

上海市出租汽车出行时空分布规律研究

Spatiotemporal Distribution of Taxi Services in Shanghai

邓中伟¹, 季民河^{1,2}

(1.华东师范大学地理信息科学教育部重点实验室,上海 200062;2.华东师范大学中国东西部合作研究中心,上海 200062)

Deng Zhongwei¹, Ji Minhe^{1,2}

(1. Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200062, China; 2. China East-West Cooperation Research Center, East China Normal University, Shanghai 200062, China)

摘要: 为测定大样本视角下出租汽车出行时空规律,采用探索性空间数据分析(ESDA)技术进行相关研究。从上海市出租汽车调度系统的GPS数据提取单个运营日9 921辆出租汽车的OD样本,按街道和时段统计出租汽车OD的时空发生量,定义并计算各区域各时段的载客空间密度指标,运用ESDA技术分时段计算、分析全局和局部莫伦值(Moran's *I*)。全局分析结果表明,上海市出租汽车出行在全局分布上具有较强的空间正自相关性,这与城市中心商业区高度集中的活动分布有关。局部分析结果表明,上海市出租汽车出行分布呈“句”字形聚集结构。全局结构和局部结构在时间序列上变化较小,表明上海市出租汽车出行分布的时空结构明显且稳定。

Abstract: Through a large sample survey, this paper uses Exploratory Spatial Data Analysis (ESDA) to investigate the spatiotemporal distribution of taxi services in Shanghai. A GPS-based taxi service dataset is first retrieved to derive a total of 9 921 taxis' OD, which are then tallied by individual street districts. An index known as Traffic Ratio Density is computed to characterize the level of taxi services for each street district and to facilitate the

0 引言

出租汽车在城市居民出行中占重要地位,以上海市为例,出租汽车承担了居民出行总量的23%^[1]。因此,非常有必要研究出租汽车的使用情况及其对城市交通的影响。由于出租汽车没有固定的起讫点和运行路线,其定量分析问题一直是城市出租汽车交通分布预测、营运站点规划及相关政策制定的瓶颈。目前,相关研究大多以智能交通为背景,探讨出租汽车调度系统的具体实现技术及数据收集和处理方法^[2-4],或以出租汽车为浮动车数据源估算交通量、车速,并对道路网进行评价^[5]。对出租汽车使用情况的研究主要分为两类:1)通过构建或修改已有的理论模型,研究其出行分布,预测其出行分担率及拥有量^[6-7]。2)利

mapping of its spatiotemporal variation. In the end, the method of Exploratory Spatial Data Analysis is used to identify spatial clusters of taxi services over time. Both global and local Moran's *I* values are computed for Shanghai as a whole and for individual street districts. The positive values of the global index strongly suggests high and stable concentration pattern across all the time-periods. The local index shows that the taxi OD pattern has a high-density cluster in the CBD area, versus the low-density cluster in the suburban regions, and between them there is a stochastic distribution. No noticeable temporal variation at either global or local

level is identified, which indicates a rather stable spatial and temporal distribution of taxi service.

关键词: 出租汽车OD; 出行时空分布; 探索性数据分析; 莫伦值; GPS
Keywords: taxi OD; spatiotemporal distribution; Exploratory Spatial Data Analysis(ESDA); Moran's *I*; GPS
中图分类号: U491

文献标识码: A

收稿日期: 2010-10-04

基金项目:国家自然科学基金项目“被动式GPS跟踪数据中出行信息的智能化自动提取”(40771138)

作者简介:邓中伟(1984—),男,河南通许人,在读博士研究生,主要研究方向:交通地理信息系统、时空轨迹数据挖掘、时间地理学。

E-mail:dengwei555006@163.com

用经典统计方法分析其出行时长、出行距离、载客量等,认识其一般运营规律^[8-10];但是,由于经典统计方法以样本独立为基本假设,无法测度由城市空间结构引起的交通需求分布的结构性规律,因此,在对影响变量和观测结果的关系进行建模时会出现估计偏差。

