

基于空间信息多级网格的 CA 模型研究

危双丰¹, 黎景良¹, 邵振峰²

WEI Shuang-feng¹, LI Jing-liang¹, SHAO Zhen-feng²

1. 武汉大学 遥感信息工程学院, 武汉 430079

2. 武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079

1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

2. State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

E-mail: weishuangfeng2001@yahoo.com.cn

WEI Shuang -feng, LI Jing -liang, SHAO Zhen -feng. Study of cellular automata model based on SIMG. Computer Engineering and Applications, 2007, 43(9): 4-7.

Abstract: The integration of GIS and CA (Cellular Automata) with spatio-temporal attribute has accelerated GIS's ability to simulate geographical process greatly. This paper introduces briefly the principle of the Spatial Information Multi-Grid (SIMG), a new representation method for spatial data and spatial information, which can not only easily run at grid computing environment, but also properly consider the differences of natural and social characteristics in earth space as well as the different levels of economical development in different areas. This paper discusses the correlation between SIMG and CA aiming at the determination in cell shape and states, cell space, neighbourhoods and transition rules, discrete time. The authors put forward a new CA extended model based on SIMG called SIMGCA. Finally a general application framework of SIMGCA for simulating the process of land use and cover change is brought forward.

Key words: Spatial Information Multi-Grid (SIMG); cellular automata; Land Use Cover Change (LUCC); SIMGCA

摘要: 具备时空计算特征的元胞自动机(CA)模型与GIS集成极大促进了GIS对地理过程的模拟能力。论文简要介绍了空间信息多级网格(SIMG)——一种既能适合网格计算环境又充分考虑到地球空间的自然特征和社会属性的差异性 & 经济发展不平衡的特点的空间信息表示新方法。充分研究了SIMG与CA之间的联系,分别讨论了在SIMG上CA元胞及状态的确定、元胞空间的确定、规则的定义、时间粒度确定等,提出了空间信息多级网格元胞自动机模型(SIMGCA),并提出了SIMGCA模型在土地利用/覆被变化中的应用框架。

关键词: 空间信息多级网格;元胞自动机;土地利用/覆被变化;空间信息多级网格元胞自动机模型

文章编号: 1002-8331(2007)09-0004-04 **文献标识码:** A **中图分类号:** TP301.1

1 引言

随着计算机网络技术的不断发展,网格技术被应用到众多领域,李德仁最近提出了广义空间信息网格和狭义空间信息网格两个层次的概念^[1]。广义空间信息网格指的是在网格技术支撑下空间数据获取、更新、传输、存储、处理、分析、信息提取、知识发现到应用的新一代空间信息系统。狭义空间信息网格则指在网格计算环境下新一代地理信息系统,是广义空间信息网格的一个组成部分。文献[1]中提出空间信息多级网格(SIMG, Spatial Information Multi-Grid)的建议,并将它定义为狭义空间信息网格,应该说网格技术浪潮为空间信息网格带来了机遇,因此重新来思考地球(地理)空间数据在计算机中的表达形式,提出空间信息多级网格,并指出SIMG的三大潜在职能和所面临的四大挑战,其中针对空间基准的不一致付迎春在博士

论文[2]中进行了研究;对于数据格式不一致和语义不一致引起的问题,也进行了研究^[3];但是对于时间基准不一致的问题尚处在初步研究阶段;由于具备时空计算特征的元胞自动机模型(CA)非常适合于网格计算,它天生就具有分布并行计算特征。近年来CA在地理信息系统中被广泛应用,CA (Cellular Automata)与GIS集成极大促进了GIS对地理过程的模拟能力。为此本文充分研究了SIMG与CA之间的联系,分别讨论了在SIMG上CA元胞及状态的确定、元胞空间的确定、规则的定义、时间粒度确定等,提出了空间信息多级网格元胞自动机模型(SIMG-CA, Spatial Information Multi-Grid based Cellular Automata Model),并提出了SIMGCA模型在土地利用/覆被变化中的应用框架。

基金项目: 国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2003AA132080)。

作者简介: 危双丰(1979-),男,博士研究生,主要从事地理信息系统和空间数据挖掘技术方面的研究;黎景良(1978-),男,博士研究生,主要从事国土资源遥感与土地利用评价研究;邵振峰(1976-),男,博士,副教授,主要从事遥感与地理信息系统方面的教学与研究工作。

