

# Beijing City Lab

Long Y, 2013, Applied urban modeling and its opportunities in the “big data” era.  
Beijing City Lab. Working paper # 13

# 大数据时代的城市模型研究及其机遇

龙瀛，北京市城市规划设计研究院，北京 100045

（《时空行为研究前沿》章节，修改稿，应编者北京大学柴彦威教授约稿提交）

## 1、城市模型发展历程

纵观城市科学的发展历史，从对城市现象的记载、描述，到对其进行归纳、总结，再到对城市事物之间的关系描述，最后发展到用系统的观点看待城市，其发展历程经历了一个从定性到定量的过程。现阶段，定量化程度已经越来越成为衡量该学科发展程度的标志。城市模型”（Urban Model）是在对城市系统进行抽象和概化的基础上，对城市空间现象与过程的抽象数学表达，是理解城市空间现象变化、对城市系统进行科学管理和规划的重要工具，可以为城市政策的执行及城市规划方案的制定和评估提供可行的技术支持。

城市模型研究始于 20 世纪初期，20 世纪初到 50 年代中期是城市模型发展的初级阶段，主要经历了一般概念模型、数学（或分析）模型和计算机模拟模型等几个阶段（图 1）。从 1990 年代开始，随着计算机硬件技术和 GIS 技术的日益成熟，GIS 在城市模型研究中的应用及其与城市模型的集成已经成为侧重于计算机模拟的城市模型发展的重要趋势。根据城市模型是否具有时间维，可以把城市模型分为静态模型和动态模型两大类。静态城市模型不考虑时间维，主要分为两类，一是城市统计模型，二是空间相互作用模型（如劳瑞模型）。从建模方法看，应该分为自上而下和自下而上两种建模方法，前者主要是以系统动力学（System Dynamics）为代表的城市模型，一般是基于微分方程的形式，往往从宏观的空间尺度出发，研究对象也往往是对城市居住区、商业区等的机械划分及其相互作用，或区位选择，无法反映造成城市动态性、自组织性和突变性等城市微观结构和理性人的个体行为；后者是以离散动力学（Discrete Dynamics）如元胞自动机（Cellular Automata, CA）、多主体系统（Multi-agent Systems）、分形理论（Fractal）和自组织理论（Self-organization）等为代表的城市模型。目前 MAS 为代表的动态城市模型是研究的重点和近期的趋势。

