

基于 Markov 链地统计模型的区域土壤性质研究进展^①

王伟鹏^{1,2}, 李晓鹏¹, 刘建立^{1*}

(1 中国科学院南京土壤研究所, 南京 210008; 2 中国科学院研究生院, 北京 100049)

摘要: 土壤水力学性质的空间变异对于区域土壤水分溶质循环模拟研究至关重要。基于 Markov 链的条件模拟是一种能融合多源信息技术的地统计学模拟方法, 与传统插值法和基于变异函数的条件模拟相比有诸多优势。目前, 该方法在土壤水力学性质空间变异性领域的研究并未全面展开。实现区域土壤水力学模型参数的随机模拟, 对于实现区域土壤水分运动和溶质运移的随机模拟, 分析土壤水力学性质空间变异性对土壤水分运动和溶质运移模拟结果的影响, 特别是参数采样点变化对土壤水分运动和溶质运移结果影响的不确定性等研究都有重要意义。本文旨在综述基于 Markov 链的地统计学模拟在土壤学相关领域的研究进展, 以期为区域模拟中面临的参数获取难题提供帮助, 为区域农业生产管理, 水分高效利用, 农田生态环境保护提供科学依据。

关键词: Markov 链; 地统计学模拟; 土壤参数; 空间变异

中图分类号: S271

随着计算机模拟技术和优化管理科学的飞速发展, 区域尺度的农田物质循环模型和优化管理决策工具的开发与应用已取得了长足的进步, 地统计学原理的推广使得人们对土壤性质与其空间位置之间关系的认识更加深入^[1-4]。AGNPS^[5]、SWAT^[6-7]、ANSWERS^[8-9]、WNMM^[10]等模型都已得到了较为广泛的应用。由于长期受到人为和自然因素的强烈影响, 土壤的空间分布既具有结构性, 也呈现出一定的随机性^[11-15]。不同质地土壤构成的层次可能延伸数公里, 也可能在局部相互叠加, 出现砂、黏土相间或出现其他障碍层。水力学性质差异巨大的土壤层次组合, 会对土壤水分转化和溶质(肥料、污染物、盐分等)迁移过程(水分的保持、污染物的优势迁移等)产生重要影响^[16-17]。区域尺度管理模型的土壤参数, 不仅要求其能体现出发育在大尺度上的空间结构, 还必须表达出某一区段或田块尺度的变异特征。目前, 土壤属性数据的采集与空间尺度扩展的常用方案仍然为: 研究区域内网格布点-现场采样-室内分析-地统计学插值。由于缺乏科学布点依据以及插值法的固有缺陷, 其结果往往不能满足计算机模型参数的精度需求, 这使得模型(特别是大尺度模型)的应用受到了很大限制^[18]。

为解决上述问题, Carle 和 Fogg^[19]提出了一种基于转移概率——Markov 链的多元指示地统计学模型。

与传统的基于交叉变异函数(cross-variogram)的多元地统计学模型不同, 该模型采用转移概率(transition probability)来描述区域变量的空间变异性, 并利用序列指示模拟或模拟退火算法, 根据实测数据来生成随机场的“实现”。该方法在模拟过程中可以方便地结合有关的地统计信息, 反映模拟变量的空间分布特征和相互转换规律, 而且能体现出实际分布中的不对称性和各向异性等特征。转移概率的引入也使得该方法与基于变异函数的传统模拟相比更为简单方便和易于理解^[20]。采用基于 Markov 链的地统计模型, 用转移概率来描述区域化变量的空间变化, 描述能力强, 易于理解, 且有很好的随机理论支持。

1 Markov 基本原理概述

1.1 Markov 性与 Markov 过程

Markov 性: 过程(或系统)在时刻 t_0 所处的状态为已知的条件下, 过程在时刻 $t > t_0$ 所处状态的条件分布与过程在 t_0 时刻之前所处的状态无关, 这种性质称为无后效性。**Markov 过程:** 设随机过程 $\{X(t), t \in T\}$ 的状态空间为 I 。如果对时间 t 的任意 n 个数值 $t_1 < t_2 < \dots < t_n, n \geq 3, t_i \in T, i=1, 2, \dots, n$, 在条件 $X(t_i) = x_i, x_i \in I, i=1, 2, 3, \dots, n-1$ 下, $X(t_n)$ 的条件分布函数恰好等于在条件 $X(t_{n-1}) = x_{n-1}$ 下的条件分布函数:

①基金项目: 国家自然科学基金项目(40871105)和中国科学院知识创新工程重大项目(KSCX1-YW-09-05)资助。

* 通讯作者(jiliu@issas.ac.cn)