2 空间信息多级网格和元胞自动机

2.1 空间信息多级网格

空间信息多级网格(Spatial Information Multi-Grid, SIMG)是一种既能适合网格计算环境又充分考虑到地球空间的自然特征和社会属性的差异性及其经济发展不平衡的特点的空间信息表示新方法。其核心思想是:按不同经纬网格大小将全球、全国范围划分为不同粗细层次的网格,每个层次的网格在范围上具有上下层涵盖关系。每个网格以其中心点的经纬度坐标(网格中心点)来确定其地理位置,同时记录与此网格密切相关的基本数据项(如经纬度、全球地心坐标、各类投影参数下的坐标)。落在每个网格内的地物对象(细部地物)记录与网格中心点的相对位置,以高斯坐标系或其他投影坐标系为基准。根据实际地物的密集程度确定所需要的网格尺度(分层密度),如地物稀疏的地方只需要粗网格,而地物密集的地方(如城市)则按细网格存储空间与非空间数据^[1]。

2.2 元胞自动机

2.2.1 CA概念和原理

元胞自动机(Cellular Automata,简称CA)最早由著名数学家冯·诺依曼(Von Neumann)在20世纪40年代提出,而J.H. Conway在1970年编制的“生命游戏”(the Game of Life)是最为著名的一个在计算机上实现的元胞自动机模型。元胞自动机是时间、空间、状态都离散,(空间的)相互作用及(时间上)因果关系皆局部的网格动力学模型^[2]。不同于一般的动力学模型,CA模型没有明确的方程形式,而是包含了一系列模型构造的规则,凡是满足这些规则的模型都可以算作是CA模型。因此,确切地讲元胞自动机是一类模型的总称,或者说是一个方法框架。CA系统中的所有元胞是相互离散的,构成一个元胞空间。在某一时刻一个元胞只能有一种状态而且该状态取自一个有限集合。一个元胞下一时刻的状态是上一时刻其邻域状态的函数,这是元胞自动机的原理。CA系统的时间也是离散时间,它不具有物理意义,和时态GIS中的系统时间具有相似的含义。用集合的语言可以将CA模型描述为: $S_{t+1}=f(S_t, N)$,式中 S 为有限集合,代表元胞状态; N 代表元胞邻域; t 表示时间; f 为局部转换规则。

2.2.2 CA组成与特征

一个标准的CA模型有5个基本构成要素:元胞空间、元胞及其状态、邻位的定义(邻域)、转换规则和离散的时间。元胞空间即元胞所分布的空间网点集合;元胞又可称为单元,是CA的最基本的组成部分,是一种存储“状态”的元素;邻域是元胞周围按一定形状划定的元胞集合,它们影响该元胞下一个时刻的状态;转换规则是CA的核心,他就是个状态转移函数。根据元胞当前状态及其邻居状况确定下一时刻该元胞状态的动力学函数;CA是一个动态系统,它在时间维上的变化是离散的,即时间是一个整数值,而且连续等间距。一个元胞在 $t+1$ 时刻只决定于 t 时刻的该元胞及其邻居元胞的状态。

CA模型以其框架的简单、开放和可以模拟十分复杂的系统行为而具有很强的生命力,从CA的构成及其规则上分析,它具有空间性、离散性、同步性、局部性、高维数、时空动态性、简单性等特点^[10]。

