

基于局部化转换规则的元胞自动机土地利用模型

宇林军^{1,2}, 孙丹峰¹, 彭仲仁³, 李 红⁴

(1. 中国农业大学资源与环境学院, 北京 100193; 2. 中国科学院遥感与数字地球研究所, 北京 100101;
3. 美国佛罗里达大学区域与城市规划系, 美国 盖恩斯威尔 32611-5706; 4. 北京市农林科学院农业综合
发展研究所, 北京 100097)

摘要: 传统土地利用元胞自动机 (Cellular automata, CA) 模型基于空间同质性假设, 使用全局性模型建立元胞转换规则, 忽略了土地利用变化驱动因素的驱动作用在空间上的变化。以美国佛罗里达州的橙县 (Orange County) 2003-2009年土地利用变化为例, 提出了基于局部化转化规则的CA土地利用模型, 其中元胞的土地利用类型适宜性由地理加权多项logit模型 (Geographically weighted multinomial logit, GWML) 获得。结果表明: GWML模型较传统全局性多项logit (Multinomial logit, MNL) 模型有更高的数据解释能力。基于GWML模型的土地利用CA模型能反映局部土地利用变化模式, 因而较基于MNL模型的CA模型具有更高的模拟精度。所得结论对未来国内地区的研究有借鉴意义。

关键词: 土地利用变化; 元胞自动机模型; 空间异质性; 多项Logit模型; 地理加权多项Logit模型

元胞自动机 (Cellular automata, CA) 是一种时间、空间、状态都离散的网格动力学模型, 已广泛应用于土地利用变化模拟。元胞的转化规则是元胞自动机模型的核心, 决定了其模拟能力^[1-3]。传统CA土地利用模型的转换规则建立在基于空间同质性假设的全局性模型基础上, 假设土地利用演变与驱动因素间的关系在空间上恒定。然而, 土地利用演变存在很强的空间异质性, 即驱动因素对土地利用的作用随着空间位置的变化而变化。特别是在大尺度的研究中, 空间异质绩效应更为显著。

土地利用演变的空间异质性体现为基于距离的空间变化和基于位置的空间变化两个方面。据Tobler的地理学第一定律^[4], 基于距离的空间变化可表达为: 元胞间的相互作用随着元胞间的距离增加而减小。基于位置的空间变化指驱动因素对土地利用演变的作用随着空间位置变化而变化。例如, 在某一区域 (位置) 邻域作用为土地利用变化的主要驱动力; 而在其他地区 (位置), 其他因素 (如可达性) 则可能为主要驱动因素。

已有的CA土地利用模型采用多种方法来反映基于距离的空间变化, 如扩大的邻域^[5-7]及空间加权的邻域^[8]等。类似地, Huang等^[9-11]在逻辑回归模型中入了位置权重, 体现了基于距离的空间变化, 消除了空间自相关的影响, 但仍未体现基于位置的空间变化。模型局部化是解决基于位置的空间异质性的一种有效方法, 其中地理加权回归 (Geographically

收稿日期: 2012-03-09; 修订日期: 2012-11-15

基金项目: 国家自然科学基金项目(41130526, 41071146)

作者简介: 宇林军(1981-), 男, 博士, 河北张家口人, 主要研究领域为土地利用与信息技术。

E-mail: yulinjun81@gmail.com

通讯作者: 孙丹峰(1971-), 男, 山东人, 教授, 主要从事土地系统分析和资源环境信息技术。

E-mail: sundf@cau.edu.cn

Weighted Regression, GWR) 是一个典型的局部化线性回归模型^[12]。由于考虑了空间异质性, GWR模型较传统的全局模型具有更高的数据解释能力^[13]、更小的误差项及更小的空间依赖性^[14]及可视化的估计结果^[15]。因此, GWR已广泛应用于空间非稳定的关系和基于空间尺度的关系的研究, 如环境正义分析 (environmental justice analysis)^[15]、灾害宣告 (disaster declarations)^[16]、森林转型 (forest transition)^[17]、城市热岛空间结构^[18]及房价^[19]等。但由于CA土地利用模型中元胞状态的转变是一个典型的离散选择过程^[20], 描述线性关系的局部化模型 (如GWR模型) 不适宜于构建模型的转换规则, 因此目前还没有将GWR模型应用到CA的研究。离散选择模型 (如MNL模型) 基于效用最大化理论, 能够体现社会经济等多种驱动因素的共同作用^[21], 因此成为CA土地利用模型的一个重要分支^[22-26]。局部化的离散模型因而可用于分析土地利用演变中的空间异质性效应。McMillen等^[27]首先将局部加权最小二乘法应用于传统离散选择模型的估计过程, 构建了一个局部化的离散选择模型, 分析了交通对土地混合利用的影响。Luo等^[28]应用类似的方法分析了驱动因素对城市扩张的影响。这些研究主要关注于分析驱动因素的影响作用在空间中的变化, 缺少对土地利用变化的模拟能力, 目前还少有研究将局部化的离散选择模型 (如MNL模型) 应用于CA中。本研究将局部化的离散选择模型应用于CA土地利用模型中, 通过创建局部化的转换规则解决土地利用动态变化模拟过程中的空间异质性效应。

1 基于地理加权多项logit模型的CA土地利用模型

1.1 GWML模型

通过将局部加权回归过程应用到模型估计, 将传统的多项logit (Multinomial logit, MNL) 模型拓展为地理加权的形式, 即GWML (Geographically weighted multinomial logit, GWML) 模型。在CA土地利用模型的背景下, GWML模型描述如下^[27]:

$$p_{i,k} = \frac{\exp(x_i \beta_{i,k})}{\sum_{k=1}^K \exp(x_i \beta_{i,k})} \quad (1)$$

