

文章编号:1671-4229(2023)05-0010-10

# 基于多源地理大数据的我国 GDP 空间建模研究

陈漾漾, 曹泳茵, 徐勇\*

(广州大学 地理科学与遥感学院, 广东 广州 510006)

**摘要:** 国内生产总值(GDP)是衡量一个国家经济状况和发展水平的重要指标,精确核算年度 GDP 对于一个国家制定发展策略非常重要。然而不同研究用于 GDP 估算的地理数据均有所差别,且哪种地理数据更有利于核算 GDP 是尚未明确的问题。文章以我国 2 848 个区县为研究区,采用多元线性回归(OLS)方法及地理加权回归(GWR)方法,结合夜间灯光遥感数据、兴趣点 POI 数据、腾讯位置大数据和城镇建设用地面积等多源地理数据模拟 2020 年区县 GDP 规模,并探讨不同变量的模拟效果。研究发现:①综合运用多源地理数据并结合空间加权回归方法可在区县级和地级市级分别实现 74% 和 87% GDP 空间模拟精度;②对比不同地理因子,发现 POI 数据最能有效反映区县级 GDP 总量,效果优于其他地理因子;③相比于 POI 因子,腾讯位置大数据能较好地反映西部地区 GDP 总量。因此,结合腾讯位置及兴趣点 POI 数据,可提高全国 GDP 建模精度。文章为快速、准确模拟我国 GDP 空间分布提供了重要数据和方法参考。

**关键词:** GDP; 夜间灯光; POI 数据; 腾讯位置大数据; 土地利用数据

中图分类号: K909

文献标志码: A

## GDP mapping in China with multi-sourced geographical open data

CHEN Yang-yang, CAO Yong-yin, XU Yong\*

(School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China)

**Abstract:** GDP is one of the most important indices in measuring the socio-economic development status of a country, and thus, an accurate estimation of GDP is vital in formulating valid development strategy for a country. However, the geographical data used for GDP simulation in various studies are different, and it is not clear which geographical data is more conducive to GDP simulation. In this study, both ordinary least squares regression (OLS) and geographical weighted regression (GWR) methods were conducted in order to simulate the GDP of 2 848 counties in China, in which, the performance and modeling capabilities of multi-sourced open data, including nighttime satellite data, points of interest (POI) data, Tencent's social user location data and built-up urban area data, were explored and assessed. The experimental results showed the following findings: ① The overall accuracies of the simulated GDP at counties and cities in China could be achieved above 74% and 87%, respectively, when multi-sourced open data, including nightlight satellite data, POI data and Tencent's user data were utilized by GWR method. ② By comparing different indicators, it is found that the point of interest (POI) performs better than the other indicators in modelling the actual GDP at the county level in China. ③ The results also indicated that Tencent's social user location data has high

收稿日期: 2023-06-01; 修回日期: 2023-08-30

基金项目: 国家自然科学基金资助项目(42271477)

作者简介: 陈漾漾(1999—),女,硕士研究生. E-mail:2112101047@e.gzhu.edu.cn

\*通信作者. E-mail:xu1129@gzhu.edu.cn

引文格式: 陈漾漾, 曹泳茵, 徐勇. 基于多源地理大数据的我国 GDP 空间建模研究[J]. 广州大学学报(自然科学版), 2023, 22(5): 10-19.

potential in modeling the GDP distribution in the western part of China, which indicated that the modeling accuracy of GDP in China could be further improved using both Tencent's social user data and POI data. The results and findings in this study could provide insights in understanding how to better simulate and model GDP distribution in China.

