

CHENGSHI YU QUYU GUIHUA YANJIU CONGSHU

城市与区域规划研究丛书

主编 ◎ 徐建刚

空间数据分析方法在人口数据空间化中的应用

闫庆武 / 著

KONGJIANSHUJU
FENXIFANGFAZAI
RENKOU SHUJU
KONGJIANHUAZHONG
DEYINGYONG

东南大学出版社

城市与区域规划研究丛书/徐建刚 主编

空间数据分析方法在人口数据 空间化中的应用

闫庆武 著

东南大学出版社
·南京·

内容提要

本书以空间数据分析(SDA)理论和方法为基础,构建了基于 SDA 的人口数据空间化的方法框架,并将空间自相关分析、空间自回归分析、空间异质性分析、区域差异分析、点模式分析、热点探测技术、面插值技术、地统计学等多项空间分析技术应用于人口分布规律的研究与识别中,实现了江苏省、徐州市域、徐州市区多级单元的人口数据空间化。

本书可以作为城市与区域规划、地理信息系统等相关专业学生的教学参考书,也可以作为地图制图学、人文地理学、人口学等领域研究人员的参考资料。

图书在版编目(CIP)数据

空间数据分析方法在人口数据空间化中的应用/闫

庆武著. —南京:东南大学出版社,2011. 3

(城市与区域规划研究丛书)

ISBN 978-7-5641-2336-9

I. ①空… II. ①闫… III. ①地理信息系统—应用—城市规划 IV. ①TU984-39

中国版本图书馆 CIP 数据核字(2010)第 139996 号

空间数据分析方法在人口数据空间化中的应用

出版发行 东南大学出版社
出版人 江建中
社 址 南京市四牌楼 2 号
邮 编 210096
经 销 全国各地新华书店
印 刷 兴化印刷有限责任公司
开 本 700 mm×1000 mm 1/16
印 张 13
字 数 220 千
版 次 2011 年 3 月第 1 版
印 次 2011 年 3 月第 1 次印刷
书 号 ISBN 978-7-5641-2336-9
定 价 35.00 元

本社图书若有印装质量问题,请直接与读者服务部联系,电话(传真):025—83792328

总 序

进入新世纪的十年以来,我国的经济继续在高速的增长,叹为世界奇迹!在人类发展史上写下了更为辉煌的一页!2010年5月,第41届世界博览会在上海的成功举办,更是将海内外世人的目光长时间的聚焦到上海。世博会的主题“城市,让生活更美好”,正好为国际著名经济学家、诺贝尔奖得主斯蒂格利茨教授的名言:“21世纪初期影响最大的世界性事件,除高科技以外就是中国的城市化。”做了华丽的诠释。然而,当下还处在转型期的中国,粗放模式主导的工业化与城市化双轮驱动下的城市及其依托的区域,不可避免地产生了众多不和谐的声音。尤其是在一些地方政府坚持唯GDP马首是瞻的发展理念下,多种不和谐声音集聚形成了倍增效应,使得在区域差异、城乡差别和群体收入差距等不断地被拉大的同时,城市空间过度开发、土地利用效率低下、公共资源分配不公、生态环境急剧恶化和文化遗产建设性破坏等严重影响经济社会可持续发展的问题扑面而来。我国这一高速发展中的诸多负面现象,引起了国内外学术界的广泛关注和深刻思考。

从地理学视角来看,改革开放三十年以来,我国的城市化主要是在广阔的农村地域上展开的,即将农业农村用地转化为城镇建设用地。从一般社会发展逻辑上讲,原先生活在被城市化土地上的农民身份理所当然地被转换为市民身份,城市应该为他们提供务工或商服等工作岗位,然而,“城中村”等现象困扰着城市空间发展。从经济全球化角度来纵观我国城市化过程,首先是从计划经济转向市场经济过程中,大量的农村廉价劳动力涌向城市,成为中国吸引世界产业转移,成为世界“加工厂”的首位要素。与此同时,我国土地的公有制使得政府不要花多大成本就能从农民手中把大量优质耕地拿到手,在我国地方土地财政和追求GDP的政绩考核体制下,地方政府为了招商引资,相互间恶性竞争,工业地价被一而再、再而三的压低,甚至于零地价被租用。尽管有如此“亲商”政策,工业区的形成还需有城市公共基础设施来支撑。因此,在做大做强“城市”策略引导下,地方政府借助城市规划,通过行政区划调整,使得我国的城市数量、城镇人口和用地规模