2.2.3 CA在地理研究中的现状

CA具有强大的空间运算能力,可以有效地模拟复杂的动

态系统。近年来,CA被越来越多地应用于城市及其他地理现象的模拟中,取得了许多有意义的研究成果。针对CA在土地利用的研究课题,White和Engelen^[6-8]将CA发展成为规划与设计的工具,针对加勒比海的St.Lucia岛进行土地利用规划,建立了一套结合了宏观(Macro)和微观(Micro)两种层次的决策支援系统。宏观层次是从经济、人口、政策层面来预测整个研究区内的土地利用需求,并且将这些需求分配到各个区域单元。微观层次中利用CA模式来计算每一个网格的土地利用潜力,将土地利用变迁数量分派到各个网格。Clarke等^[9]尝试利用CA来预测未来的土地利用变化趋势,他们的分析模式整合了地面覆盖(land cover)、坡度、交通网路、保护区等四种资料,利用过去100年的资料来建立CA的变迁法则,并利用过去100年间数个不同年度的土地利用资料来校正法则,再利用所产生的CA系统来预测未来的土地利用变迁,并纳入了自我修正的功能,用来调整急速扩张或衰退等不同阶段的都市成长,增加了CA模式的弹性。Wu F.L.集成CA模型和多因子评价模型(Multicriteria Evaluation, MCE)并在Arc/Info中应用AML和C语言,在统一的界面上实现了GIS、CA和MCE模型的集成,对广州的城市土地扩张进行了模拟^[9]。此外,受国际研究的推动,国内地理学界近年来也开始了类似的研究尝试,周成虎等于1999年出版了《地理元胞自动机研究》一书,对相关工作进行了介绍和总结,并提出了地理元胞自动机(GeoCA)的概念^[10]。黎夏、叶嘉安在对广东东莞土地利用变化系统研究的基础上,提出基于约束性的CA模型,直接利用Arc/Info Grid来开发模型,并应用于广东省东莞市城市可持续发展规划^[4]。张显峰、崔伟宏在Arc/Info Grid环境下,运用AML和VB开发地理时空动态模拟引擎,并用之于建立城市土地利用演化过程模拟预测模型(LESP模型),对包头市城市土地利用演化过程进行了模拟与预测^[12]。最近,黎夏设计了一个面向对象的元胞自动机模型用于土地利用/覆被变化^[15]。

2.2.4 CA模型在LUCC应用中存在的问题与挑战

CA模型是一种模拟城市和区域土地利用变化的比较有前景的框架,但由于CA模型框架的简单性和LUCC本身的复杂性,基于CA的模拟也面临着以下一些问题与挑战。

(1)CA模型需要的源数据质量问题。CA模拟要在深入研究区域或城市LUCC的特征、驱动机制的基础上进行,模拟需要高精度的标准化历史数据如中国城市与区域发展的历史系列数据,尤其是空间数据,很难得到或者精度很低。如何集成应用3S技术提供高精度可信数据,是模型模拟的前提。

(2)CA模型的自组织来自于系统元素的局部相互作用,状态变化取决于自身和邻居的状态组合,尽管可以模拟出各种复杂系统,但LUCC并不仅仅取决于本身局部规则的作用,而是各种尺度的多种因素的综合作用结果,宏观规划和微观决策如何整合到统一的模型中成了对CA模型的一个挑战。

(3)CA空间尺度的划分问题。不同的空间尺度下,由于模型的表现效果以及影响模型的各种外在因素作用程度的差异,系统单元表现的规律也不相同。因此,根据研究的需要,如何确定合适的空间分辨率也是一个需要考虑的问题^[10]。

(4)CA模型时间的校准问题。CA模型的时间是一个抽象的概念,那么它的一个循环如何与实际的时间尺度如年月相对应是困扰CA模型的一个难题。目前主要是利用LUCC的历史

数据来校准模型的时间概念,即计算模型运行结果与某时段LUCC在土地单元总数上相对应时模型所运行的循环次数,得到一个时间对应关系。但这种方法局限性在于一般只适用于LUCC为线性的情况。因此,如何利用社会经济指标与LUCC的相关关系,将模拟结果的土地单元统计量与社会经济指标统计量相对应,准确地判断模型某时刻应当对应的实际时间是研究的关键。

(5)CA模型与宏观经济、人口模型以及其它的微分方程模型、人工神经网络(ANN)模型、GIS空间分析方法的集成问题。如何将宏观经济、人口模型动态地嵌入CA模拟过程中,自动调整模型参数和规则定义;将模糊推理与CA时空动态模拟结合,建立智能化类似于专家系统的规则体系;如何在CA中引入随机干扰因素,如何在CA中根据不同的区域特征和土地利用发展的不同阶段,调整各种影响因素和单元演化规则,也是这类模型面临的问题。

