

引用格式: 林文棋, 陈会宴, 谢盼, 等. 基于多源数据的北京市朝阳区人口时空格局评估与预测[J]. 地球信息科学学报, 2018, 20(10): 1467-1477. [Lin W Q, Chen H Y, Xie P, et al. Spatial-temporal variation evaluation and prediction of population in Chaoyang district of Beijing based on multisource data[J]. Journal of Geo-information Science, 2018, 20(10): 1467-1477.] DOI:10.12082/dqxxkx.2018.180224

基于多源数据的北京市朝阳区人口时空格局评估与预测

林文棋^{1,2}, 陈会宴^{1*}, 谢盼¹, 李颖¹, 陈清凝¹, 李栋¹

1. 北京清华同衡规划设计研究院有限公司, 北京 100085; 2. 清华大学建筑学院, 北京 100084

Spatial-temporal Variation Evaluation and Prediction of Population in Chaoyang District of Beijing Based on Multisource Data

LIN Wenqi^{1,2}, CHEN Huiyan^{1*}, XIE Pan¹, LI Ying¹, CHEN Qingning¹, LI Dong¹

1. Beijing Tsinghua Tongheng Urban Planning and Design Institute, Beijing 100085, China; 2. School of Architecture, Tsinghua University, Beijing 100084, China

Abstract: Urban population distribution and activities are always the hot research topics. Identifying the spatial-temporal variation and predicting future trends are of great significance for estimating population accurately, making policy effectively, and warning of population booming timely. With the availability of data and the development of data processing technique, multisource data with both spatial and temporal features, such as mobile signaling data, have been used in population studies. In this paper, q-statistic was firstly applied as an exploratory analysis, then Bayesian spatial-temporal models were used to evaluate patterns of urban population and make prediction of future trends. The Chaoyang, Beijing in 2017 was selected as empirical study of this model. The spatially stratified heterogeneity was detected by q-statistic in Geodetector firstly. Then we explored the overall spatial variation, overall time trend and the departures of the local trends from the overall trend of resident population in Chaoyang by use of Bayesian spatial-temporal hierarchical model. Secondly, we applied Bayesian Gaussian predictive process to predict the resident population in December of 2017 by incorporating other relevant influential factors. The results show the perfect spatial stratified heterogeneity for resident population in Chaoyang, and the overall spatial variation demonstrates an increasing trend of population from center to the outside along the main ring road in Beijing. The overall time trend is still growing all over Chaoyang district, while the local trends, which departure from the overall trend of resident population, are different between each sub-districts in Chaoyang. Moreover, the spatial distribution of predicted resident population shows a high consistency with the observed resident population, and the prediction accuracy can be well accepted on the scale of Chaoyang district. However, prediction accuracy shows obvious difference on scale of sub-districts, with the worst prediction accuracy in the capital airport area. These findings show that Bayesian hierarchical model and Bayesian Gaussian predictive process are reliable in empirical study of

收稿日期: 2018-05-03; 修回日期: 2018-07-04.

基金项目: 国家社会科学基金项目(14BGL149); 清华大学自主科研计划项目。[**Foundation items:** National Social Science Fund Project, No.14BGL149; Tsinghua University Initiative Scientific Research Program.]

作者简介: 林文棋(1969-), 男, 福建人, 博士, 副教授, 主要从事城市研究及城乡规划实践。E-mail: linwq@tsinghua.edu.cn

*通讯作者: 陈会宴(1991-), 女, 河南人, 硕士, 主要从事城市时空分析研究。E-mail: chenhuiyan_12@163.com

population evaluation and prediction by effective application of multisource spatial-temporal data. Researches in this paper can be an excellent theoretical and practical support for mining multisource spatial-temporal data and assisting multiscale analysis with Bayesian spatial-temporal model, and provide an important basis for population controlling and early warning in urban population management.

Key words: multisource data; q-statistic; Bayesian hierarchical model; gaussian predictive process; spatial-temporal variation; population prediction

***Corresponding author:** CHEN Huiyan, E-mail: chenhuiyan_12@163.com

摘要:城市人口分布与活动呈现高度的时空动态变化,掌握人口时空变化特征并进行未来预测,对于精准人口评估、有效的政策措施制定、实时的人口预警与调控等具有重要意义。本研究利用以手机信令数据为主的多源时空数据,首先利用地理探测q统计进行探索性数据分析,其次结合贝叶斯模型进行北京市朝阳区居住人口的时空变化探究及时空预测,以期达到对朝阳区人口的动态评估与预测。首先,选用地理探测q统计进行空间异质性探测,用贝叶斯时空层次模型探究基于手机信令数据推算的北京市朝阳区居住人口的总体空间效应、总体时间效应以及局部变化趋势;其次,选用贝叶斯高斯预测过程模型,基于朝阳区各街乡的居住人口及相关人口影响因子数据进行朝阳区各街乡2017年12月的居住人口预测。时空探究表明:朝阳区居住人口在空间上存在完美空间分异,整体呈现沿环路由内向外递增的空间分布格局,整体时间趋势表现为增长,各街乡局部时间变化趋势呈现一定差异。预测的空间分布与实测空间分布整体一致,精度较高,各街乡预测精度不一。结果表明基于贝叶斯理论的时空层次模型和高斯预测过程模型可以为多源时空数据下的多尺度精准识别与人口时空模式挖掘提供有效的方法支撑。

关键词:多源数据;q统计;贝叶斯层次模型;高斯预测过程;时空格局;人口预测

1 引言

随着城市的快速扩张,“大城市病”日益突出,北京提出非首都功能疏解,其中最重要的目标之一就是人口疏解。由于城市人口分布与活动呈现高度的时空动态变化,因此首先掌握辖区内人口分布特征与动态变化情况对于提出更优更具针对性的人口疏解措施具有重要意义;其次考虑到人口疏解是一个长期的动态过程,为了提早获知目标区域的人口情况,做到未雨绸缪、提前防范,对于切实解决目标问题具有重要意义。

在城市人口领域研究中,传统研究主要依托于人口统计数据,主要包括统计年鉴、人口普查等其他人工调查数据,存在更新时效慢、空间粒度粗、获取成本高等特征。随着互联网、物联网、计算机等技术的发展,以手机信令数据为代表的多源时空大数据以其时效性高、粒度细、人力成本低等优势弥补了传统人口数据的缺陷,蕴藏的多源、多粒度时空信息可以为高精度的人口时空格局评估与预测提供可靠的数据来源,为城市地理的多尺度精准理解与时空模式识别提供成熟的条件,从而帮助深化多源大数据的时空解析,丰富城市地理的研究思路与方法^[1-2]。

