

# 基于遥感影像的中尺度人口统计数据空间化 ——以京津冀地区为例

吴健生<sup>1,2,†</sup> 许多<sup>1</sup> 谢舞丹<sup>1</sup> 彭建<sup>2</sup>

1. 北京大学深圳研究生院城市规划与设计学院, 人居环境科学与技术重点实验室, 深圳 518055; 2. 北京大学城市与环境学院, 地表过程与模拟教育部重点实验室, 北京 100871; † E-mail: wujs@szpku.edu.cn

**摘要** 基于 DMSP/OLS 夜间灯光数据、土地利用数据及其他与人口分布相关的社会经济与生态环境因子, 在中尺度区域上研究人口统计数据的空间化及应用。以京津冀地区为研究区, 首先获取 2010 年京津冀地区人口普查数据; 然后基于各个相关因子(公路、铁路、河流、坡度、土地利用和夜间灯光)计算概率系数; 最后综合各种输入变量和概率系数运用地理信息系统技术把人口普查数据分配到各个像元上。用城市人口普查数据对分配结果进行检验, 检验结果为: 京津冀地区 2010 年人口统计数据的空间化结果在城市尺度上的校验准确率可以达到 74%以上。此检验结果表明该人口统计数据空间化的方法可以很好地应用于中尺度区域人口统计数据的空间化。

**关键词** 夜间灯光; 人口统计数据; 空间化

**中图分类号** K901

## Spatialization of Demographic Data at Medium Scale Based on Remote Sensing Images: Regarding Beijing-Tianjin-Hebei as an Example

WU Jiansheng<sup>1,2,†</sup>, XU Duo<sup>1</sup>, XIE Wudan<sup>1</sup>, PENG Jian<sup>2</sup>

1. Key Laboratory of Urban Habitant Environment Science and Technology, School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518055; 2. Laboratory of Earth Surface Processes of Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing 100871; † E-mail: wujs@szpku.edu.cn

**Abstract** This study selects Beijing-Tianjin-Hebei region as the study area and attempts to realize the population spatialization with the data of DMSP/OLS, land use, socio-economic and ecological environmental factors related to population distribution at the medium scale. Firstly, the authors obtain the census data of Beijing-Tianjin-Hebei region in 2010, and then calculated the probability coefficient based on these factors (selected roads, railways, rivers, slope, land use and nighttime light), and finally assign the census data to each pixel using all related variables and their probability coefficients with the software of geographic information system (GIS). City census data was chosen to test the research result, and the results showed that the accuracy rate of the population spatialization in the Beijing-Tianjin-Hebei region is more than 74% in the urban scale. The method for population spatialization can be well applied to the spatialization of demographic data at medium scale.

**Key words** nighttime light; census data; spatialization

人口空间分布指一定的时间点人口在各地区中的分布状况, 是人口动态过程在空间上的表现形式<sup>[1]</sup>。目前, 人口统计数据一般以各级行政区作为

基本的统计单元, 导致在与基于自然单元特别是基于格网的自然生态数据的叠置分析上存在一定的困难, 并存在时空分辨率低、精度低等问题。因此,

国家自然科学基金(41130534)资助

收稿日期: 2015-01-11; 修回日期: 2015-03-20; 网络出版日期: 2015-07-13

人口统计数据的空间化极为重要。空间化的人口统计数据可以很好地与土地利用、生态环境和气候等背景数据相结合,为进行交互作用的分析提供便利条件,具有重要的研究意义和科学价值<sup>[2]</sup>,同时在犯罪管理、经济决策、灾害预警、基础服务设施建设等领域有广阔的应用前景<sup>[3]</sup>。

目前,国内外对于全球尺度及全国尺度的人口数据空间化进行了深入探索,成立了专门的人口空间化项目组,建立了不同尺度、不同研究区域的人口分布数据库,如全球人口数据库 GPW<sup>[4]</sup>、全球资源信息数据库 UNEP/GRID、美国国家橡树岭实验室全球人口项目 LandScan 以及中国科学院地理科学与资源研究所创建的 2000 年全国 1:100 万人口空间化数据库<sup>[5]</sup>。早期的人口数据空间化的具体模型及方法也较为多样,例如多元回归模型<sup>[6]</sup>、土地利用模型<sup>[7]</sup>、重力模型<sup>[8]</sup>、多源数据融合模型<sup>[9]</sup>、复合面积内插法<sup>[10]</sup>、核密度估计法<sup>[11]</sup>、作息空间法<sup>[12]</sup>等。近几年对人口统计数据空间化的研究中越来越多地运用了夜间灯光数据。

将夜间灯光数据应用于人口空间化的研究已经取得不少成果。理论上,人口密度一般是指单位面积土地上居住的人口数,但随着社会经济的不断发展,人们的活动区域已经不仅仅局限于居住地范围,也就是说,现时的人口密度应该与人类活动强度密切相关。这也是夜间灯光能应用于人口空间化研究的重要基础之一。DMSP/OLS 夜间灯光数据的分辨率为 30 弧秒,赤道处约为 1 km<sup>2</sup>。DMSP/OLS 在夜间具有极强的低光感测能力,因而具有探测地球表面人工照明的特殊能力,如人类居住区、废气燃烧、火灾以及渔船所产生的光<sup>[13]</sup>,甚至可以探测到小规模居民地、车流等发出的低强度灯光<sup>[14]</sup>。自 1992 年起,夜间灯光数据每个栅格内含有一个数字(DN),范围为 0~63,表征年均夜间灯光强度<sup>[15]</sup>。自从夜间灯光数据作为人类活动指标的潜力被确定后,国内外学者已经开展了许多有关夜间灯光数据与关键社会经济变量关系的研究。早在 1999 年,Elvidge 等<sup>[16]</sup>对于夜间灯光影像的研究就证明其与人口分布密切相关,这说明夜间灯光数据可以作为人类活动的表征,即人口统计数据空间化的重要直接指示因子。自此,越来越多的研究人员将夜间灯光影像加入到人口统计数据空间化的相关研究中。如曹丽琴等<sup>[17]</sup>基于 DMSP/OLS 数据估算了湖北省各县市的人口,卓莉等<sup>[18]</sup>选用 DMSP/OLS 数据模

