引文格式：冯心怡,刘亦婷,肖智雄,等. 多源大数据支撑下的城市交通拥堵驱动力研究[J]. 地理空间信息, 2024, 22(7): 36–40. doi:10. 3969/j. issn. 1672–4623.2024.07.008

多源大数据支撑下的城市交通拥堵驱动力研究

冯心怡[1]，刘亦婷[1]，肖智雄[1]，俞　艳[1*]，张宸赫[1]，程雨森[1]
（1. 武汉理工大学 资源与环境工程学院，湖北 武汉 430070）

摘　要：利用手机信令数据、高德实时交通态势数据、POI 数据等多源大数据，以宁波市拥堵核心区域为研究单元，对比了普通最小二乘法回归模型、经典地理加权回归模型、多尺度地理加权回归（MGWR）模型的精度；并选用最优模型定量探究了宁波市拥堵核心区不同时段的交通拥堵驱动力。结果表明：①MGWR 模型对交通拥堵驱动力的拟合效果较好，且能反映不同驱动力间的空间异质性；②实时人口密度、路网密度、土地利用混合度、职住平衡指数对交通拥堵影响存在时空分异，公司分布密度对交通拥堵影响存在圈层性。研究结果可为大数据在交通拥堵领域的应用提供一定参考，为管理者分析交通拥堵根源提供依据。

关键词：多源数据；交通拥堵；空间异质性；MGWR
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672−4623（2024）07−0036−05

Research on the Urban Traffic Congestion Driving Force Supported by Multi−source Big Data

FENG Xinyi[1], LIU Yiting[1], XIAO Zhixiong[1], YU Yan[1], ZHANG Chenhe[1], CHENG Yusen[1]
(1. School of Resource and Environmental Engineering, Wuhan University of Technology, Wuhan 430070, China)

Abstract: By using multi-source big data such as mobile phone signaling data, real-time traffic data of Amap and POI data, taking the congestion core area in Ningbo as the research unit, we compared the accuracy of ordinary least squares regression model, classical geo-weighted regression model and multi-scale geo-weighted regression (MGWR) model, and selected the optimal model to quantitatively explore the traffic congestion driving force of different time periods in Ningbo's congestion core area. The results show that ①the MGWR model fits the traffic congestion driving force better and can reflect the spatial heterogeneity among different drivers. ②There are spatio-temporal differences in the effects of real-time population density, road network density, land-use mix and job-occupancy balance index on traffic congestion. There are circles in the effects of company distribution density on traffic congestion. This study can provide some references for the future application of big data in the traffic congestion field, and a basis for managers to analyze the root causes of traffic congestion.

Key words: multi-source data, traffic congestion, spatial heterogeneity, MGWR

随着我国城市化进程的不断加快，国内城市现有交通供需矛盾持续深化、交通拥堵现象严重。因此，研判城市内部常发性交通拥堵影响因素、提升城市通行效率，成为当前城市交通与城市高质量发展的主要任务之一。传统交通拥堵驱动力研究多基于现场数据观测，在经验的基础上主观确定研究区交通拥堵的驱动力因素[1]；而现有研究多集中于空间计量经济学视角，采用各类空间统计模型方法探究交通拥堵驱动力因素[2-5]，其中地理加权回归模型（GWR）能较好地反映局部空间异质性程度，但在同一空间不同参数下，忽略了不同变量的尺度异质性，无法凸显其对空间分异规律研究的优势。为解决该问题，Fotheringham A S[6]等提出了多尺度地理加权回归（MGWR）模型，可为不同变量自动选取适配带宽，从而提高模型的精确性；沈体雁[7]等通过分析二手住宅价格影响机制，将 MGWR 模型引入地学领域。至此 MGWR 模型在地学领域开始广泛应用，且其在相关研究中对不同变量表现出了更真实的解释力度[8]。本文以宁波市拥堵核心区域为例，在把握其交通拥堵态势的基础上，对比了最小二乘法（OLS）、GWR、MGWR 模型对交通拥堵驱动力的解释精度；再基于对比结果，选用最优模型开展交通拥堵驱动力探究，更加精确地讨论考虑空间

