

城市人口分布的空间自相关分析 ——以沈阳市为例

杜国明^{1,2}, 张树文¹, 张有全^{1,2}

(1. 中国科学院东北地理与农业生态研究所, 长春 130012; 2. 中国科学院研究生院, 北京 100049)

摘要: 为探测经典城市人口密度模型在微观层面上的适用性, 本文通过计算基于格网的沈阳市人口景观密度, 利用地统计学的理论与方法来分析其空间自相关性和变异性。研究发现: 在 10 种不同采样粒度上人口密度半变异函数都符合球状模型, 表征着城市人口分布具有较强的空间自相关性和变异性; 但由于粒度不同, 人口密度的空间自相关尺度不同, 块金值不同, 基台值不同, 块金值与基台值的比值差异较大, 证明人口分布的空间自相关具有较强的尺度依赖性; 对于 700m 粒度而言, 通过计算不同方向的半变异函数曲线可以发现, 沈阳市人口密度呈现典型的带状异向性, 反映出人口分布在不同方向上具有不同的结构特征。因此, 研究城市人口分布时, 可以城市人口分布的自相关性和变异性分析为基础, 并须充分注意适宜尺度的选择和结构特征的识别。

关键词: 空间自相关; 城市人口分布; 人口密度模型; 空间变异性; 地统计学
文章编号: 1000-0585(2007)02-0383-08

1 引言

城市人口空间分布是城市地理学研究的热点问题之一, 上个世纪 50 年代以来, 吸引了西方大量学者的关注, 形成了一系列较具代表性的理论模型。如 Clark 模型^[1]、城市人口与城区面积间的异速生长模型^[2,3]、正态密度模型^[4,5]、负幂指数模型^[6]、二次指数模型^[7]等。20 世纪 80 年代末期以后, 随着西方多核心城市结构研究的兴起, 城市人口密度多核心模型被提出并成为热点研究领域^[8,9]。在我国, 近年来以冯健等为代表的学者们进行了大量的理论探讨和实证研究^[10,11]。国内外的这些研究对于认识城市人口分布规律、以及在城市人口分布演变特征基础上研究城市空间结构演变等方面贡献良多。

在统计平均的基础上, 也就是这些模型赖以成立的宏观层面上, 人口分布是连续的、光滑的。但在微观的层面上, 人口分布是否是连续的呢? 或者说人口分布在多大尺度上可以看成是连续的? 再退一步讲, 分析的尺度不同, 对人口分布的连续性有何影响? 另一方面, 人口密度模型研究经历了由单核心模型向多核心模型的转化, 也反映了城市人口分布空间结构的演变。城市人口分布多核心结构的出现, 是否会对人口分布的空间连续性产生影响? 在不同的方向(或城市的剖面)上, 人口分布的连续性是否会有所差异? 深入分析这些问题, 对于更准确地识别和理解城市人口分布的空间结构以及构建人口密度模型具有

收稿日期: 2006 03 14; 修订日期: 2006 09 23

基金项目: 中国科学院知识创新工程重要方向项目(KZCX2-SW-320-1)。

作者简介: 杜国明(1978), 男, 内蒙古宁城县人, 地图学与地理信息系统专业博士研究生。主要研究方向为地理信息系统、遥感应用。E-mail: nmgdgm@126.com

指示作用。针对上述问题,本文以沈阳市第五次人口普查数据为基本数据源,将基于统计学的经典的人口密度模型与基于区域化变量理论的地统计学结合起来,借鉴 GIS 空间分析方法和格网计算技术来计算人口的景观密度,分析城市人口分布的空间自相关性和变异性,并进一步探讨研究城市人口分布特征时应注意的问题。

2 研究区概况与研究方法

2.1 研究区概况

沈阳市国土总面积为 12980km²,截至 2000 年底建成区面积 217km²。沈阳市以平原为主,地势平坦,平均海拔 50 米左右,城市南侧有浑河流过。沈阳市是辽宁省的政治、经济、文化中心,是辽中南城市群的首位城市,中国重要的重工业城市。市辖区包括和平、沈河、大东、皇姑、铁西 5 个城区,东陵、于洪、苏家屯、新城子 4 个郊区。截至 2000 年第五次人口普查,市辖区人口 530 万人。本文的研究区是包括建成区及近郊区在内的一矩形区域,研究区面积 336km²,涉及人口普查区 981 个。