拟了中国的人口密度,Zeng 等<sup>[19]</sup>利用 DMSP/OLS 影像和土地利用数据将中国人口进行空间化,Roy Chowdhury 等<sup>[20]</sup>利用夜间灯光数据及用电量等辅助数据建立印度恒河平原的印度部分的城镇人口模型,等等。然而,纵观国内外相关的研究,基本上都仅利用灯光数据与人口统计数据之间的回归关系构建相应的人口分布模型<sup>[17-21]</sup>,即大多只构建夜间灯光数据、土地利用等社会经济因子与人口统计数据之间的关系,而忽略影响人口分布的自然环境因子。例如,海拔高度与人口密度相关性较强<sup>[4]</sup>,道路、河流以及铁路均在一定程度上影响并反映人口的分布情况。美国 LandScan 人口数据集的建立已考虑了这些会影响人口分布的自然环境及社会经济因子。

本文以京津冀地区为研究区,基于 DMSP/OLS 夜间灯光数据,同时考虑土地利用、交通、坡度和河流等社会经济和自然环境因子,采用层次分析法和专家打分法对各因子赋予相应的权重,对人口进行空间化,得到京津冀地区 1 km × 1 km 的格网人口数据。研究结果可以丰富人口数据空间化方法的研究,为人口数据在社会经济、生态环境方面中的广泛应用提供基础,便于该地区的人口管理和城市规划,具有重要的理论和实践意义。

## 1 研究区

京津冀地区包括北京市、天津市与河北省三者所处的地域范围(图1),面积 21.6 万 km<sup>2</sup>,地势西北高、东南低,由西北向东南倾斜,各种地貌类型齐全。京津冀地区河流众多,主要河流分属海河、滦河、内陆河和辽河 4 个水系。交通运输条件发达,是中国北方最大的“海洋经济”与“大陆经济”的连接枢纽,交通便利。地形、河流以及交通等对于该地区的人口分布有十分重要的影响。

2010 年全国第六次人口普查数据显示,京津冀的人口规模达到 1.04 亿,占全国人口比重的 7.79%。高密度的人口分布对城市的人口管理和城市安全规划都提出了挑战,研究人口的空间分布则可以为解决这些城市问题提供基础。此外,在世界性经济危机和国内经济形势重大变化的背景下,该地区是并将长期是北方经济的重要增长极,社会经济要素和人口空间分布具有相互促进的关系,因此,研究京津冀地区的人口空间分布及其影响要素对京津冀地区目前和将来的发展都具有重要作用。



图1 京津冀地区区位图与城市行政边界  
Fig. 1 Location map of Beijing-Tianjin-Hebei and urban administrative boundary map

## 2 数据与方法

### 2.1 数据来源

研究所用到的基础数据包括京津冀地区 2010 年的人口普查数据、数字高程图、河流路网分布图、土地利用数据、夜间灯光分布图以及市级行政边界数据,数据来源及属性如表 1 所示。其中稳定灯光数据覆盖范围为经度  $180^{\circ}\text{W}$ — $180^{\circ}\text{E}$ 、纬度  $65^{\circ}\text{S}$ — $75^{\circ}\text{N}$ ,基本覆盖全球人类活动区域。该数据为 NOAA 提供的无饱和数据,灰度值的范围为  $0\sim 1555.31$ ,0 值表示无灯光区域,灰度值越大表示

该处灯光亮度值越高。

为了便于裁切以及不同图件的叠合,本研究将所有空间数据统一为 Lambert 等积方位(Lambert Azimuthal Equal-Area)投影,栅格数据采用最邻近方法重采样,网格大小为  $1\text{ km} \times 1\text{ km}$ 。本文的研究区是京津冀地区,而有些基础图件是全国范围的,所以需要将所需图件均裁切至研究区范围,方便应用于之后的研究中。本研究共选取 6 个对人口分布有影响的相关因子,分别是夜间灯光、土地利用、坡度、道路、河流以及铁路。其中,夜间灯光和土地利用均是可以直接利用或已经过转化的栅格图,坡度数据需要通过高程数据来获取,其余 3 个相关因子的矢量图不可以直接表征其对人口分布的影响,需要处理为相应的栅格图,分辨率均为  $1\text{ km} \times 1\text{ km}$ 。

### 2.2 研究方法

本文将 DMSP/OLS 夜间灯光数据、土地利用数据及影响人口分布的社会经济和自然环境因子(坡度、道路、河流等)进行综合叠置,对京津冀地区 2010 年人口普查数据进行空间化处理。研究的基本思路如下:1) 根据研究区实际情况,选取对人口分布有显著影响的相关因子(包括夜间灯光、土地利用、坡度、道路、河流以及铁路);2) 在每一个影响因子的栅格图内,按照不同值对人口分布的影响程度,为每一个单元格赋权重;3) 结合层次分析法和专家打分法,就每个因子对人口分布的影响程度进行赋权重,然后将之前生成的各个因子的权重分布图按照该权重进行叠加分析,得到每个像元的概率系数;4) 综合各个相关因子和概率系数,运用地理信息系统技术,将人口普查数据分配到各个像元上<sup>[22]</sup>。

表 1 数据来源及属性  
Table 1 Data sources and attributes

数据名称	来源	年份	分辨率
人口普查数据	2011 年《中国城市统计年鉴》	2010	—
SRTM 90 m 空间分辨率 DEM	<a href="http://srtm.csi.cgiar.org/">http://srtm.csi.cgiar.org/</a>	2010	$90\text{ m} \times 90\text{ m}$
土地覆被数据	国家基础地理信息中心研制的全球 30 m 地表覆被产品(GlobeLand30) <a href="http://www.globallandcover.com/">http://www.globallandcover.com/</a>	2010	$30\text{ m} \times 30\text{ m}$
稳定灯光数据	美国 NOAA 国家地理数据中心	2010	约 $1\text{ km} \times 1\text{ km}$
河流路网数据	中国科学院资源环境科学数据中心	2010	—
城市行政边界数据	1:400 万地级行政界线下载于社会科学数据共享服务网	2010	—

