

基于 DMSP-OLS 和 Landsat8 数据的省域 GDP 空间化方法^{*}

——以北京市为例

李峰¹, 米晓楠², 王粉鸽¹, 刘军¹

(1. 防灾科技学院, 河北 三河 065201; 2. 山西省气候中心, 太原 030002)

摘要:随着社会经济数据的省域空间化需求的增加和夜间灯光数据应用的日渐成熟,基于北京市的 DMSP-OLS 夜间灯光数据,根据相关分析和回归分析的方法,建立综合灯光指数与第二、第三产业 GDP 的回归模型;利用北京市的 Landsat8 数据,依据 CART 决策树算法分类出高精度的土地利用类型图,构建耕地、草地、林地和水域面积与第一产业 GDP 的回归模型;通过对模拟的第一、二、三产业 GDP 共同求和的方式计算 GDP 总值,最终建立北京市的 1 km 格网 GDP 空间分布图。结果表明,模拟的第一产业 GDP 的平均残差和相对误差分别为 0.04 亿元和 3.55%,第二、三产业 GDP 的平均残差和相对误差分别为 -5.54 亿元和 3.35%,GDP 总值的平均残差和相对误差分别为 -6.43 亿元和 3.36%。生成的 GDP 密度图能较全面地反映北京市的社会经济分布特征,可为经济决策和 GDP 产值估算提供依据。

关键词:夜间灯光;Landsat 8;省域 GDP;空间化

中图分类号:TP79

文献标志码:A

文章编号:1672-6693(2016)01-0138-05

GDP 是反映社会经济发展、区域规划以及资源环境配置的重要指标之一。与传统的 GDP 统计数据相比,GDP 空间化既反映统计区域内部的 GDP 差异和空间分布特征,又使得附有空间信息的 GDP 公里格网密度值可用于空间统计计算,且不受行政边界的限制,有利于长期的可持续化研究^[1]。基于大尺度遥感影像的研究发现,夜间灯光数据与经济活动存在着密切的相关性。自第一幅全球尺度 GDP 空间分布图开发以来,DMSP-OLS (Defense meteorological satellite program's operational linescan system,美国国防气象卫星计划线性扫描业务系统)数据就被用来估计全球各国的 GDP 空间分布情况。Elvidge 等人^[2]利用 DMSP-OLS 数据分析了美国、巴西等 21 国的灯光区域面积与 GDP、电力消耗间的关系,建立的对数模型的相关系数均高于 0.85。这些研究都是基于只有灯光区存在经济活动的前提,而忽略了非灯光区的农林牧渔业对经济的贡献率^[3-5]。因此,在估算经济活动总量时,考虑进农业等活动和人口因素有助于提高 GDP 空间化的精度。刘红辉等人^[6]仅利用 1 : 10 万土地利用图,通过一、二、三产业 GDP 与土地利用类型的关联性,实现了 1 km 格网社会经济数据的空间定量模拟。韩向娣等人^[7]用土地利用图(土地利用数据)与全国第一产业数据建模,结合 DMSP-OLS 和土地利用数据对全国第二、三产业建模,生成的 GDP 密度图比单一数据的精度要高。基于国家层面的 GDP 空间化局限于行政区划的均值统计形式,简化了 GDP 的影响因素,但无法反映省域内部 GDP 分布的差异性。为了满足省级层面的经济决策和产值估算的需求,梁友嘉和徐中民^[8]建立了乡镇单元 GDP 与夜间灯光 DN 值间的模型,结合人口空间分布数据和三次产业比例,生成 GDP 的空间分布图。杨妮等人^[9]分析了第一、二、三产业 GDP 和 GDP 总量与夜间灯光指数的关系,构建广西省域 GDP 空间化模型,可以反映省级 GDP 宏观分布的状况。为了进一步提高省域大尺度 GDP 空间化的精度,更加真实反映区域的经济活动总量,本研究以北京市为例,采用 DMSP-OLS 夜间灯光和 landsat8 数据探索 GDP 空间化的新方法,为经济决策和分析提供可靠的定量依据。

1 研究数据

采用原始分辨率约为 900 m 的 2012 年北京市 DMSP-OLS 夜间平均稳定灯光数据,将之重采样为 1 km 分

* 收稿日期:2015-02-15 修回日期:2015-03-24 网络出版时间:2015-12-02 13:26

资助项目:河北省高等学校科学研究计划重点项目(No. ZD2014203);中央高校基本科研业务费资助项目(No. ZD2014203)