急速增长。在此政策背景下,工业开发区遍地开花,工业区中有大量土地被闲置,造成了很大的浪费;而数量众多的污染企业带来了废水废气的超标排放,使园区及其周边城乡环境受到了严重污染。在城市主城区,“摊大饼”式的快速扩展也带来了城市功能不完善和基础设施不健全而引发的城市交通拥堵、环境质量下降、公共服务供给不足和公共安全保障不力等突出问题。在区域层面,大江大河的水体污染触目惊心,城市饮用水源地安全屡遭威胁,耕地锐减已逼近粮食安全底线,生态危机、环境恶化正在向全国各地蔓延。

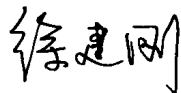
面对这一严峻现实,党中央适时地提出了“科学发展观”的新发展理念,实现了我国发展思路的重大转变。在“以人为本”思想指导下,党中央提出的五个统筹中的“统筹城乡发展,统筹区域发展,统筹经济社会发展,统筹人与自然和谐发展”四个统筹已成为新时期区域发展与城乡规划建设的行为指南。以推动城市与区域发展为己任的我国广大人文经济地理和城乡规划学者,针对这一现实背景下不断产生的城乡空间发展问题,借鉴国际学术界的先进理论和方法,以多元核心价值为导向,从经济、社会、人文、生态、地理和环境等多学科融合的视角,在产业空间集聚、城乡统筹布局、城市空间结构、城市住宅与交通、城市历史文化空间保护与复兴、城市生态网络建设与服务等方面开展了卓有成效的研究,形成了多个学科生长点与学术前沿领域。

东南大学出版社近几年陆续出版的《城市与区域规划研究》丛书,作者以国内著名学府的青年才俊为主体。从已出版的几本专著中,我欣喜地看到年轻的学者们以人文地理、区域经济、城市与区域规划领域的学术前沿问题为研究对象,通过梳理国内外相关学术理论和归纳评价前人的研究成果,提出了新的研究视角和学术切入点。尤其值得赞许的是,多位作者以专业 GIS 平台为支撑,在构建涵盖城乡海量地理数据库的基础上,将地理信息科学领域新近发展的多种空间定量分析技术,如基于格网数据结构的空间成本—距离分析法、空间自相关的地统计分析法以及城市空间组织的空间句法分析法等,巧妙地运用于城镇体系格局、城市产业空间、城市生态服务空间、城市功能空间以及城市住区与交通组织等理论与实践问题的深度剖析之中。几乎每一本书中都建立了一套有自己特色的从理论、方法到实证的研究体系,在实现研究方法创新突破的同时,完善和深化了城市地理、经济地理、区域经济、城市规划以及区域规划等学科的理论体系,并对当前城市与区域规划有直接的指导意义。

该套丛书的出版不仅丰富了城市规划、人文地理和区域经济等学科的学术宝库,而且为这些领域里的年轻学子脱颖而出提供了学术舞台。也可以说,该套丛书为我国新世纪的城市与区域可持续发展问题作出了独特的贡献。最后,我再为这些锐意进取的青年作者们送上鼓励的话语:你们学术视野开阔、创新思维活跃、研究手段先进,已初步形成了一种以科学理性分析为特色的城市与区域规划研究风格。可喜可贺!希望你们今后百尺竿头更进一步,更多的走上国际学术舞台,为我国人文地理、城市规划等领域的学术发展走向国际前沿作出自己的贡献。

是为序!

南京大学城市与区域规划系

Handwritten signature of Xu Jianwang in black ink.