一般来说,对混杂有空间效应的机理研究常采用空间统计分析方法,在模型构建和参数设定中考虑现象的空间自相关,其中空间效应的存在与否和程度高低,则可通过“探索性空间数据分析”方法(Exploratory Spatial Data Analysis, ESDA)识别^[11]。本文利用ESDA对出租汽车服务的空间分布格局进行量化分析、描述,并将其可视化,以揭示由数据的空间依赖性和空间异质性形成的空间结构,协助发现城市出租汽车需求的空间集聚和空间异常,解释并研究空间对象之间的相互作用机制。以上海市出租汽车调度系统的GPS数据为数据源进行案例研究,提取9 921辆出租汽车单个运营日的OD数据,定义载客空间密度指标,衡量出租汽车OD在时空上的分异,并对该分布结构做ESDA分析,从而探测出租汽车出行的时空分布模式和集聚规律。

1 探索性空间数据分析

探索性空间数据分析的基础是空间自相关概念,指研究区域内各地理位置表现的属性值与邻近位置属性的关系,其基本思想来源于地理学第一定律——任何事物都与周围的事物存在联系,并且与相近事物的联系更为密切^[12]。空间正自相关反映区域单元观测值的趋同聚集,即高值与高值、低值与低值之间趋于空间聚集;空间负自相关反映区域单元观测值的趋异集聚,即相似观测值之间呈分散的空间分布;空间自相关为0则表示数据之间是独立随机的。统计分析空间自相关的常用模型是莫伦统计指标Moran's I 、Geary c 比值,以及基于距离阈值范围的乘法测度空间对象相关性统计方法。

1.1 空间权重矩阵

空间权重矩阵的使用是空间统计学与传统统

计学的重要区别之一,可用于表示 N 个位置的空间邻接关系。该关系可根据邻接标准或距离标准度量,在形式上通常表现为一个二元对称矩阵:

$$W = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}, \quad (1)$$

式中: w_{ij} 表示区域 i 与 j 的邻接关系,其中 $w_{ii} = 0$ 。两种最常用的确定空间权重矩阵的规则是:

1) 简单的二进制邻接矩阵,主要用于面状分析区域。

$$w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 相邻接,} \\ 0, & \text{其他.} \end{cases} \quad (2)$$

2) 基于距离 d 的二进制空间权重矩阵,多用于点状分析对象。

$$w_{ij} = \begin{cases} 1, & \text{区域 } i \text{ 和 } j \text{ 之间的距离小于 } d, \\ 0, & \text{其他.} \end{cases} \quad (3)$$

1.2 Moran's I

Moran's I 可用来衡量相邻的空间分布对象及其属性值之间的关系,其取值范围为 $[-1, 1]$,正值表示空间事物的属性值具有正相关性,负值表示空间事物的属性值具有负相关性,0表示空间事物的属性值不存在空间自相关,即空间随机分布。计算公式为

$$I = \frac{n \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij}) \sum_i (y_i - \bar{y})^2}, \quad (4)$$

式中: n 为样本个数,即参与分析的街区数; y_i 或 y_j 为区域 i 或区域 j 的属性值,即 i, j 街区的属性值(如载客空间密度); $\bar{y}$ 为所有区域属性值的均值。

1.3 局部 Moran's I

在较大的空间尺度上,如城市或区域,其内部发展并非同质,从全局上衡量通常会掩盖空间内部的差异。因此相关文献设计了空间关联局部指标(Local Indicators of Spatial Association, LISA),比较典型的是文献[13]的G统计、文献[14]的O统计及文献[11]的局部 Moran's I 统计。其中,局部

Moran's I 统计能够有效检测区域内部由空间相关性引起的空间差异,并可判断空间对象属性取值的热点区域或高发区域,从而弥补全局空间检测的不足,较适于本文的数据分析。其计算公式为

$$I_i = \frac{y_i - \bar{y}}{S^2} \sum_{j=1}^n w_{ij}(y_j - \bar{y}), \quad (5)$$

