

元胞自动机在城市模拟中的误差传递与不确定性的特征分析

黎 夏¹, 叶嘉安², 刘 涛¹, 刘小平¹

(1. 中山大学地理科学与规划学院, 广州 510275; 2. 香港大学城市规划及环境管理研究中心, 香港)

摘要: 元胞自动机 (Cellular Automata, 简称 CA) 已越来越多地用于地理现象的模拟中, 如城市系统的演化等。城市模拟经常要使用 GIS 数据库中的空间信息, 数据源中的误差将会通过 CA 模拟过程发生传递。此外, CA 模型只是对现实世界的近似模拟, 这就使得其本身也具有不确定性。这些不确定因素将对城市模拟的结果产生较大的影响, 有必要探讨 CA 在模拟过程中的误差传递与不确定性问题。本文采用蒙特卡罗方法模拟了 CA 误差的传递特征, 并从转换规则、邻域结构、模拟时间以及随机变量等几个方面分析了 CA 不确定性产生的根源。发现与传统的 GIS 模型相比, 城市 CA 模型中的误差和不确定性的很多性质是非常独特的。例如, 在模拟过程中由于邻域函数平均化的影响, 数据源误差将减小; 随着可用的土地越来越少, 该限制也使城市模拟的误差随时间而减小; 模拟结果的不确定性主要体现在城市的边缘。这些分析结果有助于城市建模和规划者更好地理解 CA 建模的特点。

关键词: 不确定性; 元胞自动机; 城市模拟; 地理信息系统

文章编号: 1000-0585(2007)03-0443-09

1 引言

很多 GIS 文献都讨论了误差和不确定性问题。与传统方法 (如地图的人工叠置分析) 相比, 以计算机为基础的 GIS 分析能提供更为精确的结果, 但人为误差、技术限制以及自然界的复杂性使得 GIS 同样具有误差和不确定性问题。除极少数特例外, GIS 模型仅是对真实地理变化的近似^[1]。理解 GIS 中的误差和不确定性对于成功应用 GIS 技术来说是很重要的。GIS 的误差主要可分为两种: 1) GIS 数据库中的数据源误差; 2) 利用 GIS 功能进行数据操作时的误差传递。

元胞自动机 (CA) 被越来越多地用于地理现象的过程分析中。CA 模型起初用于数值计算, 现已经广泛地应用于物理、化学和生物等复杂系统的模拟。近十多年来, 学者们提出了结合 GIS 和 CA 模型来模拟复杂城市系统的方法^[2]。与复杂的数学模型相比, 城市 CA 模型有非常简洁的形式, 能很好地与栅格式的 GIS 相结合, 在模拟城市复杂系统方面有许多优势。通过适当地定义 CA 模型的转换规则, 就可以很好地模拟出城市发展的时空复杂性^[3]。在城市建模中应用 CA 模型有助于认识各种城市现象的演变过程。CA 为理解

收稿日期: 2006-05-08; 修订日期: 2007-01-15

基金项目: 基金项目: 国家杰出青年基金资助项目 (40525002); 国家自然科学基金资助项目 (40471105); 教育部博士点基金资助项目 (20040558023)

作者简介: 黎夏 (1962-), 男, 广西梧州人, 特聘教授。从事遥感和地理信息系统研究, 已发表 GIS 和遥感论文 130 多篇。Email: lixia@mail.sysu.edu.cn

城市理论提供了重要信息, 比如城市形态结构特征的形成与演变^[4,5], 同样也可以用于城市规划模型来模拟城市发展的某些规划场景^[6-8]。

尽管有关 CA 的研究相当多, 但 CA 的误差传递和不确定性研究却未引起足够重视, 仅有少数研究涉及 CA 的“敏感性”问题^[9]。城市模拟通常要用到大量的地理数据, 模拟真实城市时更是如此, 从 GIS 中获取的空间变量可以作为 CA 建模过程的输入。与其他 GIS 模型一样, 城市 CA 模型也存在着误差和不确定性问题, 这些误差在 CA 模拟时会发生传递并影响模拟结果, 故需要评价数据源误差及传递对模拟结果的影响。本文尝试研究数据源误差传递和模型不确定性对城市 CA 模拟的影响, 这将有助于城市规划者在利用 CA 模型时能对模拟的结果有更好的理解。