2.2 研究方法

2.2.1 基础数据与技术路线 本研究的基础数据包括沈阳市第五次人口普查数据、人口普查区数据、建成区及近郊区居住房屋分布图以及城市基础地理信息数据等空间数据。其中,人口普查数据由人口统计部门提供,最小统计单元为社区管委会(村),数据存储为 Oracle 数据库 dmp 格式;空间数据由相关部门测绘而成,可满足 1:5000 比例尺制图精度,数据格式为 mapinfo 的 tab 格式。本研究的主要工作内容可以分为两部分:人口统计数据的空间化处理,即人口景观密度的计算;基于地统计学的人口分布空间自相关分析。数据处理及分析的详细技术路线见图 1。

2.2.2 人口景观密度的计算 本研究所采用的人口密度为人口景观密度,类似于生态学中的种群密度,是指各采样单元内单位面积上的人口数,单位是 p/hm² (人/公顷)。由于人口数据是以行政单元(社区管委会)来统计的,因此人口景观密度计算通过以下几个步骤来实现:

第一、建立人口普查数据与普查区空间数据的关联。将人口普查数据由 dmp 格式转成 dbf 格式;将各种空间矢量数据由 tab 格式转换成 Arc/Info 的 Coverage 格式;通过普查区代码将人口普查数据与普查区空间数据相关联,获得人口普查数据分布图。

第二、采用面插值法^[15]实现人口普查数与居住单元的关联。将居住房屋底层用地矢量数据与人口普查数据分布图进行空间叠加,统计每一普查区内居住房屋底层用地面积,并依此计算各普查区内居住房屋底层用地的人口密度,获得人口实地分布图。计算公式为:

$$D_i = P_i / A_i \quad (1)$$

式中, D_i 代表第 i 个人口普查区内居住房屋底层用地的人口密度; P_i 代表第 i 个人口

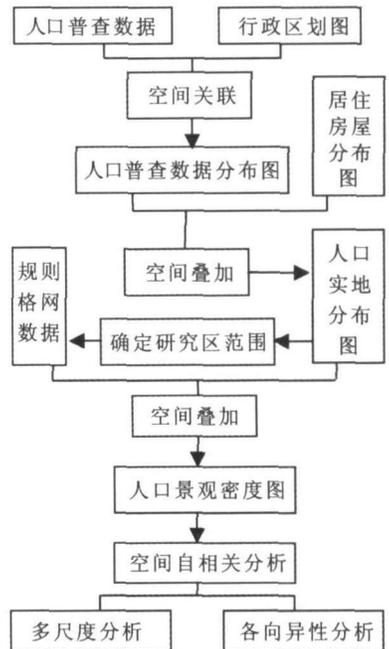


图 1 数据处理及分析技术路线

Fig 1 Route for data processing and analysis

普查区的人口数; A_i 代表第 i 个人口普查区内居住房屋底层用地的总面积。

第三、用规则格网进行采样, 计算每一个网格内的人口数和人口景观密度。根据人口实地分布图确定研究区范围, 并在 Arc/Info 环境下用 Generate 命令生成正方形格网数据, 与人口实地分布图叠加, 然后统计每一网格内的人口数和人口密度, 获得人口景观密度图。人口数的计算公式为:

$$P_j = \sum_i A_{ij} \times D_{ij} \quad (2)$$

式中, P_j 代表第 j 个网格内的人口数; A_{ij} 代表落入第 j 个网格内的第 i 个人口普查区的居住房屋底层用地总面积; D_{ij} 代表落入第 j 个网格内的第 i 个人口普查区的居住房屋底层用地的人口密度。

为反映采样尺度对城市人口分布空间自相关性的影响, 网格的边长包含 100m 至 1000m 10 种。按照景观生态学的方式, 本文把网格的边长称为粒度 (grain, G)。

2.2.3 基于地统计学的空间自相关分析 地统计学 (Geostatistics), 亦称地质统计学, 于上世纪 40 年代末由南非地质学家 D. G. Krige 和 H. S. Sichel 等提出后开始形成。地统计学以区域化变量理论为基础, 研究那些分布于空间中并显示出一定结构性和随机性的现象。地统计学的两大研究领域为: 区域化变量的空间相关及变异分析、空间插值^[12, 13]。