式中: I_i 为第 i 个空间对象的局部自相关系数;

$$S^2 = \frac{\sum_{j=1, j \neq i}^n y_j^2}{N-1} - \bar{y}^2.$$

2 数据处理与结果分析

2.1 研究区域概况与数据源

2010年,上海市实有人口2 096万人,其中常住人口1 921.32万人;全市公共交通(含轨道交通)客运量1 289.9万人次·d⁻¹,出租汽车284.8万人

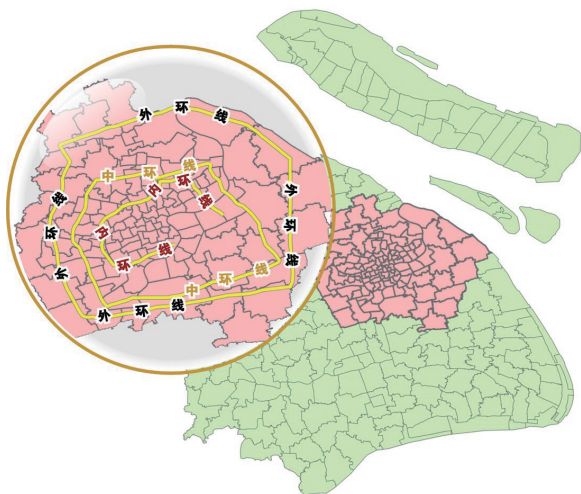


图1 研究区域示意图

Fig.1 Study area

次·d⁻¹, 共有出租汽车49 111辆^[1]。目前,上海市出租汽车一般都配有GPS和计价系统,并定期将车辆编号、时间、经纬度、运营状态等数据传送至调度中心,形成海量数据源,数据格式如表1所示。

研究区域为上海市外环线以内的区域(图1红色区域)。所使用数据来自4个主要的出租汽车运营公司,采集时段为2010年4月15日0:00—24:00,反映了出租汽车一个运营日的出行特征。原始数据中包含出租汽车18 976辆,标准记录时间间隔为20 s,共有56 597 588条位置记录。由于地面建筑物对GPS信号的遮挡及GPS设备自身的不稳定性,部分记录存在缺失、漂移、时间错乱等问题;另外,由于出租汽车驾驶人误操作,也产生了一些不完整记录。经过数据预处理,从原始样本中去除具有上述记录问题的车辆,最后形成可用于分析的车辆样本为9 921辆,总计36 858 724条记录。

2.2 出租汽车OD提取

在Oracle数据库环境下使用PL/SQL程序设计进行数据预处理,并按下述法则检索出租汽车运营状态字段,提取出租汽车OD,由此产生出行OD表。

法则一: 将标记为“2”的运营状态定为空载。这类数据占总样本的比例较小(约为3%),对统计结果影响不大。

法则二: 将运营状态从“0”或“2”转为“1”的点定为O点。

法则三: 将运营状态从“1”转为“0”或

表1 数据记录格式与含义

Tab.1 Data variables being collected by GPS

字段名	示例	备注
车辆编号	1030753	车辆的唯一标志
日期	2009-5-15	数据采集日期
时间	17:50:00	数据采集时间
经度	121.49347N	WGS84坐标系
纬度	31.310932E	WGS84坐标系
方向	7	与地理北极的夹角方向,共分为8个方向
运营状态	0,1,2	0表示空载,1表示载客,2表示状态不明

“2”的点定为D点。

处理中删除了起讫点时间间隔小于30 s和大于180 min的OD记录。前者主要是驾驶人误操作引起，后者大多因为GPS信号缺失导致很长一段时间没有数据记录，这些记录一律标记为伪记录，并从OD表中剔除。处理后，OD记录共计581 324条。最后在ArcGIS环境下将提取出的OD对与上海市街道区划图做多边形内点重叠(point-in-polygon overlay)分析，通过空间配准得到出租汽车OD记录归属的街区，并按街区统计出租汽车OD发生量。

