

广州市番禺区土地利用及空间距离 数据集（1999，2008）

张亦汉，乔纪纲*

广东财经大学 地理与旅游学院，广州 510320

摘要：研究表明，城市扩张往往与一系列的空间距离变量相关，因此本文利用 ArcGIS 获取了离区中心距离，离地铁线的距离，离地铁站的距离，离普通道路距离和离镇中心的距离。通过这些数据可构建城市扩张 CA 模型，并从采样、邻域结构和微观元胞尺度等方面研究 CA 模型的敏感性。地理空间数据云（<http://www.gscloud.cn/>）提供了免费的 Landsat 遥感数据，本文选择 1999 年和 2008 年的数据，对其进行正射校正和大气校正，再对其进行监督分类并重采样得到 1999 年和 2008 年的土地利用分类数据，作为城市扩张 CA 模型的起始影像和终止影像。利用国家地理信息公共服务平台（<http://www.tianditu.gov.cn/>）提供的地图服务，对研究区的地铁站、地铁线和道路进行数字化，并获取区中心和镇中心，形成矢量数据集。根据该矢量数据集进行欧氏距离分析，得出研究区的空间变量。为了方便比较分析，对所有的空间变量均进行归一化处理，得出城市扩张 CA 模型的空间变量数据集。数据存储格式包括.txt、.img、.xml 和.rdd，由 24 个文件组成，文件大小为 1.8 MB。该数据集的分析研究成果已发表在《遥感学报》2018 年第 22 卷第 6 期。

关键词：城市扩张；元胞自动机；参数敏感性；遥感学报

DOI: 10.3974/geodp.2019.03.09

1 前言

城市用地的扩张是土地利用变化中最受关注的部分，也是全球变化研究的核心和热点^[1-2]。部分学者关注城市扩张的空间格局，也有部分学者关注城市扩张的过程^[3-4]。城市扩张的过程往往以遥感影像等栅格数据作为数据源，这种栅格数据结构与元胞自动机（Cellular Automaton, CA）模型的离散空间的特性是相互一致的，因此有不少学者运用 CA 模型研究城市扩张的过程^[5-7]。也有学者提出了不少新的模型，但是当模型应用到实际的城市扩张中往往需要进行敏感性研究，主要包括转换规则，空间尺度和邻域空间方面的分析^[8]。CA 模型的敏感性研究中转换规则是核心，而转换规则中的参数值大小对城市扩张模拟的结果重要的影响^[9]。模型的空间尺度特征可以通过元胞大小（即栅格数据的空间分

收稿日期：2019-04-08；修订日期：2019-06-10；出版日期：2019-06-25

基金项目：广东省教育厅（2016KTSCX045）；广东省自然科学基金（2014A030313605, 2016A030313806）

*通讯作者：乔纪纲 F-7553-2019, 广东财经大学地理与旅游学院, qjg821@263.com

*数据引用方式：[1] 张亦汉，乔纪纲. 广州市番禺区土地利用及空间距离参数数据集(1999, 2008) [J]. 全球变化数据学报, 2019, 3(3): 284-289. DOI: 10.3974/geodp.2019.03.09.

[2] 张亦汉，乔纪纲. 广州市番禺区土地利用及空间距离参数数据集(1999, 2008) [DB/OL]. 全球变化科学研究数据出版系统, DOI: 0.3974/geodb.2019.03.10.V1.

辨率)来反映^[8]。邻域元胞对中心元胞的作用力,往往可以通过邻域大小及形状来体现^[10]。因此,本文通过定量获取采样,分析邻域结构,研究微观元胞尺度等以构建 CA 敏感性分析数据集。

2 数据集元数据简介

广州市番禺区土地利用及空间距离参数数据集（1999，2008）^[11]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