现有的人口时空格局研究方面,不论是基于手机信令数据的人口聚集区识别^[3]、人口分布特征分析^[4]、动态人口分布制图^[5]、群体活动时空特征识别^[6]、

昼夜空间动态分布研究^[7],还是基于计量模型进行的人口空间结构演化特征探究^[8],都没有综合考虑时间相关性、空间相关性以及时空交互作用的影响进行研究区人口在时间、空间和时空交互维度的特征探究。在人口预测方面,主要侧重于大区域的人口总量或人口结构来预测^[9-14],也有学者基于小尺度的人口空间化进行人口的模拟预测,如人口分布表面建模方法^[15]、基于智能算法的3G方法(GIS, GP-Genetic Programming-遗传规划, GA-Genetic Algorithms-遗传算法)^[16]、深度神经网络^[17]、CA-Markov建模方法^[18]。但是人口分布表面建模方法的关键在于确定合理的贡献值分布表面,而贡献值分布表面需要结合各种历史信息、专家知识等辅助资料得到,因此在辅助资料不足的情况下,该方法将难以得到较为接近真实情况的人口分布表面。智能算法虽然自动化程度高,模型构建灵活,但是结果可控性差、参数设置复杂。而CA-Markov模型基于人口密度等级的未来人口模拟是一种时间离散、状态离散的动力学模型,并不适用于连续的模拟预测。在时空大数据兴起的同时伴随着时空建模技术的发展与革新,其中尤以基于贝叶斯的时空模型能够融合不同来源的数据、信息,充分考虑各种不确定性,对时空变异规律不确定的时空数据进行建模,同时可以通过不同的先验分布来达到引入专家知识的目的,其在空间流行病学^[19-21]、犯罪学^[22-23]、空气质量研

究^[24-27]、生态学^[28]等领域都有广泛的应用。在人口研究方面,李俊明^[29]运用贝叶斯时空层次模型进行中国老龄化时空演化研究,李原^[30]运用贝叶斯层次时空模型进行省际人口流入分析的研究,指示了利用贝叶斯时空模型进行人口时空探究的可能性。其能够充分考虑人群时间随机效应和空间异质性,融合不同来源的数据、信息,使模型的估计更准确。

综上,本文采用包括手机信令数据在内的多源时空数据进行朝阳区人口时空格局评估与预测,主要包括:①采用地理探测 q 统计进行北京市朝阳区居住人口的空间分异性探索性分析,在此基础上结合贝叶斯层次模型探究其时空变化趋势;②基于手机信令数据和其他多源时空数据,并结合贝叶斯高斯预测过程模型进行北京市朝阳区居住人口的时空预测。

2 研究区概况与数据源

2.1 研究区概况

本文以北京市朝阳区为研究区域(图1)。朝阳区是北京市城六区重要组成部分之一,位于城六区东部,毗邻北京通州城市副中心。朝阳区是北京城

六区中面积最大(总面积470.8 km²)、常住人口规模最大(2016年常住人口385.60万人)的区域,现辖43个街乡(24个街道,19个地区)。人口规模过大使得北京市特别是城六区面临交通拥堵、空气污染、公共服务不足等“大城市病”。在北京市“非首都功能疏解”背景下,朝阳区作为北京市常住人口规模最大的区域面临着人口疏解与调整的迫切需求。朝阳区街乡数量较多,且各街乡之间人口特征差异较大,还与通州首都副中心存在重要的区位联系,常住人口与外来人口结构复杂。针对朝阳区巨大人口规模、复杂人口结构等特征,探究其人口时空格局并进行未来人口预测,对于精准掌握人口动态变化,有效的政策措施制定,实时的人口预警与调控等具有重要意义。

2.2 数据源

本文以北京市朝阳区为研究区,最小研究单元为朝阳区下辖的43个街乡(含24个街道和19个地区),研究时段为2017年1-12月。采用数据包括根据移动运营商手机信令数据推算的朝阳区居住人口和活动人口数以及滴滴出行数据、房屋交易数据、兴趣点POI数据。①居住人口是由移动运营商根据朝阳区覆盖范围内的基站小区所记录的每个



图1 研究区域
Fig. 1 Study area

月内有15 d及以上,每天在研究范围的非工作时间停留10 h及以上的人数,根据朝阳区街乡行政边界统计各街乡各月的居住人口数。②活动人口是移动运营商根据朝阳区覆盖范围内的基站小区所记录的每15 min出现的人数,根据朝阳区街乡行政边界统计各街乡的各月平均活动人口数。③滴滴出行数据由滴滴平台提供的出租车、快车、专车等的出行OD数据,表征属性有出行的次数和距离,根据出发地点和到达地点所在的朝阳区各街乡行政单元,去除重复后,统计在各街乡行政单元每月所发生的滴滴出行的平均次数和距离。④房屋交易数据是由链家住房网站获取的每月的住房交易量、交易价格、租房价格数据。根据各小区的经纬度坐标,统计发生在朝阳区各街乡行政单元每月的平均住房交易量、交易价格、租房价格。上述数据在空间尺度上均以朝阳区43个街乡行政单元为基准,时间尺度上为2017年1-12月各自然月,数据长度为 $43 \times 12 = 516$ 。⑤兴趣点POI数据由高德提供,包含汽车服务、餐饮、购物、生活服务、娱乐场所、医疗、酒店、风景、政府机关、教育等20个大类。根据本文应用需求,选取交通设施POI、基础设施POI、就业设施POI、居住设施POI 4类,根据各POI所在经纬度,统计计算朝阳区各街乡行政单元各类POI数量。

3 研究方法

3.1 时空格局评估

(1) 地理探测 q 统计

空间分层异质性是地理数据普遍具有的特性,地理探测 q 统计是用于挖掘空间异质性的新工具,表达式为:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \quad (1)$$

式中: $h=1, \dots, L$, 为变量 Y 的分层(strata), 即分类或分区, 本文表示分街乡; N_h 和 N 分别是层 h 和全区的单元数; σ_h^2 和 σ^2 分别是变量 Y 在层 h 和全区的方差。 $SSW = \sum_{h=1}^L N_h \sigma_h^2$ 和 $SST = N \sigma^2$ 分别表示层内方差之和(Within Sum of Squares)及全区总方差(Total Sum of Squares)。 q 值域为[0, 1], 值越大说明 Y 的空间分异越明显, q 值为0则表明 Y 的空间分布呈随机状态^[31]。