区域化变量的空间相关及变异分析以半变异函数建模为基础。假设有一组空间样本 (s_1, s_2, \dots, s_n), 则半变异函数 $\gamma(h)$ 为:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^N [Z(s_i) - Z(s_i + h)]^2 \quad (3)$$

式中, h 为样本距; $N(h)$ 为间距, 为 h 的样本对的总个数; $Z(s_i)$ 是样本空间样本 s_i 的属性值, $Z(s_i + h)$ 是距样本 s_i 距离为 h 处的样本的属性值。有些文献中将半变异函数简称为变异函数。

半变异函数有 3 个主要参数^[14]: 块金值 (nugget, C_0)、自相关尺度 (a) 和基台值 (still, $C_0 + C_1$)。用于人口密度的空间变异分析, 块金值可理解为间距小于抽样距离时的小尺度空间变异; 自相关尺度反映了人口密度的自相关的空间范围, 在自相关尺度内人口密度具有空间自相关性, 自相关尺度以外人口密度具有空间独立性; 基台值是不同采样间距中存在的半方差极大值, 表示人口密度中的空间变异性。偏基台值 (partial still, C_1), 也叫结构性方差, 即基台值减去块金值, 可以反映人口密度由空间结构特征引起的变异程度; 另外, 学者们经常要计算块金值与基台值的比值, 来比较局部性变异与总体变异间的比例关系。半变异函数云图是应用地统计学解释人口分布空间结构的基础。半变异函数云图通常可以被某些曲线方程所拟合, 这些曲线方程称为半变异函数的理论模型, 主要包括圆形模型 (Circular)、球状模型 (Spherical)、指数模型 (Exponential)、高斯模型 (Gaussian) 等。

城市人口分布已被大量研究证明可以用某些数学模型来拟合, 那么必然存在结构性特征; 结构性特征的存在, 意味着人口密度值与所在位置有关, 表现出一定的随机性。因此, 人口密度可以作为一个区域化变量, 用地统计学来分析。

3 结果与分析

3.1 空间自相关性的多尺度分析

在 ARCGIS 9.0 环境下, 利用 Geostatistical Analyst 模块来分析沈阳市人口分布的空间自相关性。为了比较其差异, 在计算半变异函数云图及模拟半变异函数理论模型时, 各种采样尺度下步长均确定为 1000m, 步长数确定为 10 组。由各粒度尺度下的半变异函数

云图 (其中, 100m、400m、700m、1000m 采样尺度的列于图 2) 可见, 半变异值随着步长变程的增加而增加, 但增加至一定程度后基本保持稳定, 因此半变异函数都可以用球状

模型来拟和。球状模型的表达式为:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C_1 \left(\frac{3}{2} \frac{h}{a} - \frac{1}{2} \frac{h^3}{a^3} \right) & 0 < h \leq a \\ C_0 + C_1 & h > a \end{cases} \quad (4)$$

式中, $r(h)$ 为半变异函数; h 为样本距; C_0 为块金值, 即 nugget; $C_0 + C_1$ 为基台值, 即 still; a 为变差距离, 即自相关尺度。与半变异函数相对应, 直接表征空间变量自相关性强弱的协方差函数为:

$$C(h) = C(0) - r(h) \quad (5)$$

式中, $C(0)$ 为先验方差, 数值上等于 $C_0 + C_1$ 。显然在自相关尺度范围内, 半变异函数和协方差函数分别是单调增函数和单调减函数。正由于 $r(h)$ 、 $C(h)$ 之间的关系, 可直接通过半变异函数来分析人口密度的空间自相关性。