2010年5月

序

地理信息系统与一般数据库的本质区别在于它的分析功能,尤其是对空间数据的分析。空间数据分析是地理学和地理信息科学领域的重要研究内容,它通过研究地理空间数据及其相应分析的理论、方法和技术,探索、证明地理要素之间的关系,揭示地理特征和过程的内在规律和机理,实现对地理空间信息的认知、解释、预测和调控。传统的统计理论是一种建立在假定观测值独立基础之上的理论。然而,在现实世界中,特别是遇到空间数据问题时,独立观测值在现实生活中并不是普遍存在的。对于具有地理空间属性的数据,一般认为距离近的变量之间比在空间上距离远的变量之间具有更加密切的关系。空间效应主要有两个来源:空间依赖性(Spatial Dependence)和空间异质性(Spatial Heterogeneity)。空间数据分析是揭示空间效应的有力工具,是广泛使用的分析空间数据的方法,尝试从凌乱的地理空间数据中,用统计方法发掘地理空间的变化规律。

目前在人口统计数据空间化的相关研究中,对于人口分布的空间统计规律考虑偏少,采用空间数据分析方法进行人口数据空间化的研究也相对较少。闫庆武博士以“空间数据分析方法在人口数据空间化中的应用”为题完成了他的博士论文,论文将空间数据分析方法运用到人口数据空间化的研究中,并在以下几个方面取得了初步的研究成果:(1)考虑到人口分布的空间依赖性与空间异质性,将空间自相关分析、空间自回归分析、空间异质性分析、人口分布的区域差异分析、热点探测技术、地统计学、面插值技术、点模式分析等空间分析技术应用于人口分布规律的研究与识别中,建立了人口空间分布的数学模型。(2)提出了衡量人口分布区域差异的锡尔系数与二阶段嵌套分解方法,并以江苏省人口分布为研究对象,研究了江苏省苏南、苏北和苏中三大地带间及三大地带内地级市之间、地级市内部人口分布的区域差异。(3)提出了基于居民点密度的人口数据空间化的方法,针对在人口数据空间化过程中有时只知道某个居民点(如镇或村)的人口数,而不知道该居民点边界的情况,提出了基于居民点密度的人口数据空间化的方法,并以江苏省为研究案例,实现了基于镇密度的县行政单元与基于村密

度的镇行政单元的江苏省人口数据空间化。(4)提出了测度人口密度分布异质性的 GPD 指标。以江苏省为研究案例,测度并分析了基于县级单元和地级市单元的江苏省人口密度分布的异质性,结合人口密度分布的空间自相关分析可以发现人口密度的异质性空间分布与人口密度空间分布的聚集性特征密切相关。

目前,人口增长已经给全球资源、环境承载能力造成了巨大压力,耕地和林地面积骤减,生态多样性破坏严重,人类生存条件日益恶化等。及时获取不同尺度上精确的人口空间分布及其变化信息对于解决这些社会、经济和环境问题,提高人口、资源和环境的综合管理能力有着重要意义。在 GIS 技术的支撑下,结合数理统计分析方法,进行区域人口空间分布的数值模拟;通过对同一区域不同时期人口空间分布状况的对比,探讨人口动态变化的主要影响因素及其作用机制,构建区域人口空间分布变化模型,有助于深入了解人口动态变化规律、人口空间分布格局形成演化机理、人文因素与生态环境间的响应机理与互动模式。同时,笔者相信类似于人口数据的其他社会属性数据的空间化研究对我们深度了解社会经济现象、揭示发生在一定空间范围内的自然、环境与社会经济变化规律必能带来有益的帮助。希望闫庆武博士能在博士论文的基础上,以 GIS 为工具,关注热点社会经济现象,深入调查研究,深度挖掘社会经济现象的本质规律,服务国家经济建设和社会发展。

中国矿业大学

卞正富

2011 年 3 月 12 日

前 言

传统的以行政区界线为标准统计表达人口分布的方法具有同一行政区内均匀分布、不同行政区间突变的缺点。人口数据空间化便是解决上述问题的主要方法之一。人口数据空间化与人口空间分布的研究,是一个多学科交叉的研究新领域,是地理信息科学中的一个重要的研究课题。

