

# 基于环境因子和联合概率方法的土壤 有机质空间预测\*

张楚天<sup>1,2</sup> 杨 勇<sup>1,2†</sup> 贺立源<sup>1,2</sup> 季世民<sup>1,2</sup> 刘颖颖<sup>1,2</sup>

(1 华中农业大学资源与环境学院, 武汉 430070)

(2 农业部长江中下游耕地保育重点实验室, 武汉 430070)

## PREDICTION OF SPATIAL DISTRIBUTION OF SOIL ORGANIC MATTER BASED ON ENVIRONMENTAL FACTORS AND A JOINT PROBABILITY METHOD

Zhang Chutian<sup>1,2</sup> Yang Yong<sup>1,2†</sup> He Liyuan<sup>1,2</sup> Ji Shimin<sup>1,2</sup> Liu Yingying<sup>1,2</sup>

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

(2 Key Laboratory of Arable Land Conservation (Middle and Lower Reaches of Yangtze River), Ministry of Agriculture, Wuhan 430070, China)

**关键词** 环境因子; 土壤有机质; 回归克里格; 土壤景观定量模型; 联合概率; 预测精度

**中图分类号** S158.2 **文献标识码** A

土壤连续属性(如土壤中养分含量、重金属含量等)的空间分布特征和定量分布信息是进行土壤质量评价和区域环境综合评估的基础。精准农业战略的实施和各种区域生态评价均需要更详细更精确的土壤属性信息作为依据<sup>[1-2]</sup>。因此,土壤属性空间预测一直是土壤学研究的热点问题。经典地统计学以各种克里格插值法为代表,是土壤属性空间预测中的常用方法。但该方法缺乏对辅助信息(如环境信息)的有效利用<sup>[3-4]</sup>,导致预测精度降低<sup>[5]</sup>。而土壤景观定量模型的理论依据就是土壤与环境的关系,但该方法忽略了采样点之间的空间相关性。因此这两种方法存在互补性。近些年来也出现了综合两种思路的方法,如协同克里格<sup>[6]</sup>、回归克里格(Regression Kriging, RK)<sup>[7]</sup>、分层克里格(Stratified Kriging)、地理权重回归<sup>[8]</sup>等。与普通克里格方法相比较,在环境因子与被预测变量相关性较强时,这些方法的预测精度均高于普通克里格方法<sup>[9-10]</sup>。但这些方法在综合利用多种不同类型的环境信息方面仍各有不足:如协同克里格虽然能

将环境因子作为辅助变量,但当辅助变量多时则在联合区域上有困难;分层克里格只是将环境信息作为一个分层或分区的依据。

为了充分利用采样数据及辅助信息,本文将普通克里格方法与土壤景观定量模型所得结果通过联合概率公式结合起来,并以所得概率分布的数学期望为预测结果,旨在提出一个利用环境信息提高土壤属性预测精度的初步方法。

## 1 材料与方法

### 1.1 研究区概况

沙洋县位于江汉平原西北腹地,地处东经 111°51' ~ 112°42',北纬 30°23' ~ 30°55'之间,国土面积 2 044 km<sup>2</sup>。属亚热带季风气候,雨量充沛、四季分明。年平均气温 16℃,年平均降水量 1 025 mm。境内西北高东南低,海拔 20 ~ 150 m。西北部为丘陵龙岗状地形,其母质为湖相沉积棕黄色、黄褐色黏土,呈酸性和微酸性;南部为汉江堆积阶地,属汉江

\* 国家自然科学基金项目(41101193)和“国家级大学生创新创业训练计划”创新训练项目(201210504073)资助

† 通讯作者,杨 勇(1980—),男,江西吉安人,副教授,硕士生导师,主要研究方向为地统计学及其应用。E-mail: yangyong@mail.hzau.edu.cn

作者简介:张楚天(1989—),男,吉林磐石人,博士研究生,主要研究方向为地统计学及其应用。E-mail: zct1989511@163.com

收稿日期: 2013-04-10; 收到修改稿日期: 2013-11-29

平原地域,其母质为亚黏土、砾土、粉砂和黏土,呈微碱性,东部主要为汉江堤防外滩地带的河漫地<sup>[11]</sup>。

## 1.2 样品采集与分析

在保证样点均匀分布的前提下,根据实地情况随机布设样点,即在东南地势平坦区域,避开湖泊等无法采样地段,随机布点,在西北地势起伏区域,在地形变化地段如坡脚布设样点。用五点混合取样法取 0 ~ 20 cm 耕层非根际土壤,共采样点 658 个,采样时间为 2007 年秋。将土样经风干、磨碎、过筛后用重铬酸钾外加热法<sup>[12]</sup>进行土壤有机质含量的测定。