(2) 贝叶斯时空层次模型

贝叶斯时空模型是在贝叶斯统计思想框架下

发展而来。基于贝叶斯理论的时空建模方法能够综合应用总体信息、样本信息和先验信息等历史知识推断参数的后验分布,既能给出一定程度的确定性,又同时保留数据的不稳定性^[32]。模型可以考虑时空过程中的空间随机过程和时间随机过程,同时加入基于时空变化的解释变量,使得对时空过程的估计更加合理^[33]。同时,基于马尔可夫链蒙特卡洛模型,可以方便快速的对高维、复杂的函数进行积分计算,使得利用贝叶斯进行统计推断更为快捷^[34]。

其中,贝叶斯时空层次模型可以将时空过程分解为总体空间效应、总体时间效应,还可以考虑时空交互效应,使得空间效应随时间推移而变化。基于此,可以利用贝叶斯时空层次模型进行北京市朝阳区居住人口时空格局探究。具体模型分为3个层次,分别定义为:

$$\text{数据模型: } y_{it} \sim P(y_{it} | \mu_{it}, \Theta) \quad (2)$$

$$\text{过程模型: } (\mu_{it} | \Theta) \sim \alpha + s_i + (b_0 t^* + v_t) + b_{it} t^* + \varepsilon_{it} \quad (3)$$

$$\text{参数模型: } \Theta \sim P(\Theta) \quad (4)$$

式中:数据模型中 Y_{it} 为街乡 $i(i=1, \dots, n, n=43)$ 在第 $t(t=1, \dots, T, T=12)$ 月的时空观测数据样本值,即 $i(i=1, \dots, n, n=43)$ 街乡 $t(t=1, \dots, T, T=12)$ 月的居住人口数; μ_{it} 为时空因变量,为 y_{it} 的时空过程转换参数。

过程模型则综合考虑空间变异组分(s_i)、时间变异组分($b_0 t^* + v_t$)和时空交互组分($b_{it} t^*$),其中, α 为截距项。 s_i 为空间变异组分,表征在整个研究时段内各街乡相对于整个研究区域(朝阳区)的总体空间效应; $b_0 t^* + v_t$ 为时间变异组分,表征整个研究区域整体时间变化趋势,包括线性趋势 $b_0 t^*$ 和时间随机效应 v_t , b_0 为整体变化系数。 $b_{it} t^*$ 为时空交互组分,是从整体时间变化趋势中分解出的局部变化趋势,相对于 b_0 而言,即为各个街乡基于 b_0 的偏差^[22]。最后, ε_{it} 用于解释不能由空间随机效应和时间随机效应所解释的部分变异^[21]。

3.2 预测模型

基于贝叶斯理论的高斯预测过程模型可以合理描述时间相关性和空间相关性,允许某一空间位置或时间点观测值的缺失,能够实现对未知观测点和后续时段观测值的预测^[35]。基于此,利用贝叶斯高斯预测过程模型可以进行朝阳区居住人口时空预测。具体的模型形式为:

$$\text{数据模型: } Y_{it} \sim P(Y_{it} | Z_{it}, \Theta) \quad (5)$$

$$\text{过程模型: } Z_{it} = \beta \mathbf{x}_{it} + \tilde{\eta}_{it} + \varepsilon_{it} \quad (6)$$

$$\text{参数模型: } \Theta \sim P(\Theta) \quad (7)$$

数据模型中 Y_{it} 为街乡 $l(l=1, \dots, n, n=43)$ 、时间 t 月的时空观测数据样本值, $N=lt$ 表示所有的观测点数。同理, Z_{it} 为 Y_{it} 的时空过程转换参数^[36]。

过程模型中 \mathbf{x}_{it} 为 p 维协变量矩阵, $\beta=(\beta_1, \dots, \beta_p)$ 为相应的系数。纯误差项 ε_{it} 服从独立正态分布 $N(0, \sigma_e^2 \mathbf{I}_n)$, 其中 σ_e^2 为误差项方差, \mathbf{I}_n 为 n 阶单位矩阵。

预测过程主要理论为: 首先在少数位置点 m 处定义随机效应 η_{mt} , 这里 $m \ll n$, 称为结点。然后用克里金预测其他位置点的随机效应, 得到全部位置点的随机效应^[37]。 $\tilde{\eta}_{it}$ 为时空随机效应项, 有 $\tilde{\eta}_{it} = \mathbf{A} \omega_{it}$; $\mathbf{A} = \mathbf{C} \mathbf{S}_\omega^{-1}$, \mathbf{S}_ω 为各点间的相关矩阵, \mathbf{C} 为 $n \times m$ 阶的相关矩阵。结点 m 处的过程模型定义为自回归过程:

$$\omega_{it} = \rho \omega_{i,t-1} + \xi_{it} \quad (8)$$

式中: ξ_{it} 为随机效应项, 假定其服从零均值的高斯过程 $N(0, \sigma_\xi^2 S_\xi)$, 初始值 ω_{i0} 服从零均值的高斯过程 $N(0, \sigma^2 S_0)$, 同理 S_0 为相关性矩阵。

为进行未观测地点 $l0$ 向后一个时间段 $t+1$ 的预测, 可知 $Z_{l0,t+1}$ 服从 $N(\beta \mathbf{x}'_{l0,t+1} + \tilde{\eta}_{l0,t+1}, \sigma_e^2)$, 同理 $Z_{l,t+1}$ 服从 $N(\beta \mathbf{x}'_{l,t+1} + \tilde{\eta}_{l,t+1}, \sigma_e^2)$ 。通过 MCMC 迭代, 最终可以从稳定的概率分布中得到预测值, 最后将标准值转换为原始值 Y 即可。