**Key words:** GDP; nighttime light; POI data; Tencent's social user location data; land use data

国内生产总值(Gross Domestic Product, GDP)是一个国家所有常住单位在一定时期内生产活动的最终成果,是国民经济核算的核心指标,也是衡量经济状况和发展水平的重要指标。目前政府核算年度 GDP 需要收集各行业和劳动者的各项经济指标,即基于行业生产过程中的增加值或报酬收入两种经济指标分别计算,最终才能确定 GDP 数值,这必然会导致数据公布在时间上出现滞后,尤其是区县级行政单元的年度 GDP 存在缺失值。快速准确预测大范围行政区的 GDP 能为政府决策提供参考,同时也是亟需解决的问题<sup>[1]</sup>。

夜间灯光遥感影像能够监测大范围的地表人造光,Elvidge 等<sup>[2]</sup>首次明确了地面人类活动造成的夜间灯光和 GDP 具有一定的线性关系,而后该遥感数据也被验证出与 GDP 具有较强的相关性,是经济相关活动的潜在指标<sup>[3]</sup>。并且该数据获取成本低,已被广泛应用于经济估算<sup>[4-7]</sup>、贫困度估算<sup>[8]</sup>、城市发展等级排名<sup>[9]</sup>和疫情复工复产<sup>[10]</sup>等社会经济领域的研究。与 DMSO OLS 夜间灯光数据相比,新一代的 NPP VIIRS 夜间灯光数据弥补了前者空间分辨率较低和灯光饱和等缺陷,对于 GDP 的拟合效果更佳<sup>[11-12]</sup>。

虽然夜间灯光对于截面 GDP 有着较好的指示性,但是灯光也会受到短暂的人类活动、自然现象(山火、云、雪、气溶胶等)和下垫面的影响<sup>[13]</sup>,有时并不能很好地反映经济发展状况。随着无线网络和便携式移动设备的普及,人类社交媒体位置数据也成为拟合人口数量<sup>[14-17]</sup>和经济状况<sup>[18-19]</sup>的一个指标。Ma<sup>[20]</sup>基于夜间灯光数据和腾讯位置大数据的关系,发现在同样光照强度不同地区的经济活动也会存在较大差异。Zhao 等<sup>[21]</sup>发现相较于稳定灯光产品,推特软件位置数据没有高估郊区和低估城市核心区的社会经济状况,可以将其视为夜间灯光的替代品来评估社会经济因素。Huang 等<sup>[18]</sup>发现腾讯用户密度数据在区县尺度 GDP 模拟中比 NPP VIIRS 夜间灯光数据更加强和可靠。这些研究表明,人类社交媒体位置数

据能够弥补夜间灯光在拟合 GDP 方面的缺陷。

类似的还有兴趣点数据(Points of Interest, POI),它能在微观尺度上反映人类经济活动,可以描述人类对于土地的细粒度开发方式,因此,POI 已经被用于提取城市和社会系统的详细信息<sup>[22-23]</sup>。同时也有学者将其运用到 GDP 预测中,如 Chen 等<sup>[24]</sup>将 POI 与夜间灯光数据耦合构建随机森林模型分别制作三大产业 GDP 空间化地图;此外,有学者结合 POI 与其他数据,通过构建线性模型来绘制 GDP 空间分布图<sup>[25-27]</sup>。但是同种类型 POI 模拟 GDP 时所占的权重相同,存在经济产出相差较大的情况,这会最终影响拟合结果,仍需补充能反映地域经济活力性质的数据<sup>[24]</sup>。

除此之外,土地利用方式也是反映经济发展水平的重要因素,其在一定程度上代表了城市化进程中人类聚居地的发展规模。Huang 等<sup>[19]</sup>将城镇建设面积纳入到 GDP 建模中,发现其对 GDP 有显著的正向作用。也有学者根据土地利用数据或将其与夜间灯光数据结合分别对不同产业 GDP 构建数学模型进行 GDP 空间化<sup>[28-31]</sup>,表明了土地利用数据具有反映 GDP 的能力。

综合前人所得的结论,夜间灯光遥感数据、POI 数据和社交媒体位置信息三者拟合 GDP 时是互相弥补的关系。尽管也有研究同时将以上地理大数据结合进行 GDP 估算<sup>[32-33]</sup>,但是鲜有研究探讨这些数据拟合 GDP 的潜力和性能,对于 GDP 空间建模尚无最优因子选择方案<sup>[18]</sup>。且中国大陆幅员辽阔,各地自然人文条件差异较大,不同地理数据在不同地区的 GDP 反映能力目前尚未明确。