### 2.2.1 指标体系建立

早在 1999 年, Elvidge 等<sup>[16]</sup>的研究就证明夜间灯光影像与人口分布密切相关, 即 DMSP/OLS 夜间灯光数据可以作为人口统计数据空间化的重要指示因子。Tian 等<sup>[23]</sup>证明土地利用数据也与人口分布有着密切的关系, 因此, 土地利用数据也是本文重要的直接因子之一。王春菊等<sup>[9]</sup>在利用多源数据融合技术对福建省人口统计数据进行空间化的过程中, 通过计算研究区人口分布与各影响因子的关系, 表明平均人口密度与海拔高度、土地利用、道路网密度、河网密度、与海岸线的距离以及居民点密度相关性较强。因此, 本研究在参考已有研究的基础上, 同时考虑空间化方法的侧重点和数据的可获取性, 选取重要指示因子(DMSP/OLS 夜间灯光、土地利用)和辅助影响因子(坡度、距最近公路距离、距最近铁路距离、距最近河流距离), 建立京津冀地区人口空间化评价指标体系(表 2)。

表 2 人口空间化因子权重评价指标体系  
Table 2 Assessment indicator system of population spatialization factors weight

A 目标层	B 准则层	C 指标层
人口空间化评价指标体系 A	重要指示因子 B <sub>1</sub>	夜间灯光强度 C <sub>1</sub>
		土地利用类型 C <sub>2</sub>
	辅助影响因子 B <sub>2</sub>	坡度 C <sub>3</sub>
		距最近河流距离 C <sub>4</sub>
		距最近公路距离 C <sub>5</sub>
		距最近铁路距离 C <sub>6</sub>

### 2.2.2 因子权重确定

权重确定是研究中至关重要的一部分, 直接影响人口空间化结果的准确性。美国 LandScan 人口数据集<sup>[24]</sup>采用专家打分法来确定因子权重。本研究考虑到专家打分法主观性强的特点, 采用层次分析法与专家打分法相结合的方式来确定所需权重。土地利用数据中各种土地利用类型的权重以及 6 个相关因子之间的相对重要性程度采用专家打分法, 对于分值的获取采用两种途径: 一是以邮件形式向该研究领域相关专家征询意见, 获取打分结果, 二是请北京大学相关实验室的老师打分, 结合每位专家或老师的打分值, 取均值作为专家打分的最终结果。6 个相关因子的权重则采用层次分析法完成。

土地利用数据的源数据有 Cultivated Land, Forest, Grassland, Shrubland, Wetland, Water Bodies,

Artificial Surfaces 和 Bareland 共 8 类, 按照专家打分法确定其权重依次为 3, 2, 1, 2, 0, 0, 9 和 1。为与土地利用类型权重的值域范围保持一致, 以及考虑最终加权计算的科学性, 本文将不同因子的权重值域均归一化至 0~9 (各栅格原值域及新权重值域如表 3 所示), 得到各相关因子对人口分布影响的权重分布(图 2)。

表 3 各相关因子原值域与新权重值域  
Table 3 Original range and new weights range of relevant factors

因子名称	原值域	新权重值域
夜间灯光	0~1555.31	0~9.00064
距最近公路距离	0~0.238630340	0.314372~9.00225
距最近铁路距离	0~1.3543396	0.0553914~9.00225
距最近河流距离	0~1.219773412	0.0615022~9.00225
坡度	0~69.49290466	0.0144475~9.00813

接着, 运用基于指数标度的层次分析法(相对重要性程度由专家打分获取)计算 6 个因子(包括夜间灯光、土地利用、坡度、公路、河流以及铁路)影响人口分布的权重, 取相对重要度  $\alpha=1.618$ <sup>[25]</sup>, 指数标度的含义见表 4, 准则层和指标层的矩阵如式(1)~(3)所示。

$$W_{B_1 B_2} = (a_{ij})_{2 \times 2} = \begin{bmatrix} \alpha^0 & \alpha^3 \\ \alpha^3 & \alpha^0 \end{bmatrix}, \quad (1)$$

$$W_{C_1 C_2} = (a_{ij})_{2 \times 2} = \begin{bmatrix} \alpha^0 & \alpha^3 \\ \alpha^3 & \alpha^0 \end{bmatrix}, \quad (2)$$

$$W_{C_3 C_4 C_5 C_6} = (a_{ij})_{4 \times 4} = \begin{bmatrix} \alpha^0 & \alpha^0 & \alpha^0 & \alpha^{-1} \\ \alpha^0 & \alpha^0 & \alpha^0 & \alpha^{-1} \\ \alpha^0 & \alpha^0 & \alpha^0 & \alpha^{-1} \\ \alpha^1 & \alpha^1 & \alpha^1 & \alpha^0 \end{bmatrix}, \quad (3)$$

由此最终计算出夜间灯光、土地利用、坡度、距最近公路距离、距最近河流距离以及距最近铁路距离 6 个相关因子对人口分布影响的权重分别为 25, 6, 1, 1, 1 和 2。

### 2.2.3 人口分布图的生成

在建立指标体系, 确定各因子权重后, 即可对人口数据进行空间化处理, 先在前两步的基础上, 得到综合权重分布图, 其计算公式如下:

$$U_{\text{总}} = \sum_{i=1}^6 \frac{U_i P_i}{\sum_{i=1}^6 P_j}, \quad (4)$$

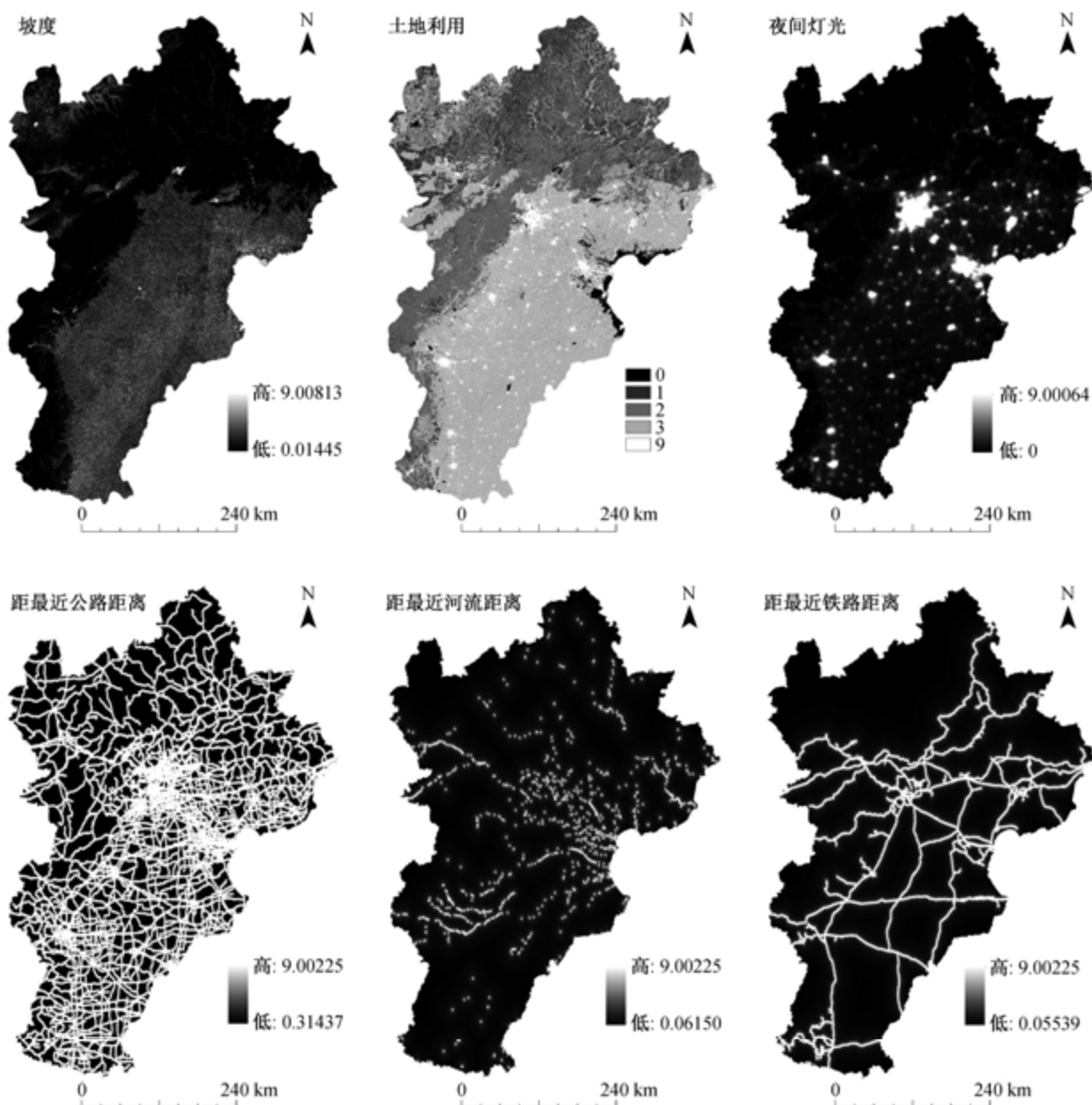


图 2 各相关因子权重分布  
Fig. 2 Weight distribution of relevant factors

表 4 指数标度的含义

Table 4 Meaning of exponent division

标度 $\alpha_{ij}$	标度定义
$\alpha^0 = 1$	$i$ 和 $j$ 同样重要
$\alpha^1 = 1.618$	$i$ 比 $j$ 稍微重要
$\alpha^2 = 2.618$	$i$ 比 $j$ 重要
$\alpha^3 = 4.236$	$i$ 比 $j$ 明显重要 <sub>4</sub>
$\alpha^4 = 6.854$	$i$ 与 $j$ 相比很重要
$\alpha^5 = 11.09$	$i$ 与 $j$ 相比绝对重要

$U_{总}$ 表示每个栅格的综合权重值,  $U_i$ 表示该栅格内第  $i$  个相关因子的值,  $P_i$ 表示第  $i$  个相关因子的权重,  $P_j$ 表示第  $j$  个相关因子的权重。需要说明的是, 该方法存在一种假设, 即灯光灰度值为 0 的区域以及土地利用类型为水体的区域没有人口分布, 因此, 需要将上面所得综合权重分布图中灯光灰度值为 0 以及土地利用类型为水体的栅格赋值为 0。

接着将人口分配到每一个栅格中去, 先利用式 (5)计算权重总和:

$$A_{\text{总}} = N \cdot U_M, \quad (5)$$

$A_{\text{总}}$ 代表综合权重分布图中每个栅格的权重总和,  $N$ 表示综合权重分布图中的栅格个数,  $U_M$ 表示综合权重分布图中每个栅格的权重平均值。

然后, 将总人口与计算得到的权重总和相除, 即可得到一个单位的权重代表的人口系数。由于该研究是探究中尺度上人口统计数据的空间化问题, 所以此处的总人口是京津冀地区的人口数量之和, 是将总人口在整个京津冀地区上进行分配。最后将单位权重人口系数与综合权重分布图相乘, 得到了最终人口空间分布图。计算公式如下:

$$S_{\text{人口}} = \frac{M}{A_{\text{总}}} \times U_{\text{总}}, \quad (6)$$

$S_{\text{人口}}$ 代表每个栅格内的人口数;  $M$ 代表 2010 年京津冀地区人口普查总数。

### 3 结果与分析

#### 3.1 人口数据空间化

最终的人口数据空间化结果如图 3 所示。从图 3 可以看出, 人口呈现东南多、西北少的空间分布。颜色较深的地方即为人口分布相对较密的地区, 人口分布的最高值分别是北京市、天津市、保定市、唐山市、石家庄市以及邯郸市等城市, 这与由人口普查数据所获得的实际人口分布情况较为相符。然而, 具体的人口分配结果是否精确, 还需进行进一步的数据检验。