2 城市 CA 模型的不确定性来源

城市 CA 模型与 Wolfram 的经典 CA 模型^[10] 存在显著差别。后者有严格的定义, 并几乎不使用空间数据, 这种 CA 模型有确定的输出结果。然而城市 CA 模型通常需要大量的空间数据作为输入来进行现实世界的模拟, 其模拟结果受到一系列来自数据源的误差和模型结构不确定性的影响 (图 1)。了解城市 CA 模型的数据误差和模型结构对模拟结果的影响是很有必要的。由图 1 可看到, CA 模型的循环结构与简单的 GIS 操作 (比如叠置) 相比有着显著的差别。简单的 GIS 操作通常能利用严格的数学方程来分析其误差传递过程, 但基于动态的 CA 模型却比较复杂。

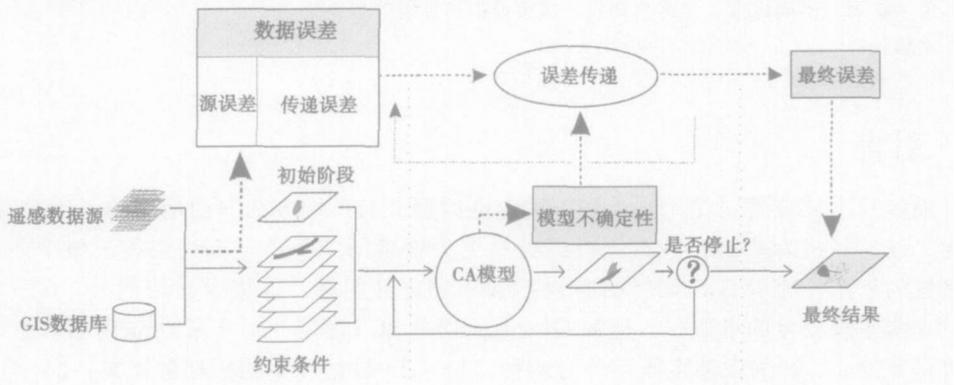


图 1 CA 模型中的数据误差、模型不确定性及误差传递

Fig 1 Data errors, model uncertainties and error propagation in cellular automata

2.1 数据源误差

城市 CA 模型使用许多空间数据, 模拟结果将受到 GIS 数据库中各种数据源误差的影响。这些误差与 GIS 数据库的数据质量有关, 具体来源于野外调查误差、制图误差和数字化误差等, 它们又可体现为两种主要形式: 位置误差和属性误差。

2.2 操作或转换误差

在模拟过程中, 使用 GIS 一般的运算或数据转换同样也会导致模拟的不确定性。为了节省存储空间, GIS 数据库往往只包含基本的空间数据, 而用户需要的特定空间信息需要通过 GIS 的有关操作或运算来获得, 如进行缓冲区分析等。为了得到 CA 的输入数据, 需要利用 GIS 的有关功能来生成一系列新的空间信息。例如, 城市发展适宜性在估算发展概率时是一个重要的空间变量^[4], 而计算发展适宜性时需要进行一系列的数据转换和叠

置运算, 这些操作都会导致模型不确定性的产生。

2.3 城市 CA 建模中的模型不确定性

模型本身的局限也带来了 CA 的不确定性。这类误差并非来自数据获取阶段, 而是由于人类的有限知识、自然界的复杂性和技术条件限制等, 即来自于模型本身。在 CA 模拟中输入误差和模型误差都会通过模拟过程加以传递。与其他计算机模型一样, 即使输入完全没有误差, CA 模型的模拟结果也不可能与现实一致。因此, CA 模型的模拟仅仅是对现实世界的一种近似模拟。现有的大多数 CA 模型是一种松散式的定义, 并没有惟一的模型。各种类型的 CA 模型是根据一定的经验提出的, 往往只是对特定的研究区域有效。不同的 CA 模型即使使用相同的数据集, 模拟也很难有完全一致性的结果。

3 城市 CA 模型不确定性的特征分析

3.1 GIS 数据源误差在模拟中的传递

评价城市 CA 模型中的误差传递对理解模拟结果是十分重要的。在城市模拟中, 初始条件、参数值和随机因素在对模拟结果的影响中扮演着重要的角色。在 CA 模拟中由于各种局部的相互作用, 可能出现难以预料的结果, 如果 CA 的行为十分不稳定且完全不可重复的话, 那么模拟结果对于城市规划人员就将变得毫无意义。但相关研究表明: 在宏观上, CA 能够产生非常稳定的模拟结果^[8], 尽管构造方式不同, CA 模拟的一般形态总是一致的, 但在微观上 CA 模拟的结果在某种程度上是无法预知的。