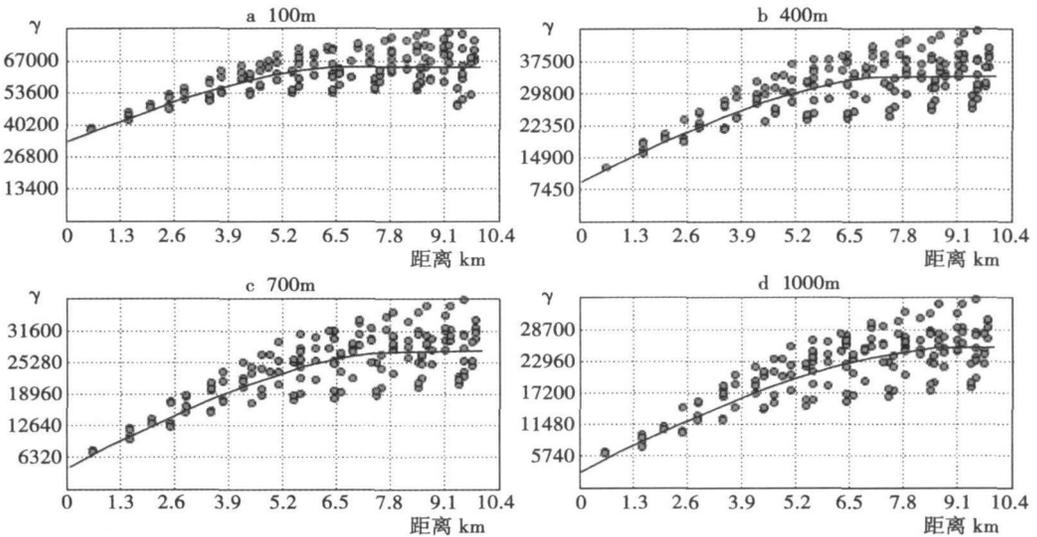


图 2 部分半变异函数云图

Fig 2 Some semivariograms

将采样各尺度下的半变异函数的主要参数列于表 2。可见, 自相关尺度由 6968 8m 至 9413. 8m 不等, 随粒度的增加而呈增加的趋势, 平均值为 8009 5m。沈阳市建成区面积 217km², 即平均城市半径为 8 31km。而本文的研究区范围东西长 21km, 南北宽 16km, 建立半变异函数时确定的步长数乘以步组数为 10km。由此可以认为, 城市人口密度的空间自相关性是客观存在的, 但其在多大范围内存在, 与计算人口密度时所采用的粒度有关, 粒度越大, 空间自相关尺度也越大。因此, 从这个角度出发, 700m 粒度就不失为研究沈阳市人口分布的适宜尺度, 其自相关尺度与城市平均半径最为接近。基台值随粒度的增加而减少, 表明人口密度的空间自相关性随粒度增加而增强。偏基台值随着粒度的增加而呈现减小的趋势, 但变化幅度相对较小。这表明粒度对人口密度的结构性特征有一定的影响, 随着粒度的增大, 由结构性特征所引起的人口密度的空间变异减小。块金值随着粒度的增大呈现明显的减小趋势, 相差达 11 倍以上。这表明, 当粒度较小时, 人口密度在小

表 2 不同粒度下的半变异函数理论模型及相关参数

Tab 2 Semivariogram models and parameters in different grains

| 粒度(m) | 模型类型 | 块金值 | 偏基台值 | 基台值 | 块金值/基台值 | 自相关尺度(m) |
|-------|-----------|---------|---------|---------|---------|----------|
| 100 | spherical | 32418.0 | 31812.0 | 64230.0 | 0.5047 | 6968.8 |
| 200 | spherical | 17405.0 | 27637.0 | 45042.0 | 0.3864 | 7183.8 |
| 300 | spherical | 11768.0 | 25684.0 | 37452.0 | 0.3142 | 7473.0 |
| 400 | spherical | 9199.6 | 24816.0 | 34015.6 | 0.2705 | 7918.5 |
| 500 | spherical | 6846.6 | 24649.0 | 31495.6 | 0.2174 | 7913.5 |
| 600 | spherical | 5214.2 | 23778.0 | 28992.2 | 0.1798 | 8218.7 |
| 700 | spherical | 3851.3 | 23542.0 | 27393.3 | 0.1406 | 8299.8 |
| 800 | spherical | 3285.5 | 23408.0 | 26693.5 | 0.1231 | 8206.1 |
| 900 | spherical | 3002.2 | 23185.0 | 26187.2 | 0.1146 | 8498.8 |
| 1000 | spherical | 2888.1 | 22796.0 | 25684.1 | 0.1124 | 9413.8 |