作者简介:李峰,男,讲师,工程师,博士,研究方向为遥感与地理信息系统的应用,E-mail:lif1223@aliyun.com

网络出版地址:<http://www.cnki.net/kcms/detail/50.1165.n.20151202.1326.018.html>

辨率,如图 1 所示,主要用于第二、三产业 GDP(GDP23)建模。该数据清除了云层的影响,包含城市、乡镇以及永久性光源地,排除了偶然灯光噪声,背景噪声已用 0 值替代,像元灰度值在 1~63 之间。因 2012 年 Landsat 5 传感器失效,选用 2013 年北京市 Landsat8 30 m 分辨率的多光谱数据划分土地覆被类型。由于 2012 年第一产业 GDP(GDP1)与 2013 年的相比变化较小,因此,可以通过 2013 年北京市土地利用类型图建立第一产业 GDP 模型。为保证投影面积的变形最小,采用 Albers 投影。北京市分区县 GDP 统计数据取自 2012 年北京市统计年鉴。

2 省域 GDP 空间化建模方法

本研究以北京市各区县行政边界为基础,以所辖 16 个区县为统计单元,GDP 空间化建模的流程如图 2 所示。为了反映全市经济和城市化发展水平,利用区域夜间灯光分布的两方面属性:区域平均灯光强度(I)和区域灯光面积(S)来构建综合灯光指数(L)和平均灯光指数(V_{CNLI})。

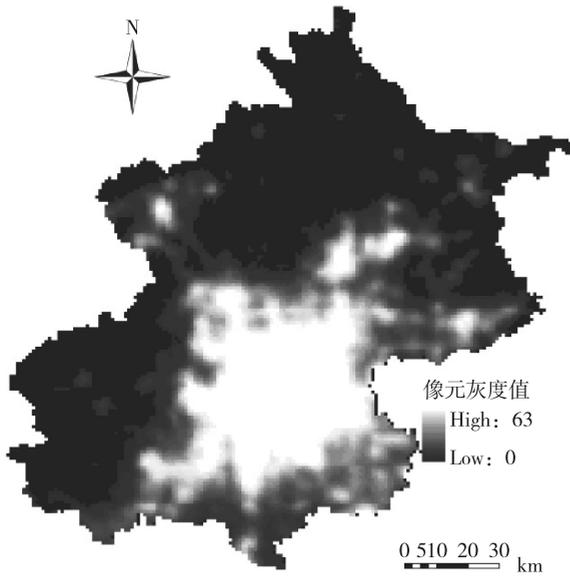


图 1 2012 年北京市 DMSP-OLS 夜间灯光数据

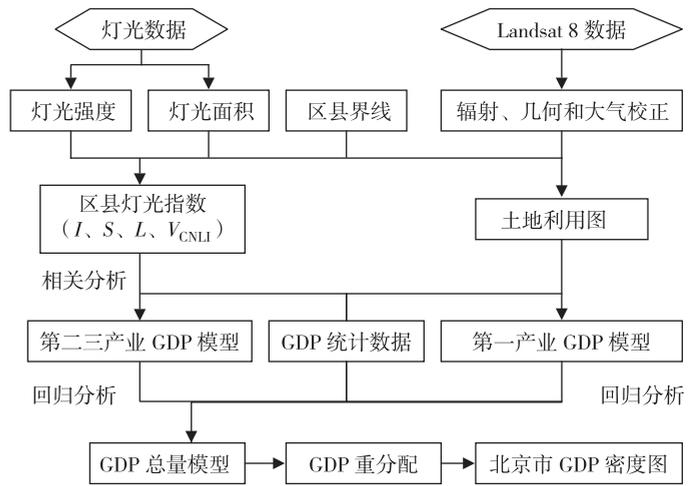


图 2 省域 GDP 空间化流程

2.1 第二、三产业 GDP(GDP23)建模方法

北京市作为东部发达地区,它的经济发展水平是产业结构、地域空间结构和人口分布等因素的综合作用的结果,作为第二、三产业的工业、建筑业和服务业主要分布在城市区域,单位产值比重很大;而 DMSP-OLS 夜间灯光强度正是第二、三产业作用结果的集中体现,很难区分各分产业产值,因而选用灯光平均强度来定量评价 GDP23 是比较恰当的^[10-11]。目前常用灯光指数 I 、 S 、 L 、 V_{CNLI} 可分别被表示为