2.3 载客空间密度计算

影响出行的外部因素主要有城市人口、居住与就业分布、土地利用格局、区位的交通可达性等。这些因素在城市内部空间的分布相对稳定，可作为预测出租汽车OD时空分布的变量。由于出租汽车的使用基于人的活动，而人的活动又受个人社会网络、家庭和社会角色及社会作息规则等方面的制约，从而使出租汽车OD呈现显著的时空变化。为了量测出租汽车OD随时间推移在空间上的变化特征，定义“载客空间密度”指标，即各街区载客空间密度与城市总载客空间密度之比。计算公式为

$$d_{ts} = \frac{c_{ts}}{A_s} \cdot \frac{1}{\sum_{k=1}^n (\frac{c_{tk}}{A_k})}, \tag{6}$$

式中： t 为时段标志； s, k 为街区标志； d_{ts} 为街区 s 在时段 t 的载客空间密度； c_{ts} 为 t 时段 s 街区中的载客次数； A_s, A_k 分别为 s, k 街区的面积； $\frac{1}{\sum_{k=1}^n (\frac{c_{tk}}{A_k})}$ 为所有街区载客次数与其面积比

值的总和。计算出的 d_{ts} 为归一化指标，取值范围为[0, 1]。以2 h为间隔将一个运营日划分为12个时段，分析不同时段载客出租汽车在空间上的变化。

2.4 出租汽车OD的空间格局分析

利用美国亚利桑那大学地理空间分析与计算中心开发的空间统计分析软件GeoDa分析上海市

出租汽车OD的时空格局^[15]。分析单元为面状街区，根据经验和通用做法，采用式(2)的邻接法构建空间权重矩阵，并以载客空间密度为变量计算上海市中心城区的全局和局部Moran's I ，分析其在空间上是否存在聚集现象和热点区域。对局部Moran's I 进行假设检验时，GeoDa软件采用的是蒙特卡罗空间排列仿真方法判断结果的显著水平，模拟次数越多，要求显著水平越高。本文采用的模拟次数统一为999次，统计检验的显著水平为0.05。

2.4.1 全局Moran's I

表2为上海市分时段全局Moran's I 计算结果。总体来看，载客空间密度存在显著的空间正自相关，表明出租汽车服务具有时空非随机性，相似密度值具有空间集聚趋势，即高载客密度区域相连成片，低载客密度区域互为邻居的空间关联结构。分时段来看，0:00—8:00的Moran's I 相对较小(0.443~0.579)，出租汽车分布相对分散，这主要是由于凌晨活动较少，客源零星分布在酒店和少数夜生活区，而出租车驾驶人一般住在郊区，这段时间更倾向于搭载与回家方向相同的乘客；6:00—8:00多为通勤和通学客流，由于城市住宅的郊区化，此时段的出行多以家为起点，导致出租汽车服务较为分散。8:00—16:00的工作时段，出租汽车服务表现了极强的空间集聚性(0.723~0.734)，这主要是因为与业务有关的所有

表2 上海市出租汽车分时段载客空间密度
全局 Moran's I

Tab.2 Global Moran's I of Shanghai taxi service distribution in 12 time slots

时段	I
0:00—2:00	0.471
2:00—4:00	0.564
4:00—6:00	0.443
6:00—8:00	0.579
8:00—10:00	0.723
10:00—12:00	0.734
12:00—14:00	0.730
14:00—16:00	0.725
16:00—18:00	0.665
18:00—20:00	0.721
20:00—22:00	0.644
22:00—24:00	0.628

出行几乎都发生在办公和商务活动集中的城市中心区域。16:00—18:00 Moran's I 的减小与 6:00—8:00 的原因相似, 主要为通勤和通学客流, 并且此时段市中心车流缓慢, 出租汽车驾驶人也倾向于向中环、外环扩散。18:00—20:00 出现了一个出租汽车服务在空间聚集的小高峰, 表明文化娱乐场所和夜生活相对集中的商业中心区吸引了某些人群(如居住在中心区周边的单身白领), 而另一部分居民(如已婚有子, 往往居住在边远区域)几乎没有外出的动机和理由。20:00—24:00, 随着夜生活的结束和整个城市休息时间的到来, 空间集聚度逐渐下降(0.628~0.644), 凌晨降至最低水平(0.471)。