误差和不确定性会在建模过程中进行传递, 初始误差可能会被放大, 也有可能被缩小。在对图层进行叠置时, 个别 GIS 图层中的误差都会体现在最终输出结果中。许多研究说明了这些误差是如何在 GIS 操作中进行传递的, 例如叠置操作中误差的传递过程^[11]。Heuvelink 等提出利用泰勒级数来推导 GIS 误差传递的方法^[12], 定量方法的优势在于不需要很大的计算量就能得到误差传递表达式。另一种分析误差传递方法是应用较广泛的蒙特卡罗方法, 这种模拟方法易于实现且普遍适用, 但缺乏有效的分析框架。

CA 模型中的误差传递与 GIS 叠置操作中的误差传递是有差别的。可以用严格的数学表达式对由 GIS 叠置分析中的逻辑“与”和“或”操作所带来的误差传递进行度量。但是 CA 模型由于使用了邻域和迭代操作, 其构造形式相当复杂。CA 模拟是一个根据简单转换规则产生复杂特征的动态迭代过程, 中心元胞的状态转换由其邻居的状态所决定。利用严格的数学方程对该动态模拟过程的误差传递进行表达几乎是不可能的。从图 1 中可以看出, 由于使用了动态的循环过程, CA 模型中的误差传递显得相当复杂。

可采用敏感性分析来研究 CA 模型中误差传递的过程。通过对原始空间变量加进一些随机误差, 然后分析该误差对模拟结果的影响。敏感性分析可用于分析数据库中误差对 GIS 分析结果的影响^[13, 14]。蒙特卡罗模拟经常用来产生随机误差, 然后使用干扰后的空间数据检验结果的精确性。Fisher 以土壤图数据为例, 提出了两个算法来干扰分类图的数据, 然后估计误差的传递^[13]。

一般的误差传递理论无法应用于包含复杂操作的模型中^[15]。当数学模型难以解决误差传递问题时, 蒙特卡罗模拟却是一种很有用的方法, 非常适合于分析 CA 模拟中的误差传递。尽管其计算量非常大, 但随着计算机技术的进步, 问题正逐步得到解决。使用蒙特卡罗方法在空间变量中加入干扰因素, 以便研究城市模拟中的干扰敏感性。当详细的误差情况无法获取时, 可人为产生简单的噪声, 而最简单的噪声实现是使用无约束条件的干

扰。可使用该方法对初始的土地利用和适宜性图等空间数据产生随机的误差分布。

本研究通过下面的实验来估计属性误差对模拟结果的影响。作为模拟的初始土地利用包含了两种主要类型：城市用地和非城市用地。它们有可能存在分类误差。大部分情况下仅有一些关于分类误差的一般信息。例如，从卫星遥感获得的土地利用分类精度通常在 80% 到 90% 之间^[16]。但在大部分情况下，这些分类误差的详细空间位置是无法确定的。

实验第一步是采用随机方法在已经分类的遥感影像上产生这些误差，其大小取决于上述遥感影像分类的误差经验。例如，可根据 20% 的误差量来生成这些随机干扰。采用一个简单的约束性城市 CA 研究这种误差的传递过程，使用过于复杂的 CA 将会使数据误差与模型的不确定性混在一起难以区别。简单的城市模型可以基于如下的转换规则^[17]：

1) 假若任意一个 $\{x \pm l, y \pm l\}$ 元胞已变为城市用地，则 $N\{x, y\} = \sum_{ij \in \Omega} D\{i, j\}$ ($\{i, j\}$ 元胞为城市用地时 $D\{i, j\} = 1$ ，否则 $D\{i, j\} = 0$)