本书以空间数据分析理论和方法为基础,将空间自相关分析、空间自回归分析、空间异质性分析、人口分布的区域差异分析、热点探测技术、地质统计学、面插值技术、点模式分析等多项空间分析技术应用于人口分布规律的研究与识别中,实现了江苏省、徐州市域、徐州市区多级单元的人口数据空间化。全书的主体内容主要由以下几个部分组成:

第一,通过采用空间统计与探索性空间数据分析的方法系统地研究了江苏省人口分布的空间统计规律;运用罗伦兹曲线和基尼系数研究了江苏省人口分布的不均衡性;采用人口分布的锡尔系数研究了人口分布的空间差异性及其地区构成;利用 Moran's I 和 Geary's C 对江苏省县级单元和镇级单元的人口分布进行全局空间相关性分析;运用局域 Moran's I 、局域 G 统计量对江苏省人口分布作局域空间自相关分析。结果表明:江苏省人口分布呈现出由东北地区,向西南地区依次递减的格局,并且,江苏省人口分布呈现出三个热点地区:环太湖地区、南京地区和徐州地区。

第二,提出了人口密度空间分布异质性测量的方法,即衡量人口密度分布异质性的 Gini 指标,基于县级单元和地级市单元对江苏省人口密度分布的异质性进行测量,结果表明:人口分布异质性大的地区主要分布在苏南地区,而异质性小的地区主要分布在江苏省东北部地区。采用地统计学的方法,实现了徐州市人口密度的空间化,并且确立了徐州市对数人口密度空间分布的变异函数模型(指数模型)及其相关参数。

第三,运用空间分析中的点模式分析技术,采用 NNI、Ripley's K 函数、样方分析、最近距离层次聚类(NNH)等方法研究了徐州市居民点的空间分布规律;分别运用核函数法与泰森多边形法计算徐州居民点的密度分布后发现,核函数法能够从总体上反映徐州市居民点密度空间分布

的地域分布差异,而泰森多边形法对居民点密度空间分布的细部差异比较敏感。

最后,提出了一种把泰森多边形法与 GCAWI 法相结合的人口数据空间化的方法,并且采用格网错置平滑的方法,以徐州市区为研究案例得到了徐州市区 2005 年 $500\text{ m}\times 500\text{ m}$ 大小格网的人口密度空间化结果。

著 者

2010 年 7 月

ABSTRACT

In social sciences, researchers are interested in the analysis and modeling the spatial data. Spatial data analysis (SDA) and GIS play a key role in modeling spatial data. The distribution of population is always influenced by spatial geographic factors. The traditional cartography of population density has a defect in discontinuity. Census spatialization is one of the methods to solve above-mentioned problems. Study on population distribution and census spatialization is a new research field of many subjects, which is an important problem in GIScience. It is a new research direction of population geography and it has broken the demographic data stemming from the boundary line of the administrative division for a long time. It offers the new thinking and method for us to count demographic data in the area that is not administrative area. So it is very important to study census specialization.

Theory and methods of spatial data analysis such as autocorrelation analysis, autoregressive analysis, spatial heterogeneity analysis, population distribution inequalities analysis, hot spots detection, geostatistics method, method of areal weighting interpolation, point pattern analysis etc are employed to analyze the quantitative rule of population distribution and census spatialization in Jiangsu, Xuzhou and the built-up area of Xuzhou.

1. Characteristic of population distribution in Jiangsu is studied.

Based on cities, counties and towns Characteristic of population distribution in Jiangsu is studied by the numbers. The results are the following: distribution of population density based on city is higher in South Jiangsu and Xuzhou-Liangyungang economic zone while lower in the Mid of Jiangsu, distribution of population density based on county is very disequilibrium and autocorrelative, distribution of population density based

on town takes on the characteristic of convergency and decentralization.

2. Cenus spatialization in Jiangu is carried out based on Spatial statistic and autoregressive model.

Spatial statistic rule of population distribution in Jiangu is found with spatial statistic and Exploratory Spatial Data Analysis(ESDA). The inequalities of population distribution in Jiangu is found with the Lorenz Curve of population distribution and Geordie index.