## 1.3 环境因子提取

地形地貌特征对区域水文、生态等过程产生影响,进而会影响到土壤养分含量等的空间分布<sup>[13-15]</sup>。参照相关研究成果<sup>[16]</sup>,本次试验包括 11 个环境因子,其中 10 个地形因子用以反映研究区地形特点,包括相对高程 ( $Hr$ )、地形起伏度 ( $QFD$ )、坡度 ( $\beta$ )、坡度变率 ( $SOS$ )、地形湿度指数 ( $\Psi$ )、地面粗糙度 ( $M$ )、垂直曲率 ( $Cv$ )、水平曲率 ( $Ch$ )、河流动能指数 ( $\Omega$ ) 和坡度坡长因子 ( $LS$ ),地形因子均由 STRM 90m 分辨率的 DEM 数据在 ArcGIS 软件平台下通过 3D Analysis 模块计算得到;归一化植被指数 (NDVI) 用来反映研究区植被覆盖状况,本试验以 LandSat-ETM + (2006 年 10 月 31 日) 影像经过多影像自适应局部回归法 (RGF)<sup>[17]</sup> 去除条带并重采样为 90m 分辨率数据后作为数据源提取并分析 NDVI 与土壤养分的相关性。

用  $EA$  表示某环境因子,设  $EA$  栅格数据行列数分别为  $row$  和  $col$ ,参与计算的  $EA$  数量为  $m$ ,栅格某网格单元为  $grid[i, j]$  ( $0 < i \leq row, 0 < j \leq col$ ), 则  $grid[i, j]$  上的各  $EA$  数据取值可表示为  $(E_{ij,1}, E_{ij,2}, \dots, E_{ij,v}, \dots, E_{ij,m})$ 。栅格大小为 90m,并作为本研究其后各步骤的尺度基准。

## 1.4 回归克里格

回归克里格方法综合考虑了地理要素的空间分布规律和其他随机因子,是借助环境因子作为辅助信息进行空间预测的常用方法<sup>[18-20]</sup>。在 SPSS19.0 中对有机质和环境因子进行相关性分析和线性逐步回归拟合,得到最优的土壤有机质线性回归模型。运用 GS + 9.0 软件对回归预测残差值进行普通克里格插值。在 ArcGIS 软件平台下,基于求得的线性回归模型使用栅格计算器计算有机质分布的线性回归插值结果,最后将线性回归插值结

果和残差克里格插值结果进行空间加运算从而得到有机质的回归克里格空间分布结果。

## 1.5 基于环境因子和联合概率方式的土壤养分空间预测

本方法将由土壤景观定量模型和普通克里格法所得的概率分布看作两个独立事件,首先基于土壤景观定量模型,获取土壤属性与环境因子之间的关系,计算得到土壤属性概率分布,然后对样点土壤属性进行普通克里格插值,将得到的连续型土壤属性概率分布离散化,最后在联合概率框架下将两个概率分布结合起来即为本方法的土壤属性概率分布,并将此概率分布的期望值作为本方法的土壤属性预测值。

**1.5.1 基于土壤景观定量模型的土壤属性概率分布计算** 设当前建模组采样点待预测土壤连续属性(记为  $CA$ ) 含量的值域范围为  $[C_{\min}, C_{\max}]$ ,在该值域范围内根据精度需求将其等分为  $n$  个含量类别,则第  $k$  ( $1 \leq k \leq n$ ) 组  $CA$  含量类别(记为  $CA_k$ ) 值域为:

$$CA_k = \left[ C_{\min} + \frac{C_{\max} - C_{\min}}{n} \times (k - 1), C_{\min} + \frac{C_{\max} - C_{\min}}{n} \times k \right] \quad (1)$$

其中值为:

$$\overline{CA_k} = C_{\min} + \frac{C_{\max} - C_{\min}}{n} \times (k - 0.5) \quad (2)$$

将采样点根据  $CA$  含量类别划分为  $n$  组。利用每个  $CA$  含量类别对应样点上的  $CA$  值及  $EA$  数据,获得该  $CA$  含量类别与单个  $EA$  的定量关系,即  $CA$  含量类别对于单个  $EA$  的模糊隶属度,最终可得到  $m \times n$  个  $CA_k$ - $EA$  的定量关系。具体而言,令  $N_{EA}$  为  $EA$  类别数,将  $EA$  值域等间距划分为  $N_{EA}$  类,扫描每个  $CA$  含量类别对应采样点所在位置,记录该位置上  $EA$  值所属的  $EA$  类别,并进行统计。因此,类别-环境因子定量关系可表示为:

$$R_{(CA_k, EA)} = \left( \left( EA_1, \frac{Count(1)_k}{Count_k} \right), \dots, \left( EA_{N_{EA}}, \frac{Count(N_{EA})_k}{Count_k} \right) \right) \quad (3)$$

式中,  $Count_k$  表示  $CA_k$  对应的样点个数,  $Count(1)_k$  表示  $CA_k$  对应样点上  $EA$  值落在  $EA$  的第 1 个类别中的个数,即用  $CA$  含量类别对应采样点上环境因子取值概率分布来表示  $CA$  含量类别对于单个  $EA$  的模糊隶属度关系,表现形式为直方图形式。