$$I = \frac{\sum_{i=t}^{63} (V_{DN_i} \times n_i)}{(63 \times N_L)}, \tag{1}$$

$$S = A_N / A, \tag{2}$$

$$L = I \times p_1 + S \times p_2, \tag{3}$$

$$V_{CNLI} = I \times S. \tag{4}$$

式中 p_1 和 p_2 为对应系数,可通过多元非线性回归法来确定; V_{DN_i} 和 n_i 分别表示行政单元内第 i 级灰度像元值和像元数; t 为去除灯光溢出的阈值,当灯光范围与城区建筑用地重合时取值^[12]; N_L 和 A_N 分别代表行政单元内位于 $[t, 63]$ 区间的像元总数和所占据的面积, A 表示行政单元面积。

将以上 4 种灯光指数分别与 GDP23 值(V_{GDP23})建立散点图并分析其中相关性,可得如下式所示的相关关系和相关系数:

$$V_{GDP23} = 29.231 \times \exp(4.213 \times I) (R^2 = 0.649), \tag{5}$$

$$V_{GDP23} = 67.803 \times \exp(2.935 \times S) (R^2 = 0.690), \tag{6}$$

$$V_{GDP23} = 1488.5 \times V_{CNLI}^{1.0596} (R^2 = 0.732), \tag{7}$$

$$V_{GDP23} = 27.771 \times L^{2.415} = 27.771 \times (3.161 \times I + 2.168 \times S)^{2.415} (R^2 = 0.750)。 \quad (8)$$

比较(5)~(8)式的相关系数知,由 L 计算的 V_{GDP23} 的相关系数最大,从而确定(8)式为 V_{GDP23} 模拟的回归模型。

2.2 第一产业 GDP(GDP1)建模方法

在城市,GDP 产值主要由第二、三产业构成,第一产业产值比例低;在农村,经济发展水平不高,灯光数据很难探测第一产业产值,因此,第一产业产值与灯光数据关系不明显。第一产业增加值和总产值均与农林牧渔业的土地面积呈现显著相关关系^[13],建立的行政单元土地利用类型与 GDP1 值(V_{GDP1})关系的模型为

$$V_{GDP1} = \sum_{j=1}^M m_j A_j + B。 \quad (9)$$

式中, m_j 为第 j 种与第一产业有关的土地利用类型分布系数, A_j 为行政单元内耕地、林地、草地、水域所占面积, M 为土地利用类型数目。考虑到“无土地利用无 GDP”的现实,截距 B 设置为 0。

采用 CART (Classification and regression tree,分类回归树)决策树算法分类 Landsat8 影像,它的基本原理是通过通过对由测试变量和目标变量构成的训练数据集循环分析形成二叉树形式的决策树结构,采用经济学中的基尼系数在每个分割处自动选择最合适的预测变量和分割阈值准则,提取出反应类别的重要特征。与其他决策树相比,CART 算法简单、结构清晰、运行速度快,能最大限度地降低各个步骤的不纯洁度,有效处理大量和高维数据^[14]。

对本研究中的土地覆被分类,测试变量采用 Landsat8 影像中提取的分类特征,包括影像中的 1~7 个多光谱波段、NDVI 值、MNDWI 值、北京市 30 m 分辨率 DEM 数据,组成一个 10 波段数据进行分类。由于草地与耕地的光谱差异极小,且北京市的草地比重仅占到总土地面积的 0.13%,所以可以将草地归入耕地类别中共同分类。选择林地、耕地(含草地)、水域、建设用地和其他(裸地和阴影)共 5 种土地类型构建 6 092 个训练样本,通过学习得到 56 个叶结点的决策树结构。根据高分辨率校正的 Google Earth 影像随机提取的 276 个检验样本建立混淆矩阵,CART 算法的总体分类精度高达 95.65%,kappa 系数为 0.94。

北京市的林地占全市总面积的 53.31%,但是并非所有的林地都能为 GDP1 贡献产值,本文通过 GIS 的分析功能选取坡向在 $90^\circ \sim 270^\circ$,坡度小于 20° 的经济林地来分析与 GDP1 的关系。根据各区县的土地利用类型面积和 GDP1,利用公式(9)建立土地利用类型与 GDP1 值(V_{GDP1})的线性回归模型为

$$V_{GDP1} = 0.0383 \times A_c + 0.1330 \times A_w + 0.0004 \times A_f (R^2 = 0.75)。 \quad (10)$$