作者简介: 王伟鹏(1985—), 男, 山西长治人, 硕士研究生, 主要从事土壤水分运动、地统计学及相关研究。E-mail: wpwang@issas.ac.cn

$$P\{X(t_n) \leq x_n \mid X(t_1) = x_1, X(t_2) = x_2 \dots X(t_{n-1}) = x_{n-1}\} = P\{X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}\}, \quad x_n \in R \quad (1)$$

则称过程 $\{X(t), t \in T\}$ 具有 Markov 性，并称此过程为 Markov 过程。Markov 过程是研究事物的状态及其转移的理论，它既适合于时间序列，又适用于空间序列。

1.2 Markov 链与转移概率

时间和状态都是离散的 Markov 过程称为 Markov 链，Markov 链是一种能用数学分析的方法，一般用来研究事件中的随机过程，它的特点是作为一种特殊的随机事件序列，其序列的所有历史信息都可通过其现在的状态来反映，这种性质称为马氏性。一般把 Markov 链记为：

$$\{X_n = X(n), n = 0, 1, 2, \dots\} \quad (2)$$

记链的状态空间为 $I = \{a_1, a_2, \dots\}$ ， $a_i \in R$ ， $I = \{a_1, a_2, \dots\}$ ， $a_i \in R$ ，称条件概率 $p_{ij}(m, m+n) = P\{X_{m+n} = a_j \mid X_m = a_i\}$ 为马氏链在时刻 m 处于状态 a_i 条件下，在时刻 $m+n$ 转移到状态 a_j 的转移概率。由转移概率组成的矩阵 $p(m, m+n) = \{p_{ij}(m, m+n)\}$ 称为马氏链的转移概率矩阵。

2 地质统计中的 Markov 链模型^[21]

基于 Markov 链的地质统计模型采用的转移概率 $t_{ij}(l)$ 定义为点 x 处是地质类型 i 的条件下点 $(x+1)$ 处是地质类型 j 的概率，即条件概率：

$$X_{ij}(l) = P\{\text{点 } x+l \text{ 是地质类型 } j \mid \text{点 } x \text{ 是地质类型 } i\} = P\{\text{点 } x+l \text{ 是地质类型 } j \text{ 且点 } x \text{ 是地质类型 } i\} / P\{\text{点 } x \text{ 是地质类型 } i\} \quad (3)$$

目前，在地统计模型领域，模型中所采用的 Markov 链主要有 3 种，分别为连续型、离散型和嵌入型。在模型的实际应用中，由于地质或区域土壤类型空间分布的最基本性质是连续性及渐变性，这就要求模拟其空间特征的地统计模型也应具有相应的特性——连续性，传统基于地统计学插值方法建立的模型是根据实测数据生成的随机场的唯一的、最优的线性估计，其结果总是比现实更为光滑，这在一定程度上过于强调地质或土壤类型的连续与渐变性而不能准确描述地质或土壤类型的空间分布的差异性，因此其在地质或土壤类型差异不大的地区模拟效果比较满意，而在地质或区域土壤类型分布差异较大的地区以及大尺度范围内土壤地质空间模拟方面具有很大的不确定性，所以插值方法的固有缺点——平滑效应，是这类基于插值方法的地统计模型的局限性。而基于 Markov 链的条件模拟地统计模型，既可以用连续的 Markov

链模型精确地模拟地质或土壤类型空间分布的连续性，又可以用转移概率的方法来描述地质或土壤类型空间分布的非连续与不对称性，这就避免了数据扩展所带来的平滑效应。因此，精确的建模过程都应该根据研究区地质或土壤类型空间分布的基本特点采用不同类型的马尔科夫链模型。

2.1 连续时间的 Markov 链

连续的马尔科夫链，在地质统计中表现为地质或土壤类型在空间上的连续分布，用空间 l 距离来代替连续时间 Markov 链的 t ，在理论上即可获得任意距离 l 的转移概率。连续 Markov 链的转移概率矩阵函数表达式^[22]为：

$$t(h_l) = e^{R_l h_l} \quad (4)$$

其中， $t(h_l) = \begin{bmatrix} t_{11,l} & \dots & t_{1k,l} \\ \vdots & \ddots & \vdots \\ t_{k1,l} & \dots & t_{kk,l} \end{bmatrix}$ 为 $k \times k$ 的转移概率矩