再利用采样点上的  $CA$  数据和  $EA$  数据,计算每

个  $EA$  与  $CA$  的决定系数,得到决定系数序列:

$$r_{(EA,CA)}^2 = (r_1^2, r_2^2, \dots, r_m^2) \quad (4)$$

对该序列进行归一化转化,使得转化后的序列和为 1,归一化公式如下:

$$r_j^2 = \frac{r_j^2}{\sum_{j=1}^m r_j^2} \quad (5)$$

对于  $grid[i, j]$ , 根据  $CA_k$ - $EA$  定量关系和栅格上的  $EA$  数据 ( $EA_{ij,m}$ ), 获得  $m$  个基于单个  $EA$  的模糊隶属度值, 即求取  $EA$  数据值 ( $EA_{ij,m}$ ) 在  $CA_k$ - $EA$  定量关系上对应的土壤属性取值概率:

$$Q_{ij,e}^k = R_{(CA_k,EA)}(EA = EA_{ij,e}), (1 \leq k \leq n, 1 \leq e \leq m) \quad (6)$$

令  $Z_{ij}$  为  $grid[i, j]$  上的  $CA$  预测含量, 考虑  $EA$  与  $CA$  的相关性强弱, 引入归一化转化后的决定系数作为权重, 计算  $Z_{ij}$  上对  $CA$  第  $k$  个含量类别的相似度:

$$S_{ij}^k = \sum_{e=1}^m (Q_{ij,e}^k \times r_e^2), (1 \leq k \leq n) \quad (7)$$

则将相似度序列归一化后,  $grid[i, j]$  上  $CA$  含量的模糊分布可表示为:

$$f_L(z_{ij}) = ((CA_1, S_{ij}^1) \cdots (CA_k, S_{ij}^k)), (1 \leq k \leq n) \quad (8)$$

### 1.5.2 基于普通克里格方法的预测结果概率分布

众所周知, 普通克里格方法在得到预测结果的同时, 可得到每个预测点的预测方差, 其预测结果符合正态分布, 即:

$$f_C(z_{ij}) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(z_{ij}-\mu)^2}{2\sigma^2}} \quad (9)$$

式中,  $\mu$  为普通克里格预测结果,  $\sigma^2$  为预测方差。为使该预测结果可与 1.5.1 所得的土壤属性概率分布结果联合使用, 将其中的  $CA$  含量类别中值序列代入式(9)中取  $n$  个概率值归一化后作为  $grid[i, j]$  上由经典地统计方法得到的离散型概率分布结果, 即:

$$f_C(z_{ij}) = ((CA_1, G_{ij}^1) \cdots (CA_k, G_{ij}^k)), (1 \leq k \leq n) \quad (10)$$

式中, 取值  $\overline{CK_k}$  时的概率值  $G_{ij}^k$  可用下述公式计算得到:

$$G_{ij}^k = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\overline{CK_k}-\mu)^2}{2\sigma^2}}, (1 \leq k \leq n) \quad (11)$$

**1.5.3 联合概率分布计算** 在本文中, 将由土壤景观定量模型和普通克里格法所得概率分布作为两个独立事件。那么, 当预测结果要顾及两个独立事件同时发生时, 采用联合概率方法计算。当预测

$grid[i, j]$  上  $CA$  含量时, 既要考虑样点数据, 又要考虑  $EA$  对它的影响, 那么  $CA$  取值  $\overline{CA_k}$  的概率为:

$$T_{ij}^k = \frac{G_{ij}^k \times S_{ij}^k}{\sum_{k=1}^n (G_{ij}^k \times S_{ij}^k)}, (1 \leq k \leq n) \quad (12)$$

所以, 将上述概率值归一化后  $Z_{ij}$  的最终概率分布可表示为:

$$f_T(z_{ij}) = ((CA_1, T_{ij}^1) \cdots (CA_k, T_{ij}^k)), (1 \leq k \leq n) \quad (13)$$

基于此概率分布, 将其数学期望作为  $grid[i, j]$  上最后的预测结果, 即:

$$E(f_T) = \sum_{k=1}^n \overline{CA_k} \times T_{ij}^k, (1 \leq k \leq n) \quad (14)$$

### 1.6 预测结果精度评价

为评价本方法的预测精度, 通过比较验证点位置上的土壤养分观测值和预测值, 采用皮尔森相关系数 (Pearson correlation coefficient,  $r$ ) 和均方误差 (Mean Squared Error,  $MSE$ ) 对预测精度进行评价, 同时以均方误差作为指标, 分析本方法相对于回归克里格方法的相对精度改进值 (Relative accuracy improvement,  $RI$ )<sup>[21]</sup>。

$$RI = \frac{MSE_{RK} - MSE_{JP}}{MSE_{RK}} \times 100\% \quad (15)$$

式中,  $MSE_{RK}$  代表回归克里格方法的  $MSE$  值,  $MSE_{JP}$  则代表本方法的  $MSE$  值。  $RI$  值为正值即本方法较回归克里格方法的预测精度高, 值越大说明预测精度提高的越多;  $RI$  值为负值则表示本方法预测精度低于回归克里格方法。