2) 假若 $N\{x, y\} > T_1$ 和 $R > T_2$ ，则 $\{x, y\}$ 元胞将转变为城市用地

其中， T_1 和 T_2 为阈值， R 为一随机变量， Ω 为 $\{x, y\}$ 的 Moore 邻域。

可以通过下面的实验来检验城市模拟中的误差传递问题。实验区为珠江三角洲的东莞市。实验中模拟了 1988~ 1993 年的土地利用发展变化，该时期东莞市发展非常迅速。本研究中参数 l 表示邻域尺寸的大小，取值为 3；阈值 T_1 和 T_2 决定每一个时间步中将有多少个元胞发展为城市区，若 T_1 和 T_2 的取值越小，获得发展的元胞将会越多。 T_2 属于 $[0, 1]$ 区间，它控制着随机变量干扰的大小，若 T_2 越大，干扰也将越大；如果是使用相同的土地消费量， T_1 和 T_2 值越小，完成模拟所需的时间就越短。因此， T_1 和 T_2 的值可以根据土地利用消费量和完成模拟的时间（步数）来定义。

第一个实验是为了检验数据源误差对模拟结果的影响。 T_1 和 T_2 分别设为 10 和 0.90，模型模拟了东莞 1988~ 1993 年间在使用 23330.5 公顷的土地开发总量下的城市演变情况。东莞的土地面积为 2465 km²，我们将其划分为 709×891 的格网，得到模拟元胞的地面分辨率为 50m²。以 1988 年作为模拟的初始情况，利用该 CA 模型模拟 1993 年的城市用地情况。模型重复运行两次：第一次以原始的遥感分类图像作为初始的土地利用输入；第二次对分类图像加入 20% 的随机干扰，即随机选取 20% 的城市用地类型和非城市用地类型，分别将它们改为非城市用地类型和城市用地类型，然后作为初始的土地利用输入。以没有误差干扰的模拟结果作为基准，将具有 20% 的误差干扰的模拟结果与其进行比较，将两次 1993 年的模拟结果与实际从遥感影像获得的 1993 年分类结果加以对比，计算得到两次重复模拟与实际情况的误差。

从图 2 中我们可以得到在模拟过程中误差传递的特点。没有加入误差干扰时，模拟结果同样有误差存在。而在初始土地利用中加入 20% 的误差干扰时，由于误差传递使模拟结果的误差有所增大，但比预期的

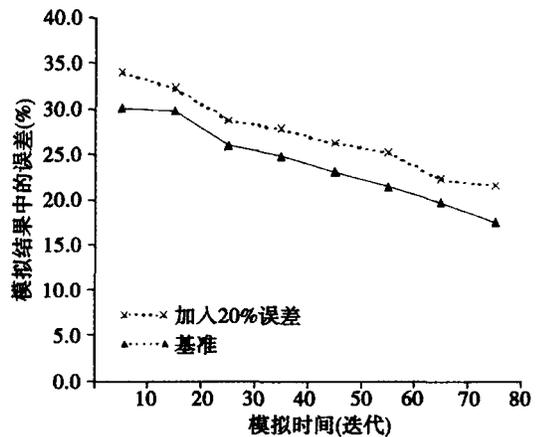


图 2 初始土地利用加入 20% 误差及在模拟中的误差传递

Fig 2 Error propagation of CA with 20% error perturbed to original initial urban areas

20% 要少得多。如图 2 所示, 如果在初始土地利用中加入了 20% 的误差, 与实际遥感影像分类结果对比, 得到模拟结果的误差是 35% ($t=10$)。与之相对应, 在没有加入误差的情况下也存在 30% 的误差 ($t=10$)。也就是说 20% 的误差干扰仅仅在结果中增加了 5% 的误差, 而不是预期的 20%。由此我们得到这样的结论: 误差干扰不会全部进入 CA 模拟的最终结果中, 并且这种数据源误差随模拟过程而减小的情形十分显著。这是由于 CA 模型中邻域的作用可以减少误差传递的程度, 而这些邻域作用通常体现在计算邻居状态的总和时在结果中引入了平均化的效果。分析表明如果模拟的时间足够的话, 所有误差都将显著减小。这是因为当城市用地增加时, 可以用于发展的土地越来越少。也就是说模拟更容易受到限制条件的影响, 而这些限制条件大大减少了误差产生的可能性。