An inequality composition method, the two-stage nested Theil decomposition method, which is an extension of the ordinary one-stage Theil decomposition method, are employed to study on the inequality of population distribution. First, the definition of inequality of population distribution index based on Theil which is called T_{pop} is put forward. And then one-stage and two-stage nested Theil decomposition of T_{pop} is brought forward. Based on the different regionalism, fractionized the inequality to find the source of the whole inequality of population distribution through decomposed the T_{pop} in Jiangu. Via the decomposition, we can find that the whole inequality of population distribution is made up of three parts, the most distinct one is the of 106 counties, in the next place, the diference in the cities also plays an important part, finally, the most inapparent one is the difference among the three belts in the province. By computing local G statistics index based on counties the following conclusion can be drawn: the inequality of population distribution in hot spots is big, the one in cool spots is small.

Moran's I and Geary's C are employed to study on global spatial autocorrelation of population distribution in Jiangu based on counties and towns. Three hot spots of population distribution in Jiangu is found by computing the Local Indicators of Spatial Autocorrelation (LISA) $G_i(d)$.

By the use of Maximum Likelihood Spatial Lag Model and the Spatial Weight(Queen), the 1 000 m \times 1 000 m grid map of population density of Jiangu is gained.

3. Cenus spatialization in Xuzhou is carried out based on geostatistics.

A index which is called Gini of Population Distribution (GPD) to

measure the heterogeneity population distribution is put up. By computing the GPD in Jiangsu based on counties and towns, the following conclusion can be drawn: the heterogeneity population distribution of south Jiangsu is big, the one of North-east Jiangsu is small.

By using statistical data of small cities and towns in Xuzhou in 2005, the methods of geostatistics and tools of ArcGIS are employed to study on the spatial distribution of population density in Xuzhou. First, with the methods of exploratory data analysis (EDA), for example, histogram, semivariogram cloud, voronoi polygons, the preliminary knowledge of population density data is gained. Second, variogram model (exponential model) is built up with curve fitting and cross-validation. Finally Log-normal Kriging(LK) is used to make isoline map and three-dimensional visualization for population density of Xuzhou.

4. Cenus spatialization in Jiangsu is carried out based on density of settlement.

Analysis of spatial distribution characteristics of settlements is an important problem in settlements geography. Using Spatial data analysis (SDA) to research on spatial distribution characteristics of settlements can result in a good accuracy of it.

Point pattern analysis technique such as NNI, *Ripley's K*, Quadrat Analysis(QA) and Nearest Neighbor Hierarchical Clustering (NNH) is used to investigate the rule of spatial distribution characteristics of settlements in Xuzhou. In this paper, the information of Xuzhou settlements distribution is acquired by TM images(2005) and city maps. Then the patterns of settlements distribution in Xuzhou are investigated by quadrat analysis(QA), nearest neighbor index (NNI), *Ripley's K* function and hot spots analysis. The following conclusions can be drawn. On the whole, settlements distribution in Xuzhou is spatial dependent and clumped. The degree of settlements agglomeration increases when the investigation scale increasing. The distribution of hot spots of settlements is characterize by spatial random in micro scale, belt structure in medium scale and district structure in macro scale.

Based on method of Kernel Estimation and Thiessen Polygons the

settlements density in Xuzhou is computed. Compared with the two methods, the method of Kernel Estimation can model the whole change tendency of distribution of settlements in Xuzhou clearly, the method of Thiessen Polygons is sensitive to local spatial difference of distribution of settlements in Xuzhou.

A method of census spatialization is put forward based on settlement density. Taking Jiangsu as an example, the 1 000 m × 1 000 m grid map of population density of Jiangsu is gained.

5. Cenus spatialization in Xuzhou is carried out based on Polygons and GCAWI

The method based on Thiessen Polygons and GCAWI to spatialize census data is put forward. With this method, the 500 m × 500 m grid map of population density of Xuzhou is gained.