阵； $R_l = \begin{bmatrix} r_{11,l} & \dots & r_{1k,l} \\ \vdots & \ddots & \vdots \\ r_{k1,l} & \dots & r_{kk,l} \end{bmatrix}$ 为 $k \times k$ 的转移强度矩阵，它

表示转移概率的变化率。(4) 式表明在连续的 Markov 过程中，任意距离 l 的转移概率可以由转移强度矩阵计算获得，并且唯一确定，与传统基于交叉变差函数的地统计模型使用的各类拟合方法来求交叉变差函数相比，使用 Markov 链地质统计模型计算连续空间上的转移概率显的简单准确，且有良好的理论支持。但是，在实际建模过程中，由于数据量的限制，更多情况下是采用离散 Markov 链和嵌入型 Markov 链。

2.2 离散时间 Markov 链

在连续 Markov 过程中，若取时间 t 、距离 l 是一些离散的点时，即为离散 Markov 过程。其转移概率有如下性质^[22]：

$$T(n\Delta h_l) = T^n(\Delta h_l) \quad T(0) = I \text{ (单位矩阵)} \quad (5)$$

公式 (5) 式表明 n 步转移概率矩阵等于一步转移概率矩阵的 n 次幂。可以看出离散 Markov 过程的概率分布完全由其单位距离的转移概率矩阵确定，因此在建立 Markov 链地统计模型时，主要是计算单位转移概率矩阵。实际建模时，可以选取恰当的空间距离，用离散 Markov 过程来近似描述地质类型的空间连续分布，这个过程的基本思路就是在选取恰当空间距离 Δh_l 的基础上，用数学方法中的微分和积分的思想将一系列不连续的小段的地质或土壤类型进行归类，然后将这些归类的小段连接起来，运用连续时间的 Markov 链模型进行处理，得到近似数据，从而解决离

散时间或距离的 Markov 过程。

2.3 嵌入型 Markov 链

目前, 部分地统计学模型应用研究实例以变异系数为研究手段, 以经验数学函数(如球形函数、指数函数等)来拟合实测的变异函数, 再依据变异函数公式来进行地统计学模拟。较为常用的地统计学模拟方法有转向带法(turning band)、序列指示模拟(sequential indicator simulation)和模拟退火算法(simulated annealing)等。但这些函数在描述地质空间或土壤类型的空间复杂连续性时有局限性且变异函数的拟合方法目前尚没有统一标准, 这也影响了模型的可靠性。实际上, 由于数据或其他原因的限制, 地质模型中应用最为广泛的是嵌入式 Markov 链。与基于时间序列的 Markov 链相比, 基于事件序列 $\{X_n\}$ 的 Markov 链称为嵌入式 Markov 链。在地质统计中, 比如在垂直方向 Z 上, 我们定义嵌入式 Markov 链的转移概率为地质类型 i 分布在下面的前提下, 地质类型 j 分布在上面的概率, 即条件概率:

$$\pi_{ij} = P\{\text{地质类型 } j \text{ 分布在上面} \mid \text{地质类型 } i \text{ 分布在下面}\} \quad (6)$$

嵌入 Markov 链的转移概率 $\pi_{ij,l}$ 和连续时间 Markov 链的转移强度 $r_{ij,l}$ 之间存在以下关系^[23],

$$r_{ij,l} = \pi_{ij} / L_{j,l} \quad (7)$$

式中, $L_{j,l}$ 为地质类型 j 在 l 方向上的平均长度。

与一般基于时间序列的 Markov 链相比, 嵌入 Markov 链事件的发生和转换更易于观察, 且误差较小。在数据量有限的情况下, 如果已知地质类型的平均长度, 并获得了嵌入式 Markov 链的转移概率, 则可通过公式(7)就可以得到连续 Markov 链转移强度。因此, 基于嵌入式 Markov 链地质统计模型具有更强的灵活性与实用性。

3 基于 Markov 链的地质统计模型的优点

随着土壤科学研究尺度的扩展, 土壤层次分布的空间变异性研究一直是农学、土壤学、水文学和地质学等相关领域的热点问题^[24-28]。各国学者从不同的角度进行探讨, 并分别形成了基于参数分区、标定理论、传统和地统计学空间插值、随机模拟等理论方法。与传统的地质统计模拟方法相比, 除了连续 Markov 链自身是一个随机过程, 用它来描述地统计信息可以很好地体现实地地质类型分布、转换的连续性和随机性外, 以转移概率作为描述区域化变量空间变化的指示变量还考虑了实际地质类型分布可能存在的不对称性, 并将地质统计信息充分融入到所建模型中, 使所建模型

