基于最大熵的CA模型及其城市扩张模拟

张亦汉¹, 刘小平^{2*}, 陈广亮², 胡国华³

1. 广东财经大学地理与旅游学院, 广州 510320;

2. 中山大学地理科学与规划学院, 广东省城市化与地理环境空间模拟重点实验室, 广州 510275;

3. 华东师范大学地理科学学院, 上海 200241

* 通讯作者, E-mail: liuxp3@mail.sysu.edu.cn

收稿日期: 2019-05-28; 收修改稿日期: 2019-09-24; 接受日期: 2019-10-17; 网络版发表日期: 2020-01-02

国家重点基础研究发展计划项目(编号: 2017YFA0604404)、国家自然科学基金项目(批准号: 41801304、41871306)和广东省教育厅特色创新项目(编号: 2016KTSCX045)资助

摘要 城市用地的扩张模拟是土地利用/土地覆盖变化(LUCC)研究的核心问题之一. 文章耦合了最大熵模型(Maxent)与元胞自动机(CA), 并提出一种新的城市扩张模型(Maxent-CA), 模型简单并可以从单期遥感影像中获取转换规则. 以2000年中国地区的遥感数据为基础数据, 结合人口密度、到城市的距离等空间变量, 构建了Maxent-CA模型并分析其训练样本、邻域空间结构和空间尺度等, 实现对模型的参数校正与验证. 研究发现, 当训练样本的采样比例与城市用地的转变量在全区的占比相匹配而且采用冯诺依曼邻域结构时, 模型将得出更好的模拟结果. 随着空间分辨率的降低模拟结果的空间结构变得简单, 但是模拟精度也会受到明显影响. 与传统逻辑回归CA模型对比, Maxent-CA能够很好地避免产生城市中心团簇的现象, 将有助于得出更接近于真实情况的模拟结果, 也能很好地满足大区域的城市扩张模拟的需求.

关键词 土地利用/土地覆盖变化(LUCC), 元胞自动机(CA), 城市扩张, 最大熵模型

1 引言

土地利用/土地覆盖变化(LUCC)是人类改造自然环境的最显著表现, 也是全球环境变化中最为剧烈的部分之一, 因此LUCC被认为是全球环境变化研究的前沿和热点(葛全胜等, 2008; 陈军和陈晋, 2018). 土地利用/覆盖的变化不仅影响着地球的能量循环和水循环等多种过程, 而且对土壤有机物分布、全球温度变化等有重要影响(Kalnay和Cai, 2003; Goldewijk和Ramanakutty, 2004; Börjesson和Tufvesson, 2011). 在土地

利用/覆盖变化的研究中, 最突出的是城市用地的扩张. 研究表明虽然城市用地的面积只占陆地面积的2%, 但约有50%人口生活在城市, 并产生了70%的二氧化碳排放物(Wang等, 2014). 而且城市用地的扩张会破坏原有的生态环境结构, 并引发多种环境问题(如城市热岛效应和大气污染)(Ding和Shi, 2013).

许多学者研究城市扩张的相关问题, 例如, 获取城市扩张的驱动力因子、动态变化过程与模拟/预测结果, 分析城市扩张对自然环境的影响, 以此制定土地利用规划和政策等(Costanza和Ruth, 1998; 何春阳等,

中文引用格式: 张亦汉, 刘小平, 陈广亮, 胡国华. 2020. 基于最大熵的CA模型及其城市扩张模拟. 中国科学: 地球科学, 50: 339–352, doi: 10.1360/SSTe-2019-0110

英文引用格式: Zhang Y, Liu X, Chen G, Hu G. 2020. Simulation of urban expansion based on cellular automata and maximum entropy model. Science China Earth Sciences, <https://doi.org/10.1007/s11430-019-9530-8>

2005). 其中, 元胞自动机(CA)具有强大的空间运算能力, 可有效地模拟复杂的动态系统, 被认为是研究城市用地变化非常方便且有效的工具(Shi和Pang, 2000; Li和Yeh, 2002; He等, 2006; 张亦汉等, 2011). 1970年, Tobler(1970)首次将CA应用到美国五大湖地区的城市扩张模拟中. Batty和Xie(1994)模拟了美国纽约州Baf-falo地区的扩张. Clarke等(1997)使用SLUTH CA模型模拟了美国旧金山地区的城市发展. Wu(2002)提出了逻辑回归CA模型并模拟了广州市的城市扩张. Li和Yeh(2004)结合数据挖掘与CA模型研究了东莞市的城市变化. Liu等(2014)整合景观扩张指数(LEI)和CA模型分析了珠江三角洲的城市群发展. 李丹等(2012)结合GPU和CA模拟了广东省的城市土地利用变化. 何春阳等(2005)运用系统动力模型和CA模型模拟了北方13省的土地利用变化. 这些研究表明CA模型能够模拟出与实际情况吻合的城市扩张特征.

在CA模型中, 往往采用一系列的环境空间变量, 这些空间变量对应着一系列的模型参数, 参数值的大小体现了空间变量的重要程度(Liu等, 2014; 林文棋等, 2018). 应用模型到真实的城市(城市群)中时, 需要对模型进行参数校正, 以模拟出接近实际的城市扩张形态. 虽然目前CA模型的参数校正方法较为丰富, 但是存在一些缺陷. 如Clarke等(1997)用计算机穷举搜索方法来寻找参数. 当空间变量较多时, 将有非常多的参数组合, 这种方法消耗的计算资源巨大. Li和Yeh(2002)提出利用训练神经网络的方法来自动获取CA的参数值. 但是神经网络属于黑箱结构, 模型参数的物理意义不明, 而且容易产生过度拟合的问题. Feng和Liu(2013)提出用模拟退火的启发式算法获取模型参数, 该方法通过不断对比模拟结果和真实结果以得到更高的模拟精度. 当研究区较大时, 该模型需要较多的计算资源. Cao等(2014)用基于Pareto最优的多目标算法纠正模型参数, 但是他们的实验需要用30h完成5000次的迭代运算. 更为重要的是, 已有的CA模型难以从单期的影像和空间变量中获取模型参数. 这些困难是大区域的土地利用变化模拟所需要着重考虑并解决的.