3.2 CA 模型中的不确定性

CA 模型本身的不确定性可以反映在转换规则、邻域结构、模拟时间, 以及随机变量等方面。

转换规则 转换规则可反映模拟过程中变量之间的关系, 对空间演变的结果有决定性的影响, 是 CA 的核心。转换规则的定义受研究者认知的影响, 定义 CA 转换规则的方式有很多, 而不同的模型结构对 CA 模拟的结果会产生影响。迄今为止, 研究者已提出各种类型的城市 CA 模型, 但都是为解决城市模拟中的某些特定问题, 模型间的差别通常取决于个人的偏好和问题的性质。在 CA 模型中定义转换规则是必须的, 但方式并不惟一。定义转换规则有许多不同的方法, 主要包括: 1) 使用五种控制因素(扩散, 增殖, 速度, 斜坡和道路因素)^[18]; 2) 基于多准则判断中的层次分析法(AHP)估计发展概率^[4]; 3) 利用模糊集定义转换规则^[19]; 4) 利用预定义的参数矩阵计算转换潜力^[20]; 5) 利用“灰度值”模拟城市转换^[6]; 6) 在城市模拟中引入规划目标^[7]; 7) 利用神经网络模拟城市发展^[21]; 8) 利用数据挖掘自动发现转换规则^[22]。

邻域结构 邻域结构对在 CA 模型中实现转换规则有较大影响。转换规则应尽可能独立于模型本身, 例如距离影响应与元胞分辨率无关。但利用计算模型来实现转换规则时, 就会出现不确定性问题。以离散空间形式表达的元胞是 CA 模型的基本单元, 该离散形式是对连续空间的一种近似, 是以损失细节为代价的。在如何选择合适的元胞尺寸和形状上存在许多问题。元胞尺寸大的话将减少数据量, 但可能会导致空间精度的降低; 邻域的形状同样也会影响 CA 模拟的结果, 研究邻域影响的一种方法是检验在不同的邻域影响下城市如何增长。一般有两种主要类型的领域——摩尔领域和冯诺依曼领域。摩尔邻域会导致城市呈现指数增长, 这种模式与实际的城市增长模式并不相同; 冯诺依曼邻域能减小增长率。但是这两种邻域都是矩形的形态, 会在城市模拟中产生边界影响。由于各向同性的原因, 圆形邻域比矩形邻域更好^[6]。

模拟时间 CA 模型使用离散的时间步数来模拟城市增长, 离散时间不同于实际的连续时间。如何决定离散时间的间隔及迭代的次数值得深入探讨。离散时间的间隔越长, 模拟的时间步数越少。对于非线性模型来说用 100 个时间步数模拟与用 10 个时间步数模拟结果是不一样的。由于在模拟中使用了近似的离散时间, 将会在 CA 模型中引入误差, 因此研究离散时间对 CA 模拟的影响是非常有必要的。

应特别注意当时间步数改变时, 如何重新计算参数值。如果使用发展概率的话, 那么根据时间步数的改变对转换概率进行调整的标准方法是这样的: 如果原来的时间步数为 t_s , 新时间步数为 $T_s (T_s = n \times t_s)$, 那么转换概率矩阵 $\|P\|$ 应该用 $\|P^n\|$ 代替。一个时间步

长对应的实际年数= 被模拟的实际年数(例如 20 年)/模拟的时间步数(例如 300 次)。由于本研究利用阈值 T_1 来决定土地利用转化,对模拟来说可以修改该值得使得土地利用的总消费量相同。阈值 T_1 可根据试验来确定。

