

Beijing City Lab

Long Y, Liu L, 2015, Four Transformations of Chinese Quantitative Urban Research in the New Data Environment. Beijing City Lab. Working paper #73

新数据环境下定量城市研究的四个变革

龙 瀛¹ 刘 伦²

1 北京市城市规划设计研究院，博士，高级工程师，longying1980@gmail.com（通讯作者）

2 剑桥大学土地经济系，博士研究生，ll454@cam.ac.uk

摘 要：本文阐述了近年来新数据环境下的城市研究变革。首先介绍了大数据和开放数据形成的新数据环境和国内外定量城市研究概况，然后围绕典型案例对当前定量城市研究的四项变革及相关实践展开讨论，最后提出相关思考。本文认为，新数据环境推动了定量城市研究的四大变革：①空间尺度上由小范围高精度、大范围低精度到大范围高精度的变革，②时间尺度上由静态截面到动态连续的变革，③研究粒度上由“以地为本”到“以人为本”的变革，④研究方法上由单一团队到开源众包的变革。在变革的同时，当前定量城市研究也面临着数据有偏、多现状研究少远景判断、多客观认识少规划启示，以及规划理论和学科发展相关问题。

关键词：大数据，开放数据，城市规划，大模型，众包

Four Transformations of Chinese Quantitative Urban Research in the New Data Environment

Ying Long¹ Lun Liu²

1 Beijing Institute of City Planning, PhD, Senior Engineer, longying1980@gmail.com
(corresponding author)

2 Department of Land Economy, University of Cambridge, PhD candidate, ll454@cam.ac.uk

Abstract: The paper provides an overview on the transformation of Chinese urban study driven by the emergence of new data environment in China in recent years. We first give a brief introduction to the new data environment, which has been made possible by the availability of big data and open data in recent years, as well as a review on the research progress both in China and abroad. It is followed by an analysis on the four major transformations in quantitative urban study, supported by typical research cases. The four transformations are (1) transformation in spatial scale from high resolution but small coverage or wide coverage but low resolution to wide coverage with high resolution, (2) transformation in temporal scale from static cross-sectional to dynamic consistent, (3) transformation in granularity from land-oriented to human-oriented, (4) transformation in methodology from conventional research group to crowd-sourcing. The paper also points out that quantitative urban research is faced with problems like data bias, lack of long term analysis, lack of linkage to planning practice, etc.

Keywords: big data, open data, urban planning, mega-model, crowdsourcing

1 引言

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，最后发展到用系统乃至复杂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^[1]。在我国城市规划逐渐由过去二三十年的“大拆大建”向精细化规划编制与管理转型的背景下，定量城市研究得到了越来越多的关注。同时近年来信息通讯技术的进步与普及为定量城市研究提供了大量新的数据来源，其中既包括严格意义上的大数据，也包括来自商业网站和政府部门的开放数据，构成了了解城市系统运行规律的重要基础，并推动了定量城市研究领域在研究方法与研究问题上的多方面变革。本文试图对当前我国定量城市研究的几项主要变革趋势做一梳理与总结，并就学科发展相关问题提出思考，希望引发国内学界的关注与进一步研究。

1.1 新数据环境

近年来，随着信息通信技术与物联网技术的发展，智能终端、射频识别（RFID）、无线传感器等装置产生的数据量与日俱增，同时，随着城市社会经济活动对互联网的依赖性不断加强，网络平台（主题网站、社交网站、搜索引擎等）也在产生着大量数据信息^[2]——据统计，如今全世界每天产生的数据量高达 2.5 艾字节（EB， 2^{60} ）——这种大容量、高速度、多样性的大数据流正在日益成为我国规划学界的关注热点。此外，各种政府和商业开放数据项目和志愿地理信息项目（Volunteer Geographic Information, VGI）也在扩充着城市研究者的数据基础，这三类数据共同形成了有别于传统调研和统计数据的新数据环境（new data environment）。

与传统数据相比，新数据环境主要呈现出精度高（以单个的人或设施为基本单元）、覆盖广（不受行政区域限制）、更新快（每月、每日、甚至每分钟更新）等特点。它不仅意味着更大的数据量，更反映了数据背后关于人群行为、移动、交流等活动的丰富信息，与新型城镇化时期“以人为本”、“存量更新”、“自下而上”等城市规划理念不谋而合，被认为是促进城市规划科学化与城镇治理高效化的有力工具，也因此为相应的学术研究、规划实践和商业咨询带来了新的发展契机。

1.2 定量城市研究

定量城市研究（quantitative urban studies）是指在一定理论基础之上，采用各种数据和技术方法，致力于探索城市发展的一般规律，并诊断城市问题、模拟城市运行、评估发展政策、寻求解决方案的科学研究方法，可应用于支持城乡规划现状分析、方案编制与方案评估等各个阶段。当前的定量城市研究并非单纯是传统意义上的规划信息化或规划新技术的应用，而更注重对城市现象客观、直观、全面的分析，并通过多种媒介将分析结果传达至政府、规划师、专家学者和城市居民，从多方面提高城市规划与相关政策制定的科学性。

在国外相关研究领域，利用新数据环境开展城市空间与人群活动分析已成为当前学术界的热点，催生的大量研究可初步归纳为七种类型，包括社交网络数据的实时描绘（real time sensing）、多种交通网络数据分析（multiple networks）、城市新型数据系统构建（new urban data systems）、新型交通模型（new models of movement and location）、城市发展路径风险分析（risk analysis of development path）、新型人群移动分析系统（new models and systems for mobility behavior discovery），以及新型交通需求管理工具（new tools for

governance of mobility demand)^[3]。相关研究呈现出三方面特点，首先是研究趋于片段化（fragmented），对城市现象的某一个具体方面的局部分析多于对城市系统的综合分析^[3]，这与新数据环境，特别是大数据数量大、精度高、但维度较少的特点有关；第二是分析算法趋于简单化，也就是通过简单的时间、空间和属性层面的统计分析，就可以得到有趣的分析结果；第三是更侧重对现状问题的识别和分析，而非对未来的预测或模拟^[1]。

目前我国利用新数据环境开展的具有一定代表性的定量城市研究包括¹：利用公交刷卡记录研究通勤出行、城市贫困、过度通勤、公共交通空间结构等问题^[4-6]；利用手机信令数据研究城市人口分布^[7]、空间结构^[8]、商圈影响力、居民出行距离等^[9]；利用出租车 GPS 数据预测拥堵地点^[10]；利用居民活动 GPS 数据分析城郊居民日常活动时空特征^[11]；利用社交网络位置数据和签到信息研究城市用地功能与混合度^[12]、城市发展边界^[13]、城市活动区域划分^[14]、城市网络信息空间结构^[15]；利用百度指数研究区域城市网络特征^[16]；利用百度、高德迁徙数据研究城镇体系、居民黄金周旅游行为等^[17]；利用大众点评数据研究餐饮业格局^[18]、餐厅选址^[19]等；利用全国 PM2.5 监测的在线数据研究 PM2.5 污染分布^[20]；利用微观尺度的人口统计数据分析中国城镇格局^[21]。

此外，由于我国此前的开放数据基础较为薄弱，数据覆盖范围受限（如政府开放的空间数据多针对一个城市），且城市与地区之间缺少统一性和横向对比性，当前定量城市研究的一个重要工作是在全国范围内对开放数据进行提取和汇总，这一方面北京城市实验室（Beijing City Lab）做了若干尝试，代表性数据和研究成果包括：（1）搜集并空间化全国超过 200 个城市的用地规划图，并初步拼合成“全国用地规划图”，进而初步评价在快速城市化的过程中，每个城市实际开发与规划的空间一致性²；（2）搜集全国各直辖市、省、市和自治区的乡、镇以及街道办事处尺度的人口普查数据，进而研究乡镇街道尺度的人口密度及其时空演变³；（3）获取 2013 年全国所有监测站每天、每小时的 PM2.5 浓度数据以及每日的 MODIS AOD（气溶胶）遥感影像，建立乡镇街道尺度的全国 PM2.5 浓度地图并评价相应的人口暴露情况⁴；（4）通过在线开放数据获得全国三百多个城市的所有公交线路和站点数据，进而开展全国范围的城市公共交通服务覆盖率评价⁵。