近年来, 一些学者提出用最大熵模型(Maxent)获取动植种群分布等(Phillips等, 2006; Phillips和Dudik, 2008; Elith等, 2006). 最大熵模型是基于Shannon的信息熵原理, 由Jaynes提出并应用于信息科学领域

的(刘振生等, 2013). 它利用单类单期的样本信息, 从各个环境变量中选择熵最大的分布作为最优分布, 构建模型并预测事物的地理分布. 由于模型简单实用且不需要多期多类数据, 其应用范围迅速扩展到其他领域中. 例如, Machado-Machado(2012)根据登革热病例位置并运用最大熵模型得出了墨西哥登革热分布情况. Couce等(2012)用最大熵模型得到全球珊瑚礁分布情况, 并分析环境因子对珊瑚分布的作用. Dellicour等(2017)调查了五种常见的欧洲大黄蜂物种, 并使用Maxent建模方法推断出现在和过去大黄蜂分布情况. Zhang等(2019)基于145个事件记录和30个环境变量, 运用MaxEnt建模和ArcGIS软件分析、预测了杨山牡丹的潜在分布. 这些研究均表明, 运用单类的样本和环境空间变量可以得出研究对象的分布概率. 因此, 本文将运用最大熵模型分析由城市用地的分布点和空间变量构成的单时段训练样本获取城市用地适宜性, 再耦合CA模型并提出新的城市扩张模型(Maxent-CA), 并以土地利用产品GlobeLand30作为基础数据(Chen J等, 2014; 陈军和陈晋, 2018), 应用到大湾区(中国)的城市用地扩张中.

2 研究方法与研究区

城市扩张是一种动态、开放和非线性的复杂过程. 应用CA模型进行城市扩张模拟时, 用户需要定义模型的结构并通过训练获取转换规则和模型参数(Li和Yeh, 2002). 在CA模型的训练和模拟中, 如何使用合适的方法从空间变量中挖掘转换规则和参数以较好地反映出城市扩张特点是最为重要的. 因此许多学者提出了不同的方法或模型, 但是它们在模型结构、转换规则和模型参数方面具有较大的差别(Wu, 2002; Li和Yeh, 2004; Liu等, 2008). 这些模型与方法往往需要采用两期甚至更多的影像和空间变量获取模型参数和转换规则(Liu等, 2010). 对于较大范围的研究区, 多期遥感影像往往较难获取, 而且使用较多的影像获取参数会有较多的误差积累(Zhang等, 2015). 文中引入最大熵模型到元胞自动机模拟中, 并提出Maxent-CA模型解决城市扩张模拟所碰到的这些问题. 模型分为三部分: 数据收集、模型训练和模型模拟(图1). 首先, 收集土地利用、城市用地和环境变量, 并输入到GIS数据库中. 其次, 将城市用地的采样数据和空间变量信息输入到最

大熵模型中, 并通过训练这些数据得出城市用地的最大熵分布. 再将最大熵分布结果输入到CA模型中, 同时考虑邻域函数、不同土地类型的易变性和随机因子, 以构建出Maxent-CA模型. 最后进行模型参数校正并利用该模型对城市扩张的动态过程进行模拟.

2.1 基于最大熵的CA模型

2.1.1 最大熵模型

最大熵模型在计算事物的概率分布时, 会充分考虑已知信息, 但对未知信息不加假设与揣测, 该模型常常用于事物的适宜性评价和预测. 城市用地的分布同样具有其适宜性, 因此也可以使用最大熵模型进行评价与预测. 该模型相对简单且仅需要两组数据, 一是待预测事物 X 的已知地理分布点, 以坐标的形式表示; 二是事物已知分布地区的范围和该地区的各种环境变量 Y (Elith等, 2006). 若待预测事物 X 的取值为 $X = \{x_1, x_2, \dots, x_n\}$, 那么其信息熵为

$$\begin{aligned} H(x) &= -\sum_i^n p(x = x_i) \log p(x = x_i) \\ &= -\sum_i^n p(X) \log p(X). \end{aligned} \quad (1)$$

事物的分布往往受环境变量 Y 的影响, 这些环境变量 Y 与预测事物 X 的坐标叠加, 便可获取由环境变量 Y 和预测事物 X 构成的训练样本数据集 $(X|Y)$. 当引入环境变量 Y 之后, Y 对 X 的分布影响概率和信息量会发生变化, 此时 X 的信息熵为

$$H(X|Y) = -\sum_i^n p(X, Y) \log p(X, Y). \quad (2)$$

模型认为满足全部已知的条件, 并对未知的情况不做假设的概率信息熵最大, 分布最为均匀, 预测的风险最小. 最大熵模型根据输入的样本数据集和目标函数(公式(3))进行训练.

$$X^* = \operatorname{argmax} H(X|Y). \quad (3)$$

模型通过随机种子生成算法不断训练并获取最大熵中的参数, 使得其计算值与实际值数据接近. 为了获取更好的结果, 将这些原始样本拆分成两类不同的样本(训练样本与检验样本)以方便交叉验证. 模型单次训练的适宜性概率往往有一定的偶然性, 多次训练的平均值作为最终结果更为合理, 将其输入到CA模型中便可构建模型进行城市扩张模拟.