式中: $p_{i,k}$ 为元胞*i*转换为土地利用类型*k* ($k=1,2,\dots,K$) 的概率; K 为全部土地利用类型; X_i 为元胞*i*的属性向量; $\beta_{i,k}$ 为在元胞*i*所在的区域上, 当因变量为土地利用类型*k*时的参数 (系数) 估计。 β 通过最大化以下似然函数 (Log-likelihood Function) 获得^[27]:

$$\ln L_i = \sum_{j=1}^n \left[w_{ij} \sum_{k=0}^K I_{j,k} \ln(P_{j,k}) \right] \quad (2)$$

式中: $P_{j,k}$ 为元胞*j*转换为土地利用类型*k*的概率; n 为观察值的个数; $I_{j,k}$ 为一个哑元变量, 当元胞*j*转变为土地利用类型*k*时, $I_{j,k}=1$, 否则 $I_{j,k}=0$ 。 w_{ij} 为在参数估计过程中赋予观察值*j*的权重。当对于所有元胞*i*及*j*, 权重 w_{ij} 都等于1时, GWML模型退化为标准的MNL模型。权重 w_{ij} 可由高斯加权函数获得^[29]:

$$w_{i,j} = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{d_{i,j}}{b_i}\right)^2\right] & \text{如果 } d_{i,j} < b_i \\ 0 & \text{其他} \end{cases} \quad (3)$$

式中: $d_{i,j}$ 为元胞 (观察值) *j* 与元胞 (回归位置) *i* 间的欧式距离; b_i 为距离带宽 (Distance Bandwidth), 距离带宽 b_i 取使位置*i*的模型达到最优的带宽。文中, GWML模型

的优劣由矫正的 Akaike 信息量准则^[30]来度量:

$$ACI_c = 2n \ln(\hat{\sigma}^2) + n \ln(2\pi) + n \left(\frac{n + \text{trace}(H)}{n - 2 - \text{trace}(H)} \right) \quad (4)$$

式中: n 为观察值的个数; H 为帽子矩阵 (Hat Matrix)^[30], 定义如下:

$$H_i = X_i (X^T W_i X)^{-1} X^T W_i \quad (5)$$

式中: $\text{trace}(H)$ 为取矩阵 H 对角线元素和的函数。 $\hat{\sigma}^2$ 为估计标准离差^[30]:

$$\hat{\sigma}^2 = \frac{\sum (y_i - \bar{y}_i)^2}{n - [2 \text{trace}(H) - \text{trace}(H^T H)]} \quad (6)$$

式中: \bar{y}_i 为观察值 i 对应的预测值; H^T 为矩阵 H 的转置矩阵; y_i 为 i 的观察值。

Akaike 信息量准则能够比较具有相同因变量的异构模型的优劣。Akaike 信息量准则值越小, 模型越接近于“真”模型。本文使用黄金分割法, 通过逐步缩小模型的 Akaike 信息量准则值的范围求最优距离带宽。为了避免在估计过程中出现奇异矩阵错误 (Singular Matrix Error), 最小带宽 (标记为 minb) 的初始值设为使以估计点 i 为中心, 能够覆盖所有类型的土地利用变化的圆的最小半径。最大带宽 (标记为 maxb) 的初始值为估计点 i 与所有观察点间的最大距离。设 $AIC_c(b)$ 为带宽为 b 时模型的矫正的 Akaike 信息量准则值。在每个循环中, 执行带宽为当前平均带宽 (meanb) 模型的估计, 其中 meanb 取当前最大带宽与最小带宽的平均值。下一循环中 minb 及 maxb 取值为:

当 $AIC_c(\text{meanb}) > AIC_c(\text{minb})$ 时, $\text{minb} = \text{meanb}$;

当 $AIC_c(\text{meanb}) < AIC_c(\text{maxb})$ 时, $\text{maxb} = \text{meanb}$ 。

当两个模型的 Akaike 信息量准则值小于 3 时, 二者对数据的解释能力没有明显差别。因此, 当 $AIC_c(\text{maxb})$ 与 $AIC_c(\text{minb})$ 绝对差小于 3 时, 循环终止。本研究在 Opus 平台^[22]的 MNL 模型源代码的基础上实现了上述 GWML 模型。

1.2 MNL 模型及 GWML 模型的变量配置及系数估计

本研究基于一个常用系统化构建方法^[31-33]来构建 MNL 模型及 GWML 模型的变量配置。第一步用所有在理论上显著的变量构建初始模型, 第二步进行模型参数估计, 第三步在模型变量配置中删除模型估计结果中最不显著的变量。以新的模型变量配置继续执行第二步和第三步, 直到模型变量配置中所有变量都显著。

系统化模型构建方法能够消除非显著变量的影响, 提高模型精度。但由于 GWML 模型为一个局部化的模型, 需要对研究空间中的不同位置进行上述系统化模型构建过程, 因此当位置数量很大时, 模型构建工作就非常耗时。本研究采用 100 m 的空间分辨率, 若对所有 91028 个位置 (可开发元胞) 进行上述系统化模型构建过程, 初步估算需要至少 3 个月以上的工作量 (基于配置为 Intel(R) Core(TM) i52.67HG 中央处理器和 4GB 内存的个人电脑)。但实际上, 由于临近的土地利用变化具有相似性, 因此没必要对所有位置进行模型空间配置。模型变量配置构建及系数估计过程可在更低的空间分辨率下施行。本研究假设同一个人口普查区 (census block group) 内的土地利用变化具有相同的模式, 可由唯一的模型配置来反映, 因此在人口普查区尺度上执行上述的系统化模型配置构建过程来构建 GWML 模型。