式中, A_c 、 A_w 和 A_f 分别为耕地(含草地)、水域和经济林地的模型回归系数。

2.3 GDP 统计数据空间化

一般认为 1 km 左右的格网大小符合 GDP 空间化过程中 GDP 渐变空间分布的规律,过小的格网会产生极少的 GDP,而过大的格网造成 GDP 统计的精度降低。因此,本研究采用 1 km 格网的单元来空间化 GDP。GDP23 的回归模型可直接将区县级 GDP 数据分配到每个 1 km 格网中。而对于 GDP1,需要将 1 km 格网的 DMSP-OLS 数据分别对应到 30 m 分辨率 Landsat8 影像生成的土地利用类型图上,统计在 1 km 范围内的各土地利用类型和 GDP1 值。

利用区县 GDP1 和 GDP23 统计数据作线性调整来纠正每个格网的 GDP1 和 GDP23 的值,分别制作出 1 km 北京市 2012 年 GDP1 和 GDP23 密度图。具体的 GDP 纠正模型如下:

$$V_{GDP_T} = V_{GDP_j} \times \left(\frac{V_{GDP_i}}{V_{GDP_{all}}} \right)。 \quad (11)$$

式中, V_{GDP_T} 为纠正后的 GDP 值, V_{GDP_j} 为模拟的每个格网的 GDP 密度, V_{GDP_i} 为该区统计数据中的 GDP, $V_{GDP_{all}}$ 为该区预测的 GDP 值。

最后,将纠正后的 GDP1 和 GDP23 的值相加即可得北京市各区县的 GDP 值。生成的 2012 年北京市各区县的 GDP 密度图如封三彩图 3 所示,该图显示的 GDP 分布符合北京市对城市功能核心区、拓展区、发展新区和生态涵养发展区的规划,GDP 密度从市中心向周边辐射过程中呈现逐渐降低的趋势。

3 GDP 空间化结果分析

以模拟的 GDP1、GDP23 和 GDP 计算得出 2012 年北京市 16 个区县的 GDP1、GDP23 和 GDP 总值,然后与年鉴统计的 GDP1、GDP23 和 GDP 总值对比,可得 2012 年北京市分区县各产业 GDP 的残差和相对误差,如表 1 所示。在 GDP1 的误差中,平均残差只有 0.04 亿元,平均相对误差为 3.55%。因东城区、西城区和石景山区作

为北京市的中心城区,第二、三产业十分发达,不存在第一产业,因此三者的 GDP1 实际值为 0,导致无法计算出它们的 GDP1 相对误差,但三者的残差均很小,分别为 0.01、0.01 和 0.03 亿元。此外,朝阳区的残差最大,达到 0.39 亿元,并且其相对误差也是最大的,高达 24.76%,这主要因为在分类土地利用类型时将朝阳区的大部分公园草地当作耕地对待而造成。如果不考虑朝阳区的误差,其余区县的平均相对误差仅为 1.78%。

对于 GDP23 的误差而言,平均残差为 -5.54 亿元,平均相对误差仅为 3.35%。东城区和西城区的残差虽然分别高达 -77.93 和 -103.55 亿元,但是相对于其实际产值而言很小,所以二者的相对误差分别仅为 5.37% 和 3.99%,类似的情况也发生在朝阳区和海淀区。最大的 GDP23 相对误差为门头沟区和昌平区,分别高达 16.33% 和 15.37%,这是因为二者作为北京市的郊区,在靠近中心城区部分的灯光溢出和过饱和现象造成了对 GDP23 产值的高估,高估值分别为 18.79 和 76.39 亿元。由于第一产业产值在北京市的 GDP 总产值中所占的比重很小,GDP 总产值主要受第二、三产业产值的影响,因此,GDP 的残差和相对误差的分布情况也类似于 GDP23 的。其中,门头沟区和昌平区的 GDP 相对误差最大,分别为 15.59% 和 15.87%,平均相对误差为 3.36%,平均残差为 -6.43 亿元。