鉴于此,本研究将夜间灯光遥感数据、POI 数据、腾讯位置大数据和土地利用数据作为构建 GDP 建模模型的候选指标,把这些指标叠加在中国大陆的区县行政范围,分别运用普通最小二乘法和地理加权回归法模拟 2020 年中国大陆区县的 GDP。此外,研究还探讨上述 4 种地理大数据在模拟 GDP 方面的能力,为最佳 GDP 建模因子的选择和建模精度的提高提供参考。

## 1 研究区域和数据

### 1.1 研究区域

截止 2020 年底,全国共有 34 个省级行政单位,333 个地级行政单位,2 843 个县级行政单位。本研究选择中国大陆的区县(除香港、台湾以及

金门县和金沙市)作为研究区域,图 1 为我国 2020 年区县 GDP 空间分布图。其中蓝色表示低值,红色则表示高值。中国区县 GDP 自东向西递减,GDP 较高的区县集中在京津地区、山东半岛、华东沿海地区、华中、成渝城市群、粤港澳大湾区,而低值主要分布在东北北部、华北北部、青藏高原地区。

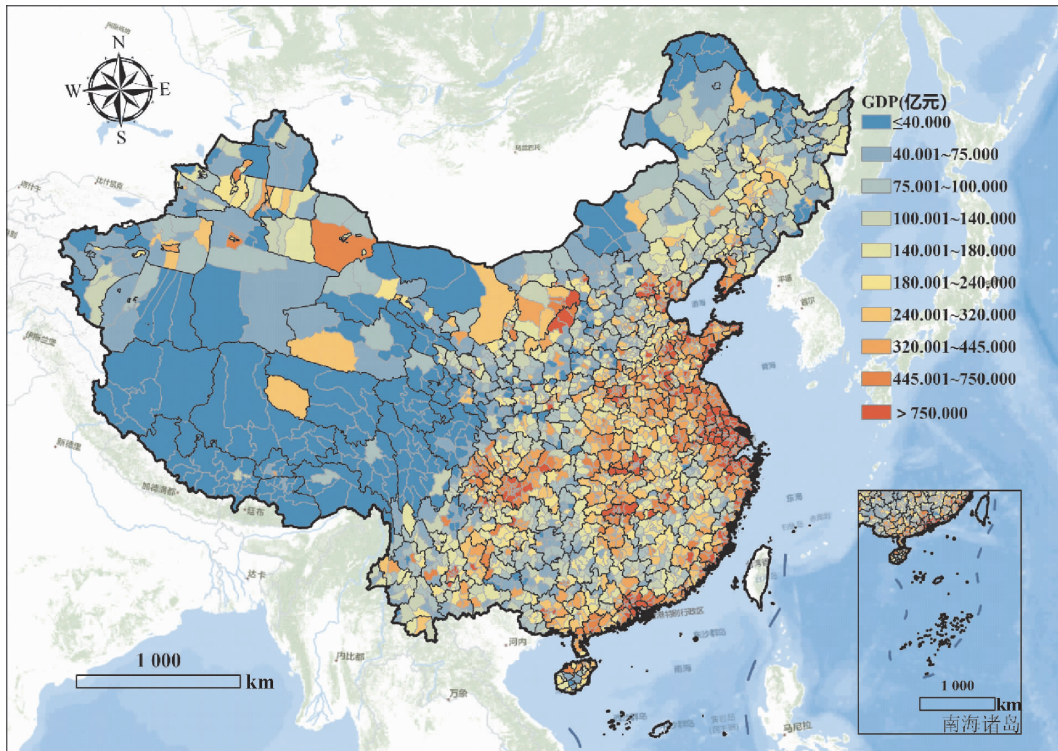


图 1 研究区

Fig. 1 Study area

注:基于自然资源标注地图服务网站 GS(2020)4619 号标准地图制作,底图边界无修改。

### 1.2 数据及预处理

#### 1.2.1 夜间灯光数据及预处理

本研究使用美国宇航局(National Aeronautics and Space Administration, NASA)基于 NPP-VIIRS 数据开发的新产品——黑色大理石(Black Marble) 2020 年年度合成数据 VNP46A4 (<https://landsweb.modaps.eosdis.nasa.gov/>) 进行分析。VNP46A4 产品的空间分辨率约为 500 m,已进行提前校正,且其分为无雪期和积雪期,为用户下载选择提供了更多的观察天数。Li 等<sup>[34]</sup>发现夜间灯光在城市不同地区和下垫面表现出不同的角度效应,因此,NASA 后续将黑色大理石月度复合产品补充为 3 个视角类别,即近星下点(天顶角 0~20 度)、侧视点(观测天顶角 40~60 度)和所有角度,本研究选

取观测天数最多的 AllAngle\_Composite\_Snow\_Free 数据作为灯光值自变量。

考虑到积雪期地面反射率增强以及积雪检测错误可能会造成亮度值波动<sup>[13]</sup>,使用 Python 遴选出无雪期的灯光亮度像元并完成影像拼接,随后重投影为 Albers\_Conic\_Equal\_Area 投影坐标系及重采样至 500 m。此外,VNP46A4 产品并未过滤火光、油气井等短暂光源干扰,为了减小影响,需要去除极大值。假设原始影像的一个地区亮度值大于同一时期的大城市最大亮度值,则将其识别成异常像素点进行校正。选取北京市、上海市、广州市、深圳市的最大亮度值作为参照阈值,当识别到大于该参照阈值的异常像素点时,将其重新赋予周围 8 个像元的中值,然后生成最终的年度夜