1.3 土地利用变化的影响因素

土地利用变化受多种驱动因素的影响。基于数据限制, 考虑的因素主要有: 邻域配置、可达性、物理特征及土地利用政策。

(1) 邻域配置特征

使用 Verburg 等^[34]提出的土地利用丰富度因子 (Enrichment Factor) 来描述元胞的邻域土地利用配置特征, 定义如下:

$$F_{i,k,d} = \frac{n_{k,i,d}}{n_{d,i}} \frac{N_k}{N} \quad (7)$$

式中: $F_{i,k,d}$ 为以元胞 i 为中心, 半径为 d 的邻域的土地利用类型 k 的丰富度因子; $n_{k,i,d}$ 为元胞 i 的邻域中土地利用类型为 k 的元胞的数量。 N 及 N_k 分别为整个元胞空间中元胞数量及土地利用类型为 k 的元胞数量。当邻域的土地利用类型丰富度小于总体平均丰富度时, 邻域丰富度因子介于 0 与 1 之间。否则, 丰富度因子大于 1。由于丰富度因子反映了局部土地利用类型密度与总体平均密度的相对程度, 能更好地描述邻域的土地利用配置特征, 因此本研究使用丰富度因子表征邻域的土地利用配置特征。

(2) 可达性

可达性是影响居民、企业、房地产开发商选址的一个重要因素, 因而常被纳入土地利用模型中^[35]。在本研究中, 元胞的可达性用一系列最短距离指标来衡量, 包括与主干道、高速公路、火车站、机场及 CBD 的最短距离。对于元胞的公交可达性, 本文采用元胞 800 m 范围内的公交线数量而不是最短距离来度量。

(3) 物理特征

元胞的物理特征决定了元胞开发为某种土地利用类型的成本。本研究考虑的物理特征包括高程、坡度和土壤质量三个指标。元胞的高程及土壤质量由收集到的数据直接获得, 坡度则为中心元胞与其 8 个邻里元胞间的最大高程差。

(4) 土地利用政策与法规

政策与法规同样是城市土地利用发展的重要影响因素。企业区划 (Enterprise Zoning) 政策给予迁入企业区划内的企业以税收优惠, 是影响工业用地选址的一个重要因素。元胞的企业区划政策由一个哑元变量描述, 定义如下:

$$ZI_i = \begin{cases} 1 & \text{如果单元 } i \text{ 在企业区划中} \\ 0 & \text{其他} \end{cases} \quad (8)$$

当元胞 i 在企业区划内, ZI_i 值为 1; 否则为 0。

1.4 基于 MNL 模型及基于 GWML 模型的 CA 土地利用模型

元胞的转换概率定义如下:

$$p_{ij} = S_{ij} * C_{ij} * \exp(\theta_{ij}) \quad (9)$$

式中: P_{ij} 为元胞 i 转换为土地利用类型 j 的联合概率; S_{ij} 为土地利用类型 j 对元胞 i 的适宜性, 由 MNL 模型或 GWML 模型获得; C_{ij} 为一个哑元变量, 指示依据规划或政策元胞 i 是否能转变为土地利用类型 j ; θ_{ij} 为以元胞 i 为中心, 6×6 邻域的土地利用类型 j 距离加权密度:

$$\theta_{ij} = \frac{\sum_{k=1}^{K=6 \times 6} I(L_k = j)}{6 \times 6 - 1} \quad (10)$$

式中: $I()$ 为一个条件函数, 当元胞 k 的土地利用类型为 j 时返回 1, 否则返回 0; $d_{i,k}$ 为元胞 i 与元胞 k 间的欧式距离。

基于 GWML 模型的 CA 土地利用模型的结构及主要模拟过程如图 1, 其中土地利用需求模型为一个外部模型, 基于社会经济模型预测各类土地利用的总需求。文中, 为了模型验证, 土地利用总需求为被模拟期间 (2003-2009 年) 的实际土地利用变化。模拟过程为

一个将全部土地需求分配到可开发元胞上的循环过程：

(1) 第一步（图1①），初始化元胞空间。

(2) 第二步（图1②），使用基于MNL或GWML模型的土地利用适宜性模型计算每个元胞的土地利用适应性。

(3) 第三步（图1③），根据公式(9)计算每个元胞的联合转换概率。

(4) 第四步（图1④），从总土地需求中，随机选择一个未被完全分配的土地利用类型*i*。

(5) 第五步（图1⑤），在所有可开发元胞中选取具有最大转换为土地利用类型*i*的联合转换概率的元胞，将其状态转变为土地利用类型*i*，并在总土地需求中移除一个元胞面积的土地利用类型*i*。

(6) 如果总土地需求不为空，则执行第6步（图1⑥）。重新计算以第五步选中的元胞为中心、6×6邻域内所有可开发元胞的联合转换概率。返回第四步，继续执行循环，直到总土地需求为空。