更好地反映地质类型和土壤类型空间分布的实际情况。与传统的地统计模型或方法相比, 基于 Markov 链的地统计模型主要有以下 3 个优点。

3.1 空间连续的描述

参数分区法是较早用于解决介质空间非均质性的方法, 在地下水流和溶质运移模拟^[29]、土壤水运动和水土保持^[30]和精准农业管理^[31]等领域都有广泛应用。该方法的优点在于操作简单易行, 缺点则是模拟精度依赖于分区数, 分区过多虽会提高模拟精度, 但又会引起结果的不确定性。由于连续时间的 Markov 链在地质统计中表现为空间连续分布, 用时间连续分布描述地质类型分布, 在理论上可以获得任意距离 h_l 的转移概率。因此, 以转移概率作为描述区域化变量空间变化, 并把转移概率作为指示变量的 Markov 链地统计模型可以很好地描述地质类型和土壤类型分布的空间连续性, 既可以保证模型的精度又有操作简单的特点, 优于传统地质统计模型中的交叉变差函数, 因为交叉变差函数常常为非连续可导, 难以刻画空间连续性。

3.2 地质分布不对称性分布趋势的考虑

现实中广泛存在由各种原因形成的不对称性的地质或土壤类型的空间分布, 如冲积或定向风力形成的土壤质地由上游至下游呈现出明显的由粗变细的趋势, 如果采用变异函数进行空间结构分析, 则不能体现出这一重要的结构特征^[32]; 这就影响了所建模型的仿真性, 从而使结果不能真实反映实际地层的分布特征。Kriging 插值虽然可解决空间变量的结构性和随机性、空间相关性和依赖性、空间格局与变异等问题, 但插值引起空间平滑效应仍不可避免, 且不能反映变量的不确定性。常见的随机模拟方法 Monte-Carlo 模拟, 该方法仅考虑变量的随机性, 却忽视了其空间不对称性, 且存在运算效率过低, 过分依赖于已有资料, 在小样本条件下误差很大等问题。另外, 由于实际地层分布常表现出一定的随机性的分布趋势, 要建立反映实际地层特性的数学模型, 必须考虑地层的不对称性的分布趋势。传统地质统计模型多采用交叉变差函数建模, 多是建立在地质类型对称分布的前提假设的基础上, 根据交叉变差函数的定义对于任意步长 h_l 恒有式(8)成立。

$$t_{ij}(h_l) = t_{ij}(-h_l) \quad (8)$$

而基于 Markov 链建立的数学模型, 理论上(8)式并不总是成立, 这就可以很好地描述地层分布的不对称性。在连续的地质统计模型中, 如果地质模型间的转换是对称的(即可逆性), 则转移概率满足式(8),

转移概率在 l 的反方向 $-l$ 上的定义为：

$$T_{ij}(h_{-l}) = (p_j / p_i) t_{ij}(h) \quad (9)$$

由 (8)、(9) 可得：

$$p_i t_{ij}(h_i) = p_j t_{ji}(h_i) \quad (10)$$

式中， p_i 、 p_j 分别表示地质类型 i 、 j 的分布比例。因此在知道了地质类型的分布比例和转移概率后，由式

(10) 即可判定地质类型的分布是否具备对称性，这就找到了在理论上区分地质或土壤类型空间分布连续与否的模拟标准，为模型的计算机自动化程度的提高和推广应用提供了良好支撑，这对于大尺度范围内的地质或土壤类型空间分布的地统计模型的建立具有十分重要的意义，不仅降低了人工调试模型参数的工作量，而且可以建立行业统一标准，利于同行的交流。

3.3 各向异性处理

在实际模拟地质或土壤类型的空间分布时，各地理要素空间分布的各向异性的处理过程非常复杂。一个各向异性的交叉变差函数，随着方向的改变，众多的模型参数都要进行实时的修正，模型将变得非常复杂，需要用多个步骤才能完成这个过程，工作量巨大。解决地质或土壤类型的空间分布多采用传统基于交叉变差函数的地统计学模拟或地统计学插值的方法，二者之间差异比较明显，传统地统计学模拟不但需要用到 Monte-Carlo 随机模拟过程，还要求保持变量的空间相关性。其与地统计学插值区别在于：地统计学模拟可以得到若干随机场的“实现”(realization)，每个“实现”都是非光滑的且具有相同的出现概率，其中任一个“实现”都无法证明比其他“实现”更接近于现实。一般来说以上地统计模拟方法都是基于一维空间某一个方向上地质特征的统计描述。在实际操作中，基于 Markov 链的地统计模型中的转移概率值通常会在不同的方向、不同的距离不断改变。转移概率理论基础是基于空间距离函数一致的各向同性，在这种情况下，可采用一维方向的统计特征来反映三维的地质模型。但在实际更为复杂的地质环境中，由于大部分实际地质或土壤类型的空间分布都是各向异性的，使得各向同性的转移强度 Markov 模型几乎没有实际意义，因此必须把一维的 Markov 模型转化成为三维的 Markov 模型。但是三维空间的方向无穷多，实际的采样数据不足以用来描述空间任意方向的分布特性。因此对于各向异性的 Markov 链的建模过程，首先要将某一方向设定为主方向，这通过结合定性定量的信息就可以做到。例如，主方向可以通过地质或土壤类型的分布特征，如土壤层次的延伸方向、土体坡向、土壤变异面等来确定，通过这些方向构造迪卡尔三维坐