2.1.2 Maxent-CA模型的构建

CA模型是一种时间、空间和状态都离散化的模型, 具有强大的空间建模和计算能力, 能够模拟出复杂动态系统的演化过程(Batty和Xie, 1994). 研究表明, 用CA模型进行地理模拟时, 主要使用一系列的影响因子, 这些影响因子以一定概率影响着城市的扩张(Wu, 2002; Liu等, 2010). 在Maxent-CA模型中, 元胞 x 在 t 时刻成为城市元胞的概率为

$$\begin{aligned} P_{(x,t)} &= RA \times P_{\maxent}(x, t) \\ &\quad \times P_{\text{neigh}}(x, t) \times P_{\text{variability}}(x, t), \end{aligned} \quad (4)$$

式中, RA 为随机项, $P_{\maxent}(x, t)$ 为最大熵模型的输出概率, $P_{\text{neigh}}(x, t)$ 为邻域的城市开发强度, $P_{\text{variability}}(x)$ 为每种土地利用类别转换为城市用地的易变性.

随机项可表示为

$$RA = 1 + (-\ln r)^w, \quad (5)$$

式中, r 为 $[0, 1]$ 间的随机数, w 为控制随机变量大小的参数(Li和Yeh, 2002).

$P_{\maxent}(x, t)$ 为最大熵概率分布, 是通过最大熵模型多次训练得出的.

邻域函数 $P_{\text{neigh}}(x, t)$ 表示 t 时刻元胞的邻域对其的影响值, 可通过如下公式表达:

$$P_{\text{neigh}}(x, t) = \frac{\sum \operatorname{con}(s_{i,j}^t = \text{urban})}{\sum \text{neighbor count}}. \quad (6)$$

易变性函数: $P_{\text{variability}}(x, t)$, 其值域范围为 $0 \sim 1$, 若该状态为利于城市扩张(如开发用地等), 则其取值越接近于1; 否则, 其值越接近于0.

Maxent-CA模型将采用约束性灰度CA的方法对城市扩张进行模拟, 因此需要引入两个约束参数(Li和Yeh, 2002). 一个是发展速率系数, 控制城市发展的速度; 一个是 t 时刻城市总扩张量 τ_t . 此时, 元胞 x 在 t 时刻的约束性的灰度CA模型可表示为

$$P^*(x, t) = f(\alpha, P(x, t), \tau_t), \quad (7)$$

那么, $t+1$ 时刻 $P^*(x, t+1)$ 的值可表示为

$$P^*(x, t+1) = P^*(x, t) + \Delta P^*(x, t), \quad (8)$$

式中, $\Delta P^*(x, t) \in (0, 1)$.

元胞 x 在 $t+1$ 时刻的状态为

图1 基于Maxent的CA模型

$$S^*(x, t+1) = \begin{cases} \text{developed, } P^*(x, t) = 1, \\ \text{developing, } 0 < P^*(x, t) < 1, \\ P^*(x, t), P^*(x, t) = 0. \end{cases} \quad (9)$$

模型采用Monte-Carlo的更新方式, 即每次迭代时随机选取一个元胞进行状态更新. 当达到 τ_t 的终止条件时停止迭代并输出结果.

2.2 研究区

2.2.1 研究区概况

中国位于亚洲东部, 太平洋西岸. 其国土面积居世界第三位, 拥有约 $960 \times 10^4 \text{ km}^2$ 陆地面积和约 $470 \times 10^4 \text{ km}^2$ 水域面积, 在东部和南部有约 $1.8 \times 10^4 \text{ km}$ 的海岸线. 中国有23个省、5个自治区、4个直辖市、2个特别行政区, 是世界上人口最多的发展中国家. 近几十年, 特别是改革开放以来, 中国经济保持高速增长, 并已成为世界第二大经济体和世界经济最大的贡献者. 根据统计数据, 2000~2017年, 中国人口年

均增长率仅为0.5%, 但是国内生产总值年均增幅达到9.3%. 伴随经济的发展, 中国的城市建成区面积不断增加, 城镇化水平也不断提高. 从2000~2018年, 城市建成区面积从 21524.54 km^2 增加到了 56225.4 km^2 , 全国城市建成区的年均扩张速度达到5.48%. 其中部分地区的扩张速度更快, 包括京津冀地区中的北京市和天津市, 其城市扩张速度分别达到了6.22%和6.05%, 长江三角洲的江苏省和浙江省的扩张速度达到7.06%和6.78%, 珠江三角洲的广东省扩张速度则达到了7.42% (中华人民共和国国家统计局, 2018). 从总体上看, 各个地区具有不同程度的扩张, 也呈现出多极增长模式, 这种研究区常常被用于验证并测试新的模型 (Liu等, 2014).