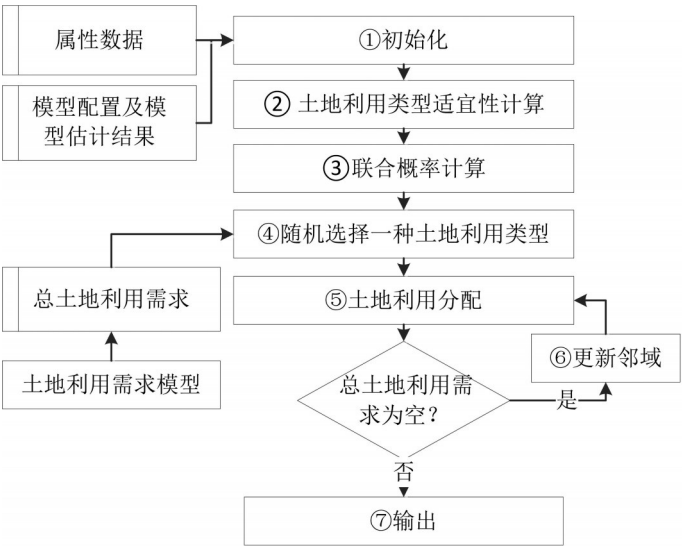


图1 系统构架及模拟过程

Fig. 1 System framework and the simulation process

2 案例研究——以美国佛罗里达州橙县为例

2.1 研究区概况与数据来源

以美国佛罗里达州橙县为研究区。橙县位于佛罗里达州中部，以奥兰多（Orlando）市为中心，介于杰克逊维尔（Jacksonville）与迈阿密（Miami）两大城市之间，由六个行政区构成（Ted Edwards, Jennifer Thompson, Lui Damiani, S. Sott Boyd, Tiffany Morre Russell和Fred Brummer）。橙县是奥兰多地区的交通枢纽。东西向的4号洲际高速公路和南北向的佛罗里达州收费高速公路交汇于此。联邦、州及本地政府机构，大量的国际贸易及区域商业、文化、体育机构都在此设有办事处。橙县面积约为2600 km²。2003-2009年间实际土地利用变化如图3，包括691.21 km²可开发用地转换为居住用地、14.52 km²可开发用地转换为商服用地、7.52 km²可开发用地转换为工业用地及1.68 km²可开发用地转换为教育用地。如图2，新增的商服用地、教育用地及工业用地呈明显的集聚分布模式。

交通网络设施数据及规划区域边界数据来自于佛罗里达地理数据图书馆（Florida Geographic Data Library, FGDL）。该机构负责收集和发布全佛罗里达州的空间数据。数字高程数据来自美国地质测量局（US Geological Survey）1999年30 m精度高程数据集。土地利用数据为2003年及2009年包裹级（Parcel level）土地利用数据，源于佛罗里达税务部（Florida Department of Revenue）并由FGDL整理为15通用类。由于本文关注于城市

用地增长,因此将这 15 种分类进一步整理为 6 类:可开发土地、不可开发土地、居住用地、商服用地、教育用地及工业用地。可开发用地指能够被开发为居住用地、商服用地、教育用地及工业用地的用地类型。不可开发用地指不能或很难开发为居住用地、商服用地、教育用地及工业用地的用地类型,例如矿区及水域等。研究区域由 260114 个 $100\text{ m} \times 100\text{ m}$ 元胞构成,其中包括

91028 个可开发元胞,83828 个不可开发元胞及 85258 个已开发元胞。本研究关注可开发元胞的状态转变,即转变为居住用地、商服用地、教育用地或工业用地。

2.2 模拟结果及讨论

首先分别基于传统 MNL 模型及 GWML 模型构建土地利用适宜性模型,其选项由“保持不变”,“居住用地”、“商服用地”、“教育用地”及“工业用地”构成。为了测试所提出的 CA 土地利用模型的解释能力和预测能力,将 50% 的数据用于模型参数估计,另外 50% 的数据的用于模型验证。测试了基于四个距离范围的土地利用丰富度因子:0-1.6 km (0-2 mile), 1.6-3.2 km (2-4 mile), 3.2-11.2 km (4-7 mile) 及 11.2 km (7 mile) 以上。结果表明只有基于距离为 0-1.6 km 的丰富度因子显著。该结果与 Zhou 等^[20]的发现一致,即 2 英里是邻域所能影响的范围。

2.2.1 MNL 模型的变量配置及系数估计结果 表 1 显示了基于传统 MNL 模型的土地利用适宜性模型的构成及参数估计结果。所有显著变量的系数估计值都具有所期望的符号。丰富度因子是土地利用发展的一个重要影响因素。对于居民用地因变量,居民用地丰富度因子变量及商服用地丰富度因子变量的系数分别为 0.426 和 -0.363,说明具有更多居民用地及更少商服用地的地区对居民用地有更高的吸引力。对于工业用地因变量,工业用地丰富度因子变量的系数为正,表明工业用地趋于集聚分布。商服用地丰富度因子变量对商服用地因变量的系数及教育用地丰富度因子变量对教育用地应变量的系数为正,表明集聚效应也存在于商服用地及教育用地中。对教育用地来说,商服用地丰富度因子变量的系数值为负,说明教育用地趋向于避开商服用地。

可达性同样在土地利用模式中起到重要作用。由表 1,不同土地利用类型对不同类型的可达性敏感。例如工业用地及商服用地更喜欢趋近于火车站及交通枢纽,居民用地周围则通常具有较少的公交车站,而商服用地更趋向于选择公共交通发达的地区,教育用地则对所有可达性不敏感。