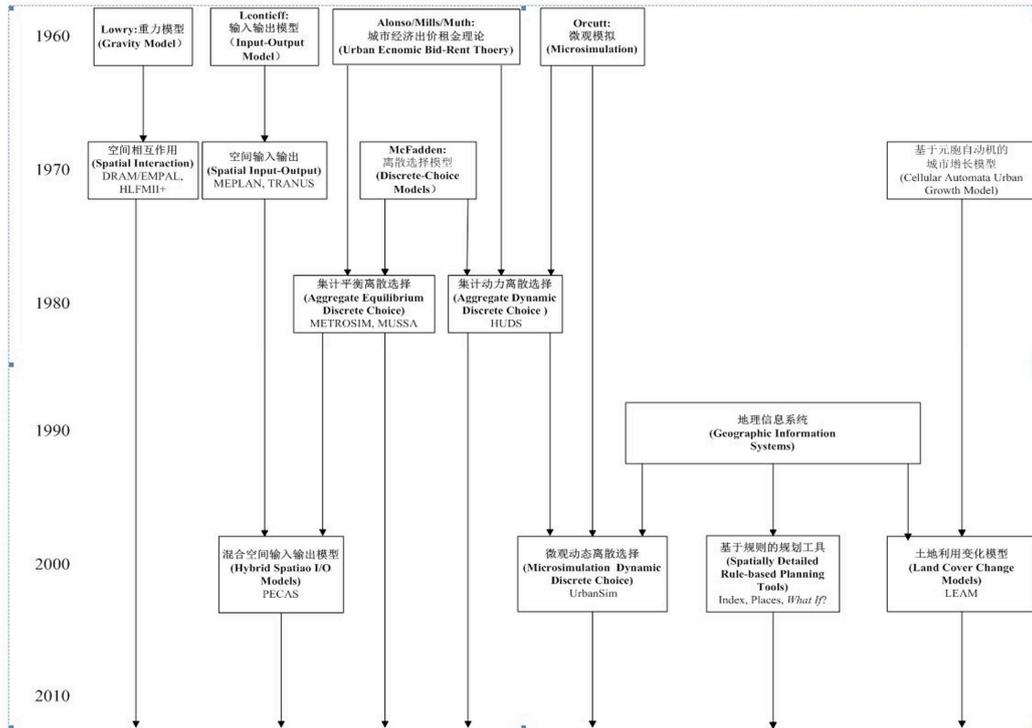


图 1 城市模型发展历程一览<sup>1</sup>

## 2、城市模型的基本分类

动态城市模型从不同角度可以有不同的分类。

从建模的方法看，常用的方法有基于空间相互作用理论（Spatial Interaction）的重力模型<sup>2</sup>（Gravity Model）、最大熵理论模型（Entropy Maximizing），来自经济学的 Alonso/Mills/Muth 地租理论（Rent Models）、离散选择模型（Discrete Choice Model）、空间投入产出模型（Spatial Input-output Model）、回归分析（Regression），来自复杂科学的元胞自动机（CA）、基于个体建模（Agent-based Modelling, ABM）<sup>3</sup>，以及微观模拟（Microsimulation Model, MSM）和地理信息系统（GIS）等技术（Pagliara 和 Wilson, 2010）。

从模型应用的具体领域看，有区域模型、城市土地模型、土地使用与交通模型、土地使用—交通—环境模型等（郑思齐等, 2010）。值得一提的是土地使用与交通模型，它是从城市交通模型发展而来的，在大城市遇到交通导致的严重的城市问题背景下，研究者开始着手研究交通对土地使用的反作用机理，从而开始了土地使用与交通模型的研究（如 Cube Land），基于传统的交通模型新增的土地模型研究内容一般包括居住区位选择、企业选址、房地产开发和土地开发等。目前，土地使用与交通模型是动态城市模型的主要存在形式之一，近几年，部分西方国家的代表城市或区域完成了一些用于实践的土地使用与交通整合模型。

从模型的空间尺度上看，又可分为宏观模型<sup>4</sup>和微观模型，宏观模型（或分区模型）

<sup>1</sup> 本图参考了 Paul Waddell 关于 UrbanSim 的介绍材料（Dynamic Microsimulation: UrbanSim, Webinar 5 of 8-part TMIP, Webinar series on land use forecasting methods）。

<sup>2</sup> 或劳瑞模型。

<sup>3</sup> 也有研究将 ABM 模型称为多主体系统或多智能体系统（Multi-agent System, MAS），本文沿用 ABM 的说法。

<sup>4</sup> 或集计模型（Aggregated Models）。

的研究尺度是地理网格<sup>5</sup> (Grid) 或小区 (Zone), 一般小区可以是交通分析小区 (Traffic Analysis Zone, TAZ), 也可以是统计小区 (Census Tract), 在这种情况下, 城市活动主体一般选用小区内的居民、家庭或企业的统计特征, 即以一类活动主体作为分析对象, 而不是个体。而微观模型, 一般基本空间单元较小, 如街区、地块或建筑, 相应的, 城市活动主体一般对应居民、家庭和企业的个体, 其原理与方法更加清晰直观。