2.2.2 数据获取与处理

首先获取了高分辨率全球30m地表覆盖 (GlobeLand30) 数据集 (<http://www.globeland30.org>). 该数据包括耕地、林地、草地和城市用地等10大类地表覆盖信

息, 精确地测定了2000~2010年地表覆盖各类型的空间分布及变化情况(Chen J等, 2014). 其中2000年数据作为模型的起始数据, 2010年数据作为验证数据以检验模型模拟的准确性. 研究表明, 城市用地的适宜性与当地的地形情况有很大的关系(Clarke等, 1997), 使用了由美国太空总署(NASA)和美国国防部国家测绘局(NIMA)联合测量的SRTM数据, 主要包括数字高程模型(DEM)数据和从DEM中获取坡度数据. 考虑到城市扩张与植被覆盖有较大的关系, 采用了地理空间云(<http://www.gscloud.cn>)归一化植被指数(Liu等, 2014). 由于城市扩张往往与城市中心、交通和人口分布有非常密切的联系, 获取了城市分布数据、道路数据和人口密度数据及其生成的距离空间变量(包括到城市中心和道路的欧氏距离)(Liu等, 2010). 这些不同来源的数据具有不同的范围、投影和量纲, 需要对所有数据设置相同的坐标系统和范围, 并进行归一化处理, 最终得到中国的DEM、到城市的距离、到道路的距离、NDVI、人口密度和坡度等归一化环境变量数据(图2).

2.3 模型验证方法

在地理模拟或者遥感分类的评价中往往会使用混淆矩阵并计算其各种指标. 最常用的指标是总精度, 它等于被正确模拟的总面积除以研究区总面积(Li和Yeh, 2004).

$$p = \text{sum}(S_{\text{right}}) / S, \quad (10)$$

式中, S_{right} 为模拟结果与真实结果一致的面积, S 为研究区面积.

总精度只计算了混淆矩阵中的对角线上的值, 仅反映了正确的情况, 因此一些学者采用了Kappa系数, 它同时考虑了对角线及对角线以外的各个值(Liu等, 2010).

$$\text{Kappa} = (p - p_e) / (1 - p_e), \quad (11)$$

式中, p 为总精度, $p_e = (a_1 \times b_1 + a_2 \times b_2 + \dots + a_n \times b_n) / (S \times S)$, 其中 n 为类别个数, $a_1, a_2, \dots, a_n$ 为真实结果中每一类土地的面积, $b_1, b_2, \dots, b_n$ 为模拟结果中每一类土地的面积.

地理模拟中没有变化的区域在整个研究区中的占比往往较高, 而真实变化与模拟变化的部分相对较少.

总精度和Kappa系数均将真实未变化和模拟未变化部分纳入计算范围, 使得其计算的值偏高, 因此一些学者采用品质因数(FoM)以避免这种情况(Chen Y M等, 2014).

$$\text{FoM} = B / (A + B + C + D), \quad (12)$$

式中, A 表示在真实情景中非城市用地转变为城市用地, 而在模拟结果中却保持不变的面积; B 表示在真实情景和模拟结果中均为非城市用地转变为城市用地的面积; C 表示在真实情景中非城市用地转变为城市用地, 但在模拟结果中发生的转变类型与真实不相同的面积; D 表示在真实情景中没有发生变化, 而在模拟结果中却发生变化的面积.

总精度、Kappa和FoM是从点对点的角度分析模拟结果, 一些学者们倾向于分析模拟结果的空间形态, 其中景观指数指标被认为能够很好地反映模拟结果的空间形态(彭建等, 2006; Liu等, 2014), 选用了景观指数中的斑块数(NP)、斑块密度(PD)和分形维度(FRAC). 其中斑块数和斑块密度可以直观地反映出城市用地总体统计情况, 分形维度可以较好地反映斑块形状复杂程度, 这些指标可以帮助理解和评价该地区的景观现状和城市土地利用格局.

斑块数指某一土地利用类型的斑块总个数, 常被用来描述整个景观的异质性, 其计算如下:

$$NP = \sum N_{ij}, \quad (13)$$

式中, N_{ij} 表示第 i 个类别中的第 j 个斑块.

斑块密度往往用来反映某一土地利用类型整体的破碎化状况, 其计算公式如下:

$$PD = N_i / \text{Area}, \quad (14)$$

式中, N_i 表示第 i 个类别的斑块个数, Area 表示整个区域的面积.

分形维度常用来描述某一土地利用类型在空间结构上复杂变化的情况, 其计算公式如下:

$$\text{FRAC} = 2 \ln(P) / \ln(AP), \quad (15)$$

式中, P 为斑块周长, AP 为斑块面积.

3 Maxent-CA模型验证与分析

3.1 训练样本的采样

研究表明, CA模型需要经过参数校正过程才能较

图 2 模型数据与空间变量归一化数据

好地应用到研究区的模拟和预测中(Wu, 2002; Liu等, 2008). 参数校正是一种获取最优(如: 最符合现实)模型参数的过程, 往往需要运用训练样本进行训练. 训练样本的数量(或采样比例)对模型参数的精度有很大的影响, 而参数的精度又影响着模拟结果的精度. 为了能够得到更好的模拟结果, 对研究区的采样比例进行了敏感性分析. 运用蒙特卡罗随机采样方法对城市用地的分布位置与空间变量进行随机采样, 采样比例分别为1%、3%、5%、10%、15%、20%、30%和40%.

将不同采样样本的训练结果分别输入到构建的Maxent-CA模型中, 并设置起始影像、模拟转换量等便可以得到多个模拟结果. 通过对这些模拟结果进行随机抽样精度评价发现, 精度均达到90%(图3), 其中模拟精度在采样比例为3%时达到最高. 通过观察还可以发现精度先随着采样比例的增加而提高, 随后降低. 这个趋势与其他CA模型的训练趋势(精度随采样比例增加而增加)有所区别(Li和Yeh, 2002). 这主要是因为传统的模型训练需要两类(非城市与城市)甚至多类的训练目标, 训练样本越多越能获取到更多的城市用地信息, 因此精度会随采样比例增加而提高. 而最大熵的训练类别只有一种, 即城市用地. 这种模型在训练时对样本的依赖程度大, 需要更纯净的训练样本. 当采样样本中包含的非纯净样本进入到模型中时, 模型会拟合这些样本并产生误差. 同时, 通过统计可以发现在2000~2010年, 新增城市用地的占比大约为3%, 因此我们认为训练样本的采样比例与新增城市用地的占比相匹配时将有助于得出更好的模拟结果.