对预测模型来说, 可以通过预测值与实测值之间的偏差比较来对预测结果进行评价, 常用的评价指标有均方根误差 (Root Mean Squared Error,

RMSE)、平均绝对误差 (Mean Absolute Error, MAE)、相对偏差 (Relative Bias, rBIAS)。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^* - y_i)^2}{n}} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^n |y_i^* - y_i|}{n} \quad (10)$$

$$rBIAS = \frac{\sum_{i=1}^n |y_i^* - y_i|}{n\bar{y}} \quad (11)$$

式中: n 为研究单元数量; y_i 为实测值; y_i^* 为预测值; \bar{y} 为实测值的算术平均。

基于本文所用数据的贝叶斯高斯预测过程模型的具体流程如图2所示。

4 结果与分析

4.1 人口时空格局评估

时空数据的探索性分析利用地理探测 q 统计。通过探测, 得到基于街乡水平的 q 统计量为 0.96, 表明朝阳区各街乡存在完美的空间异质性。基于朝阳区的完美空间异质性进而利用贝叶斯时空层次模型分别探究其总体空间效应 (图3)、总体时间效应 (图4) 和时空交互效应 (图5)。

总体空间效应表征了基于时间稳态的空间结构, 表达了各街乡居住人口相对于朝阳区街乡整体居住人口的高低比例关系, 是一个相对数值高低的概念, 并不表征具体的人口比例。如果某街乡空间

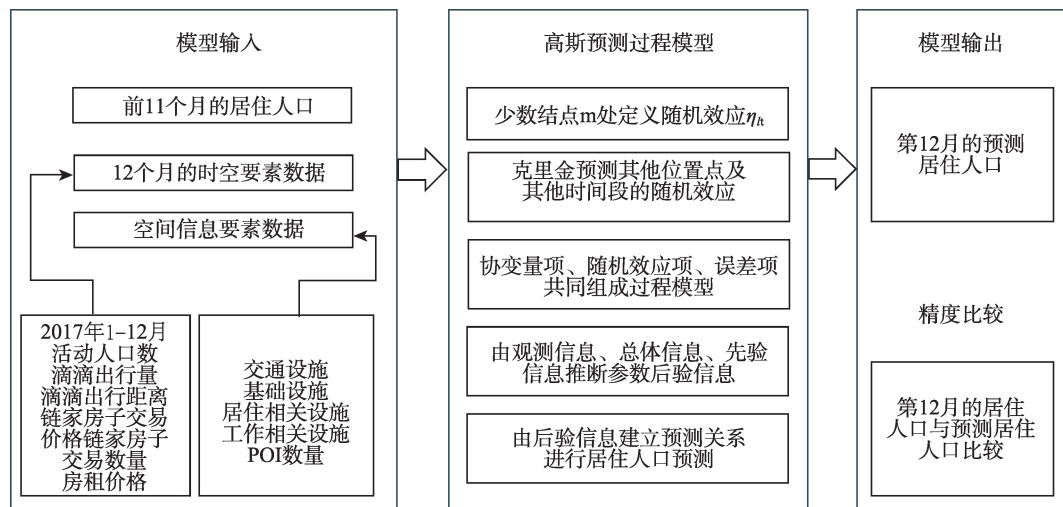


图2 流程图

Fig.2 Workflow

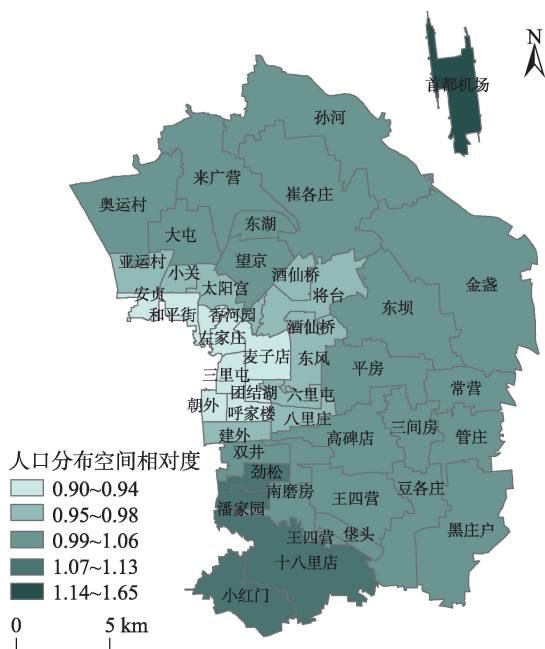


图3 朝阳区居住人口总体空间效应

Fig. 3 Overall spatial variation of resident population in Chaoyang

相对效应值为 p ,则表明该街乡的居住人口是朝阳区街乡水平平均居住人口数的 p 倍,因此 $p>1$,则居住人口高于平均水平,反之低于平均水平。图3为朝阳区2017年各街乡居住人口数总体空间效应的估计结果。在研究时段内,超过半数(25/43=58%)的街乡居住人口高于全区总体水平,包括首都机场、小红门、十八里店、潘家园、劲松、王四营、南磨房、管庄、金盏、三间房、黑庄户、双井、来广营、常营、垡头、豆各庄、高碑店、奥运村、崔各庄、孙河、东坝、平房、望京、东湖、大屯(按空间效应降序排序),在空间上主要分布在五环至六环所在街乡、四环至五环的部分街乡以及二环至四环的南部街乡。总体表现为,朝阳区居住人口在空间水平上存在完美的空间分异,呈现沿环路由内向外递增的空间分布格局。注意到各街乡行政区划面积呈现由内环至外环递增的趋势,揭示了区域面积对居住人口承载力存在空间差异。

总体时间变化趋势表征了朝阳区2017年各月居住人口数随时间的变化特征。从朝阳区居住人口数总体时间变化趋势上看(图4),2017年居住人口呈现增长态势,但也存在3个明显阶段:1~3月呈增长趋势;3~9月呈稳中有降态势;9月以后出现居住人口的急剧反弹。根据北京市交通委2017年春节数据^[38],2017年春节假期(1月27日至2月2

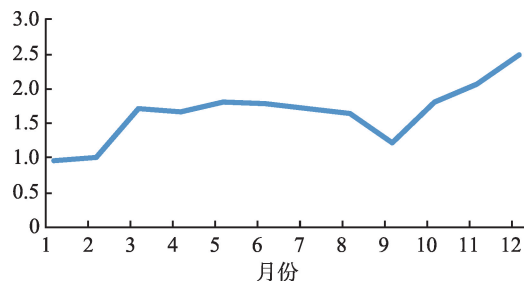


图4 朝阳区2017年居住人口数总体时间变化趋势

Fig. 4 Overall time trend of resident population of Chaoyang in 2017

日)间,在京人口减少约848万,约占全市常住人口的39%。春节结束后的一段时间人口回流,考虑到本次研究所用数据为月度数据,春节的人口减少及之后的人口回流趋势也与本次估计的1~3月的增长趋势较为一致。另外,根据中共北京市委北京市人民政府2017年9月印发《关于率先行动改革优化营商环境实施方案》的通知^[39],北京市提出了人才引进政策,在人才培养和激励等方面都有所规定,这也可能是引发9月之后人口增长的一个原因。