从上述近三年来国内定量城市研究的进展可以看到，新数据环境的出现与成熟带来了城市研究领域关注内容和研究方法的较大变化，促使城市规划与其他相关学科的进一步融合，以及在研究范式、研究方法与内容上的革新。当前我国定量城市研究的整体框架如图 1 所示，相关研究以传统数据和新数据环境的结合为基础，采取适当的数据处理与研究方法，针对一系列城市规划核心问题展开。在这样的背景下，我们观察到当前我国定量城市研究出现了主要四个方面的变革：空间尺度上由小范围高精度、大范围低精度到大范围高精度的变革，时间尺度上由静态截面到动态连续的变革，研究粒度上由“以地为本”到“以人为本”的变革，研究方法上由单一团队到开源众包的变革。

¹考虑到相关研究数量较多、本文篇幅有限，这里仅列出 2012 年（含）之后的国内代表性研究

²项目具体见 <http://www.beijingcitylab.com/projects-1/2-urban-growth-boundaries/>

³项目具体见 <http://www.beijingcitylab.com/projects-1/4-population-china/>

⁴项目具体见 <http://www.beijingcitylab.com/projects-1/13-pm2-5/>

⁵排名具体见 <http://www.beijingcitylab.com/ranking/>，#1 Bus Coverage of Chinese Cities

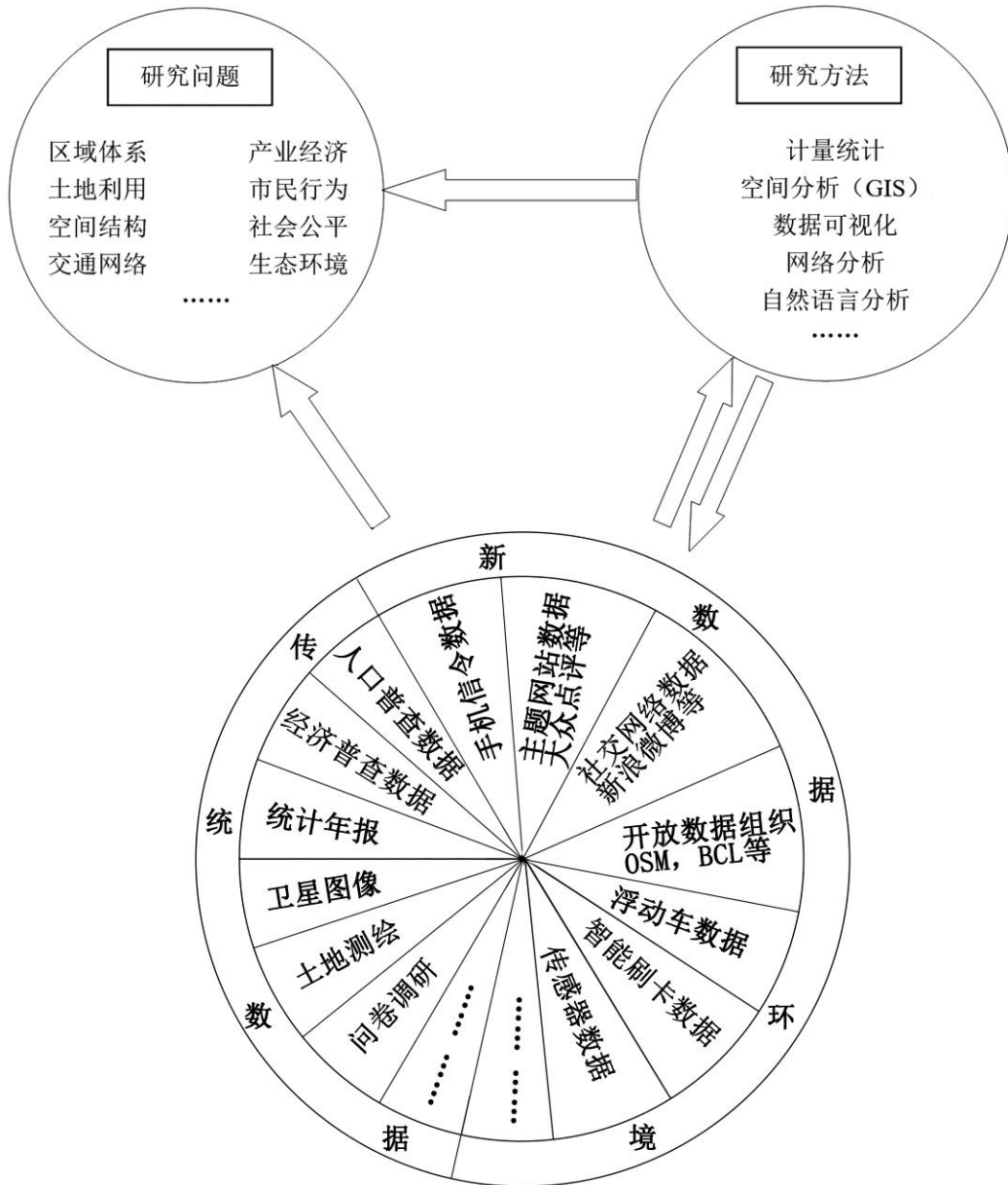


图 1 定量城市研究框架图

2 四个变革

2.1 空间尺度的变革：从小范围高精度、大范围低精度到大范围、高精度

研究范式

在传统数据环境下，受数据收集方法的限制，城市和区域研究在研究覆盖范围和精细度上往往很难做到两者兼顾——大范围的研究通常以牺牲精细度为代价，而精细度高的研究往往覆盖范围较小。因此，传统数据环境下的城市和区域定量分析可主要分为两种，一是针对单一城市做较为深入的研究，如研究广州城市贫困问题^[22]、北京公共服务设施的配套水平^[23]等；二是覆盖全国或多个省市地区的区域分析，如一些宏观经济研究，多以县、市或省为单元，研究单元较大，难以反映小尺度信息。

而新数据环境为在较大空间范围内收集高精度数据提供了可能，如社交网络和各类商业网站数据往往覆盖全国且以人、车、商户等个体为基本单位，可充分满足精细化的分析需求，而对某些传统数据的有效整合也有利于拓展数据的广度与精度。如以往人口密度研究主要在区县尺度，属于宏观分析的范畴，在新数据环境下则可将研究尺度缩小至乡镇街道级别，不仅促进了研究范围和精度的提高，且有助于呈现以往难以发现的新的问题。国外相关研究包括，Ratti 等^[24]通过英国全境一个月内的 120 亿条通话记录重新划定了英国的城市和区域范围；Becker 等^[25]利用带地理标签的 Twitter 数据进行了人群活动类型识别，并结合多主体建模方法构建了英国利兹市（Leeds）城市动态模型；Rozenfeld 等^[26]利用高精度的英（200×200 米）、美（街区尺度）两国全国人口调查数据研究了人口聚集规模；Sagi 等^[27]利用手机信号数据和 Flickr 照片数据研究了多个欧洲城市市民和游客的时空行为，等。

针对这一数据特点，龙瀛等^[28]提出了“大模型”研究范式，试图通过这一范式在城市模型研究中兼顾覆盖区域乃至全国的研究范围与精细化的城市模拟单元。大模型是由大规模数据驱动的大样本量城市研究工具，主要以全国或区域为研究对象，一般采用简单直接的建模路径，既考虑城市内部在精细尺度上的发展动态，也考虑城市间的要素流动和网络联系，较好地在一个研究范式下兼顾城市内和城市间两类城市研究。我们期待大模型这种“自下而上”和“自上而下”相结合的研究视角，以及精细化、量化、全面化的研究方法，将为城市研究带来新的视角和发现。随着新数据环境的不断成熟，传统城市研究的“研究地盘”的概念也将会逐渐弱化，异地的学者也有望对千里之外的城市进行深入的城市研究。

在传统数据环境下，中小城市和县、镇与大城市之间在数据基础和研究水平方面存在明显差距，城市研究多关注具有代表性的大城市而忽略了二三线城市和等级更低的城市^[12]。大模型通过关注全国绝大多数城市，有望在一定程度上对中小城市发展给予更多研究关注、消除技术差异，并系统探讨国家和区域城市化进程中各类城市的互动关系。此外，大范围、高精度的研究趋势有利于探索我国城市发展的一般规律。目前我国共有六百多座城市，不同学者针对不同城市采用不同方法开展的研究往往难以进行横向比较，而以大模型的为代表的研究范式可为一致性研究提供可能，从而在大样本城市研究的基础上探索我国城市发展的一般性和特殊性规律。除了笔者和合作者的研究外，我们也注意到其他学者在这方面的努力，如高晓路等^[29]。