3.2 邻域空间结构

在元胞自动机的邻域空间结构中, 较多地采用摩尔邻域. 标准摩尔邻域是指以某元胞为中心, 其上、下、左、右、左上、右上、左下和右下等8个元胞构成的邻域范围. 另一种邻域结构为冯诺依曼邻域, 标准冯诺依曼邻域考虑的范围是以当前元胞的中心为圆心, 以元胞边长为半径的区域. 若其邻域元胞中心点在圆内(包括在边界上), 则视为其邻域. 有些学者考虑到非邻近的邻域元胞对中心元胞也有一定的影响(Li和Yeh, 2002), 因此采用扩展邻域, 如5×5邻域、7×7邻域等(图4).

实验发现不论是采用摩尔邻域还是冯诺依曼邻

图3 样本采样比例与模拟精度的关系

域, 模拟精度几乎均随着邻域范围的增加而降低(表1), 这表明离中心元胞越近的邻域对其影响越大, 这与地理学第一定律是相符合的. 在相同条件下, 冯诺依曼邻域的模拟结果略好, 在3×3、5×5、7×7、9×9、11×11和13×13邻域中, 采用冯诺依曼邻域的模拟精度比采用摩尔邻域的模拟精度分别高0.08%、2.63%、2.34%、1.9%、1.37%和1.71%. 为了探究微观尺度上邻域元胞对中心元胞的影响情况, 统计了新增城市用地5×5摩尔邻域对中心元胞的贡献(图4). 从3×3邻域的范围看, 中心元胞的上(4.30%)、下(4.31%)、左(4.39%)和右(4.31%)四种元胞的贡献均高于4.3%, 形成了第一层内圈, 也是贡献最大的圈层. 这四种元胞的贡献也是高于左上(4.27%)、左下(4.28%)、右上(4.25%)和右下(4.11%)的贡献. 再从5×5邻域范围看, 依然可以看出第二个圈层内的贡献是比圈层外的贡献大. 从本实验中可知随着邻域的增加, 外围的元胞对中心元胞的影响越来越弱.

3.3 空间尺度分析

地理元胞自动机的空间尺度选择在模型应用中至关重要, 它会直接影响到元胞自动机模型的运行效果(李丹等, 2012). 先对原始数据进行重采样, 得到空间分辨率(元胞大小)分别为1km×1km、2km×2km、3km×3km、5km×5km、8km×8km、10km×10km的数据. 在此基础上, 运用Maxent-CA模型分别在6个不同的尺度上对研究区进行城市扩张模拟, 模拟结果如图5所示. 由图可知, 随着空间分辨率的降低, 模拟结果中城市用地的团簇现象越明显, 这与实际有一定的偏差. 可以从模拟精度(表2)这一指标上看出, 模拟精度是先下降后上升. 其中, 当空间尺度从1km增加到3km时, 其模拟精度变化较大, 分别从92.04%降至86.21%, 再下

图 4 邻域结构及其贡献程度

表 1 不同邻域结构的模拟精度(%)

邻域大小	3×3	5×5	7×7	9×9	11×11	13×13
摩尔邻域	91.96	86.67	84.56	84.21	84.40	84.07
冯诺依曼邻域	92.04	89.3	86.9	86.11	85.77	85.78

降到83.85%。但是当空间尺度从5km增加到10km时,模拟精度有所提高,相比起3km的模拟精度,分别提高了0.27%、0.92%和2.65%。

为了更好地描述模拟结果的空间形态特征和评价该地区的景观现状和城市土地利用格局,计算了斑块数、斑块密度和分形维度(表2)。从表中可知,随着空间分辨率的降低,斑块数和斑块密度均逐渐下降。其中斑块数从64531下降到633,斑块密度从0.0068下降到0.0001。这表明分辨率从1km变化到10km过程中,较小的斑块合并到邻近的大斑块中,斑块个数减少,斑块面积增加。与非城市用地相比,城市用地的数量相对较少。在降尺度的过程中,往往会将零散的城市用地斑块归并到非城市用地斑块中,因此城市用地斑块个数会明显减少。分形维度是反映内部结构复杂性的一个指标。从表2中可以看出随着空间分辨率的降低,其分形维度值从1.0126减少到1.0067,表明低分辨率的结构变得简单。这主要是因为受到降尺度的影响,零散城市用地消失,城市用地变得更为聚集和简单。例如,在图5a中的1km模拟结果中,中国西北部新疆地区城市用地的分布是比较零散的,但是到8km(图5e)和10km(图5f)的模拟结果中,这些用地却消失了,这表明部分零散分布的城市用地已受到降尺度的影响已归并到非城市用地并消失。

3.4 模型对比与分析

为了验证模型的有效性,设计了对比实验,即对比传统的逻辑回归CA模型。首先,准备了两个模型的训练样本。由于训练方式不同,这两个训练样本略有不同。逻辑回归CA模型的训练样本需要各个空间变量(Y)与结果(X),而结果X变量必须是两类的(如城市与非城市),还需要将其转换为1与0以方便进行模型参数的训练。而Maxent-CA模型的训练样本需要各个空间变量(Y)与结果(X)中的一类(如城市)即可。将训练样本输入到对应的模型中训练可以得到各空间变量的贡献程度。其中,逻辑回归CA模型中各变量的贡献率由大到小分别是:到城市中心的距离、人口、坡度、到道路的距离、DEM和NDVI。而Maxent-CA模型中各个空间变量的贡献率由大到小分别是:人口、DEM、到道路的距离、到城市中心的距离、NDVI和坡度。这主要是因为两个模型训练样本中结果(X)的不一致引起的。Maxent-CA模型仅使用结果(X)中城市用地这一样进行训练,而逻辑回归的训练样本中结果(X)还包括了非城市用地这一类。