局部变化趋势描述了相对于朝阳区总体居住人口变化趋势下各街乡局部变化趋势(图5)。如果某街乡的局部变化趋势值大于0,则表明该街乡的居住人口数有快速增长的变化趋势;反之,具有缓慢增长的局部变化趋势。由图5可知,虽然首都机场的居住人口空间相对度最高,但其增长趋势并不

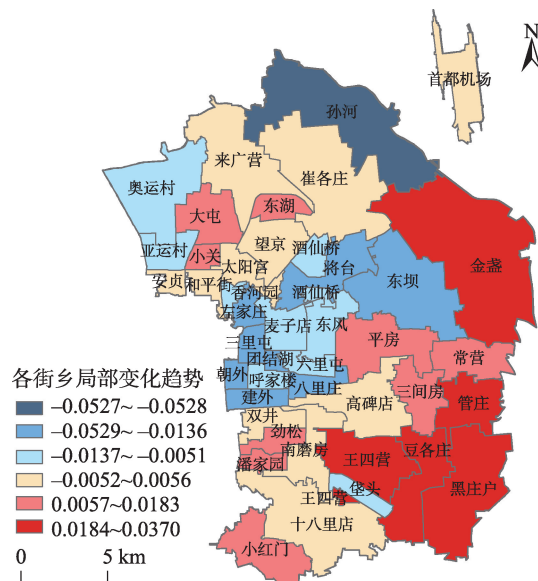


图5 朝阳区各街乡居住人口局部变化趋势

Fig. 5 The departures of the local trends from the overall trend for each subdistricts' resident population in Chaoyang

是最快的,而是保持与朝阳区总体一致的增长趋势。在居住人口相对较高的劲松、潘家园、十八里店、小红门等街乡中,十八里店居住人口增长趋势基本与朝阳区总体保持一致;而劲松、潘家园、小红门居住人口增长趋势明显快于朝阳区总体。可见,朝阳区人口疏解的重点区域可以考虑人口增长较快的劲松、潘家园、小红门等街乡。同时,金盏、管庄、豆各庄、黑庄户、王四营等街乡虽然人口空间相对度小于首都机场,但增长趋势明显,也可以成为人口疏解和管理的重点关注街乡。对于增长速度较慢的孙河等街乡,可以考虑成为人口疏解对接区域。

4.2 人口预测结果

本部分利用贝叶斯高斯预测过程模型对朝阳区居住人口进行时空预测。模型以朝阳区43个街乡,11个月(2017年1月~2017年11月)的居住人口数据作为基础,对2017年12月各街乡居住人口进行预测,并与实测的2017年12月居住人口进行比较验证。

对43个街乡、11个月的数据进行模型拟合,模型将对影响因子显著性进行检验,逐步剔除不显著的影响因子,直到所选影响因子均通过显著性检验。最终通过显著性检验的影响因子包括活动人口、滴滴出行量、房租价格、基础设施POI、就业相关设施POI、居住设施POI。其中活动人口、滴滴出行

量、就业相关设施POI与因变量居住人口呈正相关,就业相关设施POI与因变量居住人口呈正相关说明朝阳区就业设施相对完善,吸引了一部分来此工作的人居住于此。此外,房租价格、基础设施POI、居住设施POI与因变量居住人口呈负相关,与基础设施POI数据呈负相关,可能原因为朝阳区居住人口分布与朝阳区基础设施POI分布相对不均衡所致,说明朝阳区基础设施需要进行基础设施完善或人口的调控,使基础设施对于居住人口的服务更为均衡。同时,与房租价格、居住设施POI呈负相关,一方面房租价格越高,选择来居住的人越少,另一方面居住相关设施越完善,小区的房租价格越高。

根据最终拟合模型,预测2017年12月各街乡居住人口数(图6(b)),与2017年12月实测居住人口分布(图6(a))比对可知,预测人口空间分布与实测人口空间分布整体格局一致,少部分街乡预测值与实测值存在差异,总体而言,基于贝叶斯高斯预测过程模型的人口时空模拟能力较好。

考察模型整体预测精度,均方根误差(RMSE)为11 896,平均绝对误差(MAE)为9071,相对偏差(rBIAS)为0.0043。相比于2017年11月居住人口总量,12月预测值也呈现增长趋势,整体而言与贝叶斯层次模型所估计的总体时间效应中的整体增长趋势一致。

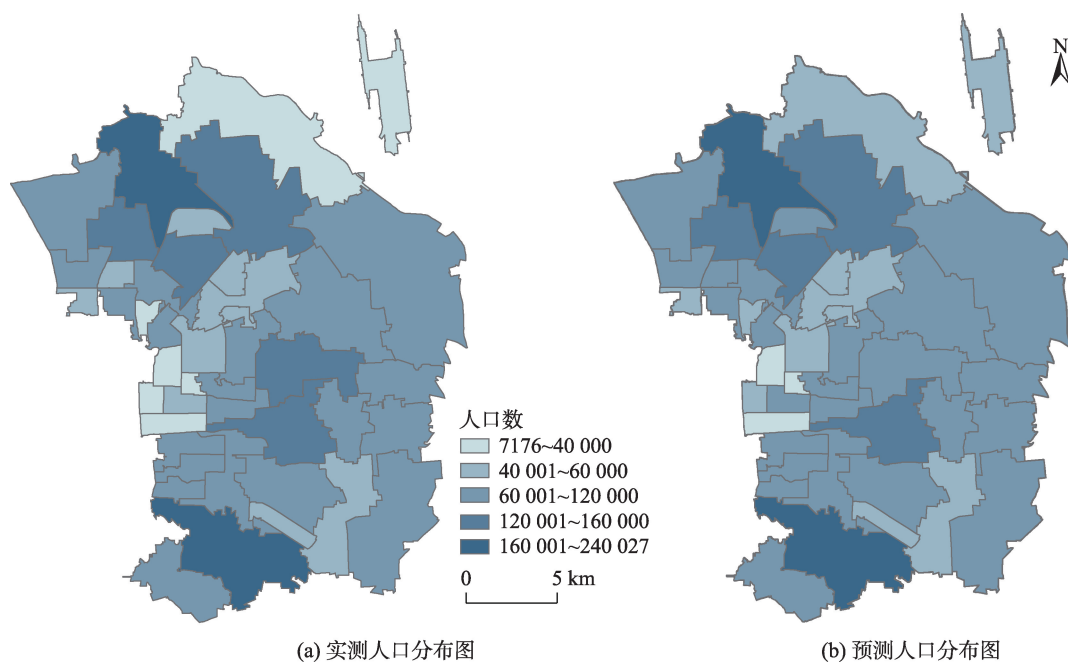


图6 朝阳区居住人口观测与预测空间分布图

Fig. 6 Spatial distribution of observed and predicted resident population in Chaoyang