研究案例

(1) 覆盖全国所有城市的地块尺度城市增长模型MVP-CA

利用海量矢量地块元胞自动机 (Mega-Vector-Parcels Cellular Automata Model, MVP-CA) 对全国 654 个城市的用地扩张过程进行地块尺度的模拟, 可用于预测每个城市未来五年内在不同政策情景下的城市扩张情况^[30], 模拟结果见图 2。

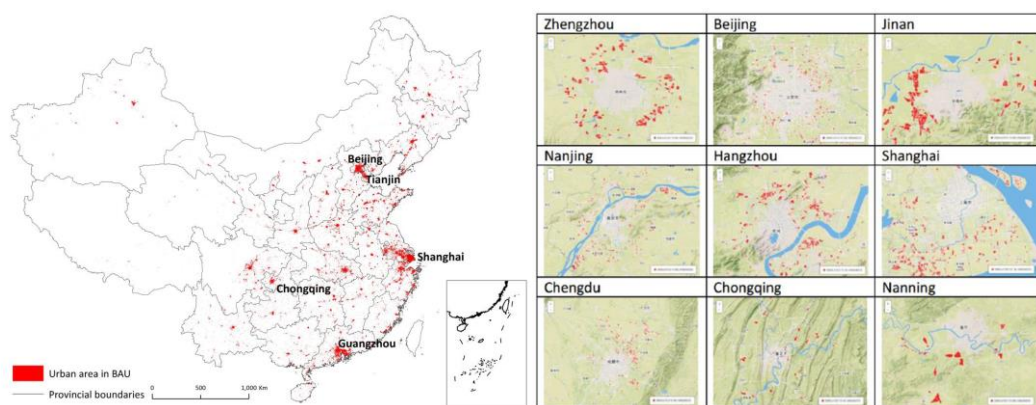


图2 全国所有城市不同情景下城市增长模拟

图片来源: 龙瀛, 吴康, 王江浩, 刘行健. 大模型: 城市和区域研究的新范式[J]. 城市规划学刊, 2014, (6): 55-63.

(2) 利用开放街道地图 (OpenStreetMap、OSM) 和兴趣点 (point of interest、POI) 数据生成全国 297 个城市的用地现状图

针对国内城市用地数据开放度不足的问题, 利用全国开放道路地图(OSM)和兴趣点(POI)数据, 对全国 297 个城市的土地使用布局、用地功能、城市建成区范围、建设密度、功能混合度等特征进行识别, 并公开发布并提供免费下载, 在一定程度上为相关领域学者提供更好的研究基础^[12, 13]。作为后续研究, Long 和 Shen^[31]还基于导航地图和兴趣点对中国所有城市的城镇建设用地范围进行了识别。

(3) 城市规划的开发控制评估

国内多个学者在北京、上海、广州和深圳等开展实证研究 (主要是用的城市扩张与规划比对的结果), 基本上都发现了总体规划之外的不容忽视的非正式开发^[32-35]。目前多数的规划评估工作局限于单个城市或区域, 城市间横向的实施评估和对比还处于空缺状态。在新数据环境下, 我们开展了全国范围内横向对比和评估规划控制成效的研究, 这一正在进行的研究首先收集了全国超过 300 个城市的正在实施的城市总体规划图 (属于政府开放数据), 从中提取城镇建设用地范围即规划城市增长边界 (urban growth boundaries、UGBs), 之后将各个城市的规划建设用地分布与来自遥感观测到的 2000-2010 的城市扩张进行对比, 发现在“与规划的一致性”指标方面, 202 个研究覆盖的城市中 27 个低于 50%, 122 个低于 80%, 后续我们还将对该指标的影响因素进行深入分析。

(4) 全国城市公交站点服务覆盖率评价

目前, 对于中国城市公交站点覆盖率的研究多针对某个具体城市, 而针对中国大多数城市寻找一般规律的研究还较为匮乏: 一方面由于全国范围的研究涉及大量基础数据; 另一方面, 在微观尺度研究大量城市时涉及到尺度转换问题。我们基于来自地图网站的全国 313 个主要城市精细化的公交站点数据, 计算 313 个城市城镇建设用地区域内公交站点覆盖率,

其中全国 281 个地级及以上城市的公交站点覆盖率平均值为 64.4%。基于公交站点覆盖的空间特征，我们将 313 个城市聚合为五类，试图找到中国城市公交服务的一般模式和规律，进一步，基于 Flickr 照片、位置微博和兴趣点数据，对公交站点 500m 服务范围的人的活动及设施情况进行分析，结果显示，该服务范围内包括了 94.4% 的设施和超过 92% 的人类活动，即我国城市公交站点布局，满足了大多数人的活动需要和设施需求。

2.2 时间尺度的变革：从静态截面到动态连续

研究范式

新数据环境所提供的另一重要突破是体现了不同时间尺度上的城市动态。传统城市研究的数据来源多为政府部门统计年鉴或抽样调查，以静态数据为主，只能反映某一时刻或一段时间内城市所处的状态（如年鉴对应一年、出行调查多对应一日），且由于数据取样的局限性，只能覆盖有限的空间范围。而包括公交刷卡、出租车轨迹、信用卡交易记录、在线点评以及位置微博和照片等在内的新数据环境则可以反映个人乃至整个城市短至每秒、长至多年的动态变化，且具有连续性高、覆盖面广、信息全面等优势^[36, 37]。例如利用精确到秒的信用卡交易记录，可以对城市每小时的销售情况进行可视化，进而识别商圈；积累多年的信用卡交易记录，则可以体现出人们生活与消费方式的变化，如传统书店的萎缩和在线购物的繁荣。假设未来一千年之后的考古，发现人类的电子记录和足迹，则有望超越目前的考古发现，对此时的人类社会进行更为全面的重现。国外相关研究已大量展开，如 González 等^[38]通过研究 10 万人在六个月内的手机信号位置轨迹发现，人们的出行行为实际上具有极强的规律性；Reades 等^[39]通过 100 万手机用户在三个月内的 350 万条信号记录研究了罗马的空间结构；Phithakkitnukoon 等^[40]通过研究波士顿 130 多万条手机信号记录发现工作地点相近的人出行模式也更为相似；Roth 等^[41]通过研究一周内两百万伦敦地铁持卡人的 1100 多万次地铁出行重新识别了伦敦的城市结构并发现了明显的多中心现象；麻省理工大学感知城市（SENSEable City）实验室与哥本哈根市合作，为五千块垃圾贴附地理标签，并在三个月内追踪垃圾流向，分析垃圾回收效率；Masucci 等^[42]通过研究近 200 年来伦敦路网的变化分析了伦敦的城市增长规律。笔者利用新数据环境开展了若干的研究工作，体现了时间尺度的变革。

研究案例

（1）利用一周公交刷卡记录评价北京通勤形态

龙瀛等^[37]利用 2008 年公交刷卡数据（854 万持卡人连续一周共 7797 万次出行），对北京市通勤出行进行了识别，得到 22 万余人的居住地、就业地和通勤出行，进而对通勤形态的时空分布进行了评价，并对典型居住区和就业地的通勤出行进行了重点分析，如图 3 所示（如发现 CBD 的通勤辐射范围远超上地和金融街，通州居民每日平均通勤时间超过一小时等）。需要说明的是，利用一日刷卡数据可以对公交系统的运行进行可视化，而难以可靠地推测持卡人的居住地和就业地以及通勤出行，而通过利用一周数据所包含的持卡人的周期性活动，则有望解决这个问题。