间灯光影像。

### 1.2.2 POI 数据

POI 数据来自于北京大学开放数据平台 (<https://doi.org/10.18170/DVN/WSXCNM>), 收集的时间截止到 2018 年 9 月 30 日, 覆盖全国。数据预处理包括清除不完整及重复的 POI 数据, 最终保留 34 955 806 份有效的记录, 而后将高德坐标转化为 WGS84 坐标再进行重投影。

POI 精细分类可以最大程度保证类似产值的 POI 聚合在一起, 因此, 参照《国民经济行业分类》

(GB/T4754—2017)<sup>[35]</sup> 对 POI 数据重新分类 (表 1), 并进行空间核密度分析。在核密度分析中, 不同搜索半径得到的结果不同。将搜索半径从 100 m 逐次增加 100 m 直至 1 000 m 以获取最佳核密度搜索半径, 发现当搜索半径在 400 m 及更大时, POI 与 GDP 的相关性基本稳定且基本涵盖所有 POI 小类别, 这也与前人的研究结论相似<sup>[36]</sup>。为了最大程度减小共线性的影响, 最终将 300 m 作为搜索半径, 以 100 m 为输出像元大小, 最后重采样至 500 m。

表 1 POI 分类表  
Table1 POI classification

POI 类别	原始 POI 类别	点个数	相关系数
工厂	产业园区、工厂、农林牧渔基地	386 728	0.690
医疗卫生	动物医疗场所、疾病预防机构、诊所、专科医院、综合医院等	599 715	0.724
政府及社会团体	政府机关、社会团体、工商税务机构等	1 943 963	0.833
科教文化	文化宫、传媒机构、图书馆、学校、科研机构等	2 022 706	0.790
房地产	房地产中介机构、商务住宅、住宅区、宿舍、别墅等	1 217 110	0.786
交通物流仓储	火车站、公交车站、机场、地铁站、物流速递、邮局等	1 357 292	0.790
公司	公司、企业、商务写字楼、事务所、人才市场、旅行社、搬家公司等	3 046 166	0.816
金融保险	银行、保险公司、自助取款机	1 038 732	0.839
居民生活服务	摩托车/汽车服务、维修站点、美容美发店、洗浴推拿场所、洗衣店、摄影冲印店等	4 314 874	0.735
零售业	摩托车/汽车销售、专卖店、便利店、加油站、医药保健销售店、服装鞋帽皮具店、花鸟鱼虫市场等	12 112 557	0.761
体育休闲	彩票彩券销售点、风景名胜、公园广场、影剧院、度假村、疗养院、棋牌室、网吧、健身中心等	1 365 006	0.849
餐饮住宿	餐厅、宾馆酒店、住宿服务相关	5 550 957	0.779