标系 x 、 y 、 z 。第二步是求出主方向上 Markov 链的转移强度矩阵。第三步是建立将任意方向 Markov 链的转移强度转化成主要方向转移强度的数学表达式。设 x 、 y 、 z 是三维 Markov 链的 3 个主方向，任意方向 l 上 Markov 链的转移强度一定满足式：

$$|r_{ij,l}| = \sqrt{\left(\frac{h_x}{h_l} r_{ij,x}\right)^2 + \left(\frac{h_y}{h_l} r_{ij,y}\right)^2 + \left(\frac{h_z}{h_l} r_{ij,z}\right)^2} \quad \forall i, j \neq \beta \quad (11)$$

式中， $r_{ij,x}$ 、 $r_{ij,y}$ 、 $r_{ij,z}$ 分别为主方向 x 、 y 、 z 上的 Markov 链的转移强度； β 表示建立模型过程中用来作为背景色的地质(土壤)类型。负方向 $-l$ 上 Markov 链的转移强度则定义为：

$$h_l = \sqrt{h_x^2 + h_y^2 + h_z^2} \quad r_{ij,-l} = \left(\frac{p_i}{p_j}\right) r_{ij,l} \quad (12)$$

通过上述 3 个步骤，就可以比较准确地解决在实际模拟地质和土壤类型的空间分布时，各地理要素空间分布的各向异性的问题。

3.4 结合其他手段

目前，区域土壤尺度上的高精度、低成本、快速的介质属性获取方法缺乏，大规模采样的野外工作量巨大、成本高昂，而插值法影响模拟结果的不确定性等问题，一方面需要引进新的技术手段，如探地雷达(GPR)、GIS 和遥感技术等，以尽可能多地获取研究区的实测资料，通过增加确定点数量来降低模型预测结果的不确定性；另一方面，改进预测方法，尽量从有限的采样数据中挖掘更多的信息，在现有采样条件下获取尽可能接近于现实土壤(质地或其他属性)的三维空间分布特征。与其他基于变异函数的地统计学模型相比，虽然地统计学插值法比较更接近于现实，但在实际应用中仍存在明显的缺陷：①变异函数的拟合需要大量的空间实测数据来拟合，对于土壤水力学特征参数来说，广泛的田间采样意味着将要花费大量的资金和劳动力；②研究过程中已有的主观判断信息、经验资料往往无法得到直接利用^[33]。而基于 Markov 链的条件模拟能较大幅度地融合传统采样、GPR 扫描等技术所获取的信息资料，利用条件模拟实现区域尺度土壤空间结构和有效属性(以土壤水力学性质为例)的三维重现；利用区域尺度水文模型对物质循环过程进行模拟，评价预测结果的不确定性并进行风险评估，再进一步地确定采样点的合理密度和布置方案。研究中选择土壤水力学性质(土壤水分特征曲线和饱和导水率)作为具体研究对象。因为土壤水力学性质是影响水分和溶质迁移的关键的因素，是土壤物质循环模拟中必不可少的参数；而且其与土壤质地、结构间存