考察街乡尺度模型预测精度,各街乡预测误差的空间分布如图7所示。预测相对偏差较大的街乡主要集中于朝阳区北部街乡以及南部的十八里店、中部的八里庄。其中,预测相对偏差最大的为首都机场,究其原因可能为机场活跃人口较多且人员流动性高,手机信令数据定义的居住人口可能与真实值之间差异较大,其他影响因子也可能受人口流动性影响与真实值存在差异;同时,首都机场地理位置上相对于朝阳区其他42个街乡距离最远,空间相关性较弱,会影响模型的预测精度。而十八里店被朝阳区政府列为人口疏解提升区,存在大量的“城

中村”、棚户区;北部的奥运村、来广营、孙河、崔各庄、金盏辖区面积较大,均处在朝阳区与北京市其他区县接壤位置,此外来广营、孙河、崔各庄、金盏也存在大量棚户区等使得人口构成复杂、流动频繁,使模型对此类规律的识别较弱。而八里庄辖区面积较小,人口数较少,与朝阳区其他街乡四面接壤,也使得模型的预测精度较差。

将模型预测结果与贝叶斯层次模型所估计的局部变化趋势进行比对分析(图8)。孙河是贝叶斯层次模型中增长趋势最缓慢的街乡,其余14个街乡增长趋势依次递增,分别为平房、潘家园、小关、大屯、劲松、东湖、小红门、常营、三间房、豆各庄、管庄、金盏、黑庄户、王四营。其中,平房、小关、东湖、小红门、常营、豆各庄、管庄7个街乡基于贝叶斯高斯预测过程模型的12月预测人口数高于11月实测人口数,表现为一定程度的增长趋势。大屯、劲松、金盏、王四营4个街乡在贝叶斯层次模型中表现出增长趋势,但12月实测值和预测值均低于11月,考虑到贝叶斯层次模型所估计的局部变化趋势是一个长期的变化趋势,并不代表一次或偶尔的绝对高或者低,因此,可以认为大屯、劲松、金盏、王四营4个街乡的预测值在合理的范围之内。相对11月人口数,孙河、潘家园、三间房、黑庄户、4个街乡的12月预测人口和实测人口数表现为不同的趋势。其中,潘家园、三间房、黑庄户12月实测人口数高于11月人口数,而12月预测人口数低于11月人口数,孙河则相反,12月实测人口数低于11月人口数,12月预测人口数高于11月人口数,这4个街乡的预测趋势偏差较大。

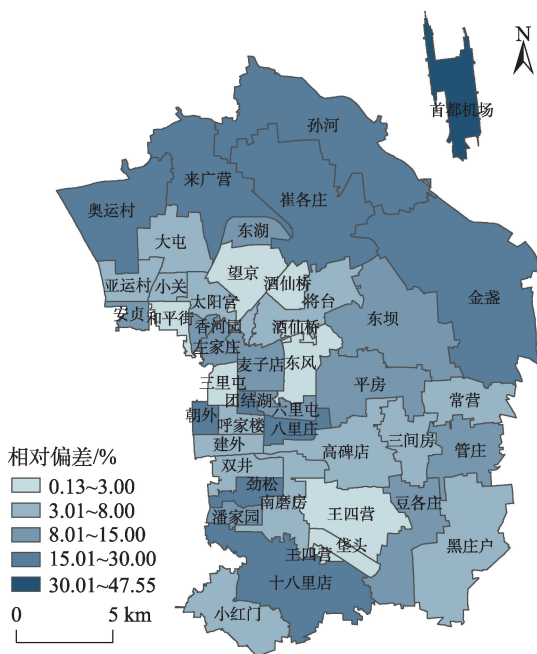


图7 预测误差空间分布

Fig. 7 Spatial distribution of prediction errors

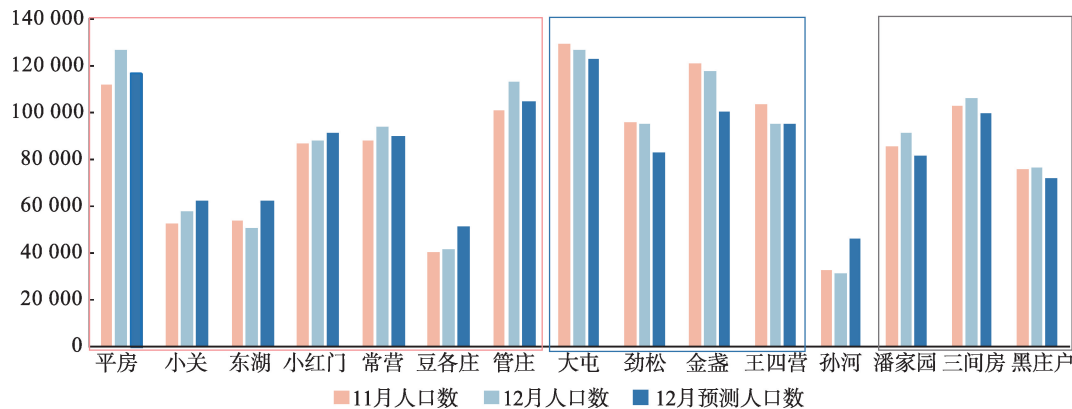


图8 朝阳区特定街乡预测结果与局部趋势比较

Fig. 8 Comparison of prediction results with the departures of the local trends from the overall trend for selected subdistricts in Chaoyang

5 结论和讨论

人口相关问题的建模与预测是一个复杂的课题,随着多源时空数据的可获得性,可以建立起包含时间变异和空间变异的时空模型。基于贝叶斯的时空模型,不仅结合了贝叶斯概率推断的优势,融合当下易获取的多源数据和知识,还能充分考虑空间随机过程和时间随机过程,使对时空过程的估计更加合理。