综上, 上海市出租汽车服务的时空分布变化既与城市功能区的分布格局有关, 也与城市作息制度以及居民的社会经济属性相关, 而前者的影响较大。

2.4.2 局部 Moran's I

利用 GeoDa 对运营日 12 个时段分别做 LISA 分析, 采用 95% 的置信水平对局部 Moran's I 进行筛选, 得到分时段局部 Moran's I 的模拟结果, 见图 2。

以红、蓝、浅红、浅蓝表示下述四类局部空间关联模式: 1) 高-高, 高观测值的区域单元被高值区域所包围; 2) 低-低, 低观测值的区域单元被低值区域所包围; 3) 低-高, 低观测值的区域单元被高值区域所包围; 4) 高-低, 高观测值的区域单元被低值区域所包围。从 LISA 所表征的空间意义来看, 四种关联模式分别代表了出租汽车服务分布的聚集区(高-高)、稀疏区(低-低)、空心区(低-高)和孤岛区(高-低)。聚集区是居民活动频繁的区域, 也是出租汽车出行的主要发生区域; 稀疏区与聚集区相反; 空心区的出租汽车服务分布密度低于周围区域, 一般是居民活动相对较少的地区, 如中年已婚有家和退休居民占多数的区域; 孤岛区与相邻区域相比出租汽车分布密度较高, 多为新开发且公共交通使用不便的区域。白色区域代表该研究单元与周围区域不存在空间自相关现象, 即处于随机状态。

12 个时段的出租汽车出行分布在整体上均呈

“句”字形聚集结构, 虽随时段不同在空间形态上略有变化, 但基本形状固定。“句”字中间的“口”形区域为出租汽车出行的聚集区, 表明上海市居民活动的时空结构具有明确的单核中心, 该中心与居民印象中的市中心相吻合; “句”字形外围的半包围区域为出租汽车出行的稀疏区; 中间区域的空隙区域则表现为不稳定的随机状态, 该状态与出租汽车本身的随机性有一定关系, 也可能是研究尺度(乡镇)过大弱化了区域间的趋同趋势。

“句”字形结构与居民印象中的“回”形或环形结构不同, 是不完整的圈层结构。未闭合的区域为外环线西南部的华漕镇、虹桥路、虹桥镇、七宝镇、莘庄镇、梅龙镇、龙华乡, 这些区域是受益于城市轨道交通和城市扩张而快速城镇化的郊区, 其中华漕镇、七宝镇是旅游休闲的景区, 人流量大; 莘庄镇毗邻虹桥机场, 是上海市对外交通的出入口, 出租汽车出行发生量和出行变化较大, 未出现低-低集聚。在随机区域, 宜川新村、曹杨新村、真如镇的载客空间密度比例明显高于周边区域, 形成了孤岛区, 其原因是这几个乡镇是所在区域的商业中心(如曹杨新村)或是上海市对外交通的货运中心(如真如镇), 客运量较大, 出租汽车的出行发生量明显高于周边区域。

3 结语

探索性空间数据分析技术为交通与城市时空结构关系研究提供了新的视角。全局 Moran's I 的结果表明上海市出租汽车分布具有明显的集聚结构, 但集聚程度随时间略有不同, 非工作时间相对分散, 工作时间相对集中, 这种居民活动空间分布特征与上海市城镇化过程中的职住分离有关。局部 Moran's I 的结果表明, 整个运营日内市中心区域出租汽车服务均处于高集聚状态, 形成居民活动分布的单核中心; 中环区域大多处于随机状态, 没有形成稳定的时空集聚结构, 但在部分新兴的区域商业中心及对外货运中心, 出租汽车出行发生量明显高于周边区域, 形成孤岛区; 居民沿外环区域的活动较少, 出租汽车的分布多