#### 3.2 结果校验

借助人口分布的影响因素得出的模拟人口分布图可以大体上反映京津冀地区的人口分布, 使得对于与人口分布相关的研究不受行政区的限制, 可以任意选取感兴趣的区域进行分析。然而, 人口数据模型或人口空间化结果的检验始终是一个比较困难的问题, 主要由于人口分布不同于一般位置固定的地理要素。如土壤、植被、土地利用等地理要素由于位置固定, 可以在一定程度上细分, 而且结果的可靠性比较高。空间分辨率过高的人口密度数据很难获得, 即使获得了也不一定科学或不具有任何实际意义。因此, 很难对人口分布模型或是人口空间化结果进行充分性检验, 一般只能进行必要性检验。由于受到数据来源的限制, 本研究只能利用研究区的亚级数据对研究结果进行校验。在上述过程中使用京津冀地区的人口总数进行整个区域内的人

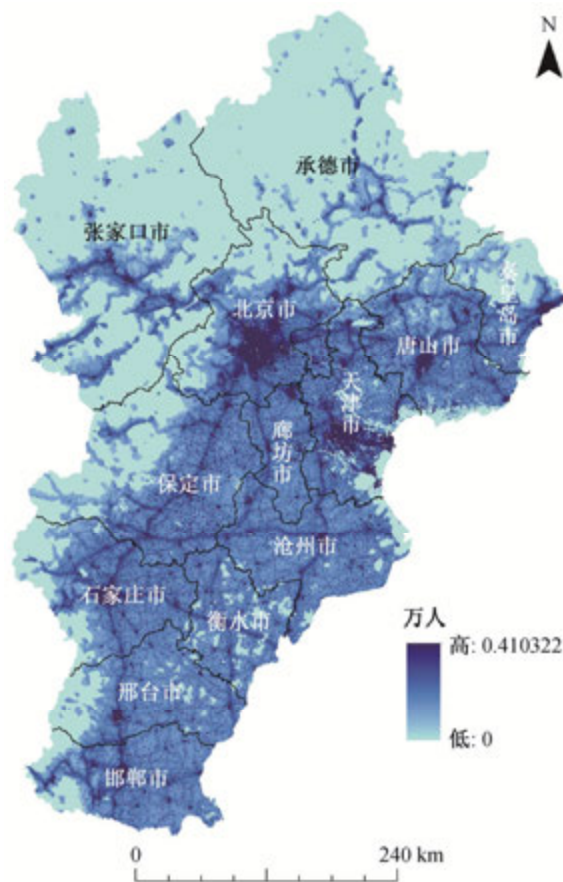


图 3 京津冀地区 2010 年人口分布  
Fig. 3 Population distribution of Beijing-Tianjin-Hebei in 2010

口分配, 因此, 在研究结果校验时可以用每个城市的实际人口统计数据与分配结果比较, 结果如表 5 所示。

按照统计数据总人口分配后的结果加和统计后, 总人口为 9693.98 万, 比统计数据总人口 9540.69 万多出 153.29 万, 因此需要对分配结果按照式(7)进行误差补偿:

$$P = P_0 - \frac{P_0}{P_{\text{总}}} \times \Delta P, \quad (7)$$

$P$ 为经过误差补偿后的城市人口,  $P_0$ 为误差补偿前的分配城市人口,  $P_{\text{总}}$ 为误差补偿前分配总人口,  $\Delta P$ 为误差补偿前的分配总人口与统计数据总人口的差值。

经过误差补偿后的城市人口分布情况如表 5 所示。所有城市的分配结果统计值与人口统计数据的差值百分比均在 26% 以内, 即对于中尺度区域来说, 该研究方法基本上可以控制各城市人口分配结

果的准确率在 74% 以上。各城市差值百分比的平均值为 15.58%，这与以往的人口模拟精度相近<sup>[17,21]</sup>，甚至高于一些研究结果的精度<sup>[26]</sup>。然而不同城市之间计算误差的绝对值差别很大，范围在 1.32%~25.67% 之间，即一些城市的人口分配结果相对较为准确，如承德市、天津市、保定市、北京市和邢台市，误差百分比均控制在 10% 以下；而另一些城市的人口分配结果存在较大误差，如张家口市、沧州市、唐山市以及邯郸市，误差均达到 21% 以上。需要对这种不同城市精度差异较大的原因进行进一步分析。

### 3.3 误差分析

研究区内所有城市的人口统计数据空间化结果在城市尺度上的校验准确率在 74% 以上，但与一些更为精细的人口分布模型相比，误差仍然较大。根据数据来源、研究方法以及整个操作流程分析，可以大致判断误差主要由以下几个因素引起。

1) 利用遥感影像提取影响因子信息时，分辨率对于提取结果影响很大，即分辨率越高，所包含的信息越丰富，计算结果也越精确<sup>[27]</sup>。本研究采用的数据分辨率均为 1 km，这在一定程度上降低了人口分配结果的精度。若获得高分辨率的遥感影像，会提高结果精度。

2) 从表 5 的比较结果来看，除承德和秦皇岛

外，其余分配结果统计数据的分布较为均匀，人口数量由多到少的城市之间的人口差值梯度相近，这与灯光分布情况十分相似。由于考虑到夜间灯光数据能更直观地反映人口的分布情况，我们人为地加重夜间灯光的权重(这是必要的)，但在一定程度上会削弱其他因子对人口分布实际情况的影响程度，导致最终人口分布情况与夜间灯光分布情况相近。

3) 夜间灯光影像具有一定程度的“灯光溢出”现象，即夜间灯光影像上探测到的有灯光区域与实际城区范围并不完全一致，探测到的灯光分布范围往往大于其对应居住的空间范围<sup>[28]</sup>。有些栅格受到临近栅格的“灯光溢出”影响，导致在夜间灯光影像上反映出来的灰度值过高，最终人口分配结果偏高。

4) 由于受到数据来源限制，在选取相关因子时，其他对于京津冀地区人口分布有影响的因子没有考虑在内，如某些社会经济要素也可以很直观地反映人口的实际分布情况。在下一步研究中，如果加入这些对人口分布有影响或可以较为直观地反映人口分布实际情况的社会经济要素，一方面可以更加精确地进行人口分配，另一方面也可以削弱夜间灯光数据单因子的高权重对其他因子的抑制程度，达到更好的人口分配效果。