收稿日期：2023−01−17。
项目来源：大学生创新创业训练计划资助项目（S202210497328）。
第一作者简介：冯心怡（2001—），本科，研究方向为地理大数据应用与城市生态遥感，E−mail：302363@whut.edu.cn。
通信作者：俞艳（1976—），博士，研究方向为智能地学计算在土地资源管理和国土空间规划中的建模与应用，E−mail：yyhrose@whut.edu.cn。

异质性情况下的城市交通拥堵驱动力；最后结合驱动力分析结果，提出了缓解研究区常发性拥堵的建设性意见。

1　研究区概况与数据来源

1.1　研究区概况

宁波市地处长江三角洲南翼，是浙江省副省级城市和东南沿海主要港口城市之一。随着城市化进程的高速推进，宁波市不可避免地存在潮汐交通拥堵等交通问题，且常发于北仑区、鄞州区等区域。因此，本文以宁波市交通拥堵核心区域为研究对象，涵盖镇海区、北仑区、江北区、海曙区、鄞州区，并基于路网形态[2]将其划分为61个研究单元。对该区域存在的交通拥堵问题开展量化分析，可帮助管理者从多角度合力治理交通拥堵，且对研究同类区域交通拥堵具有一定示范意义。

1.2　数据来源与预处理

为更合理地探究研究区交通拥堵驱动力，本文采用中国联通手机信令数据、高德实时交通拥堵态势数据等实时可靠的数据源支撑后续研究。本文采用的行政区划数据和路网数据来源于OpenStreetMap网站，时间为2021年；基于Python爬虫程序，利用高德开放平台API直接请求交通态势信息，获取2021年4月12—18日宁波市含交通拥堵等级的离散点，再通过坐标纠偏、数据清洗、属性链接等手段，重新构建含交通等级的路网，具体交通状态对应的交通拥堵等级见表1；选取宁波市2021年11月中国联通手机信令数据作为分析基础，包括脱敏处理后的用户ID、驻留开始与结束时间、位置编号、所属区县以及是否为核心用户等，市域范围内平均每天有4 194 726.033条手机信令数据，共125 841 781条；基于高德开放平台提供的API接口，参考高德POI分类编码，采集科研院所、餐饮服务、公共设施、风景名胜、文化场所等对交通拥堵存在可能影响的POI数据，并进行坐标纠偏、数据清洗，最后共获得30 417条POI数据。

表1　交通拥堵延时指数等级划分标准

指标	畅通	缓行	拥堵	严重拥堵
拥堵延时指数	<1.5	1.5~1.8	1.8~2.0	>2.0
交通拥堵等级	1	2	3	4

2　研究方法

2.1　技术路线

本文基于高德实时交通态势数据、行政区域数据、手机信令数据，评估宁波市工作日和双休日交通午间、晚高峰时段的拥堵情况，以便更加精准地把握宁波市交通拥堵核心区。根据路网结构划分研究单元，对比建立OLS、GWR和MGWR模型并分析精度；选用最优模型进一步定量探究宁波市交通拥堵核心区域的拥堵驱动力及其空间分异特征。具体技术路线见图1。

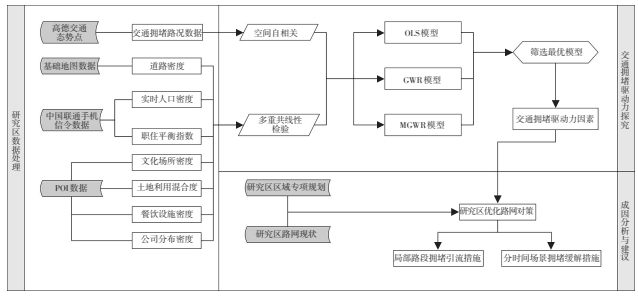


图1　交通拥堵驱动力研究技术路线

2.2　变量选取与模型构建

2.2.1　交通拥堵变量选取

对宁波市包含交通拥堵等级属性的路网进行核密度估计；结合路网分布情况，进一步将拥堵等级密度值大于2且路网较密集的区域确定为交通拥堵核心区；分别计算该区域工作日、双休日13：00和18：00交通拥堵等级密度平均值，并根据距离赋值到道路最近的研究单元内，得到宁波市交通拥堵核心区的分时段拥堵强度情况，等级越高、交通拥堵越严重。将宁波市拥堵核心区内各研究单元拥堵强度作为因变量，参考国内外相关研究成果[2-4]，并考虑数据可获得性、多重共线性和自变量空间自相关性等问题，最终选取路网密度、实时人口密度、职住平衡指数、土地利用混合度、餐饮服务场所密度、公司分布密度、文化娱乐场所密度等因素作为自变量开展对应研究（表2）。

