农田土壤质地空间分布的三维随机模拟及其不确定性评价①

陈 冲¹、胡克林 ^{1*}、贺 勇 ²

(1中国农业大学土壤和水科学系,教育部植物-土壤相互作用重点实验室,北京 100193;

2 加拿大农业部半干旱草原农业研究中心,斯威夫特卡伦特,萨省,加拿大 S9H 3X2)

摘 要:定量描述土体三维构型对于土地利用及农田水肥管理研究极其重要。本研究根据华北山前冲积平原区一块农田内的 109 个土壤剖面观测数据,运用马尔科夫地质统计学方法构建了土壤质地种类的三维空间分布模型,在 100 次随机模拟的基础上,分别得到了土壤质地种类的优化分布图及其概率分布图。结果表明,一维嵌入转移概率模型能很好地描述农田水平和垂直方向上各土壤质地种类的空间连续性及毗邻转移趋势。优化分布图虽能直观反映土壤质地种类的空间分布特征,但存在明显的平滑效应,不能刻画土壤质地种类空间分布的不确定性。而采用概率分布的方式来描述土壤质地种类空间分布的不确定性,能够有效地克服该缺点。

关键词:土壤质地;马尔科夫;地质统计学;空间分布;不确定性

中图分类号: S153; Q948

土壤质地属于多种类变量 ,传统地质统计学描述 种类变量空间相关性的主要工具是指示变异函数。由 于实测数据的稀缺以及变异函数内在的对称性,指示 地质统计学很难获得多种类变量之间复杂的空间关 系[1]。为解决上述问题,近年来马尔可夫链理论被引 入到地质统计领域,该方法采用转移概率来描述种类 变量的空间变化,描述性好,易于理解,且有很好的 随机理论支持[2]。20世纪60年代以来,一维马尔科 夫链已被广泛用来模拟地层序列[3]、土壤粒级的空间 分布^[4]和土壤质地层次的垂直变化特征^[5]。Elfeki 和 Dekking^[6]在一维马尔科夫链模型的基础上提出了双 重马尔科夫链模型。Li 等[7]将双重马尔科夫链模型 进一步扩展到三重马尔科夫链。目前这些模型可用来 描述二维种类变量空间分布,但不能用来描述三维土 体,其实际应用受到了一定限制。Carle 和 Fogg[1-2] 构建了三维马尔科夫链模型,实现了种类变量空间分 布的三维随机模拟,该方法已被广泛地用来模拟冲积 扇[8-9]、含水层[10-11]、土壤质地[12]和土壤黏粒含量[13] 等。

在未采样点土壤属性取值的不确定性对于农田水肥管理及环境风险评价是十分重要的。随机模拟是不确定性评价常用到的方法,近年来在土壤盐分^[14]、土壤饱和导水率^[15]、重金属含量^[16]等方面得到了广

泛应用。Zhang 和 Li^[17]对土壤质地种类空间分布的不确定性进行了报道,但该研究主要限于二维空间情况,对于土壤质地三维空间分布不确定性评价的研究目前还少见报道。

因此,本文以华北山前冲积平原区的一块农田为例,利用马尔科夫地质统计学方法对土壤质地三维空间分布特征进行了分析,并对各种质地种类发生的不确定性进行评价,旨在为精确农业和环境风险评价研究提供定量化的信息。

1 材料与方法

1.1 样品采集与处理

试验点位于北京市海淀区东北旺乡 属于华北山前冲积平原区,海拔 50~m,地下水埋深 $18\sim20~m$ 。年均气温 11.5~° ,年均降雨量 540~mm , $70%\sim80%$ 的降雨主要发生在 6-9~月,是典型的温带大陆性季风气候。土壤类型为草甸褐土,该区冬小麦-夏玉米轮作。研究区为一块面积为 $55~m\times60~m$ 的农田,将其划分成 $5~m\times5~m$ 的网格,采用荷兰 Eijkelkamp 公司的柱状圆筒采样器取样,采取 1m 的原状土样,根据土壤发生层次分层取样,记录各层厚度。采样时间为 2001~年 4~月,实际取样 109~个(图 1)。