2.3 基于空间信息多级网格的元胞自动机模型

2.3.1 SIMGCA 构成

从CA的特点可以看出,CA模型采用“自下而上”的构模方式,而且没有一个既定的数学方程,只是一个建模原则,因此具有很好的开放性和灵活性。这和运用微分方程或物理模型从宏观上描述空间现象的传统方法是对立的,前者更符合人们认识复杂事物的思维方式。CA模型是一个基于微观个体相互作用的时空动态模拟模型,将地理实体的空间和时间特性统一在模型中,通过划分研究对象的细胞空间和研究初始状态及状态转换规则,CA模型就可以自行迭代运算,模拟系统演化过程了,而GIS则不具备迭代运算的能力。CA模型将空间和时间离散化,适合于建立计算机模型和并行计算特征,因为计算机对客观世界的表示是离散的。CA模型具有不依赖比例尺的概念,元胞只是提供了一个行为空间,本身不受元胞空间测度和时间测度的影响,时空测度的影响通过转换规则体现。因此CA模型可以用来模拟局部的、区域的或大陆级的演化过程。从数据模型的角度看,CA模型中的元胞和基于栅格GIS中的栅格一样,所以CA模型易于和GIS、遥感数据处理等系统集成^[10]。由上述可知,CA模型较适合空间信息的时空动态分析,尤其是时空动态过程的模拟,为GIS中时空动态分析提供了一个框架思路和建模方法。

但是由于地理系统的复杂性,CA模型仍然存在一定的局限性。即简单性与真实性的矛盾,空间划分问题,时间对应问题、转换规则定义问题、与GIS集成的问题。因此标准的CA模型需要扩展和改进,才能满足地理时空模拟的需要,更加真实地模拟地理实体的演化进程。为此本文从空间信息多级网格所具有的特点出发,针对以往CA模型存在的问题,分别从元胞定义及状态、元胞空间的确定、元胞邻域、元胞转换规则、时间粒度等五个方面系统讨论SIMG与CA之间的紧密联系,提出了空间信息多级网格元胞自动机模型(SIMGCA)。

以往对于元胞自动机的划分都是基于单一的均匀网格,而空间信息多级网格是一个变网模型,因此在经济社会中,对于普遍存在的经济发展不平衡现状,应根据不同的经济发展水平采用不同的元胞自动机,并确定不同的转换规则和时间粒度,来建立预测模拟模型,这样既可以提高模拟的真实性,又可以保证模拟结果的精确性,本文旨在提出一种基于变网的CA

模型即SIMGCA。SIMGCA模型以现有的SIMG支撑技术为基础,对于SIMG的不同级格网数据,按照SIMG编码规则将空间实体的空间和属性信息按照不同的时间建立空间数据仓库,根据不同地区的自然、经济发展水平等因素确定不同区域的网格单元的粗细程度即CA的定义和元胞空间,在此基础上利用诸如神经网络、模糊数学、云理论等空间数据挖掘方法来确定元胞转换规则,利用数据推理法或宏观的预测模型如统计回归、灰色模型等方法获取时间粒度。基于空间信息多级网格的元胞自动机模型的构成如图1。

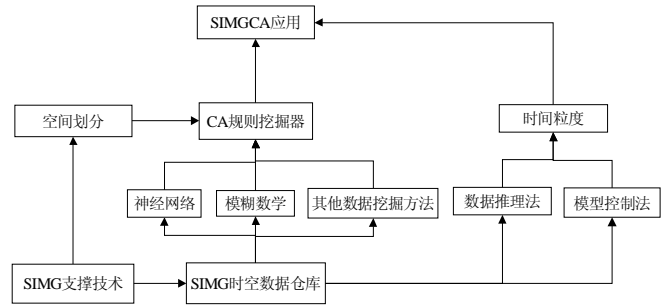


图1 SIMGCA模型结构

2.3.2 元胞及状态

根据不同的区域所具有的不同经济发展水平,来确定元胞大小。对于经济欠发达的地区可以采用较大网格单元即粗网格,反之对于经济发达区域采用细网格。由于元胞的状态与元胞大小紧密相关的,粗网格元胞状态与细网格元胞状态也不同,比如在研究土地利用与覆被变化时,欠发达地区在农村用地向城市用地的转化过程中,状态定义为发展和未发展两种状态,状态集量化表示为{1,0}。发达地区则由于发展中某些细微的变化,不能简单地划分为上述的两种状态,而应该划分得更加细,比如高度发展,中度发展,初步发展,未发展,状态集量化表示为{1,0.66,0.33,0},其中高度发展可以理解为该区域早已发展为城市用地,经过多年的经济发展,经历了多级重建翻新,经济水平发展到前所未有的高度;初步发展表示由农村用地转化为城市用地不久;中度则介于高度发展与初步发展之间;未发展表示仍然是农村用地。