### 3、典型城市模型及其基本情况

目前典型的城市模型基本信息如表 1 所示。这些模型研究城市土地使用为主, 部分结合了交通模块形成了土地使用与交通模型 (“名称”列中粗体的)。模拟的基本空间单元多为小区和网格, 仅有 UrbanSim、ILUTE 和 Agent iCity 属于微观模型, 其中 UrbanSim 可以用于多尺度的模拟。据项目组了解, 目前 Alex Anas 的研究组正在洛杉矶区域使用 Relu-Tran 建立公共政策分析的虚拟实验室 (Virtual Co-Laboratory for Policy Analysis in the Greater L.A. Region), 属于分区尺度。MEPLAN 模型目前由英国 WSP 公司在很多城市开展了实践应用。

表 1 典型城市模型一览

序号	名称 <sup>6</sup>	所在国家	研究尺度 <sup>7</sup>	开发年份	代表性开发人员/机构	主要方法	时间基础	代表性文献
1	POLIS	美国	小区	1960 年代	旧金山湾区政府协会	空间相互作用、离散选择	静态	Association of Bay Area Governments, 2009
2	DRAM/ EMPAL	美国	小区	1970 年代	Stephen H.Putman	空间相互作用、离散选择	静态平衡	普特曼 (Putman), 1995
3	<b>TRANUS</b>	委内瑞拉	小区	1982 年	Modelistica	空间投入产出	动态平衡	莫得利斯特卡 Modelistica, 1995
4	<b>MEPLAN</b>	英国	小区	1984 年	Marcial Echenique	空间投入产出	动态平衡	埃切尼克等 (Echenique 等), 1990
5	<b>TLUMIP</b> <sup>8</sup>	美国	小区	1990 年代	Tara Weidner	空间投入产出	动态平衡	韦德纳等 (Weidner 等), 2007
6	<b>IRPUD</b>	德国	小区	1994 年	Michael Wegener	离散选择	动态	韦格纳 (Wegener), 1996
7	CUF	美国	DLU <sup>9</sup>	1994 年	John Landis	基于规则建模	动态	兰迪斯 (Landis), 1994
8	<b>DELTA</b>	英国	小区	1995 年	David Simmonds Consultancy	离散选择	动态	西蒙兹 (Simmonds), 1996
9	Metrosim	美国	小区	1995 年	Alex Anas	离散选择	动态平衡	阿纳斯 (Anas), 1994
10	UrbanSim	美国	<b>多尺度</b> <sup>10</sup>	1996 年	Paul Waddell	离散选择、微观模拟、基于	动态	沃德尔 (Waddell), 2002

<sup>5</sup> 或元胞 (Cell)。

<sup>6</sup> 粗体表示该模型也属于土地使用与交通模型。

<sup>7</sup> 该表统一以小区 (Discrete Zone) 代表分区模型的研究尺度。

<sup>8</sup> 该模型是在 TRANUS 和 UrbanSim 基础上实现的。

<sup>9</sup> DLU (Developable Land Unit), 可开发用地单元, 为非规则多边形, 类似地块 (矢量格式)。

<sup>10</sup> 空间单元可以是小区、网格或地块, 城市活动主体可以是类别 (Categorical) 层次, 也可以是个体 (Individual) 层次。

序号	名称 <sup>6</sup>	所在国家	研究尺度 <sup>7</sup>	开发年份	代表性开发人员/机构	主要方法	时间基础	代表性文献
						个体建模		
11	SLEUTH	美国	网格	1997年	Keith C. Clarke	元胞自动机	动态	克拉克等 (Clark 等), 1997
12	CUF-2	美国	网格	1998年	John Landis 和 Ming Zhang	基于规则建模	动态	兰迪斯和张 (Landis 和 Zhang), 1998ab
13	ILUTE	加拿大	地块、居民、家庭	2004年	Eric J. Miller	微观模拟、基于个体建模	动态	米勒等 (Miller 等), 2004)
14	Relu-Tran	美国	小区	2007年	Alex Anas	离散选择	动态平衡	阿纳斯和刘 (Anas 和 Liu), 2007
15	PECAS	加拿大	小区	2005年	John Douglas Hunt 和 John E. Abraham	空间相互作用、空间投入产出	动态	亨特和亚伯拉罕 (Hunt 和 Abraham), 2005
16	BUDEM	中国	500m 网格	2009年	龙瀛	元胞自动机	动态	龙等 (Long 等), 2009
17	MUSSA II <sup>11</sup>	智利	小区	1996年	Francisco Martinez	离散选择	动态平衡	马丁内斯 (Martinez), 1996
18	GeoSOS	中国	多尺度	2011年	黎夏	元胞自动机、基于个体建模	动态	李等 (Li 等), 2011
19	Agent iCity	加拿大	地块、居民、家庭	2012年	Suzana Dragicevic	基于个体建模	动态	朱巴和德吉塞维克 (Jumba 和 Dragicevic), 2012
20	BLUTI <sup>12</sup>	中国	小区	2012	张宇	离散选择	静态平衡	张宇等, 2012