### 1.2.3 腾讯位置大数据

腾讯位置大数据展现的是某个点位及其四周一定范围内所有使用 QQ 和腾讯 (社交讯息软件)、京东和美团等应用程序的定位数量。由于 2020 年以后腾讯位置大数据的接口限制, 其空间分辨率约为 5 000 m, 无法满足研究需要。因此, 本研究选取 2018 年全年 0:00—24:00 的腾讯位置大数据<sup>[37]</sup> 作为建模因子, 其空间分辨率约为 1 000 m。对其进行核密度分析, 以 1 000 m 作为搜索半径, 输出像元大小为 500 m。

### 1.2.4 土地利用数据

本研究选取 2020 年 30 m 分辨率的中国土地利用遥感监测数据, 来源于中国科学院资源环境科学与数据中心数据平台 (<http://www.resdc.cn>)。该数据分为耕地、林地、草地、水域、城乡工矿居民用地和未利用土地 6 个一级分类, 25 个二级分类, 将其重采样为 500 m 分辨率, 统计城镇面积和工交建设用地面积总和作为城镇建设用地自变量。

cn)。该数据分为耕地、林地、草地、水域、城乡工矿居民用地和未利用土地 6 个一级分类, 25 个二级分类, 将其重采样为 500 m 分辨率, 统计城镇面积和工交建设用地面积总和作为城镇建设用地自变量。

### 1.2.5 社会经济统计数据

根据行政单元收集区县 GDP 数据, 获取了 2020 年 2 848 个 GDP 数据作为训练数据集。另外还收集了 2018—2019 年地级市和部分区县 GDP 作为验证数据集。

## 2 研究方法

相关性分析被用来检测 GDP 和一些潜在因子

的共变趋势,为选取敏感因子提供参考。以区县年度夜间灯光、腾讯位置大数据和 12 类区县 POI 核密度总值作为自变量,区县年度 GDP 为因变量进行相关性分析。结果表明,夜间灯光、腾讯位置大数据和城镇建设面积与 GDP 的相关性分别为 0.789、0.765、0.700。除了工厂 POI 外,其余 POI 类别核密度与 GDP 的相关性均大于 0.7(表 1)。在此基础上,通过向后筛选法选择较优 POI 类别,最终根据逐步回归法选取了夜间灯光、公司 POI、政府及社会团体 POI(以下简称“政府 POI”)、腾讯位置大数据和城镇建设用地面积作为最终模型的建模因子,这些建模因子的最大方差膨胀因子小于 7.5,表明自变量之间没有数据冗余。

### 2.1 普通最小二乘法回归

普通最小二乘法(Ordinary Least Square, OLS)是分析两个或多个变量关系的最常用的方法,采用最小二乘法量化 GDP 和地理因子的关系。为所有变量统一量纲,使用 Zcore 标准化到相同的范围。OLS 模型如下:

$$y_i = \beta_0 + \sum_{i=1}^n X_i \beta_i + \varepsilon_i, \quad (1)$$

其中, $y$  是因变量,表示第  $i$  个区县的 GDP; $\beta_0$  为模型的截距; $X_i$  对应模型的第  $i$  个区县的解释变量, $\beta_i$  是回归系数,它可以反映出每种变量对因变量的影响程度; $\varepsilon_i$  为随机误差项。