基金项目:国家重点基础研究发展规划项目(2009CB118607)和教育部新世纪优秀人才支持计划项目(NCET-07-0809)资助。

作者简介:陈冲(1986—),男,山西晋城人,博士研究生,主要从事土地利用与信息技术的研究。E-mail: chenchongsxnd@163.com

^{*} 通讯作者 (hukel@cau.edu.cn)

土壤质地采用吸管法进行测定,按美国制进行分类,结果如图 2 所示。可看出该研究区土壤质地主要分布在砂壤、壤土和粉壤的区域,位于其他质地类型区的样点很少,根据就近原则将其合并到相邻的区域,土壤质地最终划分为 3 种:砂壤、壤土和粉壤,分别用 sl、l、zl 来表示。

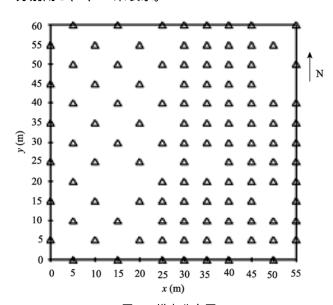


图 1 样点分布图 Fig. 1 Sampling locations

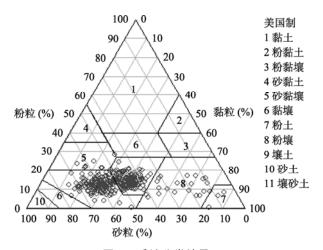


图 2 质地分类结果 Fig. 2 Soil texture classification

1.2 马尔科夫地质统计学

马尔科夫地质统计学是解决种类变量空间关系的有效方法 ,它的基本工具为转移概率矩阵 ,用 $t_{jk}(h_{\phi})$ 来表示 , 可定义为 :

$$t_{jk}(h_{\phi}) = \{k \times x + h_{\phi}$$
处发生 $\mid j \times x$ 处发生 $\}$ (1) 式中 j , $k = 1$, 2 , $\cdots K$ 定义为不同的种类变量 , x 为空间位置 , h_{ϕ} 为在方向 ϕ 上的间隔步长 , $t_{jk}(h_{\phi})$ 可以看作条件概率 , 即点 x 处种类 j 发生的条件下 , 点 $x + h_{\phi}$

的位置处种类 k 发生的概率。假设随机变量空间平稳, $t_{jk}(h_{\phi})$ 仅仅依赖于空间步长 h,而不依赖于 x 的位置[1]。

马尔科夫地质统计学模型常用的有连续型、离散型和嵌入型。连续步长的转移概率模型可以用矩阵的指数形式来表示^[2],

$$T(h_{\phi}) = \exp[R_{\phi}h_{\phi}] \tag{2}$$

其中, $T(h_{\phi})$ 为 $K \times K$ 阶转移概率矩阵,其中 K 为种类变量的数目, R_{ϕ} 为转移强度矩阵,矩阵中的元素 $r_{jk,\phi}$ 为在 ϕ 方向上从种类j 到种类 k 每单位长度转移概率变化的强度。转移概率与转移强度的关系可用下式表示。

$$\frac{\partial t_{jk}(h \to 0)}{\partial h_{\phi}} = r_{jk,\phi} \tag{3}$$

离散型马尔科夫链模型可用 $T(n\Delta h_{\phi}) = T^n(\Delta h_{\phi})$ 来表示,式中 n 步转移概率等于一步转移概率的 n 次幂,当 $h_{\phi} = 0$ 时 T(0) = I,其中 I 为单位矩阵 T(0) = I ,其中 T(0) = I