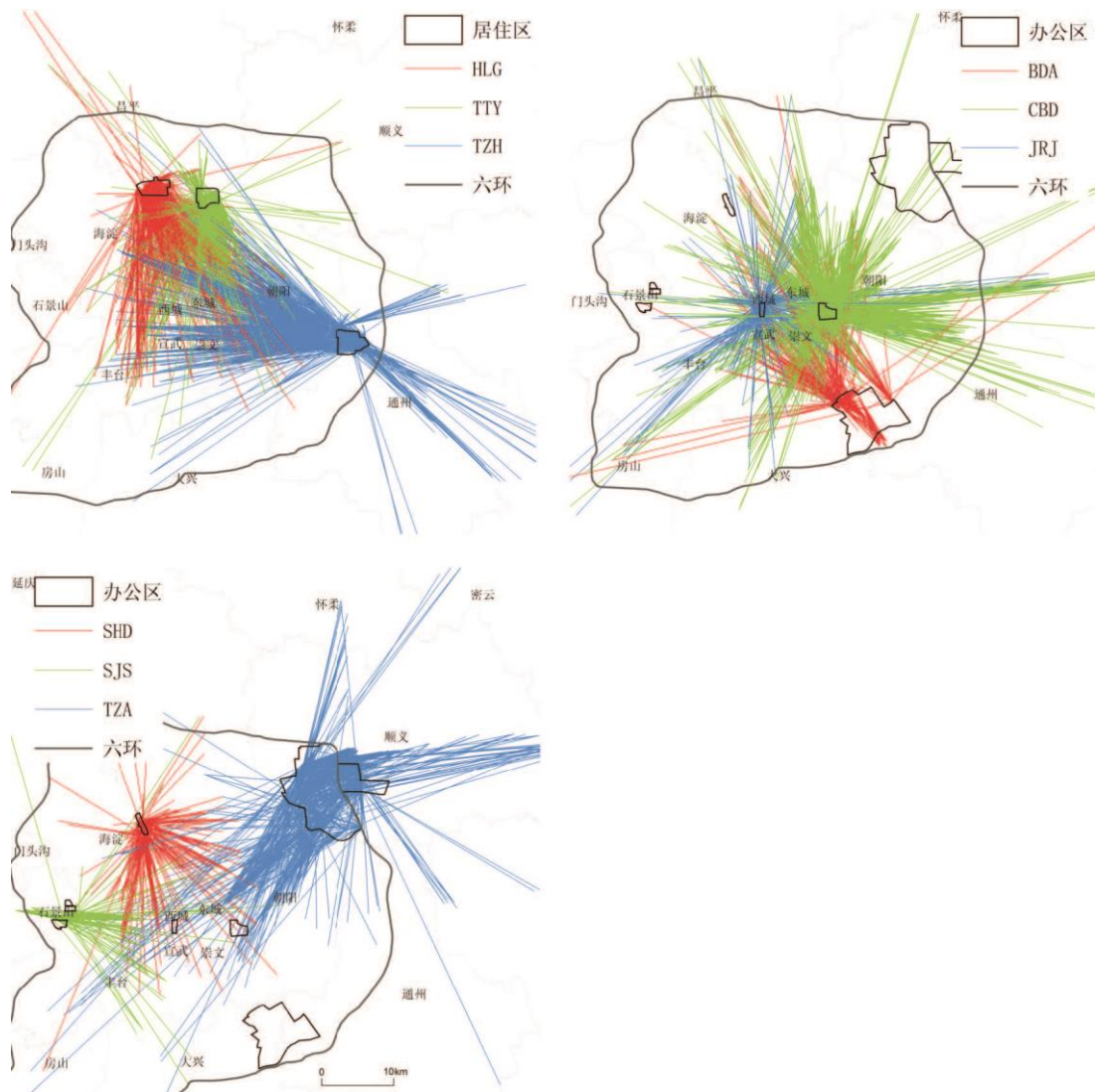


图3 典型地区的通勤形态

图片来源：龙瀛，张宇，崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行 [J]. 地理学报, 2012, 67(10): 1339-1352

(2) 利用多年公交刷卡记录研究城市贫困问题

Long 等^[5]利用 2008 年和 2010 年北京市上千万持卡人的刷卡记录，发现了 11.2 万在研究时间段内频繁地在北京使用公共交通的持卡人，并将他们识别为潜在城市贫困群体。这些人是谁？在干什么？工作在哪里？居住在哪里？家庭状况如何？深入了解这些问题，有助于城市管理者科学制定公交线路优化、公交补贴、产业布局、时间规划等政策。笔者基于龙瀛等^[5]利用每年一周的智能公交卡数据识别公交卡持有者的居住地、就业地和通勤出行，发现 80% 的持卡人更换了住址（多为搬离市中心），只有 13% 的人未更换工作。这说明，智能公交卡数据除了可以定位大规模的城市贫困群体，还可以了解他们的出行、通勤，以及居住地、就业地方面的多年变化，这是以往研究无法做到的。该研究成果可以用于支持福利申请，如申请保障性住房时可对申请人的公交卡进行评估，识别出极端贫困的申请人，在其命中概率或者租金价格等方面，予以照顾和补贴；也可将研究成果用于公交系统补贴，将目前粗略的群体性补贴转为基于精细识别的个体补贴。

(3) 基于大规模历史资料重建两百年前的中国土地利用

新数据环境的另一个方面是对大规模历史资料的数字化、整理和空间化。Long 等^[43]基于现代耕地格局，在对大量与耕地相关的历史数据进行分析的基础上，建立了历史耕地重建模型，对江苏省 1812 年的耕地的数量和空间分布进行了重建，该研究跨越了二百多年的时间尺度，超过了一般城市研究的时间尺度。

作为 Long 等^[43]的后续研究，林忆南等^[44]以典型历史断面下的建设用地为研究对象，以历史文献记载、现代统计数据 and 自然环境数据等为支撑，采用“结合数量重建进行空间重建”的思路，尝试构建一套适用于清代中期建设用地（城镇用地和农村聚落用地）的重建方法。其中，城镇用地主要基于城垣周长、城池形态进行数量重建，以历史治所为中心，以城池形态为控制边界，进行空间配置；农村聚落用地主要基于农村人口、人均居住用地需求进行数量重建，以潜在农村居民点重心为中心，考虑用地宜居性和连片性等特征进行空间重建，江苏省的初步结果见图 4，下一步研究计划将研究范围扩展到传统农区范围。

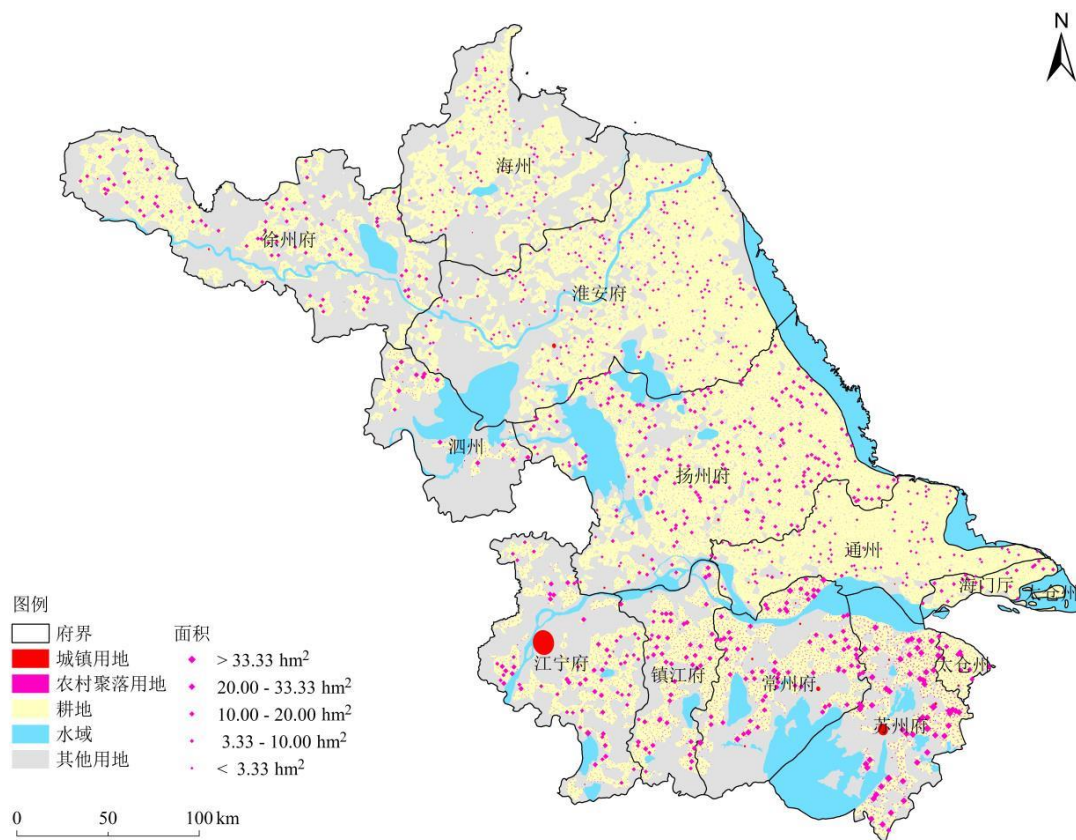


图 4 1820 年研究区建设用地空间格局

图片来源：林忆南，金晓斌，杨绪红, et al. 清代中期建设用地数据恢复与空间网格化重建：方法与实证 [J]. 地理研究, 2015, 已接受.