### 2.2 地理加权回归分析

OLS 是假定全局的参数是稳定的,其计算的系数在空间上没有显著差异,是解释变量在所有区县平均意义上的参数估计值。但是中国经济发展存在巨大的空间差异和不平衡性,这些差异会引起 OLS 精度和解释因子出现一定的扭曲,采用地理加权回归(Geographically Weighted Regression, GWR)模型能有效减少误差。地理加权回归模型在传统的线性回归模型基础上考虑了空间对象的局部效应,运用局部加权最小二乘法来解释不同空间区位自变量和因变量之间的关系,可以用于探讨地理要素的空间异质性问题。模型的公式如下:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^n \beta_j(u_i, v_i) x_{ij} + \varepsilon_i, \quad (2)$$

其中, $y$  是因变量,表示第  $i$  个区县的 GDP; $\beta_0(u_i, v_i)$  为该区县的截距; $x_{ij}$  对应模型的第  $i$  个区县的解释变量, $\beta_j$  是回归系数; $\varepsilon_i$  为随机误差项。经过

不同空间核函数和带宽确定方法的组合测试,选择 Adaptive bi-square 作为空间核函数, AICc 确定最优带宽。

### 2.3 精度评价和验证

对回归模型和结果分别采用相对误差(RE)、相对平均绝对误差(%MAE)和均方根误差(RMSE)进行评价。公式如下:

$$RE = (y - \hat{y})/y, \quad (3)$$

$$\%MAE = \frac{\frac{1}{n} \sum | \hat{y} - y |}{\frac{1}{n} \sum y}, \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y - \hat{y})^2}, \quad (5)$$

其中, $y$  是 GDP 统计值, $\hat{y}$  是估计值。

## 3 结果分析

### 3.1 中国区县 GDP 空间拟合结果

对区县 GDP 数据和建模指标进行回归系数求解,得到每个区县的 GDP 估计值。从各项指标对比发现,GWR 回归结果的模型拟合度更好(表 2)。相较于 OLS 模型的各项指标,GWR 模型的总体拟合度高出 12%;RSS 值不足 OLS 模型的一半,表明模型拟合误差更小;GWR 模型的 AICc 值与 OLS 模型的 AICc 值差距远远大于 3,这说明 GWR 模型更可靠。优秀的模型需要满足回归残差在空间上随机分布的条件。通过空间自相关测试发现残差在空间上随机分布,没有呈显著集聚或分散,也证明了 GWR 模型的可用性。因此,选择使用 GWR 模型进行后续分析(图 2)。

将估计值与统计值进行对比,发现 GDP 整体空间分布很接近,但是也存在 GDP 高估或低估的情况,在胡焕庸线以西附近地区尤为明显。另外,分别计算得到全国 2 848 个区县、368 个地(县)级市的 %MAE 值分别为 25.59% 和 12.58%,即相应的准确率分别为 74.41% 和 87.42%,模型总体的拟合准确率较高。

表 2 OLS 模型与 GWR 模型的对比

项目	OLS 模型			GWR 模型		
	Adj. R <sup>2</sup>	RSS	AICc	Adj. R <sup>2</sup>	RSS	AICc
指标值	0.75	716.03	4 164.20	0.87	299.29	2 711.14