马尔科夫地质统计学提供了 3 个参数(分布比例、平均长度和毗邻转移趋势)来直观地描述种类变量的空间分布特征。首先,分布比例反映的是变量在研究区域所占的体积百分比。例如,研究区域有 K种质地类型,各质地类型的分布比例记为 $p_k(k=1,2,\cdots K)$,则有 K种质地种类的分布比例总和为 1。假设平稳,分布比例在转移概率模型对应于模型的"基台值"(图 3),用下式表示:

$$\lim_{h \to \infty} t_{jk}(h_{\phi}) = p_k \tag{4}$$

其次,平均长度常用来表征变量的空间连贯性,某一变量 k 在 ϕ 方向上的平均长度 $\overline{L}_{k,\phi}$ 等于其在方向上的总长度 $L_{k,\phi}$ 除以发生的总次数 $N^{[1]}$ 。实际应用中常常通过自转移概率模型过(0,1)点的切线与横坐标的交点来得到(图 3),用下式表达:

$$\frac{\partial t_{kk}(0)}{\partial h_{\phi}} = -\frac{1}{\overline{L}_{k,\phi}} \tag{5}$$

最后,毗邻趋势表达的是在方向 ϕ 上种类变量之

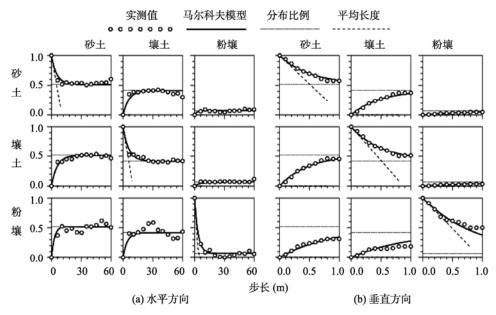


图 3 水平和垂直方向上实测转移概率函数和马尔科夫链模型

Fig. 3 Matrix of horizontal (X,Y)-direction and vertical (Z)-direction transition probabilities and Markov chain model

间是否具有明显的转移趋势,这种趋势可以通过嵌入转移概率矩阵来描述^[8]。精确地刻画这种毗邻转移趋势可以通过比较实测转移强度 r_{jk} 与最大无关转移强度 $\hat{r}_{jk,\phi}$ 来得到,如下式:

$$r_{jk,\phi} = \alpha_{jk,\phi}(\hat{r}_{jk,\phi}) \quad j \neq k \tag{6}$$

式中, α_{jk} 为体现变量间空间毗邻转移趋势的系数,值越大表示在 ϕ 方向 j 向 k 毗邻转移的趋势越强。在得到平均长度 $\overline{L}_{k,\phi}$ 和分布比例 p_k 后,通过(7)式可得到 $\hat{r}_{jk,\phi}$ 。

$$\hat{r}_{jk,\phi} = \frac{p_k}{\overline{L}_{k,\phi}(1 - p_j)} \tag{7}$$

实际应用时,马尔科夫地质统计学方法的具体步骤如下: 计算水平和垂直方向多种类变量之间的离散转移概率; 建立水平和垂直方向的一维连续马尔科夫链模型,进而得到三维马尔科夫链模型; 用马尔科夫链模型代替序贯指示模拟中的变异函数模型进行条件模拟,具体过程与序贯指示模拟相同,详见参考文献[15,18]; 用退火模拟算法改进条件模拟。

1.3 模型拟合效果评价

模型模拟的土壤质地种类转移概率值与实测值的吻合程度,采用均方根误差(RMSE)来评价。

RMSE =
$$\sqrt{\sum_{i=1}^{n} \frac{(P_i - O_i)^2}{n}}$$
 (8)

式中 O_i 表示实测值 P_i 表示和观测值对应的模拟值 ,

n 表示数据对数。

1.4 不确定性评价

令 Z(x)为采样区内空间任一位置x处的随机模拟结果, Z_k 为某种土壤质地类型,则空间位置x处为该土壤质地种类的概率 $P\{Z(x)=Z_k\}$ 按下式计算:

$$P\{Z(x) = Z_k\} = \lim_{l \to \infty} \frac{N\{Z(x) = Z_k\}}{L}$$
 (9)