尺度范围内(例如本文的样本间距 1000m)具有较强的变异性;随着粒度的增加,被“留在”外面的变异减少,人口密度的局部变异迅速降低,空间自相关性显著提高。但这也折射出另外一个问题,在小尺度范围内,人口分布是否存在一种“微结构”而不是纯粹的变异或者随机性?特别是对于人口分布具有多核心结构的城市,每一核心所“控制”的范围显然不尽相同,如果仍用单核心结构和模型区分析,那么其余的“核心”就会被淹没。由块金值与基台值的比值可见,当粒度为 100m 时,人口密度由小尺度的变异和结构性原因所造成的变异几乎参半;随着采样尺度的增加,小尺度变异占总变异的比例逐渐降低;当采样尺度为 1km 时,该比值为 0.1124,已接近 10%,小尺度变异的影响已比较微弱,所显示出来的主要是人口分布的宏观结构。

3.2 空间自相关性的各向异性分析

粒度为 700m 时半变异函数的自相关尺度为 8299.8m,与沈阳市建成区平均半径 8.31km 最为接近,因此选择粒度 700m 来分析人口密度空间自相关的各向异性。空间变量自相关的各向同性是相对的,各向异性是绝对的^[14]。分析人口密度空间自相关的各向异性,有助于认识人口分布的特征。根据 700m 粒度的人口密度分布图(见图 3),选择 0、45、90、135 度(以正北方向为 0 度,角度顺时针旋转)来对比分析各向异性(半变异函数曲线见图 4)。

由图 4 可见,各个方向上的半变异函数曲线走势均具有一定的规律性,反映出人口密度在各个方向上都具有一定的空间自相关性,即人口分布具有一定的结构特征和内在联系。另外可以发现,不同方向上半变异函数曲线的最小值(初始值)非常接近,说明各个方向上,人口密度在小尺度范围内的变异性程度接近。但不同方向上,半变异函数曲线的变化趋势不尽相同。0 度方向上,半变异函数曲线基本符合球状模型,并且在 5.5km 左右达到最大值;45 度方向上,半变异函数曲线基本符合有基台值的线性模型,在 7.8km 左右达到最大值;90 度方向上,半变异函数曲线走势平缓,但却呈现出线性模型与球状模型的复合结构,左右两端较符合线性模型,在中段具有球状模型的特征;135 度方向上,半变异函数曲线与 45 度方向上的走势相似,较符合有基台值的线性模型,在 7.8km 左右达到最大值。45 度与 135 度方向上的半变异函数曲线趋势接近,说明这两个方向上人口密度的半异性相仿,人口分布的空间结构相似。事实上,对于任何半变异函数都是非

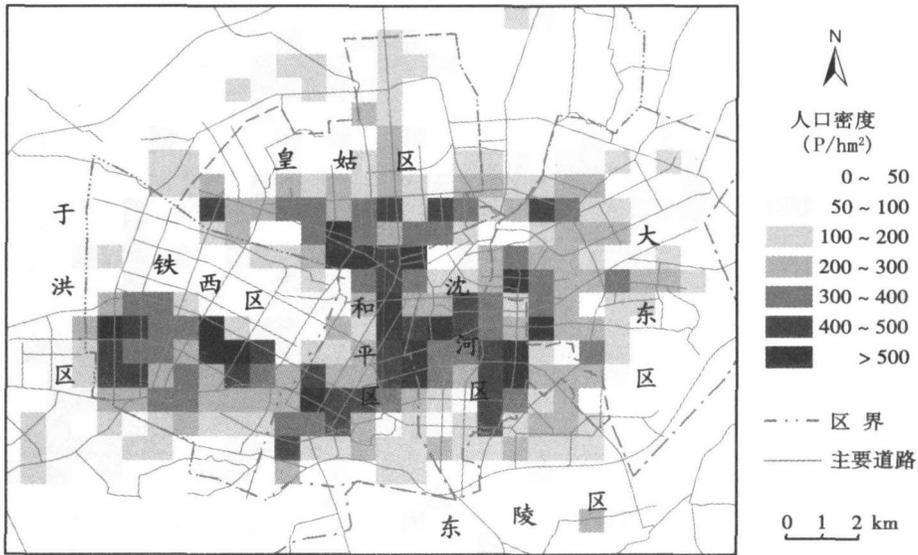


图3 沈阳市700m粒度人口密度图

Fig 3 Population density map of 700m grain in Shenyang city

降函数, 因此可以将0度及135度方向上半变异函数曲线右端下降看作人口分布在超过一定距离以后所表现出的随机性和偶然性, 人口密度之间的自相关关系已经不存在了, 这也正是要探讨人口密度空间自相关尺度的一个重要原因。如果城市范围足够大, 且半变异函数曲线呈现波状起伏, 则可说明人口分布在空间上具有一定的周期性。从上面的分析可以看出沈阳市人口密度空间分布呈现典型的带状异向性, 说明人口分布在不同方向上具有不同的结构特征。