2.3.3 元胞空间

在GeoCA中,元胞空间的概念可以很自然地转化为笛卡尔坐标系下的地理空间。在数据模型层次上,对于常用的二维元胞空间可用栅格(Grid)数据模型来表示。元胞空间被赋予了空间尺度的概念,元胞大小对应于空间分辨率。不同的尺度空间,会进一步影响整个模型的其他方面如规则、时间粒度等,例如在土地利用变化动态模拟中,当模型中元胞的大小10m×10m与1km×1km时的考虑因素、转换规则和时间粒度就会有很大的不同。基于SIMG,元胞空间可以根据元胞大小的确定,与空间数据库、多分辨遥感影像库进行匹配。对于发达地区采用大比例尺遥感影像,反之,对于欠发达地区采用小比例尺遥感影像。

2.3.4 邻居

元胞、元胞状态、元胞空间的确定直接影响元胞邻居的确定。在SIMG模型中,对于经济发达地区,采用细网格作为元胞单元和大比例尺遥感影像作为元胞空间,由于网格单元较小,其影响因素也较多,因此其元胞邻居可以采用Moore 8邻域;

采用粗网格作为元胞单元和小比例尺遥感影像作为元胞空间, 由于网格单元较大, 其影响因素相对来说要少, 其元胞邻居可以采用 Von. Neumann 4 邻域。

2.3.5 转换规则

对于元胞自动机模型, 在模拟地理现象的过程中, 关键是如何定义转换规则。转换规则的定义往往是十分繁琐的。学者们主要是采用启发式的方法来定义转换规则, 例如包括矩阵、多准则判断和灰度等方法。这些方法受主观因素影响很大, 在形式上相互有明显的差别。而且, 这些转换规则大多数都是隐含的, 是通过数学公式来表达, 如何确定公式中的参数十分困难。上述这些方法都是通过数学公式来表达转换规则, 但数学公式在反映复杂的关系时有很大的局限性^[14]。为此, 黎夏提出了利用神经网络的方法来自动获取 CA 的参数值, 从而减少 CA 的不确定性。但其缺点是神经网络方法属于黑箱结构, 用户不能清晰地知道模型运行的机制, 对模型参数的具体物理意义很难理解。

数据挖掘技术已经被应用于地理学领域, 可以获得与地理有关的空间分布规律等。因此可以通过数据挖掘的方法从 GIS 空间数据仓库自动生成 CA 的转换规则, 无需使用数学表达式来定义转换规则, 并能在生成转换规则的同时对模型自动进行纠正。这样能更方便和准确地描述自然界中的复杂关系。目前利用数据挖掘技术来建立 CA 模型的研究报道相对较少, 黎夏利用 See5.0/C4.5 挖掘转换规则, 对珠江三角洲城市土地利用进行了模拟, 取得了较好的效果^[14]。F. C. Richards 将遗传算法运用到 CA 模型规则获取中^[14]。然而数据挖掘的方法有很多, 对于其他的数据挖掘方法如模糊集理论、云理论、粗糙集理论等却很少有人用于 CA 转换规则的挖掘。在 SIMG 模型中拟展开对基于模糊集理论和云理论的数据挖掘方法进行研究, 探讨对于 CA 规则挖掘的可行性, 并与实例相结合验证比较两种方法的模拟精度。

2.3.6 时间

考虑不同地区的自然、经济发展水平等综合因素, 发达地区与欠发达地区发展速度也不同, 因此时间粒度确定因不同地区而异。发达地区城市变化迅速, 时间间隔小; 欠发达地区城市变化缓慢, 时间间隔大。在 SIMG 模型中, 基于不同元胞空间, 可以利用数据推理法或宏观的预测模型如统计回归、灰色模型等方法获取时间粒度。