表2　变量描述与检验情况表

影响因素	变量名称	方差膨胀因子	全局莫兰指数
人口分布影响	职住平衡指数	1.122	0.320***
	实时人口密度	2.231	0.471**
土地利用与开发	路网密度	1.578	0.582***
	土地利用混合度	1.507	0.136*
服务设施	餐饮服务场所密度	2.424	0.354**
	公司分布密度	1.367	0.433***
	文化娱乐场所密度	1.201	0.043**

注：*表示 $p<0.05$；**表示 $p<0.01$；***表示 $p<0.001$。

1）实时人口密度。基于前人研究[9]，本文主要以预处理的手机信令数据为基础，以8 h工作制为标准判定用户居住地和工作地，得到2021年11月宁波市交通拥堵核心区以2 h为间隔的实时人口密度、工作人口密度和居住人口密度。参考谢栋灿[10]等对职住平衡指数的

研究，利用工作人口密度与居住人口密度得到宁波市交通拥堵核心区各研究单元的职住平衡指数。

2）公共场所密度。本文基于清洗后的文化娱乐、公司企业、餐饮服务等POI数据，采用核密度估计方法得到宁波市交通拥堵核心区各类公共场所的密度。

3）路网密度。基于已获取的OSM路网数据，采用一定距离内道路长度与研究单元面积的比值[11]计算得到宁波市交通拥堵核心区内各研究单元的路网密度值。

4）城市土地利用混合度。信息熵的高低可反映城市土地利用的均衡程度[12]。参考相关学者对城市功能混合度的计算方法[13-14]，在研究区建立6 m×6 m的渔网，采用信息熵模型进行城市土地利用混合度的计算。其计算公式为：

$$H(X)=-\sum_{i-1}^{N}P_i\times\log P_i \tag{1}$$

式中，$H(X)$为随机变量熵值，即最终计算得到的城市土地利用混合度；P_i为X取X_i的概率，即各地块内不同类型POI的分布比率。

2.2.2 交通拥堵模型构建

1）GWR模型是空间统计学科的主要理论模型之一[15]，计算公式为：

$$y_i=\beta_0(u_i,v_i)+\sum_{k=1}^{p}\beta_k(u_i,v_i)x_{ik}+\varepsilon_i \tag{2}$$

式中，y_i和$x_{i1},x_{i2},\cdots,x_{ip}$分别为因变量$y$和自变量$x_1,x_2,\cdots,x_p$在样本点$i(u_i,v_i)$处的观测值；$\beta_0(u_i,v_i)$为第$i$个样本点的常数项；$\beta_k(u_i,v_i)$为第$k$个自变量在样本点$i$处的回归系数；$\varepsilon_i$为随机误差。

2）MGWR模型为不同空间尺度的自变量分配合适的带宽，能凸显自变量的空间异质性，计算公式为：

$$y_i=\sum_{k=1}^{p}\beta_{bwk}(u_i,v_i)x_{ik}+\varepsilon_i \tag{3}$$

式中，β_{bwk}为第k个变量回归系数使用的带宽。

3 研究结果与分析

3.1 交通拥堵现状分析

通过观测研究区不同时段的交通拥堵实时态势可知（图2），从整体拥堵强度来看，工作日高于双休日，且存在一定规律性，工作日各时段交通拥堵在空间上较集中，拥堵频发路段也相对稳定，双休日交通拥堵情况较离散，部分次干道会出现缓行特征；拥堵大多集中于鄞州区、北仑区等人流量较集中的区域，因此针对鄞州区、北仑区等宁波市交通拥堵核心区开展研究，能够有的放矢地探究其拥堵驱动力。