式中, $N\{Z(x)=Z_k\}$ 为 x 处一系列随机模拟值中土壤质地种类为 Z_k 的次数 ;L 为随机模拟的次数 ,实际应用中要求 $L \ge 100^{[14]}$ 。

2 结果分析

2.1 土壤质地种类的空间变化特征

分布比例是转移概率模型的一个重要参数。对于研究区 109 个采样剖面,垂直方向上以 0.05 m 的间隔进行离散,将每个离散位置处的土壤质地种类数据转换成指示数据(0 和 1),表示某土壤质地种类在该位置处是否出现。统计结果表明,砂壤、壤土和粉壤的分布比例分别为 0.515、0.416 和 0.069,说明研究区最多的土壤质地类型为砂壤,其次是壤土,最少的是粉壤。

假定水平方向(X 和 Y)为各向同性,将水平方向和垂直方向的间隔分别设为 5 m 和 0.05 m,先计算水平方向和垂直方向上的转移概率,然后用马尔科夫链模型进行了拟合,结果如图 3 所示。通过式(8)计算得到了水平和垂直方向上的 RMSE 分别为 4.8%和

2.8%, 说明土壤质地种类转移概率的模拟值与实测值 具有很好的一致性, 本文所建立的马尔科夫链模型可 用来描述该研究区土壤质地种类的空间关系。

$$\Pi_{x,y} = 1 \begin{bmatrix} \overline{L} = 10.5 \text{ m} & 0.72 & 0.28 \\ 0.89 & \overline{L} = 9.9 \text{ m} & 0.11 \\ 0.62 & 0.38 & \overline{L} = 3.7 \text{ m} \end{bmatrix} (10)$$

$$\begin{array}{c} \text{sl} & 1 & \text{zl} \\ \text{sl} & \overline{L} = 1.05 \text{ m} & 0.91 & 0.09 \\ 0.95 & \overline{L} = 0.88 \text{ m} & 0.05 \\ 0.56 & 0.44 & \overline{L} = 0.97 \text{ m} \end{bmatrix} (11)$$

本文中水平方向和垂直方向上的马尔科夫链模型均采用嵌入转移概率矩阵[[来描述。从[[x,y](10 式)和[]; (11 式)对角线的平均长度可以看出,在水平方向上,砂壤和壤土的平均长度分别为 10.5 m 和 9.9 m,较粉壤的平均长度(3.7 m)要大,说明水平方向上砂壤和壤土要比粉壤要连续;在垂直方向上,3 种质地种类的平均长度分别为 1.05、0.88 和 0.97 m,相差并不大,接近于采样深度,这说明垂直方向上砂壤、壤土和粉壤分布都比较连续。总体来看,各质地类型在水平方向的平均长度大于垂直方向,说明水平方向各质地类型的连贯性要大于垂直方向,这与许多研究者得到的结果是一致的[8,12]。

由(10)式和(11)式非对角线的转移概率可以看出,水平方向砂壤和壤土相邻出现的概率(0.72 和0.89)最大,而壤土与粉土相邻出现的概率(0.38 和0.11)最小。垂直方向从下到上,砂壤到壤土的转移概率(0.95)和壤土到砂壤的转移概率(0.91)都比较大,这说明垂直方向砂壤和壤土的联系比较紧密。另外,粉壤到砂壤和壤土的转移概率分别为0.56 和0.44,而砂壤和壤土到粉壤的转移概率分别为0.09 和0.05,两者明显不同。通过计算垂直方向从下到上的毗邻转移趋势系数(α值),发现砂壤到粉壤和壤土到粉壤的