## 2 结果

### 2.1 土壤有机质含量的基本统计特征

根据三倍标准差法, 剔除采样点中的异常值。经检验, 研究区 658 个采样点中异常值为 3 个。将剔除异常值后的土壤样点在 ArcGIS 地统计模块中利用子集要素工具划分为空间分布均匀的建模组和验证组, 其中建模组包括 510 个样点, 验证组包括 145 个样点, 两组样点分布图见图 1 所示:

对两组样点的土壤有机质数据进行统计分析, 结果如表 1 所示。由表中可知, 建模点和验证点有机质含量变异系数分别为 0.223 和 0.234, 为中等变异性。

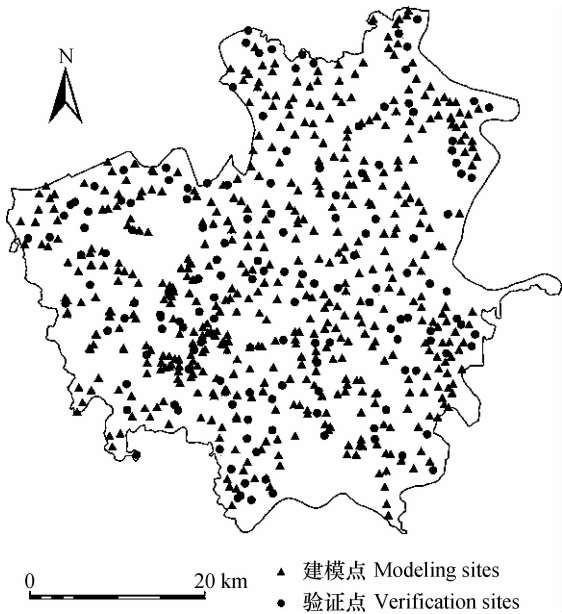


图 1 建模组和验证组样点分布图

2.2 土壤有机质与环境因子的相关性

表 2 为研究区有机质含量与环境因子之间的相关性。可以看出,有机质与  $Hr$ 、 $QFD$ 、 $\beta$ 、 $SOS$ 、 $NDVI$  五个环境因子(下文合称  $HQBSN$ )具有显著相关性。其中,有机质含量与  $Hr$  显著负相关,与  $QFD$ 、 $\beta$ 、 $SOS$ 、 $NDVI$  之间表现为显著正相关。由于某点相对高程  $Hr$  为区域内高程最大值与该点高程实测值的差值, $Hr$  越大,该点海拔越低,因此该结果可解释为有机质含量与海拔呈正相关。

进一步分析  $HQBSN$  之间的相关性,结果表明,  $Hr$  与  $QFD$ 、 $\beta$ 、 $SOS$ 、 $NDVI$  均呈显著负相关,  $QFD$ 、 $\beta$ 、 $SOS$  之间则存在显著的正相关关系,原因是这三个环境因子从不同侧面表现研究区域的地形起伏特点。  $NDVI$  与  $QFD$  呈显著正相关,即地形变化会影响植被覆盖分布。

表 1 有机质含量的基本统计特征

组别	样本数 (个)	最大值 ( $g\ kg^{-1}$ )	最小值 ( $g\ kg^{-1}$ )	平均值 ( $g\ kg^{-1}$ )	标准差 ( $g\ kg^{-1}$ )	变异系数
建模点	510	34.56	7.85	23.16	5.16	0.223
验证点	145	34.66	7.51	23.38	5.48	0.234

表 2 有机质与环境因子的相关性

$Ch$	$Cv$	$Hr$	$LS$	$M$	$NDVI$	$QFD$	$SOS$	$\beta$	$\Psi$	$\Omega$
-0.016	-0.009	-0.331 **	0.073	0.032	0.159 **	0.211 **	0.163 **	0.131 **	-0.037	0.076

注:  $Ch$  表示水平曲率;  $Cv$  表示垂直曲率;  $Hr$  表示相对高程;  $LS$  表示坡度坡长因子;  $M$  表示地面粗糙度;  $NDVI$  表示归一化植被指数;  $QFD$  表示地形起伏度;  $SOS$  表示坡度变率;  $\beta$  表示坡度;  $\Psi$  表示地形湿度指数;  $\Omega$  表示河流动能指数; \* \* 表示在  $p < 0.01$  水平上显著相关

表 3  $HQBSN$  之间的相关性

	$Hr$	$NDVI$	$QFD$	$SOS$	$\beta$
$Hr$	1				
$NDVI$	-0.185 **	1			
$QFD$	-0.512 **	0.135 **	1		
$SOS$	-0.380 **	0.044	0.582 **	1	
$\beta$	-0.325 **	0.069	0.584 **	0.516 **	1

注:  $Hr$  表示相对高程;  $NDVI$  表示归一化植被指数;  $QFD$  表示地形起伏度;  $SOS$  表示坡度变率;  $\beta$  表示坡度; 下同。 \* \* 表示在  $p < 0.01$  水平上显著相关