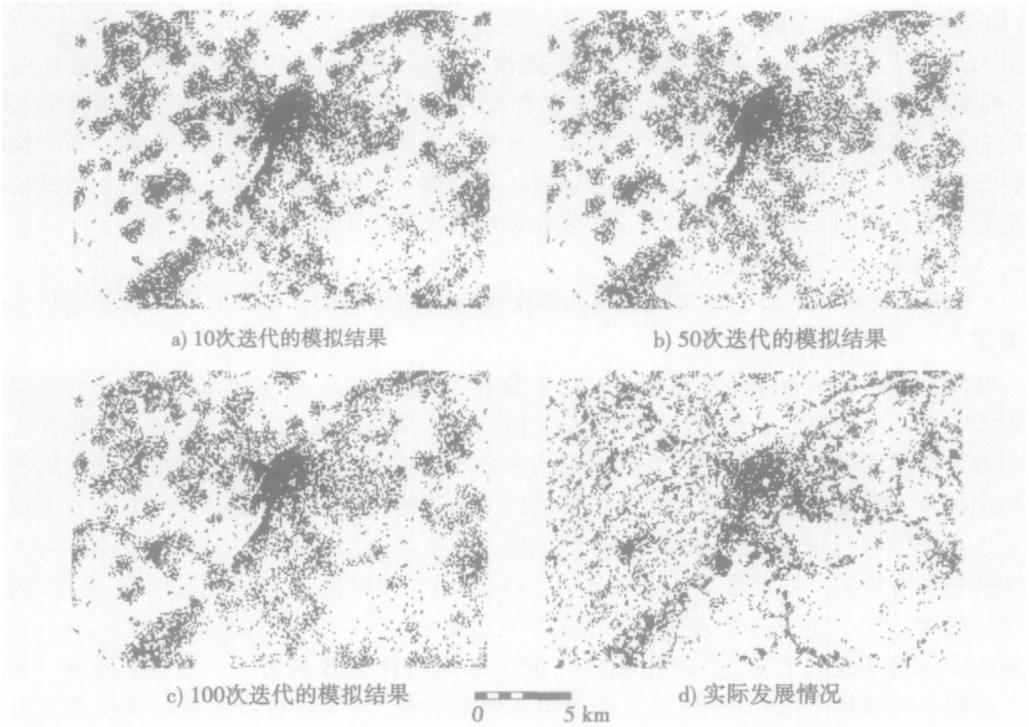


图 3 离散时间对模拟结果精度的影响

Fig 3 The influences of discrete time steps on simulation accuracies

通过试验研究了利用不同时间步数对模拟结果的影响(图 3)。图 3a 利用 10 个时间步数来得到模拟结果,其结果与从遥感影像(图 3d)上获得的实际城市用地差别很大。因为局部相互作用对产生实际的城市形态来说是非常重要的,太少的步数使得模拟过程中的空间细节无法涌现出来。增加时间步数有助于产生更加精细的模拟结果(图 3b 和 3c)。这个特点不同于那些不依赖于时间步数选取的线性模型。

随机变量 大部分城市 CA 模型在模拟复杂城市系统时并非确定性的,确定性模型在表示许多地理现象时都存在一定的问题。由于自然界的复杂性,这些复杂现象所显示的无法预料特征还不能完全通过变量加以解释。几乎任何一种计算模型都还不能准确地预测未来的特征。因此,城市 CA 模型常常需要结合随机变量来表示自然界这种不确定性。通过使用随机变量将一些“噪音”人为地加入到城市 CA 模型中,以产生更“现实的”模拟结果^[20]。在具体计算中,是通过对比发展概率与随机值来决定是否发生状态的转换^[3],在一定程度上也使城市模拟加入一定的随机性。但当这些模型被用于城市规划时问题又产生了。由于相同的输入会产生不同的模拟结果,规划者将不知道哪个结果更适合于规划。当 CA 用于城市规划时,城市模拟的可重复性应该是一个至关重要的问题,但其不确定性的大小应该只占较小的比例,否则模拟就变得毫无意义。

我们进行如下实验来研究随机 CA 模型中的可重复性及不确定性的空间分布。利用

CA 模型重复运行 10 次来模拟城市用地的空间分布, 然后利用叠置分析来检查模拟结果的可重复性。在叠置分析中, 将城市用地编码为 1, 非城市城市用地编码为 0。如果 CA 是完全确定的话, 任何两次模拟中的城市用地和非城市用地出现的位置应该是相同的, 在叠置分析中应该完全重叠。叠置分析将会产生两个值, 2 表示城市用地, 0 表示非城市用地。但随机 CA 不会产生相同的模拟结果, 在叠置分析中将产生 3 个数值。数值 2 代表城市用地, 数值 0 代表非城市用地。数值 1 则代表不确定的区域, 可能在其中一次模拟中是城市用地, 而在另一次模拟中又成为非城市用地。图 4 显示模型运行 10 次其模拟结果重叠的情况。数值 10 表示在所有的 10 次模拟中均为城市用地, 以此类推。因此, 数值越大, 在模拟中被模拟为城市用地的重复性越好。可以看到, 在城市的核心(中心)地带, 模拟结果的重叠性是很好的。