表 1 2012 年北京市 16 区县 GDP 结果模拟精度

| 区县 | 残差/亿元 | | | 相对误差/% | | |
|------|-------|---------|---------|--------|-------|-------|
| | GDP1 | GDP23 | GDP | GDP1 | GDP23 | GDP |
| 东城区 | 0.01 | -77.93 | -77.93 | - | 5.37 | 5.37 |
| 西城区 | 0.01 | -103.55 | -112.32 | - | 3.99 | 4.33 |
| 石景山区 | 0.03 | 2.25 | 2.55 | - | 0.67 | 0.75 |
| 朝阳区 | 0.39 | 57.76 | 57.83 | 24.76 | 1.59 | 1.59 |
| 丰台区 | -0.06 | 16.63 | 16.19 | 5.00 | 1.80 | 1.75 |
| 海淀区 | 0.15 | -50.57 | -52.30 | 7.54 | 1.44 | 1.49 |
| 房山区 | 0.15 | -4.25 | -5.35 | 0.95 | 0.98 | 1.19 |
| 通州区 | -0.17 | -1.20 | -2.15 | 0.88 | 0.28 | 0.48 |
| 顺义区 | 0.15 | 2.29 | 2.79 | 0.60 | 0.21 | 0.25 |
| 昌平区 | -0.18 | 76.39 | 80.36 | 2.00 | 15.37 | 15.87 |
| 大兴区 | 0.26 | -29.28 | -29.46 | 1.29 | 2.44 | 2.42 |
| 门头沟区 | -0.02 | 18.79 | 18.25 | 1.10 | 16.33 | 15.59 |
| 怀柔区 | -0.08 | 2.70 | 0.98 | 1.01 | 1.55 | 0.54 |
| 平谷区 | 0.06 | 0.35 | -0.50 | 0.35 | 0.26 | 0.33 |
| 密云县 | -0.11 | 1.42 | -0.37 | 0.61 | 0.89 | 0.21 |
| 延庆县 | -0.01 | -0.37 | -1.40 | 0.08 | 0.50 | 1.68 |
| 平均值 | 0.04 | -5.54 | -6.43 | 3.55 | 3.35 | 3.36 |

(0.04 亿元)和平均相对误差(3.55%),发现如果将城区的公园草地从耕地中分离,GDP1 模拟值的精度可以得到大幅提高。利用 DMSP-OLS 夜间灯光数据构建与第二、三产业产值的回归模型,分析模拟的 GDP23 残差(-5.54 亿元)和相对误差(3.35%)较小,GDP23 模拟值的精度较高。北京市作为经济发达地区,它的第一产业产值占总产值的比重很小,而第二、三产业的产值主要影响着 GDP 总值的大小。模拟的 GDP 平均残差和平均相对误差仅分别为 -6.43 亿元和 3.36%,GDP 总产值的模拟结果精度较高。生成的北京市 GDP 密度图的分布也符合全市 GDP 的实际分布状况。综上所述,利用 DMSP-OLS 夜间灯光数据和 Landsat 数据分产业进行 GDP 空间化的方法模拟的 GDP 的精度较高,适合预测北京市 GDP 的分布情况;但是城市中心灯光溢出和过饱和现象会造成少数区县 GDP 产值被高估的现象,从而限制了 GDP23 和 GDP 产值的模拟精度,因此如何尽量减少城市中心灯光溢出和过饱和现象来提高 GDP 产值的模拟精度有待于深入研究。

参考文献:

- [1] 韩向娣,周艺,王世新,等. 夜间灯光遥感数据的 GDP 空间化处理方法[J]. 地球信息科学学报,2012,14(1):128-136.
Han X D, Zhou Y, Wang S X, et al. GDP spatialization in China based on nighttime imagery[J]. Journal of Geo-information Science, 2012, 14(1): 128-136.
- [2] Elvidge C D, Baugh K E, Kihn E A, et al. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption[J]. International Journal of Remote Sensing, 1997, 18(6): 1373-1379.
- [3] Sutton P C. An empirical environmental sustainability index derived solely from nighttime satellite imagery and ecosystem service valuation[J]. Population and Environment, 2003, 24(4): 293-311.
- [4] Sutton P C, Costanza R. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation[J]. Ecologi-

