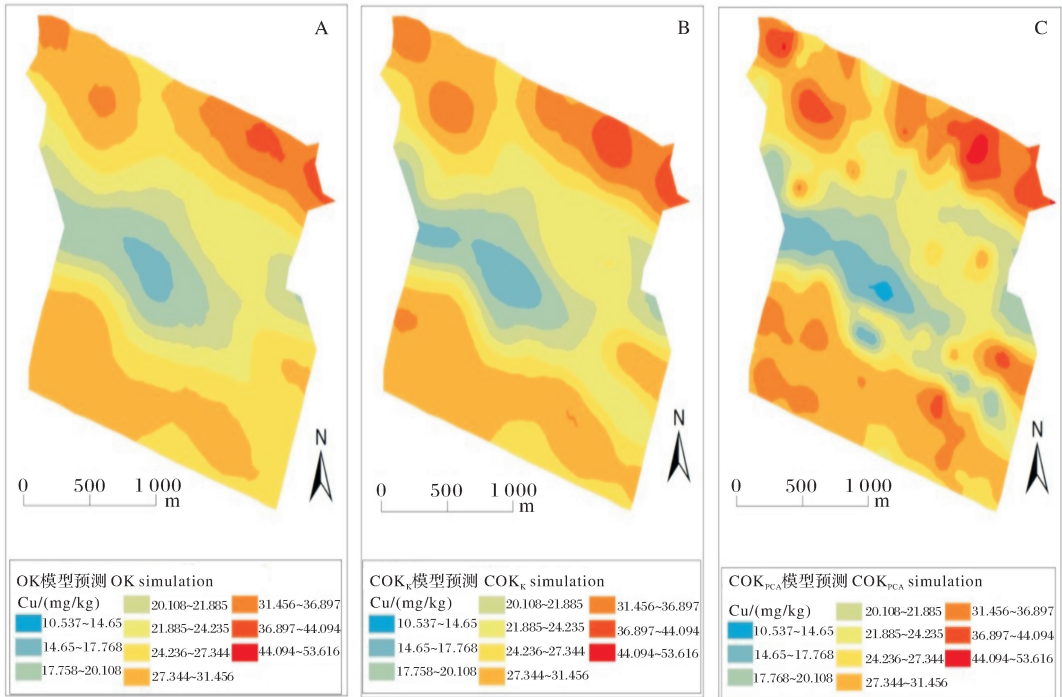
A: OK; B: COK_K; C: COK_{PCA}.

图 2 不同空间插值模型下土壤 Cu 的半方差函数图

Fig. 2 Semi-variograms of Cu in soil based on different spatial interpolations

2)土壤铜空间分布图。图 3A、B、C 分别表示 OK、COK_K、COK_{PCA} 插值得到的土壤铜含量空间分布情况。由图 3 可知,全铜含量在采样区域呈条带状和岛状分布相结合的特点。采样区的中间部分含量最低,分别向南北方向增大,而在采样区 4 个边角部分含量最高,取值范围为 10.537~53.616 mg/kg, 这些地区主要在村庄密集处和水域附近,人为活动和水污染因素较高,故铜含量较高。图 3A、B 的相

似度较高,但图 3B 在局部细节处略精确于图 3A,这主要是因为普通克里格插值只需要考虑全铜的空间信息,而协同克里格插值法可以利用与全铜相关性较好的变量全钾来提高铜含量的估值精度,图 3C 的空间分布结构更具有空间随机性和异质性,在局部结构上更能体现出土壤铜含量空间分布的随机性和依赖性,模拟效果相比较 OK、COK_K 2 种插值形式来说更具有真实性和可应用性。



A:OK 模型预测 OK simulation; B:COK_K模型预测 COK_K simulation; C:COK_{PCA}模型预测 COK_{PCA} simulation.