#### 4、传统城市模型遇到的发展趋势

基于上述对城市模型发展历程的分析可以看出,基于离散动力学的动态城市模型是目前的研究热点和未来的发展方向;国际上典型的城市模型多为宏观尺度,以地理网格或小区作为基本研究单元,将城市活动主体进行分类,这方面理论和实证都有较多研究;随着研究尺度的需要和微观数据可获得性的增强,近年来国际上微观模型发展迅速,但在真实城市中全面应用的案例仍然有限;在国内,土地使用和交通模型和侧重于城市扩张模拟的城市模型都有一定研究,都属于宏观模型范畴,微观模型的研究较少。

根据上述对城市模型发展趋势和已有典型模型的分析,精细化的城市模拟(动态的、基于离散动力学的、微观的城市空间模型)将成为未来的研究热点。其中,精细化(Fine Scale)对应模拟尺度,一方面,在物理空间,分析和模拟的基本空间单元是地块,对应城市总体规划重点地区和城市详细规划的工作尺度。另一方面在社会空间,分析和模拟的基本单元是居民、家庭和企业个体等。“精细化”作为模拟尺度,是相对宏观模拟的研究尺度提出的,宏观模型一般以统计小区、交通分析小区、行政区、行业、共同特征的人群等作为基本研究对象。

<sup>11</sup> 目前称为 Cube Land。

<sup>12</sup> 使用 Cube 软件基于北京市宏观交通模型 BMI Model 基础上开发。

精细化城市模型是一种时空动态的微观模型，多采用自下而上的研究方法，其以地块、居民、家庭、企业等微观个体作为基本模拟对象，主要用于研究城市的土地开发、居民的居住区位选择、企业的区位选址、城市活动的时空分布等空间问题，用于支持空间政策的制定和评估。多种微观模拟的研究方法都可以用于精细化城市模拟，如微观模拟、元胞自动机和基于个体建模。国际上已有少量精细化城市模型在真实城市的实践应用，如 UrbanSim、ILUTE 和 *Agent iCity*；国际国内的已有相关 ABM 研究一般都对应地块尺度，也考虑了众多城市活动主体，这些方法对精细化城市模型的建立和开发具有借鉴作用；国内目前还没有精细化城市模型用于真实城市的报道。在数据层面，国际上的精细化城市模型也不能获得全样本的微观数据，一般采用数据合成的方法（如 *iterative proportional fitting*, IPF）合成居民或家庭全样本。鉴于国内的精细化城市模型研究刚刚起步，也同样面临数据稀缺的问题，除笔者外少有研究探索微观个体数据的合成（*micro data synthesis*）问题。