2.3 土壤有机质回归克里格预测模型

2.3.1 回归预测模型 采用回归克里格法,利用研究区建模点数据拟合有机质含量与环境因子的变化趋势模型。结果如下:

$$y = -0.1Hr + 6.2NDVI + 32.061$$

$$(R^2 = 0.119, p < 0.05) \quad (16)$$

从拟合的方程来看,11 个环境因子中只有  $Hr$  和  $NDVI$  参与了有机质的空间分布预测,这可以解

释为环境因子间的交叉相关性较显著,逐步回归过程最终只选取了与有机质含量相关性最强最具代表性的环境因子参与预测。且方程的决定系数仅为 0.119,与前人的研究结果类似<sup>[14-22]</sup>。该结果可能受本研究的尺度、土地利用类型、土壤类型复杂程度、所用 DEM 数据分辨率等因素的影响。

**2.3.2 回归残差插值半方差模型** 在 GS + 9.0 中对回归克里格模型残差半方差模型进行选择,通过比较不同模型产生的基台效应(块金值/基台值)、决定系数来选择最优的理论半方差模型。表 4 为土壤有机质多元回归预测方程残差值半方差模型参数。从表中可以看出,有机质残差的基台效应值为 0.374,具有中等的空间自相关性。变程较大,这是因为研究区处于平原地区,地势较平坦,选取的环境因子在较大范围内均表现出相似性。

表 4 土壤有机质多元回归残差值的半方差模型参数

理论模型	块金值	基台值	块金值/基台值	变程	决定系数
高斯模型	11.90	19.01	0.374	9 245	0.956

## 2.4 土壤有机质联合概率预测模型

**2.4.1 计算参数的确定** 本方法中用直方图的形式表示  $CA_k$ - $EA$  定量关系,因此  $EA$  值域区间数  $N_{EA}$  必须适当选择。如果  $N_{EA}$  过大,会出现许多样本频数为零的区间,不符合统计基本要求;相反如果  $N_{EA}$  太小,则数据的分布过于集中,制作的直方图会很平坦,不利于观察数据的分布特征和规律。本文采取 Sturges 提出的经验公式<sup>[23]</sup>来确定  $N_{EA}$ :

$$N_{EA} = 1 + 3.32 \log N_{\text{sample}} \quad (17)$$

式中,  $N_{\text{sample}}$  为建模组样点数量。因此根据此公式得到合理的  $EA$  值域区间数  $N_{EA}$  为 10。

设定  $N_{EA} = 10$ ,依据本方法预测结果的相关系数  $r$  和均方误差  $MSE$  值随  $CA$  含量类别数  $n$  变化的情况来确定最佳的  $CA$  含量类别数,如图 2 所示。依据相关系数  $r$  和均方误差  $MSE$  的定义,  $r$  越大或者  $MSE$  越小则预测精度越高<sup>[24]</sup>。从图中可以看出,本方法预测结果的精度会随着  $CA$  含量类别数  $n$  的变化而上下波动。当  $n$  取 8 时,相关系数  $r$  达到极大值,而均方误差  $MSE$  达到极小值,此时认为本方法的预测精度达到最佳。

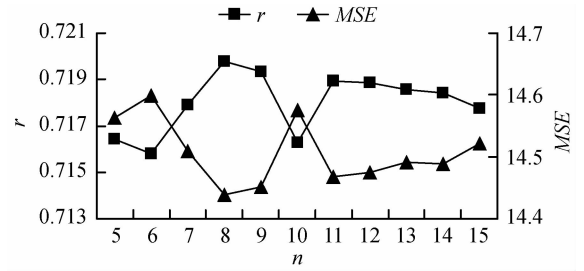


图 2  $r$  和  $MSE$  随  $CA$  含量类别数  $n$  的变化曲线图

综上所述,选取  $n = 8, N_{EA} = 10$  作为本研究的计算参数。

**2.4.2 普通克里格半方差模型** 表 5 为选取的土壤有机质普通克里格插值最优半方差模型参数。从表中可以看出,有机质的基台效应值为 0.461,与有机质残差插值结果一致,也具有中等的空间自相关性。基台值为 11.227 较大,说明由于采样误差、短距离变异或分析引起的各种正基底效应影响较大<sup>[25]</sup>。变程为 11 459m,说明研究区有机质量在较大范围内具有空间自相关性。

表 5 土壤有机质半方差模型参数

理论模型	块金值	基台值	块金值/基台值	变程	决定系数
高斯模型	11.23	20.83	0.461	11 460	0.969

**2.4.3 主要影响因子的确定** 从土壤有机质与环境因子的相关性分析可知,  $HQBSN$  之间具有显著的交叉相关性,为了最大程度融合  $HQBSN$  的信息以减小信息冗余。本文对  $HQBSN$  进行主成分分析 (PCA),计算得到融合  $HQBSN$  信息的新变量作为辅助变量(记为  $JP\_PCA$ )。同时,为从多角度对比本方法与回归克里格方法的精度,本文使用另外两种方式选取环境因子:(1)依据回归预测模型,选取  $Hr$  和  $NDVI$  作为辅助变量(记为  $JP\_HN$ );(2)依据相关性分析结果,将  $HQBSN$  同时作为辅助变量(记为  $JP\_HQBSN$ )。