图 3 基于不同插值方法的土壤铜含量空间分布情况

Fig.3 Distribution of Cu in soil based on different spatial interpolations

2.3 模型精度比较

表 6 通过评价指标显示了验证数据集在内插方法中的模拟预测能力及其最佳理论变异函数模型。 R^2 是用来衡量模型拟合度程度,取值范围为 0~1,其值越大,表明模型越精确。AIC 是模型的精确指数,用来比较不同回归模型,在考虑模型的复杂性情况下,模型的 AIC 值越小,其拟合度越高, RMSE 值越小表明残差越小。分析发现, COK_{PCA} 的 AIC、RMSE 数值最小,分别为 87.203、6.309, R^2 值最大,为 0.852,说明该模型具有最好的精度;其次为

COK_K 模型, R^2 、AIC 和 RMSE 的数值依次为 0.743、96.908 和 6.907;而 OK 模型的精度最差,该模型的 R^2 、AIC 和 RMSE 的数值依次为 0.716、101.591和 7.044。同时,对比 T_{R^2} 、 T_{AIC} 和 T_{RMSE} 的 3 个指标数值可知,COK_{PCA}、COK_K 相对 OK 精度有显著性的提高,COK_{PCA} 模型最高,COK_K 模型次之,即 COK_{PCA} > COK_K > OK。因此,通过 COK_{PCA} 模型得到的土壤全铜含量的空间分布图与其真实空间分布特征具有较强的相似性和一致性,可将其来分析土壤全铜含量的空间分布属性及其特征。

表 6 验证数据内插精度指标

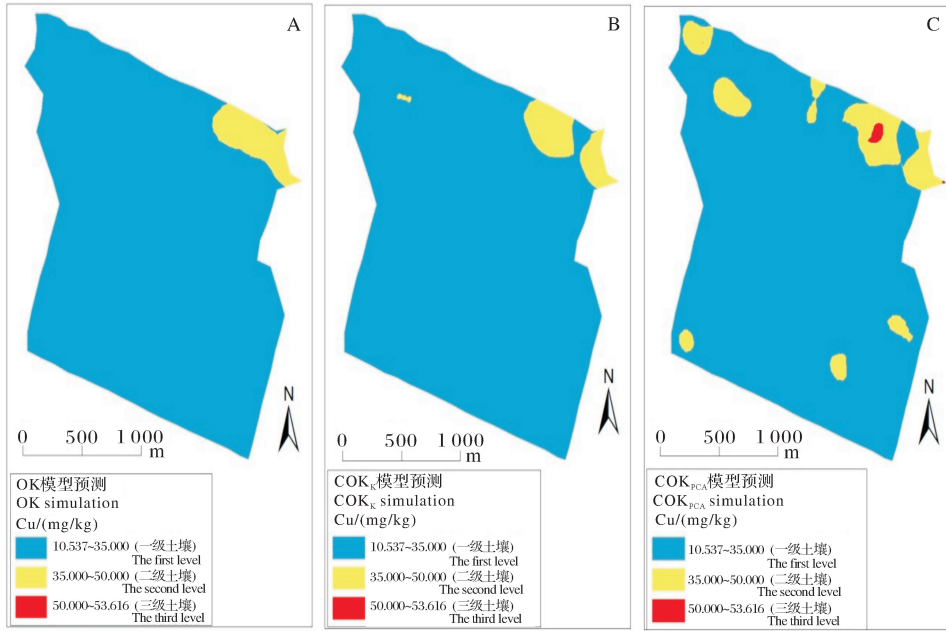
Table 6 Interpolation accuracy indexes of validation data

内插方法 Interpolation method	变量 Variable	R^2	$T_{R^2}/\%$	AIC	$T_{AIC}/\%$	RMSE	$T_{RMSE}/\%$
OK	Cu	0.716	0.00	101.591	0.00	7.044	0.00
COK _K	Cov(Cu, W_1)	0.743	3.77	96.908	4.61	6.907	1.94
COK _{PCA}	Cov(Cu, W_2)	0.852	18.99	87.203	14.16	6.309	10.43