- cal Economics, 2002, 41(3): 509-527.
- [5] Doll C N H, Muller J P, Morley J G. Mapping regional economic activity from nighttime light satellite imagery[J]. Ecological Economics, 2006, 57(1): 75-92.
- [6] 刘红辉, 江东, 杨小唤, 等. 基于遥感的全国 GDP 1 km 格网的空间化表达[J]. 地球信息科学, 2005, 7(2): 120-123.
Liu H H, Jiang D, Yang X H, et al. Spatialization approach to 1km grid GDP supported by remote sensing[J]. Geo-Information Science, 2005, 7(2): 120-123.
- [7] 韩向娣, 周艺, 王世新, 等. 基于夜间灯光和土地利用数据的 GDP 空间化[J]. 遥感技术与应用, 2012(3): 74-83.
Han X D, Zhou Y, Wang S X, et al. GDP spatialization in China based on DMSP/OLS data and land use data[J]. Remote Sensing Technology and Application, 2012(3): 74-83.
- [8] 梁友嘉, 徐中民. 基于夜间灯光辐射数据的张掖市甘州区 GDP 空间分布建模[J]. 冰川冻土, 2013, 35(1): 249-254.
Liang Y J, Xu Z M. Modeling the spatial distribution of GDP based on night light radiation: a case study in Ganzhou district, Zhangye municipality[J]. Journal of Glaciology and Geocryology, 2013, 35(1): 249-254.
- [9] 杨妮, 吴良林, 邓树林, 等. 基于 DMSP-OLS 夜间灯光数据的省域 GDP 统计数据空间化方法[J]. 地理与地理信息科学, 2014, 30(4): 108-111.
Yang N, Wu L L, Deng S L, et al. Spatialization method of provincial statistical GDP data based on DMSP/OLS night lighting data: a case study of Guangxi Zhuang autonomous region[J]. Geography and Geo-information Science, 2013, 35(1): 249-254.
- [10] 陈晋, 卓莉, 史培军, 等. 基于 DMSP-OLS 数据的中国城市化过程研究[J]. 遥感学报, 2003, 7(3): 168-176.
Chen J, Zhuo L, Shi P J, et al. The study on urbanization process in China based on DMSP/OLS data: development of a light index for urbanization level estimation[J]. Journal of Remote Sensing, 2003, 7(3): 168-176.
- [11] 卓莉, 史培军, 陈晋, 等. 20 世纪 90 年代中国城市时空变化特征-基于灯光指数 CNLI 方法的探讨[J]. 地理学报, 2003, 58(6): 893-902.
Zhuo L, Shi P J, Chen J, et al. Application of compound night light index derived from DMSP/OLS data to urbanization analysis in China in the 1990s[J]. Acta Geographica Sinica, 2003, 58(6): 893-902.
- [12] 杨眉, 王世新, 周艺, 等. 基于 DMSP-OLS 影像的城市化水平遥感估算方法[J]. 遥感信息, 2011(4): 100-106.
Yang M, Wang S X, Zhou Y, et al. A method of urbanization level estimation using DMSP/OLS imagery[J]. Remote Sensing Information, 2011(4): 100-106.
- [13] 易玲, 熊利亚, 杨小唤. 基于 GIS 技术的 GDP 空间化处理方法[J]. 甘肃科学学报, 2006, 18(2): 54-58.
Yi L, Xiong L Y, Yang X H. Method of pixelzing GDP data based on the GIS[J]. Journal of Gansu Sciences, 2006, 18(2): 54-58.
- [14] 陈云, 戴锦芳, 李俊杰. 基于影像多种特征的 CART 决策树分类方法及其应用[J]. 地理与地理信息科学, 2008, 24(2): 33-36.
Chen Y, Dai J F, Li J J. CART based decision tree classifier using multi-feature of image and its application[J]. Geography and Geo-information Science, 2008, 24(2): 33-36.

Provincial GDP Spatialization Technique Based on DMSP-OLS and Landsat 8 Images of Beijing

LI Feng¹, MI Xiaonan², WANG Fenge¹, LIU Jun¹

(1. Institute of Disaster Prevention, Sanhe Hebei 065201; 2. Shanxi Climate Center, Taiyuan 030002, China)

Abstract: With the incensement of requirement of provincial socio-economic data spatialization and the mature of nighttime light images' applications, a regression model between comprehensive light index and the secondary & tertiary industry was constructed based on correlation and regression analysis with DMSP-OLS nighttime light data of Beijing. And then, another regression model between areas of farmland, grassland, woodland and water body and the primary industry was established according to precise land-use map produced by CART decision-tree algorithm using Landsat 8 images. The total GDP were calculated throughout the sum of the simulated primary, secondary & tertiary industry so that a GDP spatialization map with the resolution of 1 km² was generated. The results show that average residual errors and relative errors of the primary, secondary & tertiary industry and total GDP were separately 0.004 billion Yuan, 3.55%, -0.554 billion Yuan and 3.35%, -0.643 billion Yuan and 3.36%. The created GDP density map would be used to reflect roundly the socio-economic distribution characteristics and provide some suggestions of economic decision-making and GDP output estimation for Beijing.

Key words: nighttime light; Landsat 8; provincial GDP; spatialization

(责任编辑 方 兴)