5) 本研究对京津冀地区总人口进行空间分配

表 5 各城市人口统计数据与分配结果统计对比

Table 5 Statistical comparison of urban population statistics and allocation results

城市名称	人口统计数据	分配结果统计	差值百分比/%	误差补偿后	
				分配结果统计	差值百分比/%
北京	1257.80	1158.50	-7.89	1140.18	-9.35
保定	1161.01	1106.61	-4.69	1089.11	-6.19
天津	984.85	965.80	-1.93	950.53	-3.48
石家庄	989.16	839.31	-15.15	826.04	-16.49
邯郸	963.50	737.63	-23.44	725.97	-24.65
唐山	735.00	933.00	26.94	918.25	24.93
沧州	730.89	923.88	26.41	909.27	24.41
邢台	732.03	671.14	-8.32	660.52	-9.77
张家口	465.97	594.98	27.69	585.58	25.67
衡水	440.20	538.57	22.35	530.06	20.41
廊坊	419.02	511.73	22.13	503.64	20.19
承德	372.96	373.95	0.27	368.04	-1.32
秦皇岛	288.30	338.86	17.54	333.50	15.68



时,无论对小城市还是特大城市,都使用相同的权重,没有根据城市级别的不同考虑权重差异,降低了人口空间化结果精度。下面基于 Zhuo 等<sup>[29]</sup>的研究探讨城市分类赋权重是否可以减小误差。首先,以中国科学院计算机网络信息中心国际科学数据镜像网站下载得到 2010 年 MYDND11M 中国 250 m NDVI 7 月份合成产品,该数据由 MYDLT1D 产品经过合成计算得到,时间为 2010 年 7 月份,计算方法为取月平均值,分辨率为 250 m。然后,利用京津冀地区城市边界,将 NDVI 图像裁切至研究区范围,如图 4 所示。根据已有研究,NDVI 值低于 0.1 的区域一般代表沙漠、冰川、岩石或水体,而 NDVI 值高于 0.9 的区域一般代表浓密的森林<sup>[29]</sup>,这些区域一般都不适宜人类居住。因此,需要判断每个栅格的居住适宜性。利用 ArcGIS 中的重分类功能将 NDVI 图层划分为适宜居住区和非适宜居住区,如图 5 所示。接下来,利用式(8)判断每个城市所属类型,结果如图 6 所示。

$$RA = \frac{LPA}{HA} \times 100\%, \quad (8)$$

RA 代表区域内有光区占适宜居住区的比例, LPA 代表区域内有光区的栅格总数, HA 代表区域内适宜居住区的栅格总数。

在 Zhuo 等<sup>[29]</sup>的研究中, Type1 类型的区域人口估计是基于灯光密度与区域总人口之间的相关性进行的, Type2 类型的区域人口估计是基于灯光密度与区域农村人口之间的相关性进行的。也就是说,这两种区域的差别在于人口属性(全部人口/农村人口)存在着差异。考虑到这一点,对于两种不同类型区域的因子权重赋值应该随之发生变化。Type1 区域各项权重仍为原值;在 Type2 区域,权重值发生变化的主要为土地利用数据,该区域农村人口比例较大,因此人工地表用地类型所对应的权重应有所降低。最终,将源数据分类中的 Cultivated Land, Forest, Grassland, Shrubland, Wetland, Water Bodies,

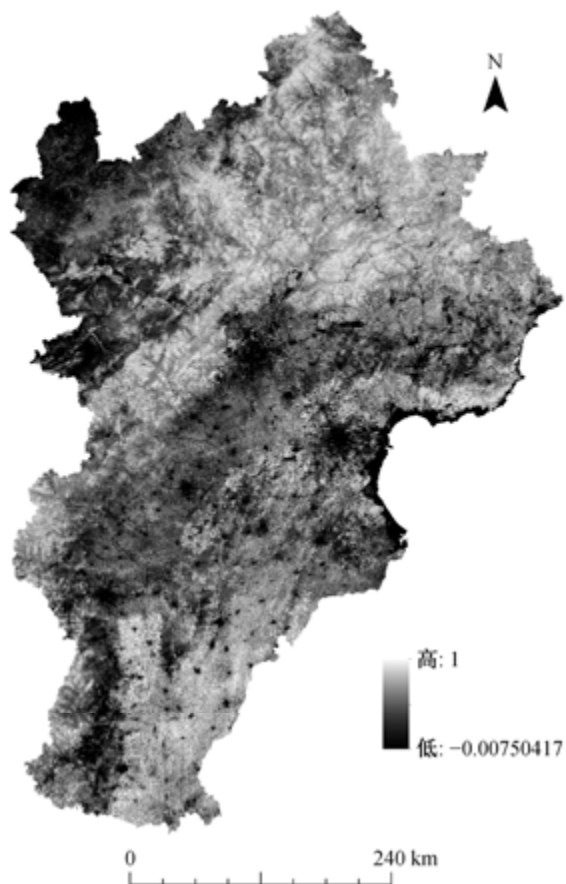


图 4 京津冀地区 2010 年 NDVI 分布

Fig. 4 NDVI distribution of Beijing-Tianjin-Hebei in 2010



图 5 京津冀地区 2010 年适宜居住区分布

Fig. 5 Livable zone distribution of Beijing-Tianjin-Hebei in 2010





图 6 京津冀地区 2010 年各城市人口分配类型  
Fig. 6 Population allocation types for all cities of Beijing-Tianjin-Hebei in 2010