表 1 广州市番禺区土地利用及空间距离参数数据集(1999, 2008)元数据简表

条 目	描 述
数据集名称	广州市番禺区土地利用及空间距离参数数据集（1999, 2008）
数据集短名	ParamSensUrbanCAmodel
作者信息	张亦汉 F-6808-2019, 广东财经大学地理与旅游学院, zyh4184@163.com 乔纪纲 F-7553-2019, 广东财经大学地理与旅游学院, qjg821@263.com
地理区域	广州市番禺区
数据年代	1999 和 2008
时间分辨率	年
空间分辨率	50 m
数据格式	.txt、.img、.xml、.rrd
数据量	1.8 MB（压缩后）
数据集组成	数据集由 2 个遥感分类影像和 5 个空间变量构成, 主要包括如下: 1、2 个遥感分类影像。包括 1999 和 2008 年的遥感 TM 影像分类结果。采用 ENVI 软件对原始影像进行大气纠正和几何纠正, 再通过查找同名点对影像进行配准。运用监督分类将全区分为城市用地、未利用地、耕地、林地、基塘、水体和园地。 2、空间距离变量。运用 ArcGIS 中的欧氏距离方法, 并根据区中心、地铁线、地铁站、普通道路和镇中心等数据, 获取了离区中心距离, 离地铁线的距离, 离地铁站的距离, 离(普通)道路距离和离镇中心的距离。
基金项目	广东省教育厅特色创新项目（2016KTSCX045）; 广东省自然科学基金（2014A030313605, 2016A030313806）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、实体数据（中英文）和通过《全球变化数据学报》（中英文）发表的数据论文。其共享政策如下:（1）“数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载;（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源;（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报》（中英文）编辑部签署书面协议, 获得许可;（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源 ^[12] 。
数据和论文检索系统	DOI, DCI, CSCD, WDS/ISC, GEOSS, China GEOSS

3 数据研发方法

3.1 研究区描述

本文以珠江三角洲的广州市番禺区作为研究区, 处于粤港澳大湾区的中心位置, 其总

面积为 529.9 km²。它位于广州中南部，东临珠江，与东莞市隔江相望；南临南沙区，与南沙自贸区相邻；西接佛山市和中山市；北近广州核心市区海珠区。番禺区管辖 10 个街道和 6 个镇。番禺区由于经济发达，吸引了不少人口居。根据广州市统计年鉴，截至 2017 年末，番禺区常住人口达到 171.93 万人。番禺区由于经济和人口不断增长，番禺区的城市扩张速度也快，其土地利用变化相对复杂。根据土地利用分类结果，在 1999 年，番禺城市用地面积为 134.6 km²；到 2008 年，用地面积增长到 255.60 km²。城市扩张的年增长率高达 6.8%。统计数据显示，整个中国城市的年扩张率仅为 3%，番禺区的扩张率明显高于全国平均水平，因此选择这个变化相对比较剧烈的典型区域作为研究区。

3.2 技术路线

地理空间数据云（<http://www.gscloud.cn/>）提供了免费的 LANDSAT 遥感数据，本文选择了 1999 年和 2008 年的数据，对其进行正射校正和大气校正，再对其进行监督分类得出 1999 年和 2008 年的土地利用分类数据^[13]，作为城市扩张 CA 模型的起始影像和终止影像。利用国家地理信息公共服务平台（<http://www.tianditu.gov.cn/>）提供的地图服务，对研究区的地铁站、地铁线和道路进行数字化，并获取区中心和镇中心，形成矢量数据集。根据该矢量数据集进行欧氏距离分析，得出研究区的空间变量^[3-4]。为了方便比较分析，对所有的空间变量均进行归一化处理，得出城市扩张 CA 模型的空间变量数据集。生成数据集的流程图如图 1 所示。

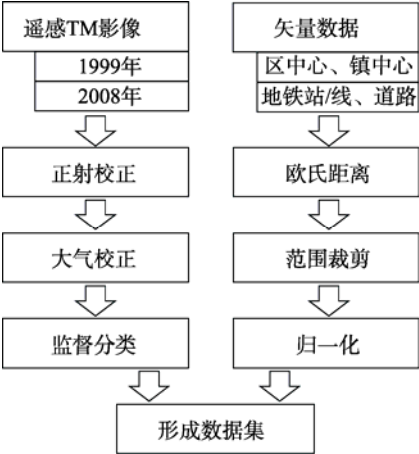


图 1 城市扩张 CA 模型参数敏感性分析数据集研发技术路线图

4 数据结果与验证

4.1 数据结果组成

本数据集包含 1999 和 2008 年的遥感土地利用分类数据，其中 1999 年的土地利用数据用于城市扩张 CA 模拟，而 2008 年的土地利用分类数据用于验证。数据集还包括离区中心距离，离地铁线的距离，离地铁站的距离，离（普通）道路距离和离镇中心的距离等空间距离变量数据（图 2）。