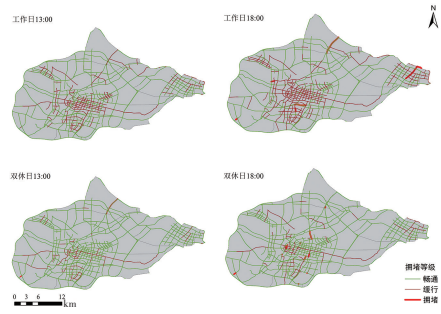


图2 研究区交通拥堵现状

审图号：GS(2020)4619

3.2 核心区交通拥堵驱动力因素分析

3.2.1 模型精度对比

本文采用标准化残差值检验模型整体效果，3类模型的标准化残差值均处于[-2.5,2.5]区间（图3）；采用空间自相关分析残差空间的分布情况，3类模型在10%的显著性水平上空间自相关程度较弱，因此可初步认为3类模型均可应用于交通拥堵驱动力研究。

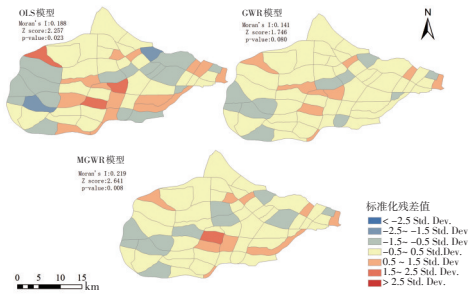


图3 3类模型标准化残差值空间分布

审图号：GS(2020)4619

进一步对3类模型进行精度对比，主要考察不同模型的拟合优度（R^2）、校正后赤池信息准则值（AICc）和残差平方和（RSS），结果见表3，MGWR模型的R^2能解释宁波市核心区交通拥堵变化的70%以上，取得了更加真实的解释力度；且MGWR的AICc和RSS均低于GWR、OLS模型，说明相较于OLS、GWR模型，MGWR模型具有更高的准确性，可

表3 OLS、GWR、MGWR的R^2、AICc和RSS对比

时段	R^2			AICc			RSS		
	OLS	GWR	MGWR	OLS	GWR	MGWR	OLS	GWR	MGWR
工作日13：00	0.650	0.664	0.707	123.014	128.642	122.772	18.853	16.375	13.691
工作日18：00	0.694	0.695	0.751	114.823	127.071	114.381	16.484	14.826	11.454
双休日13：00	0.671	0.685	0.723	119.190	123.134	119.263	17.708	15.351	12.926
双休日18：00	0.684	0.708	0.737	116.713	118.542	114.320	17.003	14.208	12.485

更好地应用于拥堵驱动力研究。

MGWR模型内不同时段不同交通拥堵驱动力的带宽不同，可更加精确地反映交通拥堵和各驱动力的作用关系；而GWR模型仅能反映交通拥堵驱动力作用尺度的平均值。对比不同变量间的带宽发现（表4），公司分布密度、土地利用混合度的截距作用尺度不同，存在一定的空间异质性；其余变量对交通拥堵的影响相对稳定。

表4　MGWR模型不同变量间带宽对比

变量	带宽	变量	带宽
职住平衡指数	60	公司密度	51
实时人口密度	60	餐饮密度	60
道路密度	60	文化场所密度	60
土地利用混合度	43		

3.2.2　交通拥堵驱动力因素分析

1）工作日午间时段交通拥堵驱动力分析。由图4可知，实时人口密度的影响占据主导地位，影响系数高值集中于镇海区与北仑区交界处，该区域既是发展较成熟的中央活动区，又邻近宁波市新兴科创园区，人口密度较大程度地影响了其交通拥堵情况；餐饮设施密度对交通拥堵呈正向影响，而文化场所密度对交通拥堵呈负向影响，作为提升居民生活质量、影响居民出行意愿的场所，服务性设施的空间位置、服务质量不均衡会导致居民长距离出行，造成交通拥堵；公司分布密度总体对交通拥堵呈负向影响，空间分布呈现自北向南的圈层性；土地利用混合度在午间时段对宁波市拥堵核心区的交通拥堵影响程度较弱且存在一定差异，在江北区、海曙区等区域土地利用混合程度越高、交通拥堵越小，而在北仑区、镇海区和鄞州区大部分区域则呈微弱正相关，过于多样性的用地功能会造成不同出行需求的涌入，进而造成拥堵；道路密度总体对交通拥堵呈负向影响，且空间分布存在随机性。