## 5、大数据时代的城市模型

城市模型，特别是大尺度模型，其发展经历了不同阶段，也遇到了数据和硬件等瓶颈。例如，1973 年李撰写的大尺度城市模型的安魂曲，大有唱衰大规模城市模型的意思，主要一个原因是计算机水平不足(Lee, 1973)。反观 40 年后的今天，计算机的软硬件水平达到了长足的发展，信息和通讯技术（*information and communication technologies*, ICT）发展迅速，社会经济活动产生的数据突飞猛进，这样的“大数据（*big data*）”规模超大，以至于超过了传统的软件工具获取、存储、管理、共享、分析和可视化的能力，例如传感器网络（*sensor networks*）、社会化网络（*social networks*）、射频识别（RFID）和通话记录（*call detail records*）等（来源：Wikipedia）。这些数据为建立城市模型提供了较好的机遇，同时应该也在建模思路和方法等方面提出新的要求。

利用大数据开展城市空间分析与模拟是目前学术界的研究热点。据不完全统计，国际主流的地理信息系统（*geographic information system*, GIS）和城市地理期刊，已经出版了多个专辑，例如江斌在 *International Journal of Geographic Information Science*（IJGIS）2011 年第 8 期上的“*Data-intensive geospatial computing*”，芦咏梅和刘瑜在 *Computers, Environment and Urban Systems*(CEUS)2012 年第 2 期上的“*Pervasive location acquisition technologies: Data and analysis*”，以及目前江斌和 Jean-Claude Thill 正在召集论文的 CEUS 专辑“*Geospatial analysis of volunteered geographic information*”。

要分析大数据时代城市模型的发展机遇，需要回过来先看大数据的特点，其主要包括几个方面。一，作为个体层面的数据，大数据一般反映了人或物的空间位置和联系（主动贡献 *volunteer* 或非主动贡献 *volunteered*），例如公交卡数据对应每个持卡人的不同时刻的空间位置，特别地，Goodchild 突破了传统的光学传感器思想，较早地提出将人作为传感器（*Human as Sensors*）以获取空间和社会信息的设想(Goodchild, 2007)；二是覆盖面广，一般能够覆盖区域或城市内的较大比例的个体，规模巨大（遍布，*ubiquitous*），例如手机数据一般能够对应一个城市内的大量手机用户；三是数据质量较高，具有较细的空间分辨率和时间分辨率，以及客观的个体反馈（如微博的内容、签到地点等）；四是在数据持有者许可的情况下，数据的获取成本较低，例如如果手机运营商许可研究人员的数据申请，则可以在短时间内获得大规模精细数据，这点是传统的调查方式难以实现的。

反观目前我国城市空间发展模型的研究，鉴于数据的限制，多数模型都是大尺度的（如乡镇或交通分析小区等）(龙瀛等, 2011)。而目前，一方面我国的大城市正逐渐由空间扩张（*urban expansion*）向内部改造（*urban redevelopment*）转变，小尺度的城市空间再开发将越来越多，为分析并预测城市空间的变化，更需要精细化的模型作支持，另一方面，城市作为复杂的自适应系统，是由作为城市空间的地块、作为城市活动主体的居民、企业等构成的，

自下而上的模拟思路在以人、地和房作为基本研究对象预测城市空间变化的同时，对规划的公众参与、社会公平等理念的需求不谋而合。这都说明精细化的城市模型更适合我国未来的发展需要。随着大规模、高质量的个体时空数据的获取正不断成为可能，遍布的个人贡献的数据（大数据）对描述和理解城市空间结构提供了新的渠道，为建立精细化的城市模型提供了较大的机遇。主要体现在，大数据对应的个体及其空间位置和移动（如人的活动和移动）是城市模型的核心，能够反映人的活动进而推算其对城市空间的需求，能够覆盖大多数城市活动主体，且具有时间动态性。