将逻辑回归训练的结果和最大熵模型训练的结果分别输入到相应的模型中,可以得到两个模型的模拟结果(图6)。从图中可以发现,在整体上,Maxent-CA的

图 5 不同尺度下的模拟结果

表 2 不同尺度下的模拟精度及其景观指数

空间尺度	1km	2km	3km	5km	8km	10km
精度(%)	92.04	86.21	83.85	84.12	84.77	86.5
斑块数	64531	9825	8484	3274	926	633
斑块密度	0.0068	0.001	0.0009	0.0003	0.0001	0.0001
分形维度	1.0126	1.01	1.0092	1.0083	1.0073	1.0067

模拟结果和实际的城市用地分布均比较零散, 团簇聚集的现象较少。但是在逻辑回归CA的模拟结果中, 京津冀、江苏南部、长江三角洲和珠江三角洲等地均出现较为明显的聚集现象。为了更好地进行对比分析, 选取了四个不同的区域进行放大。可以看出, 在京津冀(a)和珠江三角洲(d)等发达的地区, 逻辑回归CA的模拟结果有较多的新增城市用地。相比逻辑回归CA模型, Maxent-CA模型的模拟结果不论是城市用地的量还是城市用地扩张的形态均与实际有较高的吻合度。齐齐哈尔-大庆(b)和石河子-奎屯地区(c)等属于新增城市用地相对较少的地区, 逻辑回归CA的模拟结果在这些地区的城市中心也有较多的新增城市用地。同样地, Maxent-CA模型的模拟结果也能较合理地反映模拟出这些地区的新增用地情况。从整体与局部对比可以得出, 在逻辑回归CA模型的模拟结果中, 大多新增城市用地均分布在城市中心, 这与前面模型训练的空间变量的贡献度是一致的, 即到城市中心的距离贡献度最大。Maxent-CA模型的模拟结果显示新增城市用地分布到城市人群密集的地方, 即该模型中贡献率最大的是人口, 这与实际情况的模拟结果相吻合。

为了更精确地对比各个模型的精度与形态, 计算了基于点对点的指标(包括总精度、Kappa和品质因数)和景观指数。从精度上看, 逻辑回归CA模型的精度为82.99%, 而Maxent-CA模型的精度可达到92.04%。从Kappa和品质因数(FoM)上看, 逻辑回归模拟结果的Kappa和FoM分别为0.768和0.107。而Maxent-CA的模拟结果为0.870和0.134。从点对点的统计指标值中可以得出, 相比逻辑回归模型, Maxent-CA模型能够得出更高模拟精度的结果。从景观指数上看, 实际情况的斑块数(71935)是较多的, 斑块密度(0.0076)是较大的。这说明研究区新增城市用地比较分散地分布在全国各个地区, 并非集中分布。Maxent-CA模型和逻辑回归CA模型的斑块数与斑块密度分别为64531和0.0068、61571和0.0065, 均与实际情况有一定的差

异, 但Maxent-CA模型的偏差要略小。从分形维度可以看出, 实际情况(1.0120)的斑块是比较复杂的而且分布相对分散。由于逻辑回归CA模型围绕着城市中心扩张, 城市中心周边的城市元胞将连接起来, 零散的斑块变成较大的斑块。因此, 模拟结果中城市用地相对集中, 其复杂程度降低, 分形维度值减小到1.0102。而Maxent-CA模型将新增城市用地较为零散地分布到人口密度较大的地方, 因此其分形维度(1.0126)增大并与实际情况接近。

3.5 讨论

Globeland30数据集包含了10类不同的土地利用类型。为了方便, 城市用地采用了数据集中的人工地表。它包括了城市用地, 也包括了部分矿业用地、城市内部的绿地与水域(匡文慧等, 2016; 于溪等, 2016), 将在一定程度上夸大城市用地的扩张情况。这在大尺度的模拟中不会很明显, 但是多数城市扩张的研究关注较小尺度(如城市级别)的区域(Camagni等, 2002; Chen Y M等, 2014)。虽然我们放大到城市群级别并做了相应的定性分析, 但在城市尺度上仍然缺少对Globeland30数据集和模拟结果的全面定量分析与评价。在影响因子选择方面, 一些学者先收集相关所有影响因子, 然后通过非线性变换等以获取转换规则。此外也会关注影响因子在模拟过程中的变化, 如人口分布变化, 这些因子的变化会对模拟结果有影响(Li和Yeh, 1998; Liu等, 2008), 因此后续研究将考虑这些问题。从对比实验可知, 最大熵CA模型在单类的土地利用模拟中表现良好, 但是土地利用变化往往涉及多种土地利用的转化(Shalaby和Tateishi, 2007; Dadashpoor等, 2019)。从应用层面上看, 最大熵CA模型会受到一定的局限。另外还有一些学者注意到模型在模拟过程中的误差积累问题, 并采用了一些方法以减少误差积累(Verstegen等, 2014; Zhang等, 2015; Khaki等, 2018), 从这几个方面着手改进模型并控制误差传递将有助于