在 SPSS 中进行主成分分析,按照特征值 ( $\lambda$ ) 大于 1 的原则提取 2 个主成分。表 6 为提取的主成分因子载荷表,由表中可知,第一主成分主要反映了研究区的地形变化特征,第二主成分主要反映了研究区的植被覆盖情况。

表 6 因子载荷表

主成分	$Hr$	$QFD$	$\beta$	$SOS$	$NDVI$	特征值	方差百分比 (%)	累计百分比 (%)
1	-0.692	0.866	0.772	0.787	0.223	2.493	49.86	49.86
2	-0.223	-0.046	-0.196	-0.218	0.935	1.013	20.26	70.12

利用主成分特征值和 *HQBSN* 在各主成分上的因子载荷,可计算得到 *HQBSN* 的因子得分系数,进而得到主成分的回归方程<sup>[26]</sup>:

$$\begin{cases} F_1 = -0.438Hr + 0.548QFD + 0.498\beta + 0.498SOS \\ \quad + 0.141NDVI \\ F_2 = -0.222Hr + 0.046QFD - 0.195\beta - 0.216SOS \\ \quad + 0.929NDVI \end{cases} \quad (18)$$

使用该公式在 ArcGIS 软件平台中利用栅格计算得到融合 *HQBSN* 信息的新变量作为环境因子参与联合概率的预测计算。

### 2.5 土壤有机质空间分布特征

图 3 为分别使用回归克里格 (RK)、JP\_PCA、JP\_HN 和 JP\_HQBSN 四种方式进行空间插值得到的沙阳县有机质空间分布图。从图中可以看出,四种方法得到的结果中有机质的分布格局大致相似,即

研究区西部的有机质含量要大于东部,中部有一片有机质含量的高值区,有机质含量存在从该区域向四周逐渐减小的趋势,这种趋势在由西向东方向尤其明显。有机质预测值范围分别为 11.86 ~ 32.38、11.77 ~ 31.66、11.91 ~ 32.46、11.94 ~ 31.84 g kg<sup>-1</sup>,各方法基本一致。与建模点有机质含量取值范围对比发现,基于联合概率方式的三种方法的平滑效应比回归克里格方法略大。有机质含量点状分布和块状分布相结合。如图 3 a 所示,回归克里格的插值结果中,斑块的边界较基于联合概率方式的结果零碎,说明选取的辅助变量对回归克里格插值结果影响较大。图 3 b、图 3 c 和图 3 d 显示,JP\_HQBSN 方式和 JP\_PCA 方式得到的结果连续性更强,有机质含量高值区的范围更广,JP\_HQBSN 方式得到的结果斑块边界较 JP\_HN 方式和 JP\_PCA 方式得到的结果平滑。