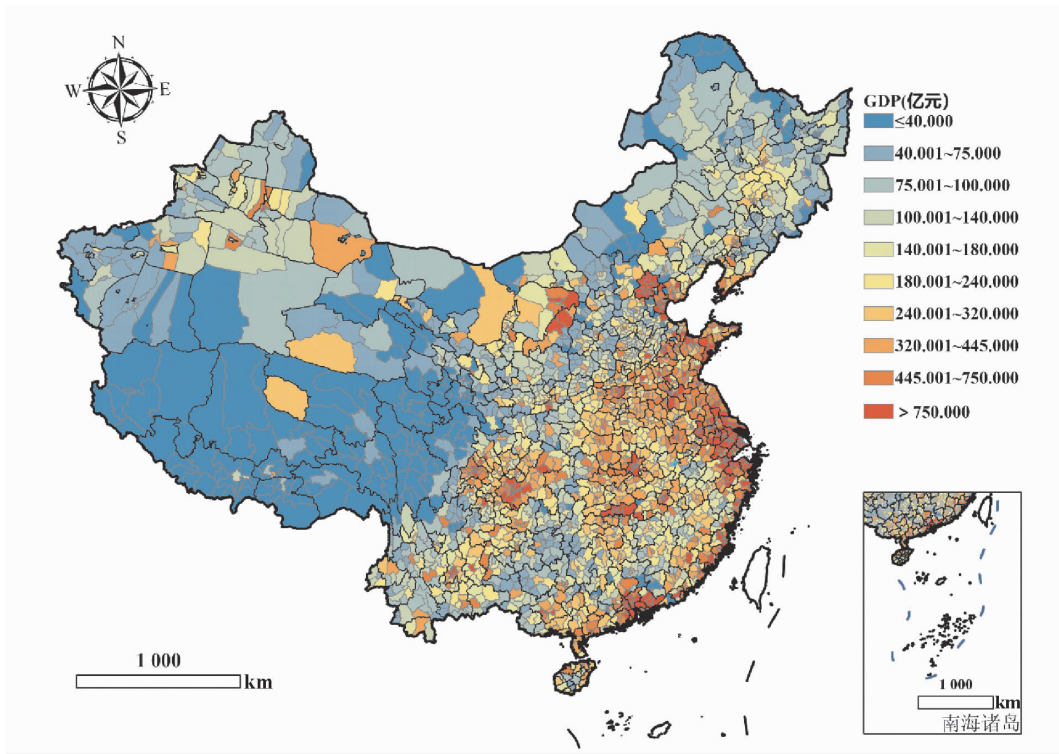


图 2 GDP 估计空间分布图

Fig. 2 The GDP distribution map of China

注:基于自然资源标注地图服务网站 GS(2020)4619 号标准地图制作,底图边界无修改。

### 3.2 不同建模因子的拟合能力

将用于 GWR 模型建模的 5 个因子分别单独进行 OLS 回归和 GWR 回归,探讨其中最有利于拟合 GDP 的因子。我国 GDP 采取分级核算制度,地级市 GDP 统计值与其下辖区县的 GDP 总和有细

微出入。由于训练数据集采用的是区县 GDP 统计值,因此,本研究将地级市 GDP 统计值作为验证数据,对估计的区县 GDP 值进行 *RMSE* 和 *% MAE* 指标验证(表 3),以期更科学合理地对不同建模因子对 GDP 的拟合能力。