4.2 数据结果验证

首先，对数据集的采样比例进行分析（表 2）。为了方便对比，本文均采用蒙特卡罗随机采样，采样比例分别为 1%、3%、5%、10%、15%、20%、25%、30%、40%和 50%。实验发现随着采样比例增加，获取的逻辑回归权重值（无量纲）逐渐稳定（图 3）。采用更多的样本进行训练，会得到相对稳定权重，但是训练的运算量也是相对较大的。若使用较小比例的样本获取权重，将会得到误差较大的权重。

其次，分析了数据集在同一空间尺度下，不同邻域结构不同邻域大小的模拟结果（表 3）。使用 3×3 摩尔邻域，其模拟结果的精度可达到 80.57%，当运用更大的 5×5 扩展邻域时，

模拟结果的精度有所降低，精度为 80.48%。当采用 7×7 和 9×9 邻域时，其模拟精度有所提高，分别为 80.51%和 80.52%。总体上，使用冯诺依曼邻域的模拟结果精度比使用摩尔邻域的精度略低。在 3×3、5×5、7×7 和 9×9 的扩展冯诺依曼邻域中，其精度分别 80.43%、80.47%、80.50%和 80.51%，并随着邻域元胞个数的增加而增加。

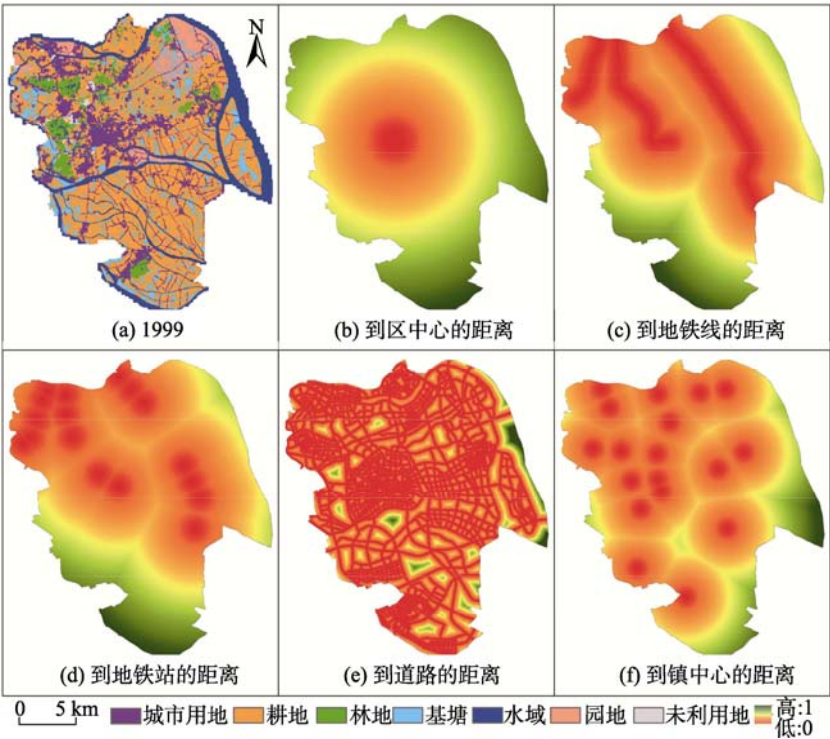


图 2 城市扩张 CA 模型的基础数据^[14]

表 2 数据集的采样检验^[14]

比例 (%)	区中心	地铁线	地铁站	道路	镇中心	精度(%)
1	-0.90	-3.58	2.23	-21.19	-2.99	80.44
3	-1.31	-4.89	3.34	-24.07	-2.17	80.65
5	-1.45	-6.93	5.57	-20.78	-2.44	80.59
10	-1.42	-5.14	3.69	-21.99	-2.41	80.64
15	-1.40	-5.14	3.83	-20.63	-2.66	80.66
20	-1.35	-5.68	4.49	-21.32	-2.72	80.62
25	-1.38	-5.43	4.04	-21.14	-2.37	80.66
30	-1.34	-5.28	3.89	-19.05	-2.64	80.67
40	-1.37	-5.35	4.04	-20.18	-2.59	80.70
50	-1.36	-5.10	3.78	-19.64	-2.69	80.64