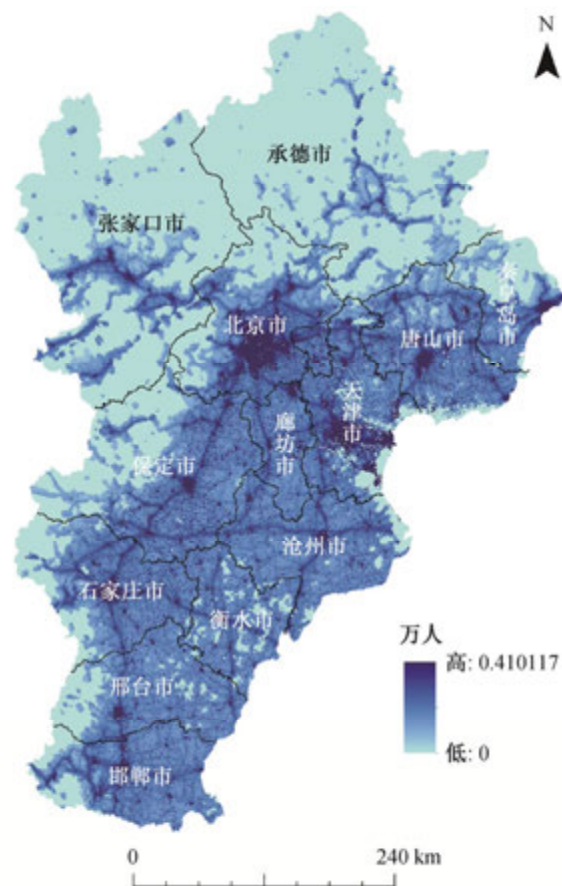


图 7 分区修正后京津冀地区 2010 年人口分布  
Fig. 7 Population distribution of Beijing-Tianjin-Hebei in 2010 after zoning amendment

Artificial Surfaces 以及 Bareland 按照专家打分法确定其权重依次为 3, 2, 1, 2, 0, 0, 6, 1。夜间灯光、土地利用、坡度、距最近公路距离、距最近河流距离以及距最近铁路距离 6 个相关因子对人口分布影响的权重仍然分别为 25, 6, 1, 1, 1, 2。然后按照 2.2.3 节的步骤得到分区修正后的 2010 年京津冀地区人口分布图(图 7)。

最后, 利用 ArcGIS 进行区域统计, 并对统计结果按照式(7)进行误差补偿, 得到的城市人口分布情况如表 6 所示。对比表 5 和表 6 的结果可知, 进行分区修正后, 人口统计数据分配结果的精度在整体上有所提高, 且修正后平均误差为 14.92%, 低于之前的平均误差 15.58%。因此, 这也是影响验证结果的因素之一。

#### 4 结论

现有人口统计数据一般以各级行政区作为基本的统计单元, 使得人口数据的自然生态数据的叠置分析存在一定的困难, 并存在时间分辨率低、空间分辨率低、精度低等问题。因此, 人口统计数据的空间化极为重要。国内外学者对于全球尺度及全国尺度的人口数据空间化进行了很多探索, 方法上相

表 6 分区修正和误差补偿后各城市人口统计数据与分配结果统计对比

Table 6 Statistical comparison of urban population statistics and allocation results after zoning amendment and error compensation

城市名称	人口统计数据	分配结果统计	差值/%
北京	1257.80	1142.64	-9.16
保定	1161.01	1090.49	-6.07
天津	984.85	971.41	-1.36
石家庄	989.16	826.27	-16.47
邯郸	963.50	726.85	-24.56
唐山	735.00	926.68	26.08
沧州	730.89	918.21	25.63
邢台	732.03	660.61	-9.76
张家口	465.97	563.23	20.87
衡水	440.20	531.16	20.66
廊坊	419.02	503.72	20.21
承德	372.96	361.97	-2.95
秦皇岛	288.30	317.45	10.11

对成熟,但目前大多数方法没有用到夜间灯光数据这一可以直接反应人类活动强度的因子,用到该数据的研究一般只构建夜间灯光数据以及土地利用等社会经济因子与人口统计数据之间的关系来模拟人口分布情况,并没有考虑影响人口分布的自然环境因子。本文将自然环境对人口分布的适宜性与限制性考虑在内,基于 DMSP/OLS 夜间灯光数据、土地利用数据及坡度、河流、交通路网等与人口分布密切相关的社会经济与自然环境因子遥感数据进行综合叠置以实现人口空间化。具体生成步骤如下: 1) 获取 2010 年京津冀地区人口普查数据; 2) 基于各个相关因子(本文选取了公路、铁路、河流、坡度、土地利用和夜间灯光)计算概率系数; 3) 综合各种输入变量和概率系数运用地理信息系统技术把人口普查数据分配到各个像元上。

在人口分配过程中所使用的人口普查数据是 2010 年京津冀地区人口总数,之后用城市人口普查数据对分配结果进行检验,检验结果为:京津冀地区人口统计数据的空间化结果在城市尺度上的校验准确率可以达到 74% 以上,与以往研究结果相近。经过分区赋权重修正后的人口统计数据分配结果精度有所提高,平均误差由 15.58% 降为 14.92%。此检验结果表明,该人口统计数据空间化方法可以很好地应用于中尺度区域人口统计数据的空间化。但是,这种方法的人口分配结果与一些更为精细的人口分布模型相比,仍然存在着较大误差,主要由于夜间灯光权重的人为加重、“灯光溢出”现象以及分析所需数据来源限制。

人口统计数据空间化研究经历了从单纯的统计数据格网化方法,到综合自然、社会等因素不同程度影响下的空间模型<sup>[30]</sup>,并随着 3S 技术的逐步成熟,加入可以直接表征人类活动强度的基础格网数据,使得人口空间化研究取得很多成果。但此类研究在实际应用中若要达到更好的效果,还需要解决以下问题。

1) 准确的、时空一致性高的人口统计数据与相关基础数据难以收集,如人口统计数据、地理要素专题图以及遥感影像的时相匹配问题,而这恰恰是提高模拟精度的基础<sup>[3]</sup>。因此应逐步建立不同尺度的、系统的、可靠的数据集<sup>[2]</sup>,这对于人口空间分布研究中基础数据的获取具有重要作用,也会在很大程度上推进研究的发展。

2) 人口结构统计数据的空间化研究还不十分

成熟。当前研究大多集中在人口数量与环境的关系方面,很少在人口结构方面研究人口分布状况,如性别、年龄、受教育程度等<sup>[31]</sup>。所以下一步的研究重点可以由人口数量的空间化扩展到人口结构统计数据的空间化。