本文选用包含手机信令数据在内的多源时空数据,进行北京市朝阳区居住人口的时空格局探究及预测。从时空格局变化看:朝阳区居住人口在空间水平上存在完美的空间异质性,整体呈现沿环路由内向外递增的空间分布格局;朝阳区居住人口整体表现出增长的时间趋势,但各街乡的局部时间变化趋势呈现一定差异。从预测结果来看:朝阳区各街乡居住人口预测值与实测空间分布整体一致,整体而言与贝叶斯层次模型所估计的总体时间效应中的增长趋势一致。各街乡预测精度不一,其中朝阳区北部街乡、南部的十八里店、中部的八里庄以及首都机场预测精度最差。

本文研究的朝阳居住人口是基于中国移动手机信令测算的数据,并没有考虑联通、电信的数据,是因为目前市面上移动、联通和电信3家运营商独立运营,手机信令数据还没有实现三家统一的渠道。为此,研究所用数据是考虑了三家运营商市场占有率、手机设备持有率等进行推算而来。另外,手机数据虽然具有动态连续性强、空间全覆盖等特征,但是依赖于基站信号强弱的定位方式存在无法避免的空间误差;手机信令数据定位精度受基站密度影响,在城市核心区定位精度高,郊区定位精度低,如何平衡不同区域的定位精度也会影响研究结果,因此需要考虑与其他数据源的研究相互佐证,进一步提高研究结论的可靠性。值得注意的是,手机信令数据与真实人口的转换目前还没有统一的标准,因此并不能代表真实的人口状况,本文更关注的是朝阳各街乡人口分布的相对状况。如果能够引入统计部门数据进行数据源的扩充和交叉校验,期望可以建立起手机信令数据与实际人口数据的关联关系,实现基于手机信令数据的真实人口推算。

总之,通过本文研究一方面可以获知朝阳区各街乡居住人口的空间差异、时间趋势和局部变化趋势,另一方面通过人口预测能够弥补传统统计口径

的人口数据空间粒度较粗、时间序列过于稀疏的局限,作为传统人口统计数据的补充和校验,丰富人口相关数据资料。在北京市人口疏解的政策背景下,可以辅助进行人口管理,据此提前制定更为精准的人口疏解政策,对未来人口增加的街乡提前关注、重点疏导,帮助有效推进人口疏解的实施。

参考文献(References):

- [1] 刘瑜,肖昱,高松,等.基于位置感知设备的人类移动研究综述[J].地理与地理信息科学,2011,27(4):8-13. [Liu Y, Xiao Y, Gao S, et al. A review of human mobility research based on location aware devices[J]. Geography and Geo-information Science, 2011,27(4):8-13.]
- [2] 郭璨,甄峰,朱寿佳.智能手机定位数据应用于城市研究的进展与展望[J].人文地理,2014,29(6):18-23. [Guo C, Zhen F, Zhu S J. Progress and prospect of the application of smart phone lbs data in urban researches[J]. Human Geography, 2014,29(6):18-23.]
- [3] Vieira M R, Inez V F I, Oliver N, et al. Characterizing dense urban areas from mobile phone-call data: discovery and social dynamics[C]. Proceedings of 2010 IEEE Second International Conference, 2010:241-248.
- [4] 李明晓,陈洁,张恒才,等.上海市精细时空尺度人口分布估计与特征分析[J].地球信息科学学报,2017,19(6):800-807. [Li M X, Chen J, Zhang H C, et al. Fine-grained population estimation and distribution characteristics in Shanghai[J]. Journal of Geo-information Science, 2017,19(6):800-807.]
- [5] Deville P, Linard C, Martin S, et al. Dynamic population mapping using mobile phone data[J]. Proceedings of the National Academy of Sciences, 2014,111(45):15888-15893.
- [6] 曹劲舟,涂伟,李清泉,等.基于大规模手机定位数据的群体活动时空特征分析[J].地球信息科学学报,2017,19(4):467-474. [Cao J Z, Tu W, Li Q Q, et al. Spatio-temporal analysis of aggregated human activities based on massive mobile phone tracking data[J]. Journal of Geo-information Science, 2017,19(4):467-474.]
- [7] 钟炜菁,王德,谢栋灿,等.上海市人口分布与空间活动的动态特征研究—基于手机信令数据的探索[J].地理研究,2017,36(5):972-984. [Zhong W J, Wang D, Xie D C, et al. Dynamic characteristics of Shanghai's population-distribution using cell phone signaling data[J]. Geographical Research, 2017,36(5):972-984.]
- [8] 孙斌栋,魏旭红.上海都市区就业——人口空间结构演化特征[J].地理学报,2014,69(6):747-758. [Sun B D, Wei X H. Spatial distribution and structure evolution of employment and population in Shanghai metropolitan area[J].