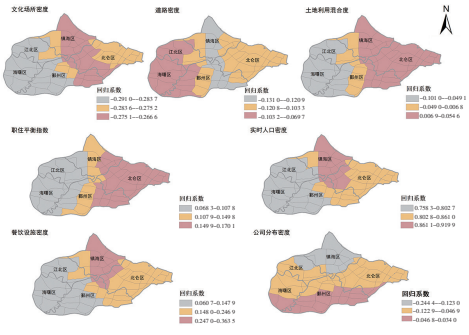


图4　工作日午间时段交通拥堵驱动力因素时空特征分布

审图号：GS(2020)4619

2）工作日晚高峰时段交通拥堵驱动力分析。由图5可知，实时人口密度在工作日晚高峰的影响系数总体呈下降趋势，且空间分布变化较明显，由于8 h工作制，居民倾向于向居住地流动，因此实时人口密度的影响高值由北仑区转向鄞州区，与午间时段的空间分布特征恰恰相反，存在一定潮汐性；道路密度对晚高峰时段的正向影响区域扩大，覆盖中央活动区全域，短期内道路密度增加，可在一定程度上为出行者提供不同出行选择，但长期来看，道路密度的增加会推动居民自发选择私家车出行，从而呈现出道路建设越多、拥堵越严重的趋势[16]；文化场所密度对交通拥堵呈正向影响且整体差异不大，空间上呈微弱的自西向东递减趋势；餐饮设施密度对交通拥堵呈微弱负向影响，总体空间分布特征与文化场所密度类似，服务性场所承接晚间高峰的大量人流，受自身功能、分布情况影响，在影响情况上与午间时段出现较大差异；公司分布密度对交通拥堵影响依旧呈现圈层性，其负向影响主要集中在鄞州区，正向影响主要集中在企业过度集中的镇海区、江北区；职住平衡指数和土地利用混合度对交通拥堵均呈现较弱的负向影响，在工作日晚高峰时段表现不明显。

3）双休日午间时段交通拥堵驱动力分析。由图6可知，实时人口密度总体差值不大，且高值聚集在邻近中心城区的北仑区；文化场所密度对交通拥堵的影响有所增强，且影响系数重心转向海曙区、江北区等居住人口聚集区；道路密度对交通拥堵呈正向影响，拥堵频发的鄞州区、镇海区和北仑区东部，路网密度过高、道路质量有待提高，交通拥堵难以疏解；相较于工作日，双休日宁波市交通拥堵核心区的职住平衡指数负向影响水平提高，表明该区域职住不平衡现象仍需解决；餐饮设施密度在影响水平上呈全域负相关，空间分布与文化场所密度类似；土地利用混合度、公司分布密度影响存在较大分异，海曙区、江北区的土地利用混合度已满足双休日居民出行需求，再增加将会带来交通拥堵，而北仑区、鄞州区等较活跃区域则需进一步提高土地多样性，以应对双休日的交

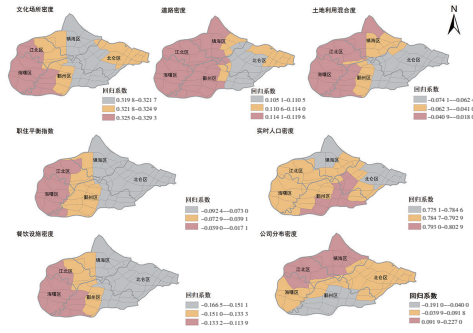


图5　工作日晚高峰交通拥堵驱动力因素时空特征分布

审图号：GS(2020)4619

通需求。此外，双休日由于企业文化影响，存在由加班导致的交通拥堵，西北部研究单元呈现对交通拥堵的正向影响。

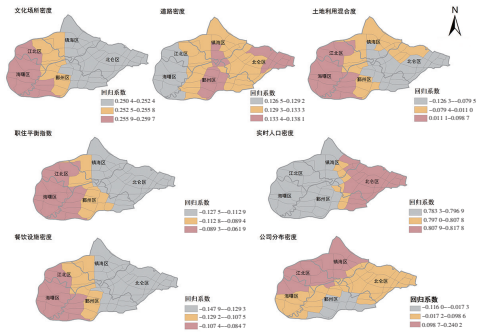


图6 双休日午间时段交通拥堵驱动力因素时空特征分布

审图号：GS(2020)4619

4）双休日晚高峰时段交通拥堵驱动力分析。由图7可知，实时人口密度对交通拥堵的影响减弱，且空间分布与午间时段存在一定相似性；文化场所密度在双休日晚高峰时段持续呈正向影响，影响系数高值仍分布在海曙区、江北区等；道路密度对交通拥堵的整体影响减弱，原因在于双休日居民出行可能存在一定随机性，且居民拥有较多的可支配时间，可选择错峰出行、公交出行；餐饮设施密度、土地利用混合度对大部分区域呈负向影响。