 α 值分别为 2.37 和 1.88 ,而粉壤到砂壤和粉壤到壤土的 α 值分别为 1.1 和 0.89 ,这说明垂直方向粉壤更倾向于发生在砂壤的上面。

2.2 土壤质地空间分布的三维优化实现

研究区的范围为 55 m×60 m×1 m,将其离散成 1.25 m×1.25 m×0.05 m的网格单元,共生成 46 305 个节点,对于每个节点随机模拟得到 100 次实现。将每个节点处出现概率最大的质地类型确定为该节点的属性值,绘制得到了土壤质地种类的三维优化空间分布图(图 4),这种处理方式与其他一些研究相同^[17,19]。由于垂直方向研究范围较小,为了获得较好的可视化效果,绘制三维空间分布图时将垂直方向的比例尺扩大了 10 倍。

从图 4 可以看出土壤质地种类呈现镶嵌分布的 特点。在地表砂壤主要集中在西部,而壤土主要集中 在东部,只有少量的粉壤分布在南部和北部。图 4b 为 3 个不同深度处(0 cm、50 cm 和 100 cm)土壤质地 的优化空间分布图。总体看来,不同层次土壤质地种 类分布的趋势是相同的(东部主要为壤土,西部主要 为砂壤),但可以明显的看出上层(0 cm 和 50 cm)土壤 质地种类的分布远远比下层(100 cm)要复杂,这可能 与人类的活动有关。在不同深度的水平方向,可以直 观地看到砂壤和壤土分布都比较连续,砂壤与壤土相 邻出现的概率比较大,在出现粉壤的地方,相邻出现 的土壤质地种类主要为砂壤。图 4c 是土壤质地三维 优化空间分布图的纵截面,结合 4a 可以看出,在垂 直方向上,很多地方土壤质地层次的排列通体都是一 样的,反映了土壤质地在研究区的垂直方向上具有较 好的连续性;上层为砂壤下层为壤土以及上层为壤土 下层为砂土的出现概率都比较大,当上层出现粉壤时 下层主要出现的质地类型是砂壤。这些结果与上文所 建立的嵌入马尔科夫链模型和刻画毗邻转移趋势的 系数 α 表述的结果是一致的。

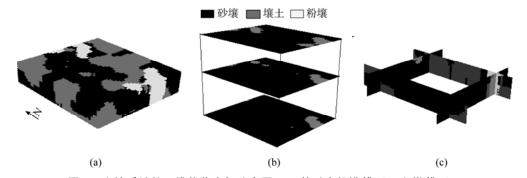


图 4 土壤质地的三维优化空间分布图(a)及其对应的横截面(b)和纵截面(c) Fig. 4 3-D optimal distribution of soil texture (a), corresponding horizontal sections (b) and vertical sections (c)

2.3 土壤质地种类空间分布的不确定性评价

表 1 给出了实测、100 次模拟和优化实现所得到的 3 种土壤质地种类的分布比例。由表 1 可知,虽然土壤质地种类的三维优化空间分布图确定了在未采样点处发生概率最大的土壤质地类型,但这样做却低估了分布较少的质地类型(壤土和粉壤),而高估了占据优势的土壤质地类型(砂壤),具有明显的平滑效应,减小了数据间的变异性,不能反映土壤质地空间分布的不确定性。但是 100 次随机实现所得到的 3 种土壤质地种类的分布比例与实测值很接近,说明该方法可以有效克服优化实现的平滑效应,因此,通过随机实现可以反映土壤质地空间分布的不确定性。

图 5 为土壤质地种类 100 次实现中随机抽取的三次,每次实现均代表了土壤质地种类空间分布的一种可能性。从图中可以看出不同次的实现,土壤质地种类的空间分布总体趋势基本一致,但局部也存在着明显不同,比图 4 要复杂得多。有些学者建议土壤质地

空间分布的这种不确定性用概率分布图的方式来表达则更有效、更精确和更清晰^[17]。

表 1 不同土壤质地类型的分布比例
Table 1 Proportions of different soil textures

质地类型	分布比例		
•	实测	100 次模拟	优化
砂壤	0.515	0.509	0.548
壤土	0.416	0.417	0.395
粉壤	0.069	0.074	0.057