在很好的关联性，便于利用土壤质地和结构属性参数进行转化。

4 基于 Markov 链的地统计模型的应用

4.1 基于 Markov 链的地统计模型在水文、土壤学模拟领域中的应用

目前，与土壤学相近的石油矿藏勘探、水文地质等领域，应用基于 Markov 链的地统计学模拟预测含油/含水介质层次的空间分布研究已得到较为广泛的发展^[34-40]。土壤学领域，已有部分学者应用基于 Markov 链模型的条件模拟对土壤质地层次的空间分布进行研究。Li 等^[4]首先将马尔可夫链模型应用于土壤质地层次的垂向结构预测，并将剖面随机模型与农田水量平衡模型结合，建立了农田水分转化的随机模型。贺勇等^[41-43]应用三重 Markov 链 (triplex Markov chain) 条件模拟和固定路径 Markov 链 (fixed path Markov chain) 条件模拟等方法进行了土壤类型空间分布预测方面的研究，并以概率向量为工具研究了土壤类型空间分布预测中的不确定性^[44]。由于基于 Markov 链的地统计模型充分考虑了地质分布存在的不对称性，简化了各向异性的成立过程，并且在建模过程中，融入了地质类型的分布比例、平均长度和相互间的转换趋势等地质统计信息，因此，可以很好地模拟含水层的空间分布，为进一步模拟地下水运动提供有力支持。美国科学家已经根据这一模型开发了相应的软件包 T-RPOGS，该软件包括 GAMEAS、MCMOD 和 CHUNK 这 3 个主要模块。其中 GAMEAS 用来计算不同地质类型间的转移概率，MCMOD 用于生成一至三维的 Markov 链模型，而 CHUNK 用来显示实际地层的三维空间分布。利用钻孔资料，通过运行 T-RPOGS 软件，即可得到含水系统的三维空间分布。

4.2 基于 Markov 链地统计模型在预测土地利用结构中的应用

殷少美等^[45]将某年地类面积数据作为初始概率矩阵，利用某确定时间段内各地类的年均转换面积得到年均转移概率矩阵，用初始状态矩阵与转移概率矩阵相乘，得到了最终状态矩阵。结果表明，利用 Markov 链的地统计学模型预测土地利用和结构不失为一种有效的手段，可以作为一种有效的方法，为土地利用规划、管理和决策提供支持。贺勇等^[46-47]将土壤质地的三维马尔可夫链模型用于指示条件模拟，得到了研究区土壤质地三维结构的随机实现，取得了较好的结果。

4.3 基于 Markov 链的地统计学模型在研究景观变化中的应用

彭月等^[48]在遥感和地理信息系统的技术支持下，将 Markov 过程模型引入到湿地景观格局分析和预测中来，利用 1990 年和 1996 年 2 个时期的卫星影像来分析四湖地区湿地景观格局的动态变化，取得了很好的效果。

4.4 基于 Markov 链地统计学模型在环境质量预测中的应用

李名升等^[49]把 Markov 链引入环境质量的预测中，将各种污染物的浓度变化过程视作 Markov 过程，通过预测各种污染物的污染负荷系数来推知其浓度值。实例中通过多元回归分析计算出 Markov 转移矩阵，以 2005 年辽宁省 5 种大气污染物的污染负荷系数作为初始向量，预测了其 5 年的变化。然后利用 GM(1,1) 模型预测出降尘的具体浓度值，由此计算出了 2006—2010 年辽宁省 5 种污染物的浓度值。从理论和应用实例看，该方法较为简单、易行，不需要很多的数据，能够解决历史数据较少、序列的完整性及可靠性低的问题。目前文献中较为流行的是完全运用 GM(1,1) 模型分别预测各种污染物，但是该方法适合指数增长的预测，而很多污染物并非是指数形式的增长，造成预测精度偏低。而运用 Markov 链进行预测时可以较好地克服这一局限性。

5 结语

基于 Markov 链条件模拟的地统计学研究，是近年来开始研究的一种地质分析新途径，与传统方法最根本的不同就在于用转移概率代替了交叉变异函数作为描述区域化变量空间变化的指示变量，以上研究仍然以探索研究为主，目的仅局限于建立土壤结构和有效属性的三维随机实现，将其与区域尺度物质循环机理模型结合进行应用，而定量评估模拟结果的不确定性方面的研究，在国内少有报道，其在土壤空间变异性研究方面的重要意义有待进一步分析验证。

参考文献:

- [1] Li W, Li B, Shi Y, Tang D. Application of the Markov chain theory to describe spatial distribution of textural layers. *Soil Science*, 1997, 162: 672-683
- [2] Li WD, Zhang CR, Burt JE, Zhu AX. A Markov Chain-based probability vector approach for modeling spatial uncertainties of soil classes. *Soil Science Society of American Journal*, 2005, 69: 1931-1941
- [3] Li WD, Zhang CR, Burt JE, Zhu AX, Feyen J. Two-dimensional Markov Chain simulation of soil type spatial distribution. *Soil*