2.3 研究粒度的变革：从以地为本到以人为本

研究范式

我国过去三十年的快速城市化进程在很大程度上是粗放的发展和扩张，相应的规划设计和政策制定往往以各种用地指标为导向，而对城市居民的真实需求和感受关注不足，导致了一系列社会、经济和环境问题。针对这些现象，我国新型城镇化提出了以人为本、集约、高效的发展模式，小型的城市更新、改造和再开发将在一定程度上取代“大拆大建”的发展模式，相应的，城市研究也亟需研究粒度的细化，回归“人”的尺度。

同时，新数据环境所提供的并非仅是扩大的数据量，还包括数据所反映的城市居民的行为特征与规律，以及人对建成环境的感觉、情感、经验、体验、信仰、价值判断等，这些以前难以量化的因素在新数据环境中都可以得到有效的表达与数理分析^[2]。国外相关研究包括，Calabrese 等^[45, 46]通过研究 100 万手机用户的位置和通信数据发现，通讯联系强的个体在实体空间中也更可能出现在同样的地理位置，从而为预测城市人群移动提供了新的途径；Schneider 等^[47]利用多日手机数据发现了城市人群活动的 17 种基本模式；Liu 等^[48]利用 Flickr 图片数据分析了人对城市空间的认知图像，重新阐释了凯文·林奇的城市意象理论，等。在我国新型城镇化的背景下，具有高粒度特性的新数据环境也为以人为本的城市研究提供了极佳的素材。在此数据平台基础上，居民行为、活动及其影响下的城市空间组织和结构的变化，社会群体特征、网络、活动等及其影响下的社会空间分异或融合等课题都可得到深入分析。

研究案例

(1) 基于乡镇街道办事处尺度人口数据的中国收缩城市研究

以往的中国范围的人口研究多为县市尺度，Long 和 Wu 发表在 *Environment and Planning A* 中的研究基于五普（2000 年）和六普（2010 年）的乡镇和街道办事处尺度的人口数据，发现中国 180 个城市发生了人口总量和密度的下降，即存在着 180 个收缩城市（图 5）。中国收缩城市可以根据原因分为四个类型：资源型城市在工业化进程中出现的局部收缩、传统工业城市工业衰落导致的局部收缩、欠发达地区人口迁移引发的局部收缩、行政区划调整导致的局部收缩。这一研究发现意味着城市规划中“人口增长”的基本假设应有所调整，特别是 180 个收缩城市的规划，亟需从城市规模扩张转向关注居民生活质量。此外，研究也在一定程度上提示规划师和决策者，人口向大城市的聚集是城市发展的客观规律，中小城市的发展促进政策是否适合需要因地制宜的客观评价，而非盲目推进。

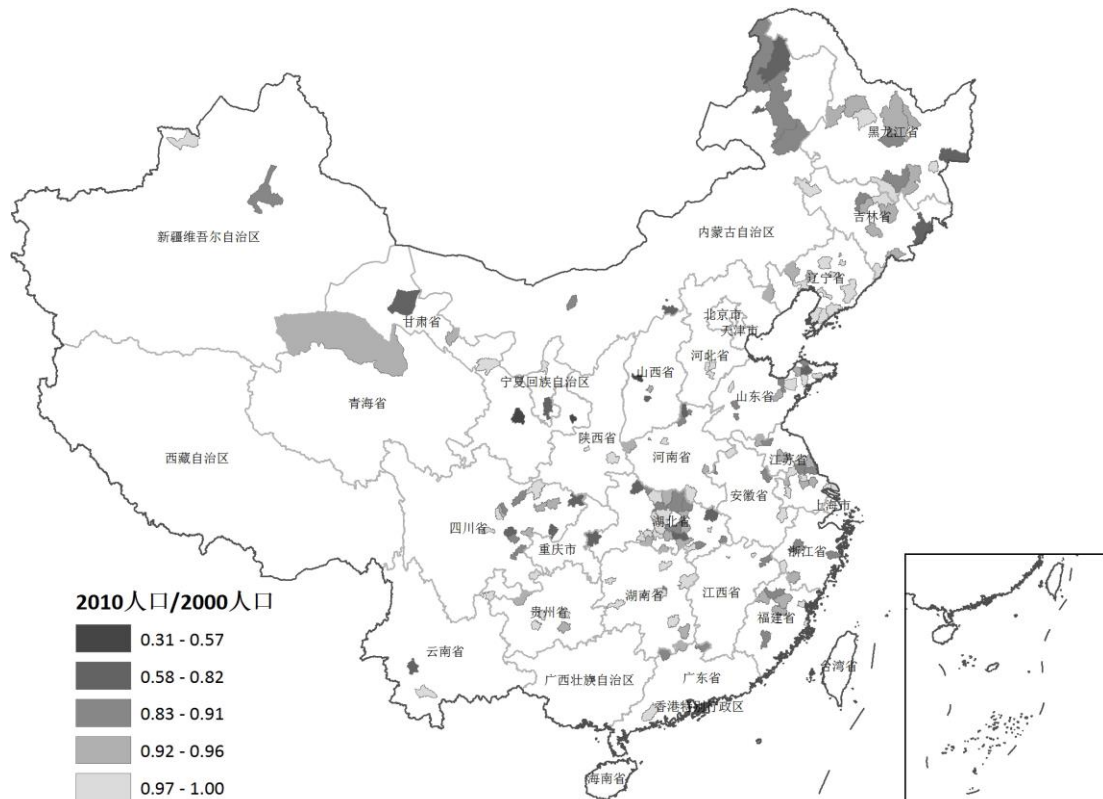


图 5 2000 年-2010 年中国收缩城市分布

(2) 基于人类活动和移动数据的规划实施评价

以往的城市增长边界 (UGBs) 评估主要利用遥感解译反映的城市扩张数据, 少有研究利用个人层面的数据。Long 等^[13]利用包括公交地铁刷卡数据、出租车轨迹、位置微博和照片等多源新数据, 评价北京 UGBs 的实施效果, 研究结果显示, 虽然有大量的非正式开发分布在 UGBs 之外, 但 UGBs 所包含的区域内容纳了 95% 以上的城市活动和移动, 与非正式开发的比例并不一致。原因是多方面的, 如非正式开发的使用程度较低、非正式开发缺少公共服务如公共交通配套等。

(3) 全国乡镇街道办事处尺度人口对 PM_{2.5} 的暴露评价

已有关于 PM_{2.5} 的研究稍有关注中国精细化尺度的人口暴露情况。在新数据环境下, Long 等^[20]收集了全国 190 个城市共 945 个空气监测站的 2013 年 4 月 8 日至 2014 年 4 月 7 日间的每日 PM_{2.5} 浓度值 (www.cnemc.cn)。此外还利用中分辨率成像光谱仪大气气溶胶厚度 (MODIS AOD) 数据对 PM_{2.5} 进行插值补充。另一方面, 结合第六次人口普查的分街道乡镇统计数据, 重点研究全国所有街道乡镇单元的人口 PM_{2.5} 暴露几率和暴露强度风险 (图 6)。评估发现, 中国人口加权的 PM_{2.5} 年浓度均值为 68.3 $\mu\text{g}/\text{m}^3$, 大大超过了国家标准的 35 $\mu\text{g}/\text{m}^3$ 。中国人暴露在 PM_{2.5} 污染中的平均天数则是 113 天。中国的 PM_{2.5} 污染集中在三个区域, 其中最醒目的污染区是覆盖我国中东部以北京、上海、广州、成都为四个顶点构成的菱形区域, 另两个污染集中区域是东北的哈尔滨-沈阳走廊, 以及新疆的乌鲁木齐一带。