企业区划与工业用地间的关联明显。正如期望的那样,企业区划变量对工业用地因变量的系数为正,说明企业在其他条件一致的情况下更趋向于选择被企业区划的地区。

2.2.2 GWML 模型的模型配置及参数系数估计结果 在人口普查区尺度上执行上述系统

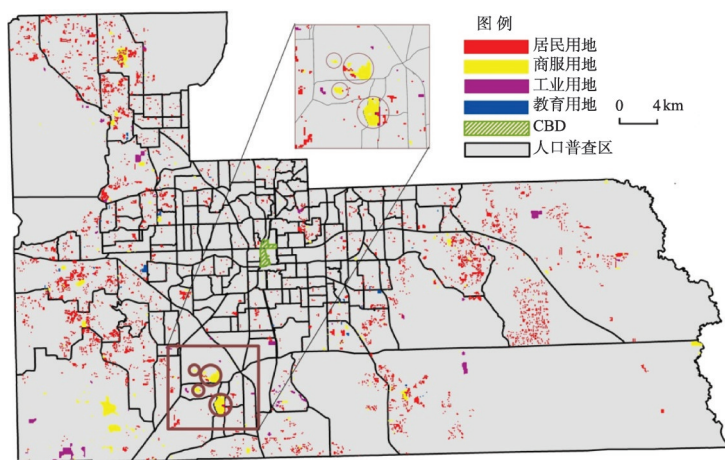


图 2 2003-2009 年间研究区实际土地利用变化

Fig. 2 Real land use changes from 2003 to 2009 in Orange County, FL, USA

表1 MNL模型的系数估计结果
Tab. 1 Estimation results of the MNL model

变量名称	“保持不变”		“居住用地”		“工业用地”		“商服用地”		“教育用地”	
	系数	T检测	系数	T检测	系数	T检测	系数	T检测	系数	T检测
常数	-	-	-	-	-	-	-3.773	-31.69	-2.831	-12.12
工业用地丰富度因子	-	-	-	-	0.1	12.86	-	-	-	-
教育用地丰富度因子	0.218	12.02	-	-	-	-	0.362	16.7	0.419	14.6
居民用地丰富度因子	-	-	0.426	12.98	-	-	-	-	-	-
商服用地丰富度因子	-0.605	-13.71	-0.363	-6.08	-	-	0.246	5.07	-0.887	-5.81
火车站或交通枢纽可达性	-	-	-	-	0.211	-11.32	0.1	-12	-	-
飞机场可达性	0.076	11.78	-	-	-	-	0.195	18.07	-	-
CBD可达性	0.155	10.3	-0.069	-4.39	-0.071	-3.93	0.06	3.06	-	-
工业区划	-	-	-	-	0.066	0.48	-	-	-	-
公交可达性	-	-	-0.143	-13.95	-	-	0.112	14.64	-	-
AIC _c : 26022.94878										
Log-likelihood ratio index: 0.818053797										

性模型配置构建方法来构建GWML模型。研究区包括369个人口普查区。人口普查区是由美国人口调查局根据区域相似性而定义的人口普查最小地理单元。城市中的人口普查区边界通常由街道、道路、河流或其他可见的物理或文化实体的边界构成。因此，一个人口普查区内的土地利用变化模式具有很强的相似性。本文因此选择人口普查区尺度上构建GWML模型。

由于GWML模型使用局部化模型反映基于位置的空间异质性效应，因而在不同位置（人口普查区）具有不同的变量配置及变量系数。表2统计了GWML模型变量系数的估计值在研究区的最大、最小及平均值。与表1结果相比，GWML模型中变量系数的平均值与MNL模型中的变量系数的符号一致，说明MNL模型作为全局性模型反映了土地利用变化在整个研究区中的总体趋势。以商服用地丰富度因子变量对商服用地因变量的系数为例，其在GWML模型中的平均值为0.319，与MNL模型结果（0.246）在符号上一致，说明商服用地总体上趋于集聚分布。此外，商服用地丰富度因子的系数存在显著的空间变化（图3），因此商服用地丰富度因子对商服用地因变量的作用在空间中具有显著的空间异质性。总体上来说，商服用地丰富度因子变量的系数在大部分人口普查区中显著，在CBD附近具有较大值，说明商服用地的积聚效应在CBD附近更为明显。GWML模型中的其他变量也表现出了显著的空间变化。总之，与基于空间同质性假设的传统MNL模型不同，GWML模型通过模型局部化表达了土地利用变化中的空间异质性效应。

由表2，GWML平均校正的Akaike信息量准则值为3537.86，高于传统MNL模型的26022.95，因此GWML模型的数据拟合优度优于传统的MNL模型。似然指数也常被用来测度离散选择模型的数据拟合优度，其值范围为0-1，表明模型由优到劣。GWML的似然指数的平均值为0.83，高于MNL模型的0.818。因此似然指数值同样表明GWML模型的数据拟合优度优于传统的MNL模型。