图 6 Maxent-CA和逻辑回归CA模型的对比

扩大其应用范围。

传统的CA模型应用到大面积的城市模拟中往往采用分区的策略以提高模拟精度, 但是如何合理地分区与聚合, 将同一类别的区域统一进行模拟往往是需要考虑的, 而且分区产生的计算量、城市接合部的城市扩张模拟不精确等问题也应注意(李丹等, 2012; Dahal和Chow, 2014)。Maxent-CA模型能够避免分区并将整个研究区参与到模拟中, 而且能够模拟出与真实情况吻合的空间格局。除此之外, 模型训练也较为简单, 它仅需要采用单期的样本便可模拟训练, 这对于大面积(如全球)的城市扩张模拟与预测是非常方便的。我们也将关注全球性的城市扩张模拟及其影响因素的差异性。不同国家和地区的决定性因素不完全一样, 因此在进行大范围模拟, 特别是全球城市扩张时, 需要对不同国家和地区的核心因子进行分析, 才能得出更合理的模拟结果。

4 结论

研究大范围的城市扩张过程具有重要的意义, 因为它不仅影响着地表覆盖变化, 也对全球环境变化有重要的影响。文中以高分辨率地表覆盖数据产品GlobeLand30为基础数据, 引入最大熵模型(Maxent)到CA模型中, 并提出基于单期遥感影像的城市扩张模型(Maxent-CA), 避免了传统CA模型需要从多期遥感影像进行参数纠正和大区域模拟时需要分区的问题。模型将2000和2010年的中国城市用地数据作为起始数据和检验数据, 以距离变量和人口密度等作为空间变量进行建模, 随后进行了相应的分析和验证。实验发现:

(1) Maxent-CA模型的模拟精度随采样比例的增加出现先上升再下降的情况。当训练样本的采样比例与研究区的城市用地转换量占比接近时, 其精度最高。

(2) 模型的模拟精度随邻域范围的增大而降低, 而且遵循地理学第一定律, 即离中心元胞越近的邻域对其影响越大。当扩展邻域的范围越大, 带入的干扰元胞越多, 其模拟精度也越低。

(3) 通过分析模拟的空间尺度可以发现, 随着空间分辨率的降低, 斑块数、斑块密度和分形维度均在逐渐下降, 城市用地的空间结构变得简单, 同时模拟精度也随着空间尺度的增加而降低。

(4) 在对比实验中可以发现, Maxent-CA模型能够

很好地避免逻辑回归CA模型团簇误差的产生。点对点精度和景观指数均表明, 提出的模型比逻辑回归CA模型更能模拟出接近真实情景的模拟结果。

致谢 感谢陈军教授、黎夏教授、乔纪纲老师、李丹老师、杨绪红老师和刘友桂老师在数据收集和论文修改中提供的帮助, 也非常感谢匿名审稿人对本文提出的宝贵意见和建议。

参考文献

- 陈军, 陈晋. 2018. GlobeLand30遥感制图创新与大数据分析. 中国科学: 地球科学, 48: 1391-1392
- 葛全胜, 戴君虎, 何凡能, 潘娜, 王梦麦. 2008. 过去300年中国土地利用、土地覆被变化与碳循环研究. 中国科学D辑: 地球科学, 38: 197-210
- 何春阳, 史培军, 陈晋, 潘耀忠, 李晓兵, 李京, 李月臣, 李景刚. 2005. 基于系统动力学模型和元胞自动机模型的土地利用情景模型研究. 中国科学D辑: 地球科学, 35: 464-473
- 匡文慧, 陈利军, 刘纪远, Xiang W N, 迟文峰, Lu D S, 杨天荣, 潘涛, 刘爱琳. 2016. 亚洲人造地表覆盖遥感精细化分类与分布特征分析. 中国科学: 地球科学, 46: 1162-1179
- 李丹, 黎夏, 刘小平, 陈逸敏, 李少英, 刘凯, 乔纪纲, 郑奕钟, 张亦汉, 劳春华. 2012. GPU-CA模型及大尺度土地利用变化模拟. 科学通报, 57: 959-969
- 林文棋, 吴梦荷, 张悦, 曾荣俊, 郑筱津, 邵磊, 赵鹿芸, 李少星, 唐燕. 2018. 中国城镇化的地区差异及其驱动因素. 中国科学: 地球科学, 48: 639-650
- 刘振生, 高惠, 滕丽微, 苏云, 王晓勤, 孔芳毅. 2013. 基于MAXENT模型的贺兰山岩羊生境适宜性评价. 生态学报, 33: 7243-7249
- 彭建, 王仰麟, 张源, 叶敏婷, 吴健生. 2006. 土地利用分类对景观格局指数的影响. 地理学报, 61: 157-168
- 于溪, 张冰琦, 李强, 陈晋. 2016. 基于30m地表覆盖数据的城市扩张测度方法. 中国科学: 地球科学, 46: 1180-1187
- 张亦汉, 黎夏, 刘小平, 乔纪纲. 2011. 基于数据同化的元胞自动机. 遥感学报, 15: 475-491
- 中华人民共和国国家统计局. 2018. 中国统计年鉴. 北京: 中国统计出版社. 1-16
- Batty M, Xie Y. 1994. From cells to cities. *Environ Plann B Plann Des*, 21: S31-S48
- Börjesson P, Tufvesson L M. 2011. Agricultural crop-based biofuels—Resource efficiency and environmental performance including direct land use changes. *J Clean Prod*, 19: 108-120
- Camagni R, Gibelli M C, Rigamonti P. 2002. Urban mobility and urban form: The social and environmental costs of different patterns of