2.4 土壤铜含量空间分布

由图 4 可知,研究区域绝大部分属于一级土壤区($Cu \leq 35 \text{ mg/kg}$),存在若干二级土壤区($35 \text{ mg/kg} < Cu \leq 50 \text{ mg/kg}$),二级土壤区的个数依次为 1、3 和 9 个,集中在东北部,而南部区域存在少量。图 4C 东北部三级土壤区($Cu > 50 \text{ mg/kg}$),以其为中心,呈同心圆形状,向四周逐步递减。根据研

究区域的土地利用类型,二级土壤区多分布在农田等人为活动较多的地方,三级土壤区村庄密集,同时离水源较近,比较发现图 4C(COK_{PCA})土壤等级空间分布情况最接近土壤铜含量实际分布,图 4B(COK_K)次之,而图 4A(OK)最为粗糙。研究区土壤级别划分后,可以更形象而又直观地根据土壤级别区空间分布情况对柑橘种植区进行相应的调整。



A: OK 模型预测 OK simulation; B: COK_K 模型预测 COK_K simulation; C: COK_{PCA} 模型预测 COK_{PCA} simulation.

图 4 OK、COK_K 及 COK_{PCA} 模型下土壤等级空间分布

Fig.4 Spatial distribution of soil grade by OK, COK_K and COK_{PCA} models respectively

3 讨论

目前地统计学中的克里格插值方法借助于原始数据的空间依赖性和半方差函数的结构性,对未知采样区域的土壤属性进行无偏最优估值,被广泛应用于土壤理化属性空间分布特征的估测和模拟中^[15]。由于普通克里格插值方法只简单地考虑被预测土壤属性的空间信息,具有易计算和少变量的特征,因此,在环境和土壤领域受到广泛的关注和应用。但是铜属于土壤中的微量元素,含量较少,其化

学实验结果容易受到误差的影响,并且在土壤的发育和构造中很容易受到其他影像因素的影响,因此,在估测土壤铜含量的过程中只考虑单一的空间信息,而忽略了土壤属性之间相辅相成的影响关系,只单纯地利用普通克里格方法进行区域尺度上的土壤铜含量的估算会使预测精度受限。同时受到采样精度和数量的影响,其预测结果具有很大的不稳定性和变化性,而增加样点的密度和数量又会大幅度地增加成本,因此,选择一种合适的模型和方法对土壤铜含量的预测将会起到很大的帮助^[16-17]。与普通克里

格插值法相比之下,协同克里格插值法可以利用与目标变量相关性较好的辅助变量来提高目标变量的估值精度,能充分考虑影响目标变量的其他土壤因子,将目标变量的空间自相关性和辅助变量间的交互相关性结合起来用于无偏最优估计,有利于提高估值的精度^[18]。本研究运用 2 种形式的协同克里格插值后,精度相比普通克里格插值提高的程度不同,基于 OK 模型,COK_K 和 COK_{PCA} 模拟精度提高依次为 4% 和 15% 左右,初步证明协同克里格对区域尺度土壤属性空间模拟的效果要优于普通克里格^[19]。其中,以往研究将与目标变量相关性最高的空间因素作为协助变量,该方法在一定程度上忽略了其他空间因素对目标变量的影响^[20]。我们分别将与目标变量相关性最高的全钾(W_1)和基于主成分分析所得的综合得分(W_2)为协助变量进行协同克里格插值。COK_{PCA} 预测精度提高的程度明显高于 COK_K,说明 COK_{PCA} 能更有效地减少原始数据信息损失、简化数据结构、避免主观随意性等,从而达到对研究区土壤铜含量空间预测与模拟的更好效果。