2.2.3 CA模型模拟结果 基于MNL模型及基于GWML模型的CA模型对2003-2009年佛罗里达州橙县土地利用变化模拟结果如图4所示。总体上，与基于MNL模型的CA土地利

表2 GWML模型的系数估计结果
Tab. 2 Estimation results of the GWML model

变量名称	系数估计			T检验值		
	最小	最大	平均	最小	最大	平均
“保持不变”						
教育地丰富度因子	0.152	0.536	0.204	-14.51	11.654	-4.067
商服用地丰富度因子	-0.732	-0.428	-0.547	-12.537	16.751	-0.189
飞机场可达性	0.046	1.581	0.167	-3.045	15.698	9.263
CBD可达性	-0.735	-0.076	-0.137	-8.082	4.613	-2.069
“居住用地”						
居住用地丰富度因子	0.262	1.325	0.622	-4.711	6.685	2.728
商服用地丰富度因子	-1.942	-0.274	-0.444	-16.772	3.148	-6.896
飞机场可达性	-0.535	-0.281	-0.377	-8.373	9.943	5.284
CBD可达性	0.015	0.843	0.059	-7.781	14.752	1.746
公交可达性	-1.232	-0.113	-0.252	-15.268	12.546	1.343
“工业用地”						
工业用地丰富度因子	0.127	1.586	0.273	-14.983	11.794	5.263
火车站或交通枢纽可达性	1.681	3.628	2.875	-6.672	13.553	4.234
CBD可达性	-0.252	-0.631	-0.502	-5.466	6.266	-2.144
工业区化	0.043	0.463	0.268	-0.427	12.682	8.116
“商服用地”						
常数	-3.203	-0.601	-3.234	-6.118	11.418	5.521
工业用地丰富度因子	0.212	1.438	0.286	-0.625	14.516	4.719
商服用地丰富度因子	0.212	0.53	0.319	-2.518	10.551	4.236
火车站或交通枢纽	-2.514	-0.026	-0.158	-7.164	4.882	-0.674
飞机场可达性	0.564	1.838	0.163	-8.276	4.503	-1.756
CBD可达性	0.552	1.675	0.249	-5.256	7.451	0.235
公交可达性	1.143	4.542	1.568	-2.154	9.748	4.156
“教育用地”						
常数	-4.264	-0.214	-2.166	-16.792	7.086	-8.625
教育地丰富度因子	0.213	0.486	0.107	-9.509	20.542	8.774
商服用地丰富度因子	-4.996	-0.519	-2.652	-24.161	1.573	-11.683
CBD可达性	1.439	3.772	2.253	-10.241	16.471	3.461
AIC _c :	最小: 478.576496; 最大: 7069.64671; 平均值: 3568.85733					
Likelihood Ratio Index:	最小: 0.818614621; 最大: 0.83721225; 平均值: 0.82976309					

用模型相比，基于GWML模型的CA土地利用模型模拟结果与图2所示的实际土地利用变化更为相似。基于GWML模型的CA土地利用模型更能反映局部土地利用格局，如图2的放大区域为例，其中圆圈标记了4个实际存在的商服用地集聚。基于MNL模型的CA土地利用模型只能识别4个商服用地集聚中的一个（图4a）。基于GWML模型的CA土地利用模型则将识别的商服用地集聚区提高到3个（图4b）。

表3总结了基于MNL的CA模型及GWML模型的CA模型的解释精度及模拟精度。模型的解释精度指所有土地利用状态转变可以被解释的元胞数量占全部元胞数量的比率。模

拟精度指模拟结果与实际土地利用变化相一致的元胞数占全部元胞数的比率。因此, 解释精度表征了模型的数据解释能力。而模拟精度用以测度CA模型的模拟能力, 其不但决定于其所使用的统计模型, 而且还取决于其所使用的模拟机制。结果表明, 基于GWML模型的CA模型的解释精度优于基于MNL模型的CA模型。71%的元胞的土地利用变化可以由基于GWML模型的CA模型解释, 而基于传统的MNL的CA模型只能解释52%的元胞的土地利用变化。GWML模型同样提高了CA模型的模拟精度。结果显示, 基于GWML模型的CA模型的模拟精度为30.7%, 大于基于传统MNL的CA模型的模拟精度(17.5%)。

GWML模型和传统MNL的解释精度均大于其对应的CA模型的模拟精度, 说明CA模型的土地利用变化需求分配方法是决定模拟精度的一个重要因素。本研究中土地利用变化需求被分配到那些具有最大联合转换概率的元胞上。例如, 一个元胞的状态转换为土地利用类型A可以被模型解释, 只要其转化为土地利用A的概率大于转化为其他类型的概率。但如果其转化为土地利用类型A的联合概率小于其他可开发元胞转换为土地利用类型A的联合概率, CA

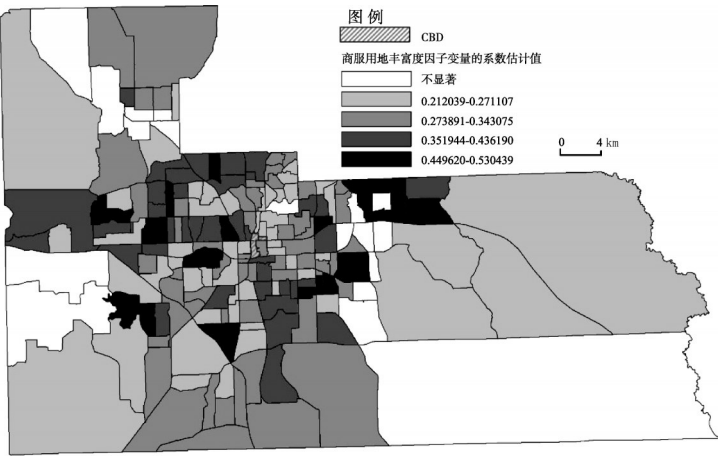


图3 商服用地丰富度因子变量对商服用地因变量的系数估计分布
Fig. 3 Spatial distribution of EF_ComSer for commercial/service land use