- urban expansion. *Ecol Econ*, 40: 199–216
- Cao K, Huang B, Li M C, Li W W. 2014. Calibrating a cellular automata model for understanding rural-urban land conversion: A Pareto front-based multi-objective optimization approach. *Int J Geogr Inf Sci*, 28: 1028–1046
- Chen J, Ban Y F, Li S N. 2014. Open access to Earth land-cover map. *Nature*, 514: 434
- Chen Y M, Li X, Liu X P, Ai B. 2014. Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy. *Int J Geogr Inf Sci*, 28: 234–255
- Clarke K C, Hoppen S, Gaydos L. 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environ Plann B*, 24: 247–261
- Costanza R, Ruth M. 1998. Using dynamic modeling to scope environmental problems and build consensus. *Environ Manage*, 22: 183–195
- Couce E, Ridgwell A, Hendy E J. 2012. Environmental controls on the global distribution of shallow-water coral reefs. *J Biogeogr*, 39: 1508–1523
- Dadashpoor H, Azizi P, Moghadasi M. 2019. Land use change, urbanization, and change in landscape pattern in a metropolitan area. *Sci Total Environ*, 655: 707–719
- Dahal K R, Chow T E. 2014. A GIS toolset for automated partitioning of urban lands. *Environ Model Softw*, 55: 222–234
- Dellicour S, Kastally C, Varela S, Michez D, Rasmont P, Mardulyn P, Lecocq T. 2017. Ecological niche modelling and coalescent simulations to explore the recent geographical range history of five widespread bumblebee species in Europe. *J Biogeogr*, 44: 39–50
- Ding H Y, Shi W Z. 2013. Land-use/land-cover change and its influence on surface temperature: A case study in Beijing City. *Int J Remote Sens*, 34: 5503–5517
- Elith J, H. Graham C, P. Anderson R, Dudik M, Ferrier S, Guisan A, J. Hijmans R, Huettmann F, R. Leathwick J, Lehmann A, Li J, G. Lohmann L, A. Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, J. Phillips S, Richardson K, Scachetti-Pereira R, E. Schapire R, Soberón J, Williams S, S. Wisz M, E. Zimmermann N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29: 129–151
- Feng Y J, Liu Y. 2013. A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing. *Int J Geogr Inf Sci*, 27: 449–466
- Goldewijk K K, Ramankutty N. 2004. Land cover change over the last three centuries due to human activities: The availability of new global data sets. *GeoJournal*, 61: 335–344
- He C Y, Okada N, Zhang Q F, Shi P J, Zhang J S. 2006. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Appl Geogr*, 26: 323–345
- Kalnay E, Cai M. 2003. Impact of urbanization and land-use change on climate. *Nature*, 423: 528–531
- Khaki M, Ait-El-Fquih B, Hoteit I, Forootan E, Awange J, Kuhn M. 2018. Unsupervised ensemble Kalman filtering with an uncertain constraint for land hydrological data assimilation. *J Hydrol*, 564: 175–190
- Li X, Yeh A G O. 1998. Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *Int J Remote Sens*, 19: 1501–1518
- Li X, Yeh A G O. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int J Geogr Inf Sci*, 16: 323–343
- Li X, Yeh A G O. 2004. Data mining of cellular automata's transition rules. *Int J Geogr Inf Sci*, 18: 723–744
- Liu X P, Li X, Chen Y, Tan Z Z, Li S Y, Ai B. 2010. A new landscape index for quantifying urban expansion using multi-temporal remotely sensed data. *Landsc Ecol*, 25: 671–682
- Liu X P, Li X, Shi X, Wu S K, Liu T. 2008. Simulating complex urban development using kernel-based non-linear cellular automata. *Ecol Model*, 211: 169–181
- Liu X P, Ma L, Li X, Ai B, Li S Y, He Z J. 2014. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *Int J Geogr Inf Sci*, 28: 148–163
- Machado-Machado E A. 2012. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. *Appl Geogr*, 33: 82–93
- Phillips S J, Anderson R P, Schapire R E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model*, 190: 231–259
- Phillips S J, Dudik M. 2008. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31: 161–175
- Shalaby A, Tateishi R. 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Appl Geogr*, 27: 28–41
- Shi W Z, Pang M Y C. 2000. Development of Voronoi-based cellular automata—An integrated dynamic model for Geographical Information Systems. *Int J Geogr Inf Sci*, 14: 455–474
- Tobler W R. 1970. A computer movie simulating urban growth in the Detroit region. *Econ Geogr*, 46: 234–240
- Verstegen J A, Karssenbergh D, van der Hilst F, Faaij A P C. 2014. Identifying a land use change cellular automaton by Bayesian data assimilation. *Environ Model Softw*, 53: 121–136

- Wang Y, Ji W, Yu X F, Xu X L, Jiang D, Wang Z, Zhuang D F. 2014. The impact of urbanization on the annual average temperature of the past 60 years in Beijing. [Adv Meteorol](#), doi: 10.1155/2014/374987
- Wu F L. 2002. Calibration of stochastic cellular automata: The application to rural-urban land conversions. [Int J Geogr Inf Sci](#), 16: 795–818
- Zhang K L, Zhang Y, Zhou C, Meng J C, Sun J, Zhou T H, Tao J. 2019. Impact of climate factors on future distributions of *Paeonia ostii* across China estimated by MaxEnt. [Ecol Inf](#), 50: 62–67
- Zhang Y H, Li X, Liu X P, Qiao J G. 2015. Self-modifying CA model using dual ensemble Kalman filter for simulating urban land-use changes. [Int J Geogr Inf Sci](#), 29: 1612–1631

(责任编辑: 岳天祥)