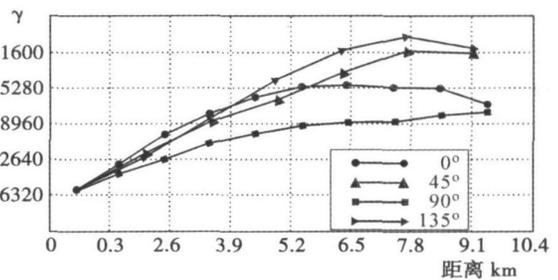


图4 不同方向上的半变异函数曲线

Fig 4 Semivariograms of different directions

4 结论与讨论

综合本研究, 得出以下主要结论:

(1) 沈阳市100m至1000m的10种粒度上, 人口密度半变异函数模型都符合具有块金值和自相关尺度的球状模型, 这表明沈阳市人口分布在一定的空间范围内具有空间自相关性, 并表现出某种结构特征, 具备了被某些人口密度模型加以抽象描述的客观基础。

(2) 由于采样粒度不同, 导致人口密度的空间自相关的各个指标, 如空间自相关尺度、块金值、基台值等有所不同, 说明人口分布的空间自相关性具有强烈的尺度依赖性, 城市人口格局会受到研究尺度的影响。在研究区范围确定的情况下, 自相关尺度与研究区半径的关系可以作为确定采样粒度(适宜尺度)的主要依据。

(3) 在特定的粒度上, 不同方向的人口密度半变异函数曲线具有不同的分布特征,

说明沈阳市人口密度呈现典型的带状异向性,这反映出人口分布在不同方向上具有不同的结构特征,单核心的人口密度模型是难以表征沈阳市人口分布的真实态势的。

基于本文及相关研究,提出以下观点和问题,与学者们商榷:

陈彦光等的研究表明,城市人口分布在信息熵趋于最大的同时,人口分布的长程相关显示明确的区域化倾向^[17]。在城市范围不减的情况下,这是否意味着城市人口分布正朝着多核心结构演化?王法辉等的研究证明沈阳市人口分布具有多核心特征^[18]。由此我们可以进一步引申,对于具有多核心的城市,如果这些人口分布核心相对于城市中心而言不是对称分布的,那么人口密度分布的各向异性是必然存在的;如果这些人口分布核心相对于城市中心而言是对称分布的,而且我们分析各向异性时选择的方向又恰恰是其中几个核心延展的方向,那么各向异性就很难得到反映。不过,这时我们可能会发现另外一个规律,如果人口密度的采样尺度以及计算半变异函数值时选择的间距适宜,那么半变异函数曲线必然会呈现出波状起伏的趋势。这也是分析人口密度分布各向异性时选择适宜方向的重要原因。再者,各向异性的存在是由于城市人口分布具有单核心不对称结构还是由于多核心结构(或者更复杂的结构)造成的应该在实际研究中加以注意和分析。冯健的研究已经说明人口分布具有多核心结构的城市,同样可以用单核心模型来拟合^[11]。对于某一时刻某一城市,人口分布状态是一种客观实在,人口密度模型是对人口分布特征的抽象表达,而模型中的参数只不过是借以表达人口分布格局的量化指标。纵使研究总要经历由相对真理趋向于绝对真理的过程,但在研究城市内部人口分布特征时,不管采用人口的区域平均密度,还是采用人口的景观密度,都必须面对尺度效应和适宜尺度的选择问题,毕竟,尺度与格局是紧密联系在一起的^[16];在建立某一城市某一时刻的人口密度模型时,应该注意人口分布的空间结构特征识别,机械地套用经典的人口密度模型,只会陷入形而上学的尴尬境地。

参考文献:

- [1] Clark C. Urban population densities. *Journal of Royal Statistical Society*, 1951, 114: 490~ 496
- [2] Naroll R S, Bertalanffy L von. The principle of allometry in biology and the social sciences. *General Systems Yearbook*, 1956, 1: 76~ 89
- [3] Stewart J Q, Warntz W. Physics of population distribution. *Journal of Regional Science*, 1958, 1: 99~ 123.
- [4] Sherrant G G. A model for general urban growth. In: Churchman C W, Verhulst M (eds.). *Management Sciences, Model and Techniques: Proceedings of the Sixth International Meeting of Institute of Management Sciences (V. 1. 2)*. Elmsford, N. Y: Pergamon Press, 1960. 147~ 159
- [5] Tanner J C. Factors affecting the amount travel. *Road Research Technical Paper No. 51*, HMSO (Department of Scientific and Industrial Research). London, 1961.
- [6] Smeed R J. The traffic problem in towns. *Manchester Statistical Society Papers*. Manchester: Norbury Lockwood, 1961.
- [7] Newling B E. The spatial variation of urban population densities. *Geographical Review*, 1969, 59: 242~ 252.
- [8] Berry B J, Kim H M. Challenges to the monocentric model. *Geographical Analysis*, 1993, 25: 1~ 4
- [9] Small K A, Song S. Population and employment densities: Structure and change. *Journal of Urban Economics*, 36: 292~ 313
- [10] 冯健. 杭州市人口密度空间分布及其演化的模型研究. *地理研究*, 2002, 21(5): 635~ 646
- [11] 冯健. 转型期中国城市内部空间重构. 北京: 科学出版社, 2004
- [12] 张仁铎. 空间变异理论及应用. 北京: 科学出版社, 2005. 13
- [13] Bongarcon D Francois. Theory of sampling and geostatistics: An intimate link. *Chemometrics and Intelligent La*

boratory Systems, 2004, 74: 143~ 148

- [14] 王政权. 地统计学及在生态学中的应用. 北京: 科学出版社, 1999, 151~ 155
- [15] 潘志强, 刘高焕. 面插值的研究进展. 地理科学进展, 2002, 21(2): 146~ 152
- [16] 李双成, 蔡运龙. 地理尺度转换若干问题的初步探讨. 地理研究, 2005, 24(1): 11~ 18
- [17] 陈彦光, 刘继生. 城市人口分布空间自相关的功率谱分析. 地球科学进展, 2006, 21(1): 1~ 9.
- [18] Wang Farhui, Meng Yar chun. Analyzing urban population change patterns in Shenyang, China 1982-90: Density function and spatial association approaches. Geographic Information Sciences, 1999, 5(2): 121~ 130

Analyzing spatial auto correlation of population distribution: A case of Shenyang city

DU Guo-ming^{1,2}, ZHANG Shu-wen¹, ZHANG You-quan^{1,2}

(1 Northeast Institute of Geography and Agricultural Ecology, CAS, Changchun 130012, China;

2 Graduate University, of the Chinese Academy of Sciences, Beijing 100049, China)

Abstract: For detecting the applicability of classical urban population density models on the microcosmic level, this paper calculates grid-based population landscape density of Shenyang city, analyzes its spatial autocorrelation and variability using theories and methods of geostatistics. It is found out that all semivariogram functions of population density fit with spherical model with nugget in ten kinds of grains from 100m to 1000m, indicating that population distribution presents structural characteristics in the spatial extension of Shenyang city. But because of different grains, the autocorrelation scales, nuggets, and sills, the rate of nugget and sill is different. This means that spatial autocorrelation of population distribution depended on the scale intensively, resulting in scale effects. By calculating semivariogram functions curves of different directions, it can be found that population density takes on a classical zonal anisotropy, which means that there were different structural characteristics in different directions for population distribution. The analysis of population density autocorrelation and variability should be taken as bases for researching urban population density distribution. The impacts of grain on population density autocorrelation and anisotropy are objective, so more attention should be paid to choosing feasible scale and identifying structure of population distribution. In the specific time for a certain city, the condition of population distribution is objective, while population density models are abstract expressions of population distribution characteristics, whose parameters are only quantitative indexes of expressing population distribution pattern. So, when researching spatial structure of population distribution in a city, the emphases should be put on how to identify spatial structural characteristics of population distribution, rather than on how to apply classical population density models mechanically.

Key words: spatial autocorrelation; urban population distribution; population density models; spatial variability; geostatistics