在大数据时代，开展城市模型或城市研究的态势主要体现在以下及各方面：一，多个领域的学者共同关注利用大数据开展城市研究，例如除了城市规划领域本身，计算机学科已经有很多专门的会议（如 LSBN 和 UbiComp）和主流的学者研究大数据挖掘的算法，进而对城市问题进行识别和诊断，提出相应的建议。而地理信息科学也将目光关注到大数据，在自愿地理信息（volunteer geographical information, VGI）利用大数据进行城市研究。而时间地理学的学者，也因为大数据时代的到来而受益，以往多采用调查方法获得日志，目前很多学者已经开始以具有时空信息的大数据作为分析的数据源。在这样的多学科百花齐放的背景下，基于大数据的城市研究蒸蒸日上。二，基于大数据的研究日趋破碎化（fragmented），英国伦敦大学学院（UCL）高级空间分析中心（CASA）的巴蒂，在 *Environment and Planning B* 的 Editorial 中提到了这点(Batty, 2012)，即这些研究往往侧重于城市现象的某一个局部方面，而鲜有综合的分析，这点可能在于大数据的特点是广、精和深，适合专业分析和挖掘。而传统的城市模型，则基于来自多个渠道的数据，实现较为综合的分析。三，大数据分析算法的趋于简单化，甚至有观点声称“大数据本身就是模型”，即通过对大数据的简单的时间、空间和属性层面的统计分析，就可以得到有趣（surprising）的分析结果。四，基于大数据的研究，多关注两个层面，在城市内分析城市的空间结构，如笔者利用公交刷卡数据分析北京的通勤特征(龙瀛等, 2012)，或在区域尺度分析区域间的联系，如甄峰等利用微博数据分析中国的城镇体系(甄峰等, 2012)。五，这类研究偏对城市的现状评价而非对未来的预测或对城市系统的模拟，这基本符合精细化城市模型的特征，即越精细化的模型越不适合对远景进行判断，而长于对现状的分析和问题的识别。

总体上，能够看到大数据的出现对精细化的城市模型研究带来了非常大的机遇，得到了多学科的关注。在这样的背景下，有几个问题还需要思考。如大数据的出现，使得传统的城市空间分析、统计、模拟和可视化的方法不能够很好适应，也改变了建模的范式，具体如何应对，是需要考虑的。此外，大数据分析和挖掘的过程中，如何充分利用已有的通过传统方式获得的数据所包含的信息，也是需要考虑的。大数据的信息全面细致，但很多大数据的信息在信息维度弱于传统数据，例如公交刷卡数据有细致的每个持卡人的上下车信息，但是没有持卡人的社会经济属性信息，这样就很难用于分析不同特征人群的交通行为和偏好，而传统的城市居民家庭交通出行调查数据，则两方面数据都非常完善。我们的经验是，利用传统的数据获取规则，用于大数据的模式识别，在龙瀛等(2012)的研究中采用的就是这个思路，即从出行调查数据中识别人的居住地的特征和就业行为的特征，以及步行距离的概率密度分布，然后将这些规则识别大数据中每个持卡人的居住地和就业地。这种监督分类的方法可以保证大数据分析的可靠性。

## 参考文献

1. A. Anas, Y Liu. A Regional economy, land use, and transportation model (Relu-Tran©): formulation, algorithm design, and testing. *Journal of Regional Science*, 2007, 47(3): 415-455.
2. A. Anas. METROSIM: A Unified Economic Model of Transportation and Land-Use.