在小尺度区域内进行 3 种插值形式研究,其中相对 OK 而言,COK_K 利用相关性最高的全钾(W_1)作为协助变量,对全铜含量空间分布预测精度在 R^2 、赤池信息准则(AIC)、均方根误差(RMSE)分别提高 3.77%、4.61%、1.94%。COK_{PCA} 是将与全铜具有显著相关性的 CEC、全 N、全 K、全 Cr 和全 Pb 等 5 种土壤因子进行主成分分析得到土壤综合得分(W_2)作为协助变量,其精度指数提高程度依次为 18.99%、14.16%、10.43%,OK、COK_K 和 COK_{PCA} 3 种模型构建的半方差函数残差值依次为 0.994、0.981 和 0.976,残差值越大,表明半方差函数的拟合度也就较差,再基于该 3 种模型的土壤等级空间分布图可以看出,COK_{PCA} 的土壤等级空间分布情况最接近土壤铜含量的实际分布,COK_K 次之,而 OK 最为粗糙,综上得到预测精度 COK_{PCA} > COK_K > OK。由此看出,COK_{PCA} 模型对土壤铜空间分布的模拟效果相对较好,而 COK_K 模型次之,OK 模型相对较差。这是由于普通克里格插值法主要考虑的是全铜的空间信息,而协同克里格插值法会考虑到协助变量对全铜空间分布情况的影响。相比仅使用与全铜相关性最高的全 K(W_1)而言,通过主成分分析计算综合得分(W_2)作协助变量的方法,更大更全面地考虑了其他土壤因子的空间信息对目标变量全铜的影

响,精度较高,较为科学。

在研究区,根据 COK_{PCA} 模型预测的土壤铜空间分布情况分析得出最适合种植柑橘地区为一级土壤区,二级土壤区次之,应尽量避免在三级土壤区种植柑橘。研究区土壤铜含量自南向北总体呈现增加趋势,结合其土地利用类型可知,该趋势主要由于自南向北水田面积明显增多,相应灌溉率高,引水量多,而旱地面积是朝反方向趋势增多,灌溉率相对较低,引水量较少,表明长期接近水源可导致土壤铜富集。中部以及南部区域土壤铜含量较低,北部区域土壤铜含量较高,应结合该空间分布情况采取重金属铜污染的治理措施,便于今后种植柑橘。

参 考 文 献

- [1] MOUSAVIFARD S, MOMTAZ H, SEPEHR E, et al. Determining and mapping some soil physico-chemical properties using geostatistical and GIS techniques in the Naqade region, Iran [J]. Archives of Agronomy and Soil Science, 2013, 59 (11): 1573-1589.
- [2] WU C, WU J, LUO Y, et al. Statistical and geostatistical characterization of heavy metal concentrations in a contaminated area taking into account soil map units [J]. Geoderma, 2008, 144(1/2): 171-179.
- [3] 林媚, 冯先桔, 张伟清, 等. 台州柑橘主栽区土壤重金属环境质量及评价 [J]. 浙江农业科学, 2012(9): 1301-1305.
- [4] 庞凤, 李廷轩, 王永东, 等. 县域农田土壤铜含量的协同克里格插值及采样数量优化 [J]. 中国农业科学, 2009, 42(8): 2828-2836.
- [5] OTTOSEN L M, HANSEN H K, JENSEN P E. Relation between Ph and desorption of Cu, Cr, Zn, and Pb from industrially polluted soils [J]. Water, Air, and Soil Pollution, 2009, 201(1): 295-304.
- [6] 高远, 赵玉杰, 杨海松. 普通克里格评估土壤污染物空间分布精度控制研究 [J]. 农业环境科学学报, 2009, 28(5): 871-876.
- [7] MARTINEZ-COB A. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain [J]. Journal of Hydrology, 1996, 174(1): 457-463.
- [8] 李楠, 徐东瑞, 吴杨洁. 土壤养分含量的协同克里格插值研究 [J]. 浙江农业学报, 2011, 23(5): 1001-1006.
- [9] 郭龙, 张海涛, 陈家赢, 等. 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 [J]. 土壤学报, 2012, 49(5): 1037-1042.
- [10] 郭笃发, 王秋兵. 主成分分析法对土壤养分与小麦产量关系的研究 [J]. 土壤学报, 2005, 42(3): 523-527.
- [11] 赵彦锋, 郭恒亮, 孙志英, 等. 基于土壤学知识的主成分分析判断土壤重金属来源 [J]. 地理科学, 2008, 28(1): 45-50.
- [12] 鲍士旦. 土壤农化分析 [M]. 北京: 中国农业出版社, 2005: 30-