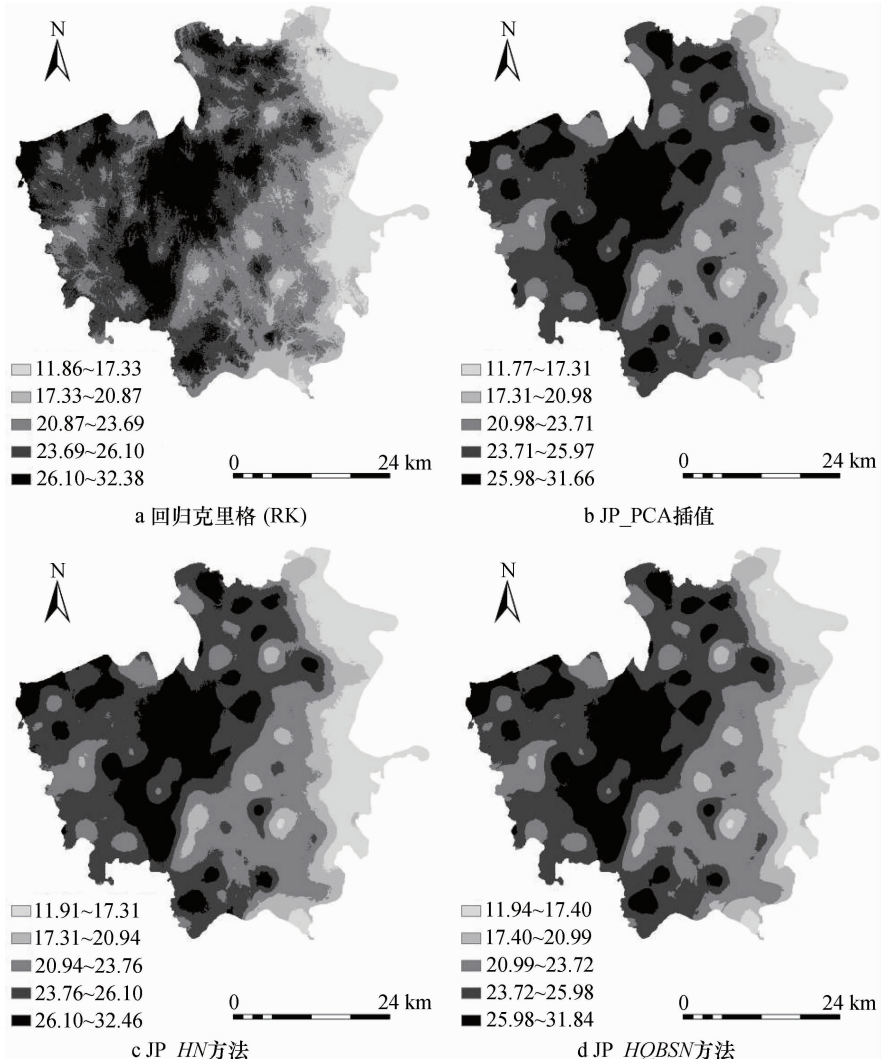


图 3 湖北沙阳县土壤有机质空间分布图

## 2.6 不同插值方法的空间预测精度

以回归克里格插值结果为参照,按照式(15)计算说明本方法对土壤有机质含量的预测精度的影响,结果见表7。从表中可以看到,在样本数一致的情况下,四种方法的预测值与验证点有机质含量值

相关系数范围为 0.697 ~ 0.722,均达到了显著相关的水平。JP\_PCA、JP\_HN、JP\_HQBSN 三种方式相对于回归克里格的精度提高值分别为 7.202%、6.472%、6.890%,由此证明本文提出的方法是提高土壤养分空间预测精度的有效方法。

表7 土壤有机质预测精度检验值

插值方法	样本数(个)	相关系数 $r$	均方误差 $MSE$	相对精度改进值 $RI$ (%)
RK	510	0.697	15.438	—
JP_PCA	510	0.722	14.326	7.202
JP_HN	510	0.720	14.439	6.472
JP_HQBSN	510	0.721	14.374	6.890

## 3 讨论

土壤养分空间分布受众多因素影响,相关分析表明土壤有机质与相对高程呈负相关,与地形起伏度、坡度、坡度变率呈正相关,即在海拔高、坡度大的地方,土壤有机质相对充足,这与相关研究结果<sup>[16-20]</sup>正好相反。研究表明,不同的土地利用方式会对土壤养分特征产生影响<sup>[27-28]</sup>,研究区沙洋县盛产水稻、油料(油菜),其东部海拔低、较平缓的区域分布着大片油菜种植区,而西部海拔较高区域则小片连续种植水稻等作物,不同作物的耕作方式、生长习性等均会对土壤有机质的含量造成很大影响。水稻田长期处于淹水状态使微生物活性有较大的提高,机械化的收割方式提高了秸秆还田率等原因可能是造成所处地块有机质含量较高的原因。

精度分析结果表明,基于联合概率方式的三种预测方法相关系数均高于回归克里格方法,且 JP\_PCA > JP\_HQBSN > JP\_HN。JP\_HN 只使用了相对高程( $Hr$ )和归一化植被指数(NDVI)作为辅助变量,融合的信息量有限,JP\_HQBSN 方式将与有机质含量显著相关的四种环境因子均作为辅助变量,获取的辅助信息更多,因此精度相对于 JP\_HN 方式有所提高;JP\_PCA 方式利用主成分分析方法从 HQBSN 中提取出主成分作为辅助变量,在最大限度融合源数据信息的同时减少了信息的冗余,因此精度在三种基于联合概率的预测方法中最高。进一步分析四种方法的均方误差( $MSE$ )可得到相同的规律,即 JP\_PCA < JP\_HQBSN < JP\_HN < RK。因此环境因子的选取和使用方式对利用辅助变量进行空间预测的方法的精度影响很大,选取与待预测空间变

量相关性强的,而交叉相关性弱的环境因子更能有效提高预测精度。而获取环境因子的源数据质量也会影响预测精度,随着 GIS 和遥感等技术的发展,高分辨率、多元化的数据获取将变得日趋方便,将进一步提高本方法预测的精度。

本文选取的 DEM 分辨率较低,获取的环境变量大部分为地形因子,没有考虑土壤类型等其他类型数据,并且仅使用普通克里格方法与土壤景观定量模型相结合来预测土壤养分单因子的含量,也即目前的研究对可获得的各种数据及方法还没有充分利用起来。因此在今后的工作中将尝试融合更多种类的数据,进一步提高土壤含量的预测精度,并尝试对土壤多因子的同时预测。

## 4 结论

本文以湖北省沙洋县土壤有机质含量为研究对象,在联合概率框架基础上,将普通克里格方法和土壤景观定量模型有机结合起来,对土壤有机质含量进行空间预测模拟。本方法灵活开放,可充分利用除采样点以外的多源数据作为辅助数据,且计算过程简单易懂,计算参数如属性分类数量等可根据研究目的和数据特点灵活调节,使得本方法能有效处理多种数据并且在其他研究区也具有适用性。相比于回归克里格方法,本方法能在一定程度上提高预测精度,且预测精度与环境因子的选取和使用方式有关。本研究方法和本案例研究结果均为初步的,在方法和验证设计上仍有需改进之处,有待今后进一步研究。