- Williamsville, Alex Anas & Associates, 1994.
3. A. Jumba, S. Dragicevic. High resolution urban land-use change modeling: Agent City approach Applied Spatial Analysis and Policy (in press). 2012.
  4. Association of Bay Area Governments, 2009, Available online at: <http://www.abag.ca.gov/planning/currentfcst/modeling6.html>.
  5. D. C. Simmonds. DELTA Model Design. David Simmonds Consultancy, Cambridge, UK 1996.
  6. E. J. Miller, J. D. Hunt, J. E. Abraham, et al. Microsimulating urban systems. *Computers, Environment and Urban Systems*, 2004, 28(1-2): 9-44.
  7. F. J. Martinez. MUSSA: a land use model for Santiago City. *Transportation Research Record*, 1996, 1552:126-134.
  8. F. Pagliara, A. Wilson. The state-of-the-art in building residential location models, in F. Pagliara et al. (eds.), *residential location choice: models and applications*, advances in Spatial Science Springer-Verlag, Berlin, Heidelberg, 2010.
  9. J. D. Hunt, J. E. Abraham. Design and implementation of PECAS: A generalized system for the allocation of economic production, exchange and consumption quantities. London, *Foundations of Integrated Land-Use And Transportation Models: Assumptions and New Conceptual Frameworks* (Lee Gosselin and Doherty, eds.), Elsevier, 2005: 217-238.
  10. J. D. Landis. The California urban futures model: A new generation of metropolitan simulation models. *Environment and Planning B: Planning and Design*, 1994, 21: 399-420.
  11. J. D. Landis. The second generation of the californian urban futures model: part 1: model logic and theory. *Environment and Planning B: Planning and Design*, 1998a, 25(5): 657-666.
  12. J. D. Landis. The second generation of the californian urban futures model: part 2: specification and calibration results of the land-use change submodel. *Environment and Planning, B: Planning and Design*, 1998b, 25(6): 795-824.
  13. K. C. Clarke, L. Gaydos, S. Hoppen. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 1997, 24: 247-261.
  14. Jr. Lee, B. Douglass. Requiem for large-scale models. *Journal of the American Institute of Planners*. 1973, 39: 163-178.
  15. M. Batty. Smart cities, big data. *Environment and Planning B: Planning and Design*, 2012, 39: 191-193.
  16. M. H. Echenique, A. D. J. Flowerdew, J. D. Hunt, et al. The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews: A Transnational Transdisciplinary Journal*, 1990, 10(4): 309-322.
  17. M. Wegener. Reduction of CO2 emissions of transport by reorganisation of urban activities. Hayashi, Y., Roy, J. (eds.): *Land Use, Transport and the Environment*. Dordrecht, Kluwer Academic Publishers, 1996:103-124.
  18. M. F. Goodchild. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 2007, 69(4): 211-221.
  19. Modelistica. TRANUS Integrated Land Use and Transport Modeling System Version 5.0. 1995. (Modelistica, Caracas, Venezuela)
  20. P. Waddell. UrbanSim: Modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association*, 2002, 68(3): 297-314.
  21. S. H. Putman. EMPAL and DRAM location and land use models: A technical overview. urban simulation laboratory, department of city and regional planning, University of Pennsylvania, Land Use Modeling Conference Proceedings, Dallas, TX, 1995.
  22. T. J. Weidner, R. Donnelly, J. Freedman, et al. A Summary of the Oregon TLUMIP model microsimulation modules. Washington D.C, 86th Annual Meeting of the Transportation Research Board, 2007.
  23. X. Li, X. Shi, J. Q. He, X. Liu. Coupling simulation and optimization to solve planning problems in a fast developing area. *Annals of the Association of American Geographers*, 2011, 101(5): 1032-1048.
  24. Y. Long, Q. Mao, A. Dang. Beijing urban development model: urban growth analysis

- and simulation. *Tsinghua Science and Technology*, 2009, 14(6): 787-794.
25. 龙瀛, 沈振江, 毛其智. 城市系统微观模拟中的个体数据获取新方法. *地理学报*, 2011, 66(3): 416-426.
  26. 龙瀛, 张宇, 崔承印. 利用公交卡刷卡数据分析北京职住关系和通勤交通形态. *地理学报*, 2012, 67(10): 1339-1352.
  27. 张宇, 郑猛, 张晓东, 等. 北京市交通与土地使用整合模型开发与应用. *城市发展研究*, 2012, 12(2): 108-115.
  28. 甄峰, 王波, 陈映雪. 基于网络社会空间的中国城市网络特征. *地理学报*, 2012, 67(8): 1031-1043.
  29. 郑思齐, 霍焱焱, 张英杰, 等. 城市空间动态模型的研究进展与应用前景. *城市问题*, 2010, (9): 25-30.