- Science Society of American Journal, 2004, 68: 1 479-1 490
- [4] Li Y, White R, Chen DL, Zhang JB, Li BG, Zhang YM, Huang YF, Edis R. A spatially referenced water and nitrogen management model (WNMM) for (irrigated) intensive cropping systems in the North China Plain. *Ecological Modelling*, 2007, 203(3/4): 395-423
- [5] Arnold JG, Engel BA, Srinivasan R. A continuous time, grid cell watershed model // *Proceedings of Application of Advanced Information Technologies for the Management of Natural Resources*. Sponsored by ASAE. June 17-19, 1993, Spokane, WA
- [6] Baily TC, Gatrell AC. *Interactive spatial data analysis*. Harlow, England: Longman Scientific & Technical, 1995
- [7] Beasley DB, Huggins LF, Monke EJ. ANSWERS: A model for watershed planning. *Trans. ASAE*, 1980, 23 (4): 938-944
- [8] Bechini L, Bocchi S, Maggiore T. Spatial interpolation of soil physical properties for irrigation planning. a simulation study in northern Italy. *European Journal of Agronomy*, 2003, 19(1): 1-14
- [9] Bouraoui F, Dillaha TA. ANSWERS-2000: runoff and sediment transport model. *Journal of Environmental Engineering, ASCE*, 1996, 122(6): 493-502.
- [10] Burrough PA. Soil variability: A late 20th century view. *Soils and Fertilizers*, 1993, 56: 529-562
- [11] Carle SF, Fogg GE. Transition probability based indicator geostatistics. *Mathematical Geology*, 1996, 28(4): 453-477
- [12] Carle SF, Fogg GE. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Mathematical Geology*, 1997, 29(7): 891-918
- [13] Carvalho HO, Cassel DK, Hammond J, Bauer A. Spatial variability of in situ unsaturated hydraulic conductivity of Maddock sandy loam. *Soil Science*, 1976, 121(1): 1-8
- [14] Clausnitzer V, Hopmans JW, Nielsen DR. Simultaneous scaling of soil water retention and hydraulic conductivity curves. *Water Resources Research*, 1992, 28(1): 19-31
- [15] Cressie N. *Statistics for Spatial Data*. Chichester, NY: John Wiley & Sons, 1991
- [16] Davis JC. *Statistics and Data Analysis in Geology*. Chichester NY: John Wiley & Sons, 1986
- [17] Deutsch CV, Journel AG. *GSLIB, Geostatistical Software Library and User's Guide*. Oxford: Oxford University Press, 1977
- [18] Elfeki A, Dekking M. A Markov Chain model for subsurface characterization: Theory and applications. *Mathematical Geology*, 2001, 33(5): 569-589
- [19] Seeboonruang U. An application of transition probability approach to geostatistical simulation: A case study in the Lower Chao Phraya Basin, Thailand. *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications*. Penang, 2006
- [20] Snepvangers JJC, Heuvelink GBM., Huisman JA. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 2003, 112(3/4): 253-271
- [21] 何芳, 吴吉春. 基于马尔可夫链的多元指示地质统计模型. *水文地质工程地质*, 2003(5):28-32:
- [22] Steven F, Carle, Fogg GE. Modeling spatial variability with one and multidimensional continuous—Lag Markov Chains. *Mathematical Geology*, 1997, 29(7): 891-918
- [23] 王志明, 侯景儒. 协同克立格法及其在矿产储量计算中的应用. *地质与勘察*, 1994, 30(3): 39-40
- [24] Fogg GE, Noyes CD, Carle SF. Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting. *Hydrogeology Journal*, 1998, 6: 131-143
- [25] Jr Garten CT, Kang S, Brice DJ, Schadt CW, Zhou J. Variability in soil properties at different spatial scales (1m- 1km) in a deciduous forest ecosystem. *Soil Biology& Biochemistry*, 2007, 39: 2 621-2 627
- [26] He Y, Hu KL, Chen DL, Suter HC, Li Y, Li BG., Yuan XY, Huang YF. Three dimensional spatial distribution modeling of soil texture under agricultural systems using a sequence indicator simulation algorithm. *Computers and Electronics in Agriculture*, 2009, doi:10.1016/j.compag.2009.06.012
- [27] Herbst M, Diekkruger B, Vereecken H. Geostatistical co-regiona- lization of soil hydraulic properties in a micro-scale catchment using terrain attributes. *Geoderma*, 2006, 132(1/2): 206-221
- [28] Issaks EH, Srivastava RM. *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press, 1989
- [29] Knotters M, Brus DJ, Oude Voshaar JH. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 1995, 67(3/4): 227-246
- [30] Lee SY, Carle SF, Fogg, GE. Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation. *Advances in Water Resources*, 2007, 30: 1 914-1 932
- [31] Li WD. A fixed-path Markov Chain algorithm for conditional simulation of discrete spatial variables. *Mathematical Geology*, 2007, 39(2): 159-176
- [32] Lin HS, Wheeler D, Bell J, Wilding L. Assessment of soil spatial variability at multiple scales. *Ecological Modelling*, 2005, 182: 271-290