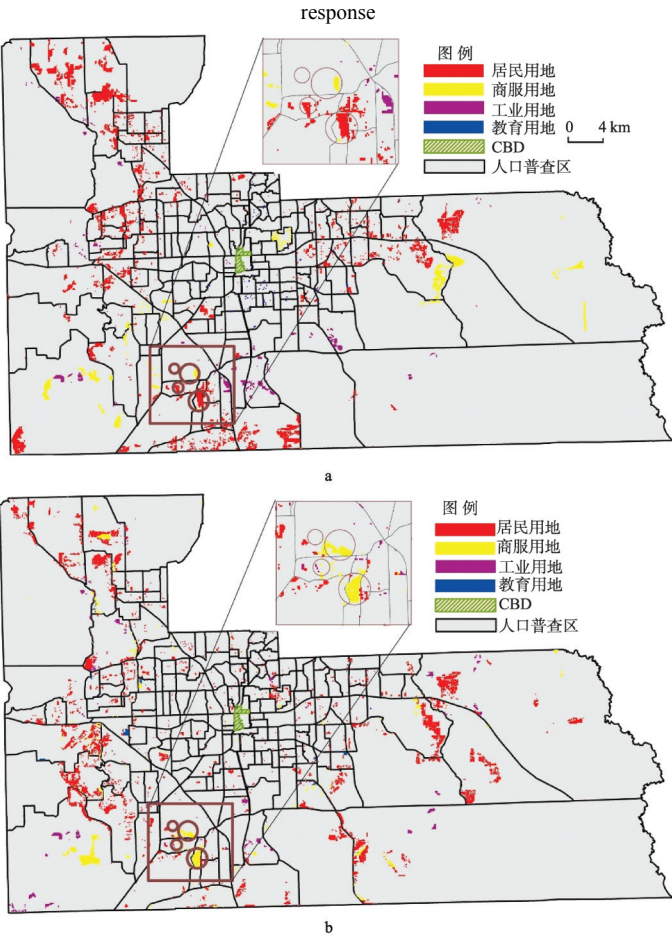


图4 模拟结果
(a) 基于MNL模型的CA模型; (b) 基于GWML模型的CA模型
Fig. 4 Simulation result. (a) Simulation result of MNL model based CA model;
(b) Simulation result of GWML model based CA model

表 3 CA 模型的数据解释精度及模拟精度
Tab. 3 Summary of data explanation accuracy and simulation accuracy

元胞状态转化	元胞数量	基于 MNL 模型的 CA 模型		基于 GWML 模型的 CA 模型	
		解释精度(%)	模拟精度(%)	解释精度(%)	模拟精度(%)
“保持不变”	81732	61.63	91.25	71.64	92.72
“居住用地”	6930	62.15	20.27	77.33	32.48
“商服用地”	1436	24.72	10.79	68.11	29.53
“工业用地”	756	20.9	7.41	51.98	14.68
“教育用地”	174	8.05	8.05	43.1	41.38
合计:					
不包括“保持不变”	9296	52	17.53	73.2	30.74
全部	91028	60.64	83.72	71.8	86.39

模型则不能将土地利用类型 A 赋予该元胞。因此，为了提高模拟精度，未来的研究需要将城市不同的增长模式整合到模拟分配过程中，如 Jantz 等^[36]提出的中心蔓延式增长、边界增长及道路影响下的增长等城市增长模式。

未变化的土地是影响度量模拟精度的重要因素。当用地类型未变化的土地被纳入计算时，模拟精度将很高。例如，在 Jantz 等^[36]的研究中，当未发生变化的区域（占整个研究区的 88%）纳入模拟精度计算时，模拟精度从 19%提高到了 93.1%。在本研究中，当所有的土地利用需求被分配到指定元胞后，模拟过程结束，剩余的元胞则设为选择默认选项，即保持原土地利用类型。在最坏情况下，如果将所有实际土地利用类型未改变的元胞（大约 10.2%）错误地预测为发生了变化，剩余的元胞仍会被默认地的赋予正确的选项（“保持不变”），从而使得模拟精度仍可达到 79.6%。在本研究中，当实际土地利用未发生改变的元胞纳入模拟精度计算时，基于传统 MNL 模型的 CA 模型的模拟精度由 17.5%提高到 83.7%，而基于 GWML 模型的 CA 模型的模拟精度由 30.7%提高到 86.4%。

3 结论

本研究将地理加权多项 Logit 模型应用于 CA 土地利用模型中，以美国佛罗里达州橙县的土地利用变化为例，结果显示，驱动因素对土地利用变化的作用在空间上存在显著的空间异质性。GWML 模型通过模型局部化，能够反映驱动因素的作用在空间中的变化，较传统的多项 logit（MNL）模型具有更高的数据拟合度。基于 GWML 模型的 CA 模型较基于 MNL 模型的 CA 模型更能体现土地利用变化在局部的变化模式，因而具有更高的模拟精度。

参考文献(References)

[1] Lau K H, Kam B H. A cellular automata model for urban land-use simulation. *Environment and Planning B: Planning and Design*, 2005, 32(2): 247-263.

[2] 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, 27(1): 19-27.

[3] Verburg P H, van Eck J R R, de Nijs T C M et al. Determinants of land-use change patterns in the Netherlands. *Environment and Planning B: Planning and Design*, 2004, 31(1): 125-150.