再次，分析了数据集的空间尺度敏感性，采用总精度、Kappa 系数和城市用地模拟精度进行点对点对比分析，以此制作出模拟精度与空间尺度之间的关系图（图 3）。从总体上看，不论是总精度、Kappa 系数，还是城市用地的模拟精度都是随着空间尺度的增加而降

低。其中，空间尺度从 50 m 增加到 200 m 时，其总精度、Kappa 系数和城市用地的模拟精度变化较小，分别从 80.57%降至 79.98%、56.11%降至 55.04%、70.58%降至 70.18%。但是当空间尺度从 200 m 增加到 500 m 时，各个指标降幅明显，总精度从 79.98%降至 76.36%、Kappa 系数从 55.04%降至 47.45%，和城市用地的模拟精度从 70.18%降至 64.94%。

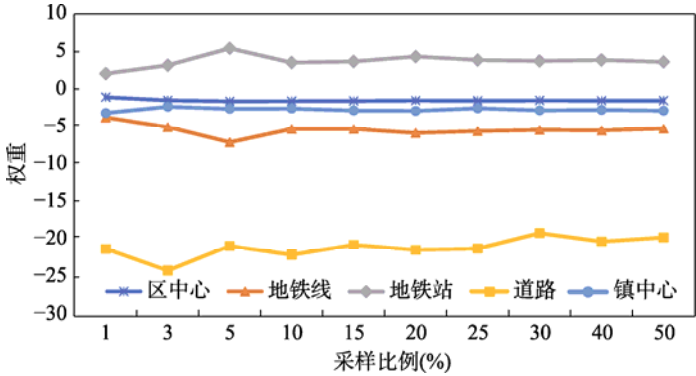


图 3 采样比例与权重的关系

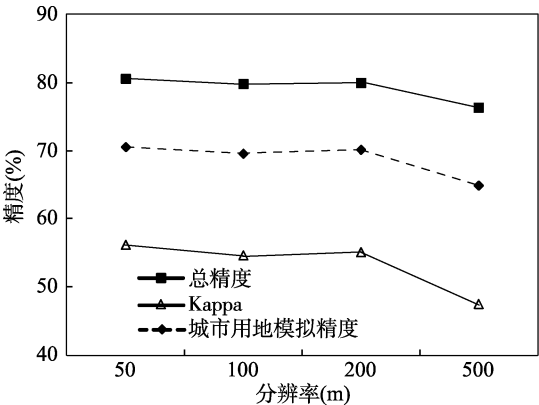


图 4 不同分辨率下的 CA 模拟精度^[14]

表 3 数据集的邻域结构检验^[14]

邻域大小	精度（%）	
	摩尔邻域	冯诺依曼
3×3	80.57	80.43
5×5	80.48	80.47
7×7	80.51	80.50
9×9	80.52	80.51

最后，为了更好地理解不同空同尺度下微观层次的检验，本文统计了 1999 年和 2008 年遥感影像在 50 m×50 m、100 m×100 m、200 m×200 m 和 500 m×500 m 空间分辨率下摩尔邻域的城市元胞占比情况（表 4）。通过统计发现，不论是 1999 年还是 2008 年，随着空间

表 4 数据集的微观层次检验-不同分辨率下摩尔邻域的城市元胞占比^[14]

个数	1999 年				2008 年			
	50 m	100 m	200 m	500 m	50 m	100 m	200 m	500 m
0	69.1	59.5	54.5	44.6	50.5	40.1	32.6	18.7
1	4.0	9.2	15.9	23.2	4.5	8.9	14.3	20.9
2	4.1	8.7	9.2	12.8	4.5	8.9	11.2	16.2
3	5.0	6.9	6.1	7.0	6.0	8.6	8.6	13.5
4	3.7	4.2	4.2	5.2	4.7	6.3	6.6	9.1
5	3.3	3.2	3.2	3.5	4.9	5.7	5.8	8.8
6	2.2	2.5	2.4	2.0	3.6	4.6	5.6	6.2
7	1.8	2.1	2.0	1.1	3.1	4.6	5.8	4.3
8	6.9	3.8	2.5	0.7	18.2	12.2	9.5	2.2