目 录

1 绪论	1
1.1 研究的背景和意义	2
1.1.1 相关背景	2
1.1.2 研究意义	4
1.2 国内外研究现状	6
1.2.1 GIS-SDA 研究现状	6
1.2.2 属性数据空间化研究进展	11
1.2.3 人口数据空间化研究进展	13
1.3 目前研究中存在的主要问题	17
1.3.1 人口数据空间化的精度与准确度有待提高	18
1.3.2 地理尺度转换方法的应用需要进一步加强	18
1.3.3 人口数据空间化过程中对于居民点密度考虑较少	18
1.3.4 对于中观尺度人口数据空间化的研究有待加强	18
1.4 研究的目标、内容与技术路线	19
1.4.1 研究目标	19
1.4.2 研究内容	19
1.4.3 技术路线	19
1.5 创新之处	20
1.6 篇章结构	21
2 人口数据空间化的理论与方法	23
2.1 人口数据空间化问题的提出	23
2.1.1 人口密度空间化的背景	23
2.1.2 人口密度及其表示方法	24
2.2 人口数据空间化的理论基础	25
2.2.1 尺度转换理论	25
2.2.2 GIS 空间分析理论	28

2.2.3	等级理论	30
2.2.4	分形理论	31
2.2.5	人口分布理论	32
2.3	人口数据空间化的方法	34
2.3.1	人口数据空间化的一般过程	34
2.3.2	无辅助信息的空间化方法	35
2.3.3	有辅助信息的空间化方法	39
2.3.4	人口密度空间化主要方法比较	42
3	研究区域概况及研究数据库的建立	44
3.1	研究区域概况	44
3.1.1	区位	44
3.1.2	地形、地貌	45
3.1.3	气象	46
3.1.4	水文	47
3.2	研究数据的准备	49
3.2.1	图形数据	49
3.2.2	人口数据	51
3.2.3	研究数据库的建立	52
3.3	研究的软件条件	55
3.3.1	ArcGIS 9.0	56
3.3.2	GeoDa	57
3.3.3	CrimeStat	58
3.3.4	SPSS	58
3.4	江苏省人口空间分布现状	59
3.4.1	市域人口分布呈现出南北两侧高、中间低的“凹”字形格局	59
3.4.2	县市(区)人口分布极不均衡,具有较强的自相关性	61
3.4.3	镇域人口的频率分布呈现出集中性和离群性双重特征	63
4	基于空间统计学的人口数据空间化研究	67
4.1	探索性空间统计分析	68
4.1.1	空间权重矩阵	68

4.1.2	全局空间自相关: Moran's I 和 Geary's C	69
4.1.3	局域空间自相关指数	71
4.2	江苏省人口分布的一般统计分析	74
4.2.1	人口分布的罗伦兹曲线	74
4.2.2	人口分布的锡尔系数	76
4.3	人口分布的空间统计分析	81
4.3.1	空间权重矩阵的建立	81
4.3.2	全局空间自相关分析	82
4.3.3	局域空间自相关分析	84
4.3.4	人口分布的热点分析	88
4.4	人口分布的空间自回归分析	90
4.4.1	空间自回归	90
4.4.2	人口数据的格网化处理	90
4.4.3	基于空间自回归模型的人口数据空间化	92
4.5	本章小结	94
5	基于地统计的人口数据空间化研究	96
5.1	地统计学方法研究的步骤	97
5.1.1	探索性数据分析	97
5.1.2	变异函数的定义和参数	98
5.1.3	克里金估值及结果的验证	99
5.2	人口空间分布的异质性测量	100
5.2.1	空间异质性	100
5.2.2	数据的异质性测量	100
5.2.3	人口空间分布的异质性测量	101
5.2.4	江苏省人口分布的异质性测量	102
5.3	徐州市人口密度空间分布的地统计分析	108
5.3.1	探索性数据分析	108
5.3.2	变异函数及其参数	111
5.3.3	克里金内插及人口密度空间分布的直观表达	112
5.3.4	徐州市人口空间分布的特点	112
5.4	本章小结	113