- 279.
- [13] 万田英, 霍庆, 祁志福, 等. 武汉钢铁公司周边地区土壤和蔬菜重金属含量分析[J]. 华中农业大学学报, 2014, 33(4): 77-83.
- [14] CAMACHO J, FERRER A. Cross-validation in PCA models with the element-wise k -fold (ekf) algorithm: practical aspects [J]. Chemometrics and Intelligent Laboratory Systems, 2014, 131(7): 37-50.
- [15] DAI Q. A competitive ensemble pruning approach based on cross-validation technique [J]. Knowledge-Based Systems, 2013, 37: 394-414.
- [16] JOURNAL A G, ROSSI M E. When do we need a trend model in Kriging? [J]. Mathematical Geology, 1989, 21(7): 715-739.
- [17] SANDEEP K, RATTAN L. Mapping the organic carbon stocks of surface soils using local spatial interpolator [J]. Journal of Environmental Monitoring: JEM, 2011, 13(11): 3128-3135.
- [18] HASSANI S, MARTENS H, QANNARI E M, et al. Degrees of freedom estimation in principal component analysis and consensus principal component analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2012, 118: 246-259.
- [19] AKBAR D A. Reserve estimation of central part of choghart north anomaly iron ore deposit through ordinary Kriging method [J]. International Journal of Mining Science and Technology, 2012, 22(4): 573-577.
- [20] BANCROFT B A, HOBBS G R. Distribution of Kriging error and stationarity of the variogram in a coal property [J]. Mathematical Geology, 1986, 18(7): 251-257.

Spatial distribution of soil heavy metal Cu content using Cokriging interpolation model combined with principal component analysis

ZHANG Qing¹ ZHANG Haitao¹ GUO Long² DU Peiyong¹
LI Linwei¹ LI Ruijuan¹ TANG Xiaofei³

1. College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China;

2. College of Resources and Environment, Wuhan University, Wuhan 430072, China;

3. Hubei Institute of Surveying and Mapping, Wuhan 430074, China

Abstract Insufficient or excess heavy metals in soils seriously affect the healthy growth of citrus. It is necessary to establish an accurate space model which can effectively reflect the spatial distribution of heavy metals content. This paper took the citrus production areas in Honghuatao county, Yidu City, Hubei Province were used to study heavy metal Cu in soil. Soil factors including total K, total Cr, CEC, total Pb and total N ($P < 0.05$) closely correlated with Cu were chosen as the spare variables through Pearson's correlation coefficient method. Total scores were calculated by principle components analysis (PCA) method and used as auxiliary variables to reduce the dimension and redundancy of original variables. Three spatial interpolation models including ordinary Kriging (OK) model, ordinary Cokriging (COK) models which took total scores (COK_{PCA}) and total K (COK_K) as auxiliary variable were constructed. The optimal space model was proposed by comparing the prediction accuracy of the models (OK, COK_K , COK_{PCA}). The most suitable model was used to predict spatial distribution of Cu in the area studied. The results showed that RMSE of OK, COK_K and COK_{PCA} models were 7.044, 6.907 and 6.309, with R^2 of 0.716, 0.743 and 0.852, and AIC of 101.591, 96.908 and 87.203, respectively. Interpolation results were comprehensively compared. COK_{PCA} was found to be the best simulating model, followed by COK_K but OK was poor. It will provide reference for accurately simulating soil heavy metals Cu in the small-scale area and a scientific basis for modern agricultural production and fine management of agriculture.

Keywords principal component analysis; Cokriging; Pearson's correlation coefficient; heavy metal Cu; microelement; spatial distribution; estimation

(责任编辑: 陆文昌)