尺度的增加, 高城市元胞占比的值均在下降。这表明通过降尺度, 城市发展密度正在减少, 即由高密度向低密度转换。从表中也可以看出, 邻域城市元胞个数为 1, 2, 3 和 4 的, 通过降尺度, 其比例不断增加。而邻域城市元胞为 6, 7 和 8 的几乎均有下降的趋势

5 讨论和总结

为了验证城市扩张 CA 模型敏感性, 获取了一个迅速城市化区域-广州市番禺区的相关数据, 并构建了空间分辨率为 50 m 的数据集。使用本数据集, 可以方便地分析采样比例、邻域结构和空间分辨率对城市扩张 CA 模型敏感性。数据集的验证也表明本数据集对于获取 CA 模型权重(参数)具有较好的稳定性; 对于验证在同一空间尺度下, 不同邻域结构不同邻域大小的模拟结果也有非常有效的; 在探讨空间尺度敏感性方面, 也能很好地分析出模拟精度与空间尺度之间的关系与不同空间尺度下的微观层次的情况。

本数据集旨在为城市扩张 CA 模型或者是土地利用变化 CA 模型提供基础的运行数据集, 为研究人员分析和理解 CA 模型提供必要的敏感性分析数据。下一步研究工作将提供更长的时间序列和更多的空间分辨率以研究 CA 模型的预测能力。

作者分工: 乔纪纲对数据集的开发做了总体设计, 并采集和处理了遥感 TM 数据; 张亦汉设计了模型和算法、做了数据验证, 并撰写了数据论文等。

参考文献

- [1] 刘纪远, 刘明亮, 庄大方等. 中国近期土地利用变化的空间格局分析[J]. 中国科学:地球科学, 2002, 32(12): 1031-1040.
- [2] 王雷, 李丛丛, 应清等. 中国 1990-2010 年城市扩张卫星遥感制图[J]. 科学通报, 2012, 57(16): 1388-1399.
- [3] Li, X., Yeh, A. G. O. Data mining of cellular automata's transition rule [J]. *International Journal of Geographical Information Science*, 2004, 18(8): 723-744.
- [4] 张亦汉, 黎夏, 刘小平等. 基于数据同化的元胞自动机[J]. 遥感学报, 2011, 15(3): 475-491.
- [5] White, R., Engelen, G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns [J]. *Environment and Planning A*, 1993, 25(8): 1175-1199.
- [6] Wu, F. Calibration of stochastic cellular automata: the application to rural-urban land conversions [J]. *International Journal of Geographical Information Science*, 2002, 16(8): 795-818.
- [7] Li, X., Yeh, A. G. O. Neural-network-based cellular automata for simulating multiple land use changes using GIS [J]. *International Journal of Geographical Information Science*, 2002, 16(4): 323-343.
- [8] 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[J]. 地理研究, 2010, 29(5): 863-872.
- [9] Liu, X., Ma, L., Li, X., et al. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata [J]. *International Journal of Geographical Information Science*, 2014, 28(1): 148-163.
- [10] 吴浩, 周璐, 史文中等. 基于正交试验设计的土地利用变化元胞自动机模拟过程的尺度敏感性分析[J]. 地理科学, 2013, 33(10): 1252-1258.
- [11] 张亦汉, 乔纪纲. 广州市番禺区土地利用及空间距离参数数据集(1999, 2008) [DB/OL]. 全球变化科学研究数据出版系统, 2019. DOI: 10.3974/geodb.2019.03.10.V1.
- [12] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策 [OL]. DOI: 10.3974/dp.policy.2014.05 (2017 年更新).
- [13] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 166-208
- [14] 张亦汉, 乔纪纲, 刘婉华. 城市扩张 CA 模型的参数敏感性分析[J]. 遥感学报, 2018, 22(6): 1051-1059.