- Acta Geographica Sinica, 2014,69(6):747-758.]
- [9] Lee R D, Tuljapurkar S. Stochastic population forecasts for the United States: Beyond high, medium, and low[J]. Publications of the American Statistical Association, 1994,89(428):1175-1189.
- [10] Karlsson A. Statistical demography and forecasting[J]. Psychometrika, 2007,72(2):273-274.
- [11] Raftery A E, Li N, Ševčíková H, et al. Inaugural article by a recently elected academy member: Bayesian probabilistic population projections for all countries[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012,109(35):13915-13921.
- [12] 顾川.山东省人口预测与分析[D].济南:山东师范大学,2017. [Gu C. Prediction and analysis of population in Shandong [D]. Jinan: Shandong Normal University, 2017.]
- [13] 仰园.全面二孩政策下中国人口总量的预测[D].合肥:中国科学技术大学,2017. [Yang Y. Prediction of China's population based on the universal two-child policy[D]. Hefei: University of Science and Technology of China,2017.]
- [14] 安亚革.基于人口生育政策变动的中国人口预测分析 [D]. 沈阳:辽宁大学,2017. [An Y G. Analysis of chinese population prediction based on the population birth policy change[D]. Shenyang: Liaoning University, 2017.]
- [15] Yue T X, Wang Y A, Chen S P, et al. Numerical simulation of population distribution in China[J]. Population & Environment, 2003,25(2):141-163.
- [16] 廖一兰,王劲峰,孟斌,等.人口统计数据空间化的一种方法 [J]. 地理学报,2007,62(10):1110-1119. [Liao Y L, Wang J F, Meng B, et al. A method of spatialization of statistical population[J]. Acta Geographica Sinica, 2007,62 (10):1110-1119.]
- [17] Zhang J, Zheng Y, Qi D, et al. DNN-based prediction model for spatio-temporal data[C]. ACM Sigspatial International Conference on Advances in Geographic Information Systems. ACM, 2016:92.
- [18] 胡曾曾,赵志龙,张贵祥.非首都功能疏解背景下北京市人口空间分布形态模拟[J].地球信息科学学报,2018,20 (2):205-216. [Hu Z Z, Zhao Z L, Zhang G X. Simulation and projection of the spatial pattern of the population in Beijing under the background of non-capital function extraction[J]. Journal of Geo-information Science, 2018,20 (2):205-216.]
- [19] 陈会宴,廖一兰,张宁旭,等.山西省原平市神经管畸形时空分析[J].地球信息科学学报,2017,19(4):502-510. [Chen H Y, Liao Y L, Zhang N X, et al. Spatial and temporal analysis of neural tube defects in Yuanping County, Shanxi Province[J]. Journal of Geo-information Science, 2017,19(4): 502-510.]
- [20] 李世元,王学梅.贝叶斯时空模型在空间流行病学中的研究进展[J].世界最新医学信息文摘,2017,17(34):55-57. [Li S Y, Wang X M. Research Progress of Bayesian spatial temporal model in spatial epidemiology[J]. World Lat-est Medicine Information, 2017,17(34):55-57.]
- [21] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk[J]. Statistics in Medicine, 2000,19(17-18):2555 - 2567.
- [22] Li G, Haining R, Richardson S, et al. Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach[J]. Spatial Statistics, 2014,9:180-191.
- [23] Law J, Quick M, Chan P. Bayesian spatio-temporal modeling for analyzing local patterns of crime over time at the small-area level[J]. Journal of Quantitative Criminology, 2014,30(1):57-78.
- [24] 王晓荷,田茂再.基于分层贝叶斯时空模型的雾霾天气过程计数分析[J].数理统计与管理,2017,36(6):970-982. [Wang X H, Tian M Z. Analysis of haze counts using hierarchical Bayesian spatiotemporal models[J]. Journal of Applied Statistics & Management, 2017,36(6):970-982.]
- [25] 梅波,田茂再.贝叶斯时空分位回归模型及其对北京PM_{2.5}浓度的研究[J].统计研究,2016,33(12):91-100. [Mei Bo, Tian M Z. Bayesian spatio-temporal quantile regression model and its application in PM_{2.5} concentration in Beijing [J]. Statistical Research, 2016,33(12):91-100.]
- [26] Bakar, Shuvo K. Bayesian analysis of daily maximum ozone levels[D]. Southampton: University of Southampton, 2012.
- [27] Sarto S D, Ranalli M G, Bakar K S, et al. Bayesian spatio-temporal modeling of urban air pollution dynamics[C]. Joint Meeting Metma Vii - Graspa, 2014.
- [28] 伍维模,王家强,曹琦,等.塔里木盆地北缘盐土有机碳含量的贝叶斯地统计预测[J].应用生态学报,2017,28(2):439-448. [Wu W M, Wang J Q, Cao Q, et al. Bayesian geostatistical prediction of soil organic carbon contents of solonchak soils in northern Tarim Basin, Xinjiang, China[J]. Chinese Journal of Applied Ecology, 2017,28(2):439-448.]
- [29] 李俊明.基于Bayesian层次时空模型的我国老龄化分析与预测[J].统计研究,2016,33(8):89-94. [Li J M. Space-time variation of Chinese aging based on Bayesian hierarchy spatio-temporal model[J]. Statistical Research, 2016, 33(8):89-94.]
- [30] 李原.贝叶斯层次时空模型在省际人口流入分析中的应用[J].统计与决策,2018(6):73-77. [Li Y. The application of Bayesian hierarchy spatio-temporal model in analyzing population inflows in provinces[J]. Statistics & Decision, 2018(6):73-77.]
- [31] 王劲峰,徐成东.地理探测器:原理与展望[J].地理学报,

- 2017,72(1):116-134. [Wang J F, Xu C D. Geodetector: Principle and prospective[J]. *Acta Geographica Sinica*, 2017,72(1):116-134.]
- [32] Haining R. *Spatial data analysis: Theory and practice*[M]. UK: Cambridge University Press, 2003: 307-377.
- [33] Bakar K S, Sahu S K. spTimer:spatio-temporal Bayesian modeling using R[J]. *Journal of Statistical Software*, 2015,63(15):1-32.
- [34] Smith A F M, Roberts G O. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods[J]. *Journal of the Royal Statistical Society Series B*, 1993,55(1): 3-23.
- [35] Dryden I L, Kent J T. 13 Bayesian forecasting using spatiotemporal models with applications to ozone concentration levels in the Eastern United States[M]. *Geometry: Driven Statistics* John Wiley & Sons, 2015.
- [36] Sahu S K, Gelfand A E, Holland D M. High resolution spacetime ozone modeling for assessing trends[J]. *Publications of the American Statistical Association*, 2007,102(480):1221-1234.
- [37] Sahu S K, Bakar K S. Hierarchical Bayesian auto-regressive models for large space time data with applications to ozone concentration modelling by Sujit Kumar Sahu and Khandoker Shuvo Bakar: Rejoinder[J]. *Applied Stochastic Models in Business & Industry*, 2012,28(5):395-415.
- [38] 北京市交通委:春节期间在京人口减少约848万[ED/OL].[2017/2/2]. http://www.xinhuanet.com/2017-02/02/c_1120400764.htm [Beijing Traffic Committee: population in Beijing declined about 8.48 million during the Spring Festival[ED/OL]. [2017/2/2]. http://www.xinhuanet.com/2017-02/02/c_1120400764.htm.]
- [39] 中共北京市委 北京市人民政府印发《关于率先行动改革优化营商环境实施方案》的通知[ED/OL]. [2017/9/16]. <http://zhengce.beijing.gov.cn/library/192/33/50/438650/1283720/index.html>. [Beijing Committee of the Communist Party of China The people's Government of Beijing City print and distribute the notice of 《Reforming and optimizing the implementation plan for business environment are the first to act.》[ED/OL]. [2017/9/16]. <http://zhengce.beijing.gov.cn/library/192/33/50/438650/1283720/index.html>.]