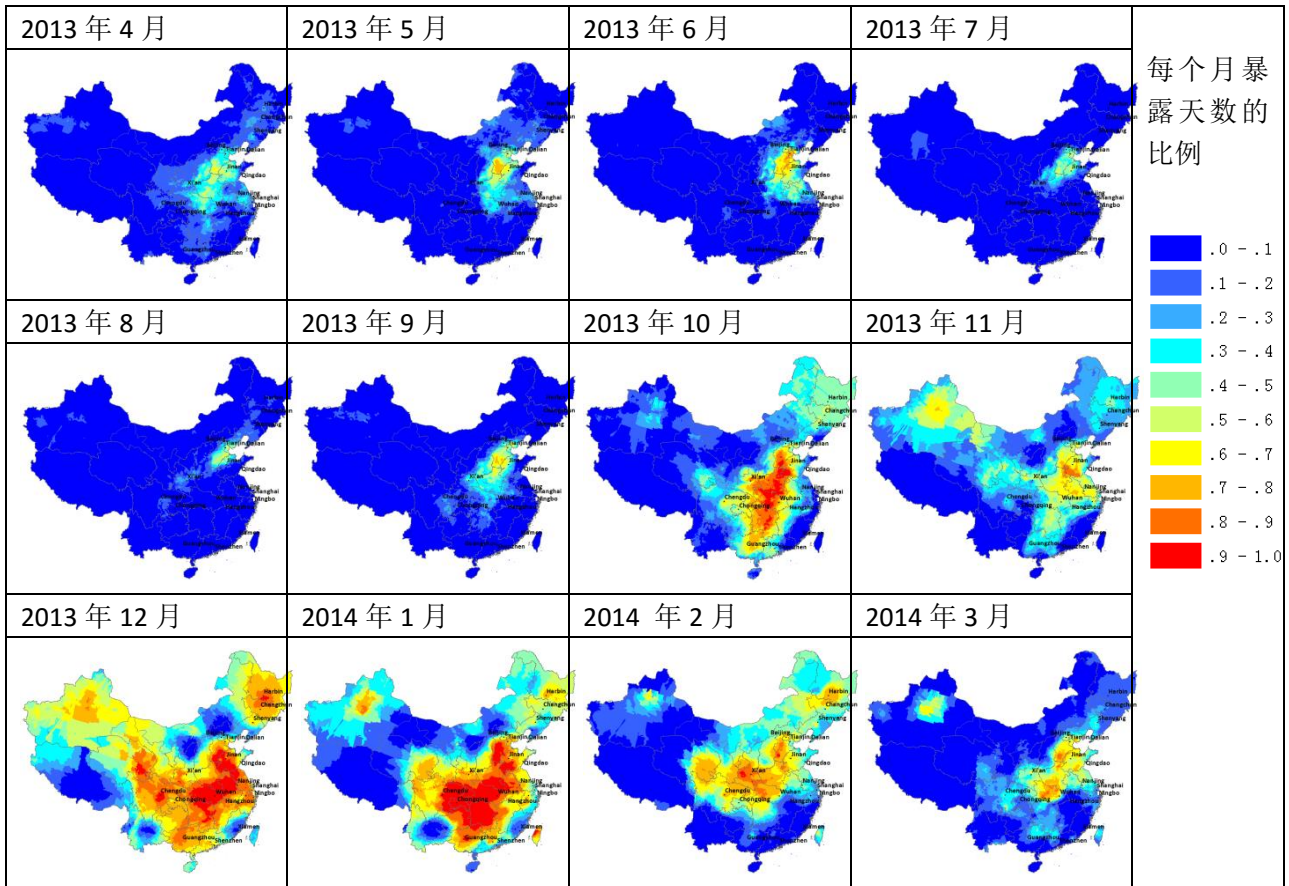


图 6 全国乡镇街道尺度 PM2.5 超标比例示意图

图片来源: LONG Y, WANG J, WU K, et al. Population Exposure to Ambient PM 2.5 at the Subdistrict Level in China [J]. Available at SSRN 2486602, <http://ssrncom/abstract=2486602> or <http://dxdoiorg/102139/ssrn2486602>, 2014

2.4 研究方法的变革：从单一团队到开源众包

研究范式

众包 (crowdsourcing) 则是互联网带来的新的生产组织形式, 即利用互联网将原先单一机构内的工作任务以自由自愿的形式分配给机构外的志愿人员 (通常为个人) 完成, 这一组织方式可以充分利用志愿者的创意和技能, 以更低的成本、更高的效率完成任务⁶。虽然开源、众包等概念听来与城市研究和城市规划领域相距甚远, 但近两年来随着数据的开放和北京城市实验室 (BCL) 等开放研究平台的成熟, 众包模式也在逐渐融入到定量城市研究和相关数据平台构建中 (如搜集数据、学术合作、验证研究成果), 并体现出优势。这种众包的城市研究方式有望突破传统的单一团队开展研究工作的模式, 例如针对中国大量存在的收缩城市现象, 探讨其背后的深层原因则需要大量实地调查, 相关工作量大大超出了一个课题组或单一机构的承担能力, 为此 BCL 于 2014 年 11 月共同发起了“中国收缩城市研究网络” (<http://www.beijingcitylab.com/projects-1/15-shrinking-cities/>), 持续跟踪收缩城市方面的国际研究并开展国内的理论和实证研究工作。总体上, 新数据环境下, 城市研究的工作方法正在由单一团队向开源众包模式转变。

⁶ <http://baike.baidu.com/view/729695.htm>

研究案例

(1) 基于开放数据建立城市用地现状图并模拟城市扩张

众所周知，用地现状图数据在我国被严格控制，外资事务所、大学、规划机构等都难以获得完备的用地现状图数据，Long 和 Liu^[12] (2015) 利用开放数据如 OSM (OpenStreetMap) 道路网络和兴趣点数据，生成了中国 297 个城市的用地现状图，包括功能布局、密度和混合使用程度等方面。研究成果的 GIS 数据通过网络完全共享，并通过志愿者征集，获得了十余个城市的真实用地现状图，进而验证了利用开放数据推导用地现状图的可行性。

在此研究基础上，Long 等^[30]基于大量的道路网和兴趣点数据，利用约束性元胞自动机方法，建立了覆盖全国 654 个城市，包括了 76 万个城市地块的城市增长模拟模型 (MVP-CA, 图 7)，对不同政策下每个城市的城市空间发展模式进行了模拟。模拟结果采用在线交互地图的形式共享在 BCL 网站上，通过微博进行宣传，得到了诸多用户的反馈，一些具有当地知识的用户对所熟悉的城市的模拟结果进行了评价，指出了模拟结果的可靠与不可靠之处，这对优化模型构建具有较大的促进意义。

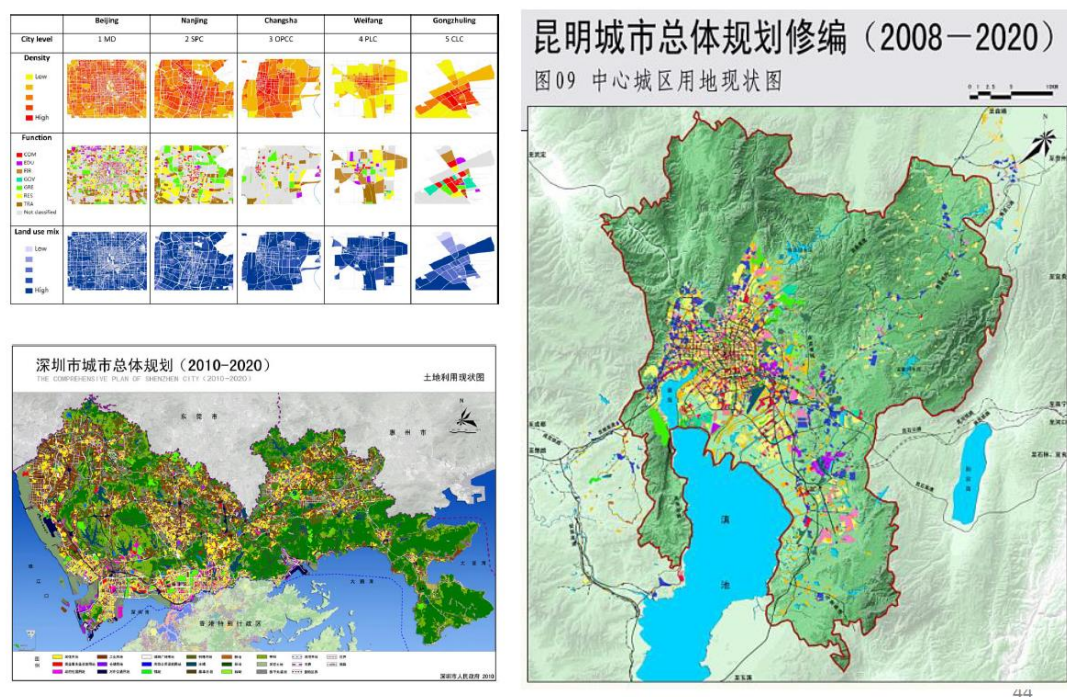


图 7 利用各地用地现状图验证开放数据推导的城市功能

图片来源: LONG Y, WU K, MAO Q. Simulating urban expansion in the parcel level for all Chinese cities [J]. arXiv preprint arXiv:14023718, 2014,