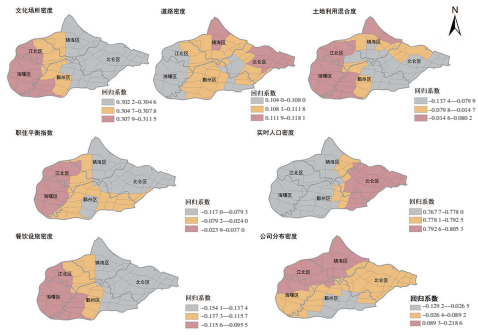


图7 双休日晚高峰交通拥堵驱动力因素时空特征分布

审图号：GS(2020)4619

4 结 语

本文基于实时交通拥堵态势数据、手机信令数据、POI数据等多源时空大数据，结合路网结构划分研究单元，通过对比OLS、GWR、MGWR模型发现，利用MGWR模型探究交通拥堵驱动力更加准确。基于该模型，本文重点分析了宁波市拥堵核心区的交通拥堵驱动力因素，主要结论为：

1）相较于OLS、GWR模型，MGWR模型的拟合性更好，在交通拥堵领域的应用效果更好。同时，MGWR模型可反映不同变量在不同空间范围内的影响程度，凸显空间异质性，对交通拥堵影响源的分析更加精细。

2）在交通拥堵驱动力因素方面，实时人口密度在各时段均呈显著正相关性，且工作日影响水平高于双休日；职住平衡指数与土地利用混合度在不同时段均存在圈层性，高值区主要集中在中央活动区附近，由于住大于职引发的长距离通勤，进一步导致职住平衡指数对中央核心区存在正向影响；路网密度对交通拥堵影响高值集中在路网较密集的区域；餐饮服务设施密度、文化场所密度对交通拥堵影响存在相反的空间分异特征，二者分别对双休日、工作日交通拥堵具有较强影响；公司分布密度影响程度较弱，高值集中于企业较集中的区域。

综合考虑上述因素，宁波市在未来仍需从城市规划、车流管理、公共服务设施管理等多角度合力开展拥堵治理。一方面，宁波市应适当建立多中心层级体系，减少通勤时间和远距离跨区流动，疏解中央活动区的人口压力和职住失衡；另一方面，在发展相对成熟的中央活动区，应适当疏解土地功能、提升道路质量、完善次干道和低等级路网建设，从而增强常发性拥堵路段的车流承载力，在道路基础设施仍需提升的甬江科创区，应提升土地利用混合度、合理增加道路密度，缓解环杭州湾高速带来的交通拥堵，餐饮、文化娱乐场所等服务设施建设应结合服务半径，合理布局、全面规划。

由于数据局限性，本文针对实时交通态势的观测时间尺度还有待进一步扩大，以减少偶发性拥堵造成的影响。此外，本文在驱动力因素的选取上还不够全面，未来将进一步从出行管理政策、公共服务设施均等化等多个视角开展深入探究。

参考文献

[1] 夏晶，黄承锋，宋融秋. 跨江城市交通拥堵的成因与治理[J]. 重庆交通大学学报(社会科学版)，2017，17(4)：41-45

[2] 牛强，杨超，顾重泰，等. 基于实时交通态势数据和地理加权回归的城市道路通畅性影响因素分析：以武汉市中心区工作日早高峰为例[J]. 现代城市研究，2022(3)：72-79

[3] 袁浩，任杰，杨睿，等. 考虑空间异质性的城市建成环境对交通拥堵状态的影响[J]. 武汉理工大学学报(交通科学与工程版)，2021，45(5)：840-845

[4] 孙斌栋，尹春. 人口密度对居民通勤时耗的影响及条件效应：来自上海证据[J]. 地理科学，2018，38(1)：41-48

[5] 张振龙，邱煜卿，蒋灵德，等. 基于实时路况的交通拥堵时空特征及其影响因素分析：以苏州古城区为例[J]. 现代城市研究，2020(1)：104-112

（下转第67页）