呈低-低集聚，但在受益于轨道交通城镇化进程较快的郊区出现了类似中环区域的随机状态。总体

来说，整个集聚结构在时间序列上差异不大，表明上海市居民生活时空结构明显且稳定。

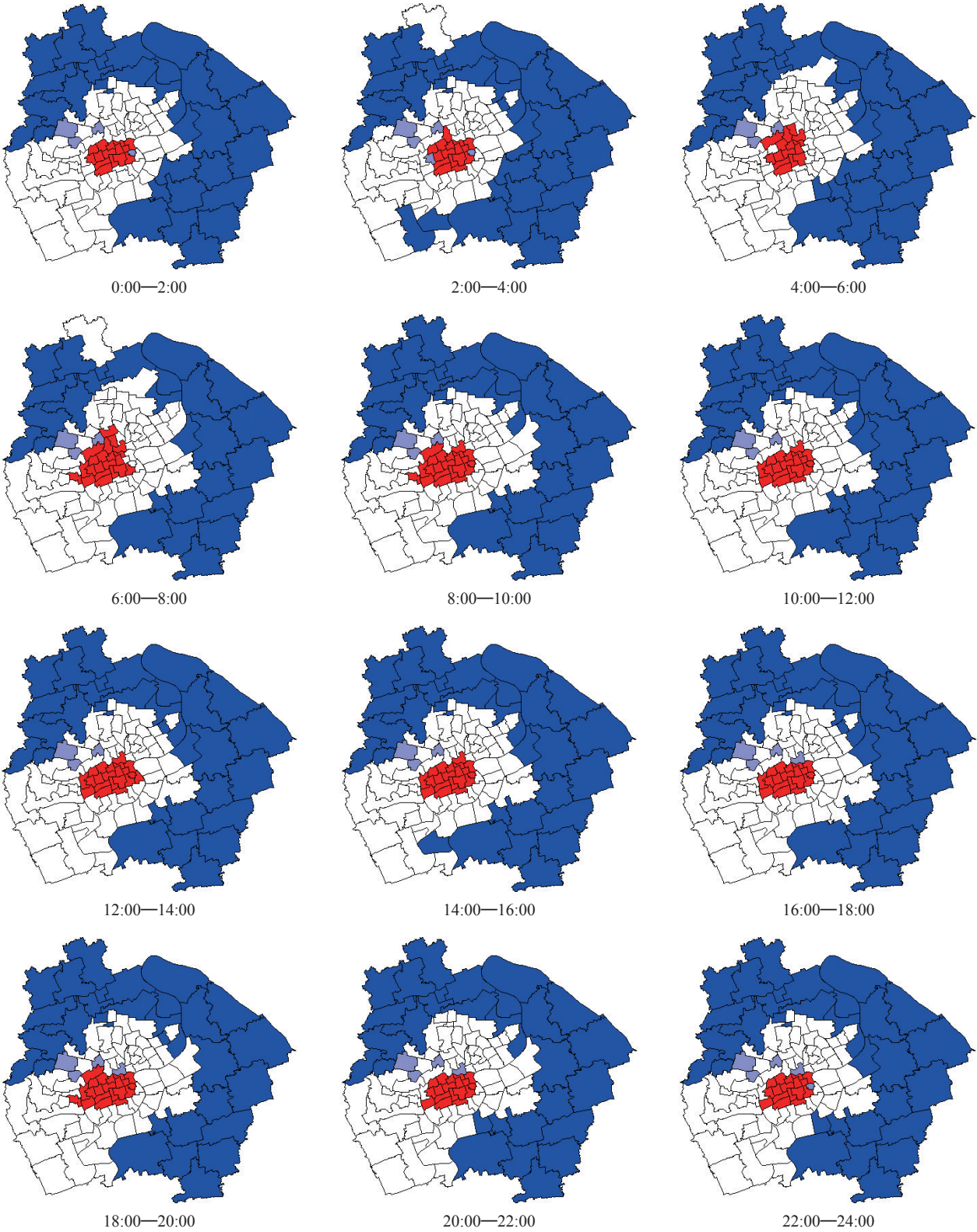


图2 上海市出租汽车分时段载客空间密度局部Moran's I
Fig.2 Local Moran's I for spatiotemporal distribution of taxi service in Shanghai

参考文献:

References:

- [1] 上海市统计局. 上海统计年鉴 2008[DB/CD]. 北京: 中国统计出版社, 2008.
- [2] 张存保, 杨晓光, 严新平. 基于浮动车的交通信息采集系统研究[J]. 交通与计算机, 2006, 24(5): 31-34.
Zhang Cunbao, Yang Xiaoguang, Yan Xinping. Traffic Data Collection System Based on Floating Cars[J]. Computer and Communications, 2006, 24(5): 31-34.
- [3] 郭继孚, 温慧敏, 陈峰. 浮动车系统功能分析与应用设计研究[J]. 交通运输系统工程与信息, 2007, 7(3): 39-44.
Guo Jifu, Wen Huimin, Chen Feng. Function Analysis and Application Design of Floating Car System[J]. Journal of Transportation Systems Engineering and Information Technology, 2007, 7(3): 39-44.
- [4] 钱寒峰, 林航飞, 杨东援. 浮动车车速处理分析系统中的数据融合技术[J]. 计算机工程与应用, 2007, 43(31): 230-232.
Qian Hanfeng, Lin Hangfei, Yang Dongyuan. Data Fusion Technique in Floating Car Processing Analysis System to Obtain Vehicle Speed[J]. Computer Engineering and Applications, 2007, 43(31): 230-232.
- [5] 辛飞飞, 陈小鸿, 林航飞. 浮动车数据路网时空分布特征研究[J]. 中国公路学报, 2008, 21(4): 105-110.
Xin Feifei, Chen Xiaohong, Lin Hangfei. Research on Time Space Distribution Characteristics of Floating Car Data in Road Network[J]. China Journal of Highway and Transport, 2008, 21(4): 105-110.
- [6] 边扬, 王伟, 陆建, 等. 城市出租车出行方式分担率预测方法研究[J]. 交通运输系统工程与信息, 2006, 6(2): 95-100.
Bian Yang, Wang Wei, Lu Jian, et al. A Forecasting Method for Urban Taxi Traffic Mode Split Rate[J]. Journal of Transportation Systems Engineering and Information Technology, 2006, 6(2): 95-100.
- [7] 王昊, 王伟, 陈峻, 等. 城市出租车交通分布预测模型[J]. 公路交通科技, 2006, 23(6): 145-156.
Wang Hao, Wang Wei, Chen Jun, et al. A Forecasting Model of Trip Distribution for Urban Taxis[J]. Journal of Highway and Transportation Research and Development, 2006, 23(6): 145-156.
- [8] 李艳红, 袁振洲, 谢海红, 等. 基于出租车OD数据的出租车出行特征分析[J]. 交通运输系统工程与信息, 2007, 7(5): 85-89.
Li Yanhong, Yuan Zhenzhou, Xie Haihong, et al. Analysis on Trips Characteristics of Taxi in Suzhou Based on OD Data[J]. Journal of Transportation Systems Engineering and Information Technology, 2007, 7(5): 85-89.
- [9] 胡小文, 冯均佳. 基于GPS数据采集的出租汽车交通运行特点研究[J]. 城市交通, 2007, 5(2): 91-95.
Hu Xiaowen, Feng Junjia. Research on Characteristics of Taxi Traffic Based on GPS Data[J]. Urban Transport of China, 2007, 5(2): 91-95.
- [10] 扈中伟. 基于浮动车技术的北京市出租车运营管理研究[D]. 北京: 北京工业大学, 2007.
Hu Zhongwei. Study on Beijing Taxi Operation Management Based on Floating Car Technology [D]. Beijing: Beijing University of Technology, 2007.
- [11] Anselin Luc. Local Indicators of Spatial Association-LISA [J]. Geographical Analysis, 1995, 27(2): 93-115.
- [12] Harvey J Miller. Tobler's First law and Spatial Analysis [J]. Annals of Association of American Geographers, 2004, 94(2): 284-289.
- [13] Getis A, Ord J K. The Analysis of Spatial Association by the Use of Distance Statics[J]. Geographical Analysis, 1992, 24(3): 189-206.
- [14] Ord J K, Getis A. Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation[J]. Journal of Regional Science, 2001, 41(3): 411-432.
- [15] Anselin L. GeoDa 0.9 User's Guide. Spatial Analysis Laboratory (SAL)[R]. Urbana: Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, 2003.