根据 100 次随机实现可以得到不同土壤质地种类的三维空间概率分布图(图 6),图中白色的区域为高概率区,土壤质地种类的三维优化空间分布图(图 4)就是由不同质地种类的高概率区组合得到的。由白色向黑色的过渡区为不确定区域。从这些图可以明显地看出在研究区内某种土壤质地种类主要出现在哪些地方,而且在这些地方发生的概率也可以直接得到。

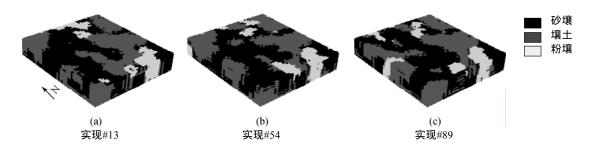


图 5 土壤质地种类的 3 次随机实现 Fig. 5 Three stochastic simulated maps of soil textural classes

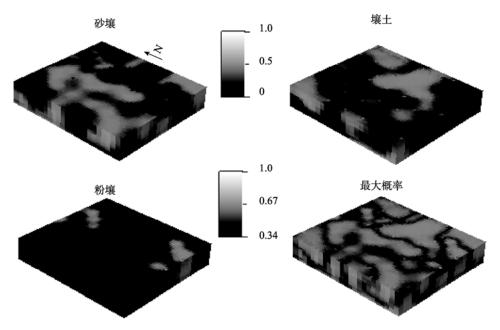


图 6 不同类型土壤质地的三维概率分布图和土壤质地种类的最大出现概率图

Fig. 6 3-D probability distribution of different soil textural classes and maximum occurrence probability maps of soil textural classes.

这些数据可以为精确农业和风险评价提供有价值的信息。土壤质地的最大出现概率图(图 6)可以用来评估优化空间分布图的质量[17],白色区域说明优化空间分布图的质量高。白色区域之间的黑色带状区域为不确定区域。从图中可以看出各节点处土壤质地的最小出现概率为 0.34(黑色区域),最大为 1(白色区域)。

3 结论

本研究通过马尔科夫地质统计学方法分析了一块面积为 55 m×60 m 农田土壤质地种类的空间变化,分别在水平和垂直方向上建立了嵌入马尔科夫链模型,定量地描述了水平和垂直方向上土壤质地种类的变化特征。土壤质地的三维优化空间分布图能够直观地反映土壤质地种类的这些空间分布特征,结果表明砂壤主要集中在西部而壤土主要集中在东部,只有少量的粉壤分布在南部和北部。由于优化分布图存在明显的平滑效应,土壤质地种类的三维概率分布图可用来定量刻画土壤质地种类空间分布的不确定性,这为精确农业和环境风险评价研究提供了有用的信息。

致谢 :感谢中国林业科学院高峻博士为本文研究 提供了试验数据。

参考文献:

- [1] Carle, SF, Fogg GE. Transition probability-based indicator geostatistics[J]. Mathematical Geology, 1996, 28(4): 453–476
- [2] Carle SF, Fogg GE. Modelling spatial variability with oneand multi-dimensional continuous Markov chains[J]. Mathematical Geology, 1997, 29(7): 891–918
- [3] Krumbein, WC. Statistical models in sedimentology[J]. Sedimentology, 1968, 10: 7–23
- [4] Burgess TM, Webster R. Optimal sampling strategies for mapping soil types: I. Distribution of boundary spacings[J]. Soil Science, 1984, 35: 641–654
- [5] Li WD, Li BG, Shi YC. Markov chain simulation of soil textural layers[J]. Geoderma, 1999, 92(12): 37–53
- [6] Elfeki AM, Dekking FM. A Markov chain model for subsurface characterization: theory and applications[J]. Mathematical Geology, 2001, 33: 569–589