(2) 北京主要街道的可步行性评价

茅明睿和储妍^[49]在志愿者的组织下，通过一个展览，邀请公众通过用不同颜色的图钉扎针的方式绘制“北京城市街道环境评价”地图，收集了 1600 多个正面、负面评价点。基于公众的评价结果，研究者对正负面评价点进行了深入分析，并利用腾讯街景地图，对每个评价点抓取了 8 个方向的街景照片，设定了“路面铺装、无障碍性、停车占道、设施占道、视线遮挡、步行尺度、可达性、魅力空间、绿化景观”9 项指标，对街道评价点进行了处理和分析，探寻公众对每个地点给予正面评价或者负面评价的原因，分析公众对哪些评价指标更敏感，以及根据不同评价指标，发现北京不同区域、不同街道存在的主要问题，并进一步对每个差评点给出有针对性的设计建议，通过匹配与其相应指标得分相反的正向评价路段及街

景图给出每个差评点改进的参考示例。

(3) 中国 1km² 网格的数据共享计划（徐霞客计划）

和城市有关的数据正日益增多，且涉及到了人们生活的各个方面。而城市研究者们正在尝试透过非传统城市数据加深我们对城市的理解。但由于很多数据过于精确，往往涉及到隐私和数据持有者的核心利益。在这样的背景下，我们发起了一个中国范围内的 1km² 网格尺度的微观数据平台 SinoGrids，提供指南和工具，协助大规模微观数据持有者将其数据汇总到 1km² 网格尺度并上传到 SinoGrids 平台，进而形成一个“众筹”的中国基础数据平台（<http://www.beijingcitylab.com/projects-1/14-sinogrids/>）。

尺度上来说，1km² 是既能够开展城市间区域分析也能够进行城市内部研究的尺度。同时，SinoGrids 将以众包的形式进行数据开放，我们将网络上公开的或者学者贡献的数据（微博签到、出租车轨迹、交叉口数量、公交站点的数、照片等）汇总到 1km² 的规模上，然后在 SinoGrids 平台上进行公开。也就是说平台将会显示每平方公里网格的微博总量、照片总量、公交站点等数据。同时，平台上将一直保持最多最全的指标和数据指南，方便公众使用。有数据资源的人一方面可以利用我们的网格来贡献自己的微观数据，另一方面可以利用其他人贡献的数据进行区域分析、城市研究、规划咨询、公共参与等，SinoGrids 将是一个公众的、开放的城市数据平台，希望能够为我国研究者的定量城市和区域研究工作提供全面且透明的数据支持。

3 结 论

定量城市研究作为一种服务于城市研究和城市规划的研究工具，在功能上较之于传统研究方法显示出了更强的科学性、客观性与延展性。随着信息通信技术、数据挖掘技术以及开放数据理念的发展，新数据环境蕴含的各种社会经济文本信息为定量城市研究的拓展提供了更多可能，也使原先的研究范式逐渐在数据获取、研究范围以及研究对象等方面不断演进，我们亟需关注、探讨并把握相关趋势，及时扩展、更新定量城市研究乃至城市研究的相关方法、理论，在与人文地理学、城市社会学、信息科学等相关学科的交叉融合中提高分析和解决城市问题的能力。虽然相关研究在过去一段时间内取得了大量进展，但我们也注意到以下几点现象，需要在日后研究中进一步讨论并提升。

(1) 数据有偏性问题

新数据的有偏性广为学界讨论，这与大部分新数据的互联网属性直接相关——新数据大量来自于互联网的积极使用者，导致了数据在人群和地域代表性方面的偏差。例如利用位置微博数据进行的城市幸福感研究，面临着多个层面的偏差，包括微博用户的口是心非、自然语言处理的能力约束、发微博人群的代表性、新浪提供微博API的黑箱。这些偏差使得研究结果的可信性受到质疑。目前应对数据有偏性问题的研究思路主要分为三种，一是充分利用数据的有偏性研究特定人群的行为特征，如利用公交刷卡数据研究城市贫困群体的出行特征、居住地和就业地的变化情况；二是采用多种数据研究同一问题，以验证研究结果的稳定性；三是在承认新数据有偏性的前提下，采用“盲人摸象”的策略，采用多组有偏数据相互补充，勾勒出城市系统的完整面貌。

(2) 与规划实践的对接问题

虽然定量城市研究在学术领域迅速发展，但相关方法在规划实践中的应用仍非常有限，

尚需要大量探索。这涉及到当前研究的两方面问题：首先，现有研究大多采用短期数据记录，多现状分析、少远景判断，部分研究仅处于可视化层次，而少有利用覆盖多年的数据进行远景判断；其次，现有研究多客观认识、少规划启示，以针对城市系统的现状评价和问题识别为主，而较少可以用于直接指导规划实践。第一方面问题和新数据的“新”有关，随着数据的不断积累，研究者将可以获得更多反映城市演变历程的时序性数据，突破这一瓶颈。而第二方面问题涉及如何利用新数据环境下的诸多定量研究成果反哺规划设计，如何利用相关技术方法更为深刻地、别开生面地、革故鼎新地服务于规划方案的设计和公共政策的制定，挖掘这一新的分析工具所应发挥的实践价值，是定量城市研究的关键难点，也是下一阶段相关研究亟需探讨的问题。

为此笔者与合作者提出了数据增强设计（Data Augmented Design、DAD）的规划设计方法论，希望以定量城市分析为驱动，通过数据分析、建模、预测等手段，为规划设计的全过程提供调研、分析、方案设计、评价、追踪等支持工具，以数据实证提高设计的科学性，并激发规划设计人员的创造力。DAD利用简单直接的方法，充分整合新旧数据源，强化规划设计中方案生成或评估的某个环节，易于推广到大量场地，同时兼顾场地的独特性。DAD的定位是现有规划设计体系（标准、法律、法规和规范等）下的一种新的规划设计方法论：它不是艺术设计的背叛者，而是强调定量分析的启发式作用的一种设计方法，致力于减轻设计师的负担而使其专注于创造本身，同时增强设计结果的可预测性和可评估性，属于继计算机辅助设计（Computer Aided Design、CAD）、地理信息系统（Geographical Information System、GIS）和规划支持系统（Planning Support System、PSS）之后的一种新的规划设计支持形式。详见龙瀛和沈尧^[50]。

基于新数据环境、特别是大数据的研究也在一定程度上被质疑规划理论的缺失，即相关研究缺少相应规划理论的支撑。但事实上，定量城市研究的理论支持充分与否与采用的数据类型并无直接关联，基于传统数据的研究也可能存在同样问题，取决于具体的研究方法和路径。

（3）与规划学科发展的互动问题

可以看到，定量城市研究、特别是新数据环境下的相关研究所要求的研究者知识背景与传统规划知识体系存在明显差异，这也涉及到当前研究与规划学科原有体系如何相辅相成的两方面问题。首先，定量城市研究在采用创新性数据与方法的同时如何为规划学科核心理论发展做出贡献，如何利用新数据揭示出更多传统数据和方法难以应对的城市运行内在规律，是相关研究应关注的问题。第二，未来规划院校的学科专业设置、课程设置、人才培养是否需要有所转变，以应对学科发展出现的新现象和新趋势，也是学界需关注与讨论的问题。目前世界上大多数规划院校都并未将定量城市研究作为规划教育的基本组成部分，调整规划教育体系并非易事，且需要若干年的时间，同时还需要许多其他转变同时发生，例如，教育体系的转变需要规划院校教师人才结构的转变，引进更多可能来自其他背景的具有相关知识的人员^[4]。

随着我国城乡规划成为一级学科，高等学校城乡规划学科专业指导委员会所编写的《高等学校城乡规划本科指导性专业规范（2013年版）》规定的城乡规划专业本科生所需掌握的知识点，增加了多项定量城市研究相关的内容，如城市系统分析方法、统计分析方法、城市模型建构方法和城市模拟技术等，我们期望这一举措，对促进定量城市研究在城乡规划与设计中的应用起到有益作用。