- [4] Tobler W R. Cellular geography. In: Gale S, Olsson G. *Philosophy in Geography*. Reidel: Dordrecht, 1979. 379-386.
- [5] Pan Y, Roth A, Yu Z et al. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 2010, 34(5): 400-408.
- [6] 柯新利, 邓祥征, 何书金. 地理元胞自动机模型的尺度敏感性及其原因. *地理研究*, 2010, 29(5): 863-872.
- [7] 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(8): 1175-1199.
- [8] Barredo J I, Kasanko M, McCormick N et al. Modelling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 2003, 64(3): 145-160.
- [9] Huang B, Zhang L, Wu B. Spatiotemporal analysis of rural-urban land conversion. *International Journal of Geographical Information Science*, 2009, 23(3): 379-398.
- [10] Hu Z, Lo C. Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems*, 2007, 31(6): 667-688.
- [11] Wu D, Liu J, Zhang G et al. Incorporating spatial autocorrelation into cellular automata model: An application to the dynamics of Chinese tamarisk (*Tamarix chinensis* Lour.). *Ecological Modelling*, 2009, 220(24): 3490-3498.
- [12] Fotheringham A S, Charlton M E, Brunsdon C. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 1998, 30(11): 1905-1927.
- [13] Brunsdon C, Fotheringham A S, Charlton M E. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 1999, 39(3): 497-524.
- [14] Tu J, Xia Z. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of The Total Environment*, 2008, 407(1): 358-378.
- [15] Gilbert A, Chakraborty J. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social science research*, 2011, 40(1): 273-286.
- [16] Schmidtlein M C, Fincha C, Cutter S L. Disaster declarations and major hazard occurrences in the United States. *The Professional Geographer*, 2008, 60(1): 1-14.
- [17] Clement F, Orange D, Williams M et al. Drivers of afforestation in Northern Vietnam: assessing local variations using geographically weighted regression. *Applied Geography*, 2009, 29(4): 561-576.
- [18] Szymanowski M, Kryza M. Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wroclaw (SW Poland). *Procedia Environmental Sciences*, 2011, 3: 87-92.
- [19] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatiotemporal variation in house prices. *International Journal of Geographical Information Science*, 2010, 24(3): 383-401.
- [20] Munroe D K, Müller D. Issues in multi-scale, spatial evaluation of statistical LUCC models: Examples from western Honduras and the Central Highlands of Vietnam. *Land Use Policy*, 2007, 24(3): 521-530.
- [21] Yüzer M A. Growth estimations in settlement planning using a land use cellular automata model (LUCAM). *European Planning Studies*, 2004, 12(4): 551-560.
- [22] Waddell P, Borning A, Noth M et al. Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, 2003, 3(1): 43-67.
- [23] 聂婷, 肖荣波, 王国恩, 等. 基于 Logistic 回归的 CA 模型改进方法: 以广州市为例. *地理研究*, 2010, 29(10): 1909-1918.
- [24] Zhou B, Kockelman K. Neighborhood impacts on land use change: A multinomial Logit model of spatial relationships. *The Annals of Regional Science*, 2008, 42(2): 321-340.
- [25] 梁友嘉, 徐中民, 钟方雷. 基于 SD 和 CLUE-S 模型的张掖市甘州区土地利用情景分析. *地理研究*, 2011, 30(3): 564-576.
- [26] 吴桂平, 曾永年, 冯学智, 等. CLUE-S 模型的改进与土地利用变化动态模拟: 以张家界市永定区为例. *地理研究*, 2010, 29(3): 460-470.
- [27] McMillen D P, McDonald J F. Land use before zoning: The case of 1920's Chicago. *Regional Science and Urban Economics*, 1999, 29(4): 473-489.
- [28] Luo J, Kanala N K. Modeling urban growth with geographically weighted multinomial logistic regression. In: Liu L, Li X, Liu K et al. *Proceedings of Geoinformatics 2008 and Joint Conference on GIS and Built Environment: The*

- Built Environment and Its Dynamics. Guangzhou, China: SPIE, 2008, 71440: 71440M-71440M-11.
- [29] Fotheringham A S, Brunsdon C, Charlton M E. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester, UK: John Wiley and Sons Ltd., 2002.
- [30] Hurvich C M, Simonoff J S, Tsai C L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1998, 60(2): 271-293.
- [31] Guo J Y, Bhat C R. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 2007, 15(1): 31-45.
- [32] Owen P D. General-to-specific modelling using PcGets. *Journal of Economic Surveys*, 2003, 17(4): 609-628.
- [33] Shaver J M. Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of management*, 2005, 31(3): 330-353.
- [34] Verburg P H, de Nijs T C M, Ritsema van Eck J et al. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 2004, 28(6): 667-690.
- [35] Waddell P, Ulfarsson G F, Franklin J P et al. Incorporating land use in metropolitan transportation planning. *Transportation Research Part A: Policy and Practice*, 2007, 41(5): 382-410.
- [36] Jantz C A, Goetz S J, Shelley M K. Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the baltimore- washington metropolitan area. *Environment and Planning B: Planning and Design*, 2004, 31(2): 251-271.

A cellular automata land use model based on localized transition rules

YU Linjun^{1,2}, SUN Danfeng², PENG Zhongren³, LI Hong⁴

(1. College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China;

2. Institute of Remote Sensing and Digital Earth, CAS, Beijing 100101, China;

3. Department of Urban and Regional Planning, University of Florida, Gainesville, FL 32611-5706, USA;

4. Institute of Agriculture Integration Development, Beijing Academy of Agriculture and Forestry Science, Beijing 100097, China)

Abstract: Spatial variations in the relationships between driving forces and land use change are ignored in traditional cellular automata (CA) based land use models which are based on assumption of spatial homogeneity. To address this issue, a geographically weighted multinomial logit (GWML) mode, developed by integrating a locally weighted regression procedure in the estimation of the multinomial logit (MNL) model, is applied in a CA land use model using a case study in Orange County, FL, USA. The results show that the GWML has higher data explanation power than conventional MNL model, and the GWML based CA land use model results in a higher simulation accuracy than the MNL based CA model because local land use change patterns are identified by the locally created transition rules. Although a region in the USA was used as a case study, the conclusion is still meaningful for future studies using regions of China as cases.

Key words: land use change; cellular automata model; spatial heterogeneity; multinomial logit model; geographically weighted multinomial logit model