3) 空间化结果的精度验证有待完善。目前的研究中使用较多的验证方法是对估算结果进行定性描述或是进行抽样验证,对估计结果在细致尺度上的空间分布精度没有很好的评估方法。未来的研究应当建立完善、合理、可操作的精度验证体系,进一步提高人口统计数据空间化结果的精确性与实用性。

此外,在人口统计数据空间化的基础上,我们还可以将研究的重点放在其他相关人文因子的空间化中。各项人文因子的空间化研究对于解释各种社会经济现象以及制定各项相关政策都具有十分重要的现实意义和指导意义。

## 参考文献

- [1] 胡焕庸. 论中国人口之分布. 上海: 华东师范大学出版社, 1983
- [2] 王雪梅, 李新, 马明国. 基于遥感和 GIS 的人口数据空间化研究进展及案例分析. 遥感技术与应用, 2004, 19(5): 320-327
- [3] 林丽洁, 林广发, 颜小霞, 等. 人口统计数据空间化模型综述. 亚热带资源与环境学报, 2010, 5(4): 10-16
- [4] 廖顺宝, 孙九林. 基于 GIS 的青藏高原人口统计数据空间化. 地理学报, 2003, 58(1): 25-33
- [5] 葛美玲, 封志明. 基于 GIS 的中国 2000 年人口之分布格局研究: 兼与胡焕庸 1935 年之研究对比. 人口研究, 2008, 32(1): 51-57
- [6] Langford M, Harvey J T. The use of remotely sensed data for spatial disaggregation of published census population counts // IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas. Rome, 2001: 260-264
- [7] Kraus S P, Ryerson J M, Senger L W. Estimating population from photographically determined residential land use types. Remote Sensing of Environment, 1974, 3(1): 35-42
- [8] 廖一兰, 王劲峰, 孟斌, 等. 人口统计数据空间化的一种方法. 地理学报, 2007, 62(10): 1110-1119
- [9] 王春菊, 汤小华. 基于 GIS 的福建省人口统计数据

- 空间化. 地理与地理信息科学, 2004, 20(4): 71-74
- [10] 刘建军, 李春来. 基于遥感和 GIS 的巢湖流域人口信息提取. 科学通报, 2002, 23(47): 1835-1837
- [11] 闫庆武, 卞正富, 张萍, 等. 基于居民点密度的人口密度空间化. 地理与地理信息科学, 2011, 27(5): 95-98
- [12] 康敏捷, 曹可, 王辉, 等. 基于作息空间思想的人口统计数据空间化方法. 地理与地理信息科学, 2012, 28(4): 60-62
- [13] Elvidge C D, Imhoff M L, Baugh K E, et al. Night-time lights of the world: 1994-1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2001, 56(2): 81-99
- [14] 何春阳, 史培军, 李井刚, 等. 基于 DMSP/OLS 夜间灯光数据和统计数据的中国大陆 20 世纪 90 年代城市化空间过程重建研究. 科学通报, 2006, 51(7): 856-861
- [15] Baugh K, Elvidge C D, Ghosh T, et al. Development of a 2009 stable lights product using DMSP-OLS data. *Proceedings of the Asia-Pacific Advanced Network*, 2010, 30: 114-130
- [16] Elvidge C D, Baugh K E, Dietz J B, et al. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing of Environment*, 1999, 68(1): 77-88
- [17] 曹丽琴, 李平湘, 张良培. 基于 DMSP/OLS 夜间灯光数据的城市人口估算: 以湖北省各县市为例. 遥感信息, 2009(1): 83-87
- [18] 卓莉, 陈晋, 史培军, 等. 基于夜间灯光数据的中国人口密度模拟. 地理学报, 2005, 60(2): 266-276
- [19] Zeng C, Zhou Y, Wang S, et al. Population spatialization in China based on night-time imagery and land use data. *International Journal of Remote Sensing*, 2011, 32(24): 9599-9620
- [20] Roy Chowdhury P K, Maithani S, Dadhwal V K. Estimation of urban population in Indo-Gangetic Plains using night-time OLS data. *International journal of remote sensing*, 2012, 33(8): 2498-2515
- [21] 梁友嘉, 徐中民. 基于 LUCC 和夜间灯光辐射数据的张掖市甘州区人口空间分布建模. 冰川冻土, 2012, 34(4): 999-1006
- [22] Dobson J E, Bright E A, Coleman P R, et al. LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering & Remote Sensing*, 2000, 7: 849-857
- [23] Tian Y, Yue T, Zhu L, et al. Modeling population density using land cover data. *Ecological Modelling*, 2005, 189: 72-88
- [24] Lo C P. Modeling the population of China using DMSP operational line scan system nighttime data. *Photogrammetric Engineering & Remote Sensing*, 2001, 67(9): 1037-1048
- [25] 许永峰. 指数标度中重要度的研究. 纺织高校基础科学学报, 2003, 16(2): 138-140
- [26] 陈晴, 侯西勇, 吴莉. 基于土地利用数据和夜间灯光数据的人口空间化模型对比分析: 以黄河三角洲高效生态经济区为例. 人文地理, 2009, 29(5): 94-100
- [27] 曾丽波, 张新武, 肖慧娟. 人口数据空间化方法研究. 林业调查规划, 2014, 39(5): 29-31
- [28] Sutton P. A scale-adjusted measure of "urban sprawl" using nighttime satellite imagery. *Remote Sensing of Environment*, 2003, 86(3): 353-369
- [29] Zhuo L, Ichinose T, Zheng J, et al. Modelling the population density of China at the pixel level based on DMSP/OLS non-radiance-calibrated night-time light images. *International Journal of Remote Sensing*, 2009, 30(4): 1003-1018
- [30] 符海月, 李满春, 赵军, 等. 人口数据格网化模型研究进展综述. 人文地理, 2006, 21(3): 115-119
- [31] 童玉芬. 国外人口与环境关系研究的理论与方法综述. 中国人口·资源与环境, 2004, 14(5): 115-119