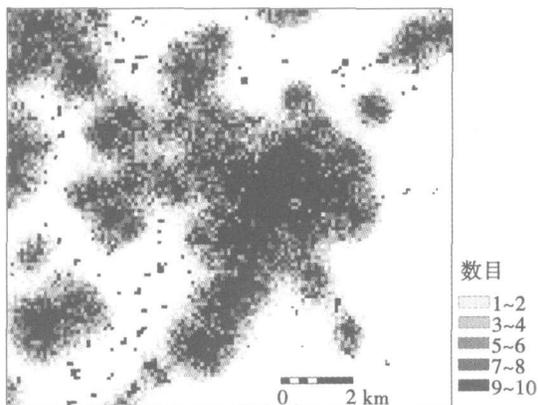


图 4 将随机 CA 模型重复运行 10 次后进行叠置的模拟结果

Fig. 4 Overlay of the simulation results by repeatedly running the stochastic CA ten times

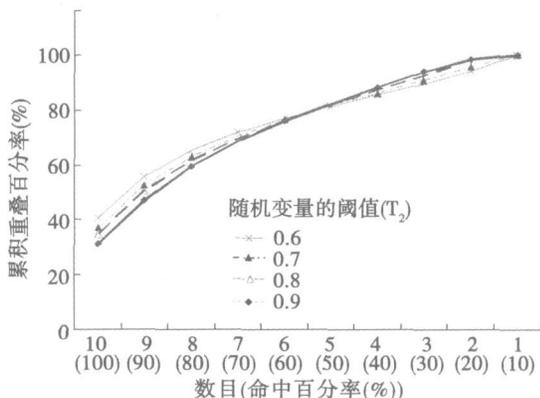


图 5 随机变量不同值与 CA 重复模拟 10 次得到的重叠累积百分率

Fig. 5 Cumulative percentages of overlapping areas for ten repeated simulations using stochastic CA with different sizes of perturbation

进一步研究随机变量 (R) 的不同值对模拟结果的影响 (图 5)。在被模拟的城市用地中, 多次模拟中具有较低重叠度的元胞所占的比例较小, 而数值大于 7 (70% 命中) 的元胞所占比例高达 71.8%。即有 70% 的机会可以重复模拟为城市用地的元胞所占比例为 71.8%。模拟结果同时也显示了在随机模拟中使用随机变量 (R) 较高的阈值将导致不确定性的增加, 即重叠性较高的元胞数目所占的比例将会更低。

上述实验对城市规划 CA 模型的有着重要的实用价值。城市规划者在利用城市 CA 模型进行模拟时, 应运行尽可能多的时间步数来获得较好的模拟结果。上述实验也将有助于找出那些具有较高发展概率的地点, 比如可以选择那些命中率高于 70% 的地方。在城市规划中, 这种方法对得到更为可靠的模拟结果是非常有用的。

4 结论

城市 CA 模型中数据误差、误差传递和模型不确定性的问题非常重要, 但又常常被忽视。本文通过使用 GIS 数据对其进行研究, 并通过实验对这些问题加以解决。许多模型误差与模型的构建有关, 例如如何定义一个适当的模型来反映城市发展的实际过程。本研究得到 CA 模型中误差和不确定性的一些初步结果:

- 1) 在 CA 模拟过程中由于邻域函数平均化的影响, 数据源误差将减小;
- 2) 随着城市用地的增长, 可用的土地越来越少, 该限制将有助于使得模拟误差随时间而减小;
- 3) 为确保 CA 模型能够模拟详细的空间细节, 足够的迭代是非常必要的;
- 4) 模拟结果的不确定性主要体现在城市的边缘。

CA 模型的不确定性特征与 GIS 模型的不确定性特征有着显著区别。本研究的发现将有助于建模和规划人员更加清楚地理解城市模拟中误差和不确定性的性质, 可避免对模型结果的误解。进一步的工作需要找到减小误差影响的方法, 以便得到更为稳定的模型结果, 提供更加高效的规划工具。本文的不确定分析只是建立在所提到的 CA 模型, 对于其他类型的 CA 模型的结果, 还需要进一步分析。另外, 城市模拟还涉及其他的不确定性, 例如政府决策者的行为等, 对于这些因素所带来的影响, 还需要进一步深入研究。

参考文献:

- [1] Goodchild M F, Sun G Q, Yang S H R. Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*, 1992, 6(2): 87~ 104.
- [2] Batty M, Xie Y. From cells to cities. *Environment and Planning B: Planning and Design*, 1994, 21: 531~ 548.
- [3] 黎夏, 叶嘉安. 基于神经网络的元胞自动机及模拟复杂土地利用系统. *地理研究*, 2005, 24(1): 19~ 27.
- [4] Wu F, Webster C J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design*, 1998, 25: 103~ 126.
- [5] Webster C J, Wu F. Regulation, land-use mix, and urban performance. Part 1: Theory. *Environment and Planning A*, 1999, 31: 1433~ 1442.
- [6] Li X, Yeh A G O. Modelling sustainable urban forms by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 2000, 14(2): 131~ 152.
- [7] Yeh A G O, Li X. A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B: Planning and Design*, 2001, 28: 733~ 753.
- [8] Yeh A G O, Li X. A cellular automata model to simulate development density for urban planning. *Environment and Planning B*, 2002, 29: 431~ 450.
- [9] Benati S. A cellular automaton for the simulation of competitive location. *Environment and Planning B: Planning and Design*, 1997, 24: 205~ 218.
- [10] Wolfram S. Cellular automata as models of complexity. *Nature*, 1984, 31(4): 419~ 424.
- [11] Veregin H. Integration of simulation modeling and error propagation for the buffer operation in GIS. *Photogrammetric Engineering & Remote Sensing*, 1994, 60(4): 427~ 435.
- [12] Heuvelink G B M, Burrough P A, Stein A. Propagation of errors in spatial modeling with GIS. *International Journal of Geographical Information Systems*, 1989, 3(4): 303~ 322.
- [13] Lodwick W A. Developing confidence limits on errors of suitability analysis in GIS. In: Goodchild M F, Gopal S (eds). *Accuracy of Spatial Databases*. London: Taylor & Francis. 1989 69~ 78
- [14] Fisher P F. Modeling soil map-unit inclusions by Monte Carlo simulation. *International Journal of Geographical Information Systems*, 1991, 5(2): 193~ 208.
- [15] Heuvelink G B M, Burrough P A. Error propagation in cartographic modeling using Boolean logic and continuous classification. *International Journal of Geographical Information Systems*, 1993, 7(3): 231~ 246.
- [16] Li X, Yeh A G O. Principal component analysis of stacked multi-temporal images for monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, 1998, 19(8): 1501~ 1518.
- [17] Batty M. *Growing cities*. Working paper, Centre for Advanced Spatial Analysis, University College London. 1997.
- [18] Clarke K C, Gaydos L, Hoppen S. A self-modifying cellular automaton model of historical urbanization in the San

Francisco Bay area. *Environment and Planning B: Planning and Design*, 1997, 24: 247~ 261.

- [19] Wu F. A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computer, Environment, and Urban Systems*, 1999, 20(6): 367~ 387.
- [20] White R, Engelen G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, 1993, 25: 1175~ 1199.
- [21] Li X, Yeh A G O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 2002, 16(4): 323~ 343.
- [22] Li X, Yeh A G O. Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science*, 2004, 18(8): 723~ 744.

Analysis of error propagation and uncertainties in urban cellular automata

LI Xia¹, Anthony Ga-On Yeh², LIU Tao¹, LIU Xiao-ping¹

(1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;

2. Centre of Urban Planning and Environmental Management,
the University of Hong Kong, Hong Kong SAR, China)

Abstract: The issues of data errors, error propagation and model uncertainties are important but often neglected in urban CA models. This paper has examined and addressed some of these issues by carrying out experiments with GIS data. Many model errors are related to model configurations, i. e. how to define a proper model to reflect the real process of urban development. This study demonstrates that some of them, however, are quite unique to CA:

- 1) data source errors will be reduced during simulation because of the averaging effects of neighborhood functions;
- 2) simulation errors will decrease with time because the availability of land suitable for urban development will be decreased in constrained urban CA as the urban areas grow in size;
- 3) the number of time steps (iterations) can cause different spatial patterns and simulation closer to actual development can be achieved with the increase in time steps; and
- 4) the major uncertainties of simulation are mainly found at the edge of simulated urban areas.

These characteristics are quite different from those of general GIS modeling. The study shows that errors and uncertainties of urban CA are less severe than what one would normally expect from a simulation model. The uncertainties of the simulation will be reduced if more amounts of land are developed and the uncertainties are mainly located at the urban fringe. The findings of the study can help urban modelers and planners to understand more clearly the characteristics of errors and uncertainties in urban CA so that they can be used more effectively in urban planning.

Key words: uncertainty; cellular automata; urban simulation; GIS