3 SIMGCA 在土地利用演化中的应用

针对 CA 在 LUCC 中面临的问题, SIMG 可以为 CA 模型提供高质量的数据源, 因为 SIMG 是多级比例尺的网格数据, 在一个数据库中存储了不同时期不同比例尺的数据, 而且可以保证数据的精度。SIMG 充分考虑到地球空间的自然特征和社会属性的差异性, 及经济发展不平衡的特点, 网格点属性项包括了自然属性、社会属性、经济属性和文化属性等, 兼顾了宏观和微观的影响因素, 因此为 CA 在 LUCC 中的转换规则的确定提供了充分的参考因子。SIMGCA 突破了以往基于单一均匀网格 CA 模拟的局限性, 对于不同区域采用变网 CA 模型, 能较好解决 CA 空间尺度划分问题。同时 LUCC 时间粒度也可以根据不同的区域而变化, 因此在理论上模拟土地利用变化是可行的。将土地利用/覆被变化和 SIMGCA 的集成框架如图 2。

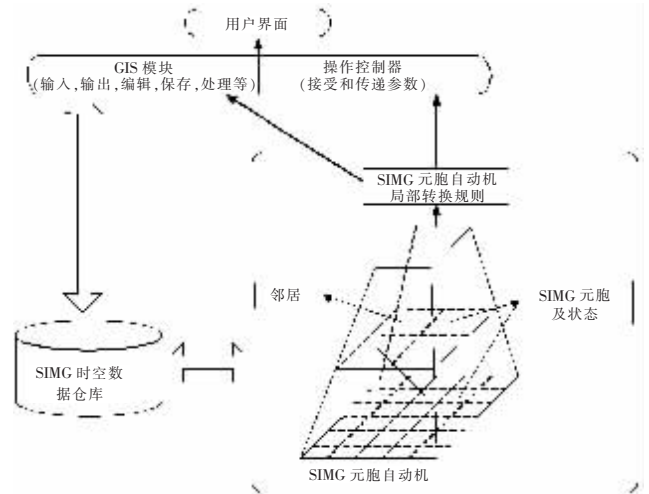


图 2 SIMGCA 在土地利用/覆被变化中应用框架

4 结论

以往很多学者所研究的 CA 模型都是以某单一比例尺的均匀网格单元作为元胞, 然而在经济社会中, 对于普遍存在的经济发展不平衡现状, 应根据不同的经济发展水平采用不同的元胞自动机, 并确定不同的转换规则和时间粒度, 建立预测模拟模型, 这样既可以提高模拟的真实性, 又可以保证模拟结果的精确性。SIMG 是多级比例尺的变网模型, 结合多分辨率遥感影像, 在经济发达地区采用细网格和高分辨率遥感影像, 欠发达地区则采用粗网格和 low 分辨率遥感影像来确定元胞及相关组成要素。本文研究了 SIMG 与 CA 之间的联系, 分别讨论了在 SIMG 上 CA 元胞及状态的确定、元胞空间的确定、规则的定义、时间粒度确定等, 提出了空间信息多级网格元胞自动机模型 (SIMGCA), 提出了 SIMGCA 模型在土地利用/覆被变化中的应用框架。下一步的工作将重点讨论如何利用数据挖掘方法在 SIMG 数据库中挖掘元胞转换规则, 并利用挖掘得到的规则模拟城市土地利用演化。(收稿日期: 2006 年 11 月)

参考文献:

- [1] 李德仁. 论广义空间信息网格和狭义空间信息网格[J]. 遥感学报, 2005, 9(5): 513-520.
- [2] 付迎春. 空间信息多级网格关键技术研究[D]. 武汉: 武汉大学, 2006.
- [3] 李德仁, 崔巍. 空间信息语义网格[J]. 武汉大学学报: 信息科学版, 2004, 29(10): 847-851.
- [4] 黎夏, 叶嘉安. 基于神经网络的单元自动机 CA 及真实和优化的城市模拟[J]. 地理学报, 1999, 54(4): 289-298.
- [5] Clarke. A self-modifying cellular automata model of historical urbanization in the San Francisco Bay area[J]. Environment and Planning B: Planning and Design, 1997, 24: 247-261.
- [6] White R, Engelen G. Cellular automata and fractal urban form: a cellular modeling approach to the evolution of urban land use patterns[J]. Environment and Planning A, 1993, 25: 1175-1199.
- [7] White R, Engelen G. Cellular automata as the basis of integrated dynamic regional modelling [J]. Environment and Planning B, 1997, 24: 235-246.
- [8] White R, Engelen G, Uljee I. The use of constrained cellular automata for high-resolution modeling of urban land use dynamics[J].