## 参考文献

- [1] 杨琳,朱阿兴,秦承志,等. 运用模糊隶属度进行土壤属性制

- 图的研究——以黑龙江鹤山农场研究区为例. 土壤学报, 2009, 46(1): 9—15
- [ 2 ] 连纲, 郭旭东, 傅伯杰, 等. 黄土高原小流域土壤养分空间变异特征及预测. 生态学报, 2008, 28(3): 946—954
- [ 3 ] Ortan T G, Lark R M. Accounting for the uncertainty in the local mean in spatial prediction by bayesian maximum entropy. *Stochastic Environmental Research and Risk Assessment*, 2007, 21(6): 773—784
- [ 4 ] Seung J L, Robert B, Patricia G. Bayesian maximum entropy mapping and the soft data problem in urban climate research. *Annals of the Association of American Geographers*, 2008, 98(2): 309—322
- [ 5 ] 檀满枝, 陈杰, 徐方明, 等. 基于模糊集理论的土壤重金属污染空间预测. 土壤学报, 2006, 43(3): 389—396
- [ 6 ] 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
- [ 7 ] Hengl T, Heuvelink G B M, Stein A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 2004, 120(1/2): 75—93
- [ 8 ] Zhang C, Tang Y, Xu X, et al. Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in ireland. *Applied Geochemistry*, 2011, 26(7): 1239—1248
- [ 9 ] Li Y. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? *Geoderma*, 2010, 159(1/2): 63—75
- [ 10 ] Zhu Q, Lin H S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere*, 2010, 20(5): 594—606
- [ 11 ] 沙洋县人民政府网. 地理环境 2013. <http://sylyj. sybm. gov. cn/html/syjk/dlhj/>
- [ 12 ] 鲍士旦. 土壤化学分析. 北京: 中国农业出版社, 2000: 30—34
- [ 13 ] 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
- [ 14 ] Ziadat F M. Analyzing digital terrain attributes to predict soil attributes for a relatively large area. *Soil Science Society of America Journal*, 2005, 69(5): 1590—1599
- [ 15 ] 秦松, 樊燕, 刘洪斌, 等. 地形因子与土壤养分空间分布的相关性研究. *水土保持研究*, 2008, 15(1): 46—49, 52
- [ 16 ] 张素梅, 王宗明, 张柏, 等. 利用地形和遥感数据预测土壤养分空间分布. *农业工程学报*, 2010, 26(5): 188—194
- [ 17 ] 中国科学院计算机网络信息中心. Landsat etm slc-off 遥感影像条带修复. 2013. <http://datamirror. csdb. cn/landsat/files/gap. jsp>
- [ 18 ] Stacey K F, Lark R M, Whitmore A P, et al. Using a process model and regression kriging to improve predictions of nitrous oxide emissions from soil. *Geoderma*, 2006, 135(1/2): 107—117
- [ 19 ] Sun W, Minasny B, McBratney A. Analysis and prediction of soil properties using local regression-kriging. *Geoderma*, 2012, 171/172: 16—23
- [ 20 ] 连纲, 郭旭东, 傅伯杰, 等. 基于环境相关法和地统计学的土壤属性空间分布预测. *农业工程学报*, 2009, 25(7): 237—242
- [ 21 ] Sumfleth K, Duttman R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators*, 2008, 8(5): 485—501
- [ 22 ] Johnson C E, Ruiz-Méndez J J, Lawrence G B. Forest soil chemistry and terrain attributes in a catskills watershed. *Soil Science Society of America Journal*, 2000, 64(5): 1804—1814
- [ 23 ] 张建方, 王秀祥. 直方图理论与最优直方图制作. *应用概率统计*, 2009, 25(2): 201—214
- [ 24 ] 姚荣江, 杨劲松, 韩建均. 海涂围垦区土壤盐分空间变异模拟的比较研究. *土壤学报*, 2012, 49(2): 275—281
- [ 25 ] 秦鱼生, 涂仕华, 冯文强, 等. 成都平原水旱轮作种植下土壤养分特性空间变异研究. *土壤学报*, 2008, 45(2): 355—359
- [ 26 ] 邓楚雄, 谢炳庚, 李晓青, 等. 长沙市耕地集约利用时空变化分析. *农业工程学报*, 2012, 28(1): 230—237
- [ 27 ] 赵瑞芬, 张一弓, 张强, 等. 不同土地利用方式对土壤养分状况的影响——以太原市为例. *中国农学通报*, 2011, 27(14): 262—266
- [ 28 ] 刘占仁, 王立志. 不同土地利用方式对土壤养分及肥力的影响. *水土保持研究*, 2012, 19(6): 72—76

(责任编辑: 檀满枝)