- [7] Li W, Zhang C, Burt JE, Zhu AX, Feyen J. Two-dimensional Markov chain simulation of soil type spatial distribution[J]. Soil Science Society of America Journal, 2004, 68(5): 1 479–1 490
- [8] Weismann GS, Fogg GE. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework[J]. Journal of Hydrology, 1999, 226(1): 48-56
- [9] Weissmann GS, Carle SF, Fogg GE. Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics[J]. Water Resources Research, 1999, 35(6): 1 761–1 770
- [10] Ye M, Khaleel R. A Markov chain model for characterizing medium heterogeneity and sediment layering structure[J]. Water Resources Research, 2008, 44, W09427, doi:10.1029/2008WR006924
- [11] 何芳, 吴吉春. 基于马尔可夫链的多元指示地质统计模型[J]. 水文地质工程地质, 2003(5): 28-32
- [12] 贺勇, 胡克林, 李卫东, 李保国. 华北冲积平原区土壤 剖面质地层次空间分布的三维随机模拟[J]. 土壤学报, 2008, 45(2): 193-200
- [13] He Y, Hu KL, Li BG, Chen DL, Suter H, Huang YF. Comparison of sequential indicator simulation and transition probability indicator simulation used to model clay content in microscale surface soil[J]. Soil Science, 2009, 174(7): 395–402
- [14] 姚荣江,杨劲松,赵秀芳,李晓明,刘梅先.沿海滩涂土壤盐分空间分布的三维随机模拟与不确定性评价[J].农业工程学报,2010,26(11):91-97
- [15] 李保国, 胡克林, 陈德立, R.E.White. 农田土壤表层饱和 导水率的条件模拟[J]. 水利学报, 2002, 33(2): 36-40
- [16] 史舟, 李艳, 程街亮. 水稻土重金属空间分布的随机模 拟和不确定评价[J]. 环境科学, 2007, 28(1): 209-214
- [17] Zhang C, Li W. Regional-scale modelling of the spatial distribution of surface and subsurface textural classes in alluvial soils using Markov chain geostatistics[J]. Soil Use and Management, 2008, 24: 263–272
- [18] Deutsch CV, Journel AG. GSLIB, Geostatistical Software Library and User's Guide[M]. New York: Oxford University Press, 1998:43–62
- [19] Park E, Elfeki AM, Yungoo Song and Kangjoo Kim. Generalized coupled Markov chain model for characterizing categorical variables in soil mapping[J]. Soil Science Society of America Journal, 2007,71: 909–917

3-D Stochastic Simulation and Uncertainty Assessment of Soil Texture at Field Scale

CHEN Chong¹, HU Ke-lin^{1*}, HE Yong²

(1 Department of Soil and Water Sciences, China Agricultural University, Key laboratory of Plant-Soil Interactions, MOE, Beijing 100193, China; 2 Semiarid Prairie Agricultural Research Centre, Agriculture and Agri-Food Canada, Box 1030, Swift Current, Saskatchewan S9H 3X2, Canada)

Abstract: Quantitative description spatial distribution of soil texture in three dimensional is very important for the studies of land use and water and fertilizer management. In this study, we constructed 3-D Markov chain model based on observation of 109 soil profiles in piedmont alluvial plain on the North China, and then optimal prediction map and occurrence probability map of soil textural classes in 3-D were obtained based on the 100 stochastic simulations. The results indicated that one-dimensional embedded Markov chain model could well describe the spatial continuity and juxtapositional tendencies of different textural classes both in horizontal and vertical directions. 3-D optimal distribution of soil texture could directly reflect the spatial distribution of soil textural classes. However, it existed obvious smoothing effect and could not describe the uncertainty of spatial distribution of soil textures. The probability distribution map was applied to analyze the uncertainty of spatial distribution of textural classes and it could effectively overcome this shortcoming.

Key words: Soil texture, Markov chain, Geostatistic, Spatial distribution, Uncertainty