致 谢：本研究受到国家自然科学基金项目（编号 51408039）的资助，王斌喆对本文语言

进行了编辑，在此一并表示感谢。

参考文献

- [1] 刘伦, 龙瀛, 麦克·巴蒂. 城市模型的回顾与展望——访谈麦克·巴蒂之后的新思考 [J]. 城市规划, 2014, 8): 63-70.
- [2] 甄峰, 秦萧, 王波. 大数据时代的人文地理研究与应用实践 [J]. 人文地理, 2014, 29(3): 1-6.
- [3] BATTY M, AXHAUSEN K W, GIANNOTTI F, et al. Smart cities of the future [J]. The European Physical Journal Special Topics, 2012, 214(1): 481-518.
- [4] ZHOU J, WANG M, LONG Y. Big data for intrametropolitan human movement studies: A case study of bus commuters based on smart card data [M]. Beijing City Lab. Working paper #50. 2014.
- [5] LONG Y, LIU X, ZHOU J, et al. Profiling underprivileged residents with mid-term public transit smartcard data of Beijing [J]. arXiv preprint arXiv:14095839, 2014,
- [6] 龙瀛, 孙立君, 陶遂. 基于公共交通智能卡数据的城市研究综述 [J]. 城市规划学刊, 2015, 003): 70-7.
- [7] KANG C, LIU Y, MA X, et al. Towards estimating urban population distributions from mobile call data [J]. Journal of Urban Technology, 2012, 19(4): 3-21.
- [8] 钮心毅, 丁亮, 宋小冬. 基于手机数据识别上海中心城的城市空间结构 [J]. 城市规划学刊, 2014, 006): 61-7.
- [9] 王德, 朱玮, 谢灿栋, et al. 基于手机信令数据的城市空间分析框架、难点及初步进展 [M]. “智能城镇化”第 70 期可持续发展沙龙. 2014.
- [10] WANG J, MAO Y, LI J, et al. Predictability of road traffic and congestion in urban areas [M]. Beijing City Lab. Working paper #55. 2014.
- [11] 郭文伯, 张艳, 柴彦威, et al. 基于 GPS 数据的城市郊区居民日常活动时空特征——以北京天通苑, 亦庄为例 [J]. 地域研究与开发, 2013, 32(6): 159-64.
- [12] LONG Y, LIU X. Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest [J]. Environment and Planning B: Planning & Design, 2015,
- [13] LONG Y, HAN H, TU Y, et al. Evaluating the effectiveness of urban growth boundaries using human mobility and activity records [J]. Cities, 2015, 46(76-84).
- [14] 王波, 甄峰, 张浩. 基于签到数据的城市活动时空间动态变化及区划研究 [J]. 地理科学, 2015, 35(2): 151-60.
- [15] 陈映雪, 甄峰, 王波, et al. 基于社会网络分析的中国城市网络信息空间结构 [J]. 经济地理, 2013, 33(4): 56-63.
- [16] 熊丽芳, 甄峰, 王波, et al. 基于百度指数的长三角核心区城市网络特征研究 [J]. 经济地理, 2013, 33(7): 67-73.
- [17] 李栋. 基于互联网春运迁徙数据的城镇体系规划思考 [M]. 大数据与未来人居研讨会. 2015.
- [18] 秦萧, 甄峰, 朱寿佳, et al. 基于网络口碑度的南京城区餐饮业空间分布格局研究 [J]. 地理科学, 2014, 34(7): 810-7.
- [19] SHENG Q, LI M, ZHANG X. Location does not matter in the informational age?—a case study on the distribution of restaurants listed in 'dazhongdianping' in Beijing [M]. Beijing City Lab. Working paper #61. 2015.
- [20] LONG Y, WANG J, WU K, et al. Population Exposure to Ambient PM 2.5 at the Subdistrict Level in China [J]. Available at SSRN 2486602, <http://ssrncom/abstract=2486602> or

<http://dxdoi.org/102139/ssrn2486602>, 2014,

- [21] 龙瀛. 中国人口密度的时空演变与城镇化空间格局初探: 2000-2010 [M]. 城乡治理与规划改革——中国城市规划年会. 2014.
- [22] 袁媛, 许学强, 薛德升. 转型时期广州城市户籍人口新贫困的地域类型和分异机制 [J]. 地理研究, 2008, 27(3): 672-82.
- [23] 谌丽, 张文忠, 杨翌朝. 北京城市居民服务设施可达性偏好与现实错位 [J]. 地理学报, 2013, 68(8): 1071-81.
- [24] RATTI C, SOBOLEVSKY S, CALABRESE F, et al. Redrawing the map of Great Britain from a network of human interactions [J]. PloS one, 2010, 5(12): e14248.
- [25] BECKER R A, CACERES R, HANSON K, et al. A tale of one city: Using cellular network data for urban planning [J]. IEEE Pervasive Computing, 2011, 10(4): 18-26.
- [26] ROZENFELD H D, RYBSKI D, GABAIX X, et al.: National Bureau of Economic Research, 2009.
- [27] SAGL G, RESCH B, HAWELKA B, et al. From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments; proceedings of the Proceedings of the GI-Forum 2012: Geovisualization, Society and Learning, F, 2012 [C].
- [28] 龙瀛, 吴康, 王江浩, et al. 大模型: 城市和区域研究的新范式 [J]. 城市规划学刊, 2014, 006): 52-60.
- [29] 高晓路, 许泽宁, 牛方曲. 基于“点—轴系统”理论的城市群边界识别 [J]. 地理科学进展, 2015, 34(3): 280-9.
- [30] LONG Y, WU K, MAO Q. Simulating urban expansion in the parcel level for all Chinese cities [J]. arXiv preprint arXiv:14023718, 2014.
- [31] LONG Y, SHEN Y. Mapping parcel-level urban areas for a large geographical area [J]. arXiv preprint arXiv:14035864, 2014.
- [32] HAN H-Y, LAI S-K, DANG A-R, et al. Effectiveness of urban construction boundaries in Beijing: An assessment [J]. Journal of Zhejiang University Science A, 2009, 10(9): 1285-95.
- [33] LONG Y, GU Y, HAN H. Spatiotemporal heterogeneity of urban planning implementation effectiveness: Evidence from five urban master plans of Beijing [J]. Landscape and Urban Planning, 2012, 108(2): 103-11.
- [34] 徐毅松, 石崧, 范宇. 新形势下上海市城市总体规划方法论探究 [J]. 城市规划学刊, 2009, 2(10-5).
- [35] TIAN L, SHEN T. Evaluation of plan implementation in the transitional China: A case of Guangzhou city master plan [J]. Cities, 2011, 28(1): 11-27.
- [36] BAGCHI M, WHITE P. What role for smart-card data from bus systems? [J]. Municipal Engineer, 2004, 157(1): 39-46.
- [37] 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行 [J]. 地理学报, 2012, 67(10): 1339-52.
- [38] GONZALEZ M C, HIDALGO C A, BARABASI A-L. Understanding individual human mobility patterns [J]. Nature, 2008, 453(7196): 779-82.
- [39] READES J, CALABRESE F, RATTI C. Eigenplaces: analysing cities using the space-time structure of the mobile phone network [J]. Environment and Planning B: Planning and Design, 2009, 36(5): 824-36.
- [40] PHITHAKKITNUKON S, HORANONT T, DI LORENZO G, et al. Activity-aware map: Identifying human daily activity pattern using mobile phone data [M]. Human Behavior Understanding. Springer. 2010: 14-25.
- [41] ROTH C, KANG S M, BATTY M, et al. Structure of urban movements: polycentric activity and

- entangled hierarchical flows [J]. PLoS one, 2011, 6(1): e15923.
- [42] MASUCCI A P, STANILOV K, BATTY M. Limited urban growth: London's street network dynamics since the 18th century [J]. PLoS One, 2013, 8(8): e69469.
- [43] LONG Y, JIN X, YANG X, et al. Reconstruction of historical arable land use patterns using constrained cellular automata: A case study of Jiangsu, China [J]. Applied Geography, 2014, 52(67-77).
- [44] 林忆南, 金晓斌, 杨绪红, et al. 清代中期建设用地数据恢复与空间网格化重建: 方法与实证 [J]. 地理研究, 2015, 已接受.
- [45] CALABRESE F, PEREIRA F C, DI LORENZO G, et al. The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events [J]. 2010, 6030(22-37).
- [46] CALABRESE F, SMOREDA Z, BLONDEL V D, et al. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data [J]. PLoS one, 2011, 6(7): e20814.
- [47] SCHNEIDER C M, BELIK V, COURONN T, et al. Unravelling daily human mobility motifs [J]. Journal of The Royal Society Interface, 2013, 10(84): 1-8.
- [48] LIU L. C-IMAGE: city cognitive mapping through geo-tagged photos [D]; Massachusetts Institute of Technology, 2014.
- [49] 茅明睿, 储妍. 规划云平台: “互联网+”规划公众参与的实践 [R]. 北京市城市规划设计研究院. 2015.
- [50] 龙瀛, 沈尧. 数据增强设计——新数据环境下的规划设计回应与改变 [J]. 上海城市规划, 2015, 2(81-7).