- [33] Neitsch SL, Arnold JG, Kiniry JR, Williams JR. Soil and water assessment tool theoretical documentation. Version 2000. USDA Grassland, Soil and Water Research Laboratory, and Blackland Research Centre of Texas Agricultural Experiment Station, Temple, Texas, USA, 506
- [34] Webster R, Oliver MA. Statistical Methods in Soil and Land Resource Survey. Oxford: Oxford University Press, 1990
- [35] Weissmann GS, Carle SF, Fogg GE. Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics. *Water Resources Research*, 1999, 35(6): 1761-1770
- [36] Weissmann GS, Fogg GE. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *Journal of Hydrology*, 1999, 226(1/2): 48-65
- [37] Willis WO. Evaporation from layered soils in the presence of a water table. *Soil Science Society of American Journal*, 1960, 24(2): 239-242
- [38] Yeh, GT. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, 1986, 22(2): 95-108
- [39] Young RA, Onstad CA, Bosch DD, Anderson WP. AGNPS: A non-point source pollution model for evaluating agricultural watersheds. *Journal of Soil and Water Conservation*, 1989, 44(2): 168-173
- [40] 自由路, 金继运, 杨俐苹, 梁鸣早. 基于 GIS 的土壤养分分区管理模型研究. *中国农业科学*, 2001, 34(1): 1-4
- [41] 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. *土壤学报*, 2008, 45(2): 193-200
- [42] 雷志栋, 胡和平, 杨诗秀. 土壤水科学研究进展与评述. *水科学进展*, 1993, 3(1): 311-318
- [43] 李长青, 邵景力, 靳萍, 崔亚莉. 平原地区水文地质结构条件模拟及其应用: 以华北平原为例. *现代地质*, 2009, 23(1): 137-143
- [44] 刘慧. 基于马尔可夫链的多元指示地质统计学模型及应用(硕士学位论文). 南京: 南京大学, 2004
- [45] 殷少美, 周寅康, 濮励杰, 赵翠微, 赵姚阳. 马尔科夫链在预测土地利用结构中的应用. *经济地理*, 2006, 26: 120-124
- [46] 徐秋宁, 马孝义, 姜宗科, 姬雅珍. 小型集水区降雨径流计算模型研究. *水土保持研究*, 2002, 9(1): 139-150
- [47] 朱耀鑫, 陆锦文, 石元春. 黄淮海平原雨季的土壤水盐运动和黏土夹层的影响. *盐渍土的水盐运动*. 北京: 北京农业大学出版社, 1986: 175-200
- [48] 彭月, 魏虹, 朱韦. 基于马尔科夫模型的土地景观动态模拟研究. *安徽农业科学*, 2006, 34(23): 6172-6173
- [49] 李名升, 佟连军, 任晓霞. 马尔科夫链在环境质量预测中的应用. *上海环境科学*, 2008, 27(1): 21-24

Markov-chain-based Geostatistics Simulation for Regional Soil Hydraulic Properties Estimation: A Review

WANG Wei-peng^{1,2}, LI Xiao-peng¹, LIU Jian-li¹

(1 *Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China;*

2 *Graduate University of Chinese Academy of Sciences, Beijing 100049, China)*

Abstract: Spatial variability of soil hydraulic properties is critical for regional soil moisture simulation. Condition simulation based on Markov chain is a geostatistical simulation method using multiple information sources, and has several advantages comparing to the traditional interpolation and variogram-based conditional simulation. At present, the method in the research of soil hydraulic properties spatial variability is in need of further research. The achievement in regional hydraulic parameters is of great importance for stochastic simulation of soil water and solute transport, impacts of soil hydraulic properties spatial variability on soil moisture movement and solute simulation, especially for the uncertainty of sample sites variation on the simulation results of soil moisture movement and solute transport. This paper gave a review of Markov-chain-based geostatistical simulation in soil science and related fields in order to solve the parameters generating problem in regional simulation and to provide scientific support for regional agricultural management, efficient water use, and agro-environment protection.

Key words: Markov chain, Geostatistical simulation, Soil parameters, Spatial variability