表 3 不同建模因子的精度评价

Table 3 Accuracy evaluation of different modeling factors

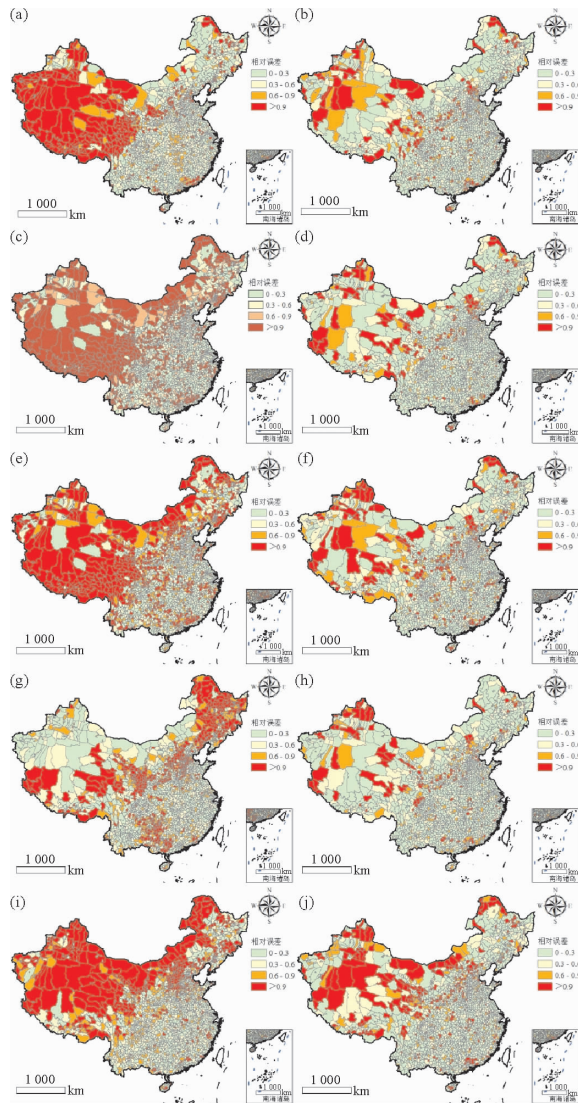
OLS 模型						
变量	夜间灯光	公司 POI	政府 POI	腾讯位置大数据	城镇建设面积	所有指标
<i>RMSE</i>	1 862.68	1 298.52	1 440.80	2 122.58	2 628.29	1 193.27
<i>% MAE</i>	29.49	24.72	27.08	38.21	43.67	21.85
GWR 模型						
变量	夜间灯光	公司 POI	政府 POI	腾讯位置大数据	城镇建设面积	所有指标
<i>RMSE</i>	979.87	808.61	892.90	810.77	1 040.91	723.38
<i>% MAE</i>	16.18	13.83	14.87	14.24	18.57	12.58

根据表 3 可知,在各个因子中,GWR 模型比 OLS 模型的准确率至少高出 10%,并且结合所有因子估计 GDP 要比单个因子估计 GDP 的准确率高。同时,每种因子对 GDP 的拟合能力也不同,在 GWR 模型中,公司 POI 对 GDP 的拟合表现最好

(相对应在地级市的准确率为 86.17%),其次是腾讯位置大数据(85.76%)、政府 POI(85.13%)和夜间灯光数据(83.82%),最后是城镇建设用地面积(81.43%)。综合两个模型的不同因子精度评价结果,POI 数据相较于其他数据更能反映

GDP;而腾讯位置大数据全局的拟合效果不如夜间灯光数据,但在局部的效果却比夜间灯光更好;城镇建设用地面积对 GDP 的拟合能力较其他数据逊色。

将 OLS 模型和 GWR 模型的相对误差链接到地图上,可以探究不同数据在不同经济发展程度地区的 GDP 拟合性能。研究将相对误差划分为  $(0, 0.3]$ ,  $(0.3, 0.6]$ ,  $(0.6 \sim 0.9]$ ,  $(0.9, +\infty]$  等 4 种等级的准确度,其中,左列为 OLS 模型的相对误差,右列为 GWR 模型的相对误差(图 3)。



(a) ~ (b) 夜间灯光; (c) ~ (d) 公司 POI; (e) ~ (f) 政府 POI; (g) ~ (h) 腾讯位置大数据; (i) ~ (j) 城镇建设面积

图 3 不同建模因子的拟合相对误差图

Fig. 3 Relative error of GDP prediction using different indicators

注:基于自然资源标注地图服务网站 GS(2020)4619 号标准地图制作,底图边界无修改。

从空间视觉上比较,无论是哪种数据和模型,胡焕庸线以西的地区 GDP 拟合误差整体上都比其他地区呈现出更多数量的高值聚集现象,表明该地区的估计误差相较东部地区来说更大。与夜间灯光数据和 POI 数据、城镇建设用地面积相比,腾讯位置大数据明显在新疆、西藏和青海等经济较落后区县的 GDP 拟合误差更小,即能够弥补上述数据在这些地区 GDP 拟合能力弱的缺陷,在东部地区,每种数据都能较好地拟合 GDP,并没有明显区别。

### 3.3 不同建模因子的空间分异

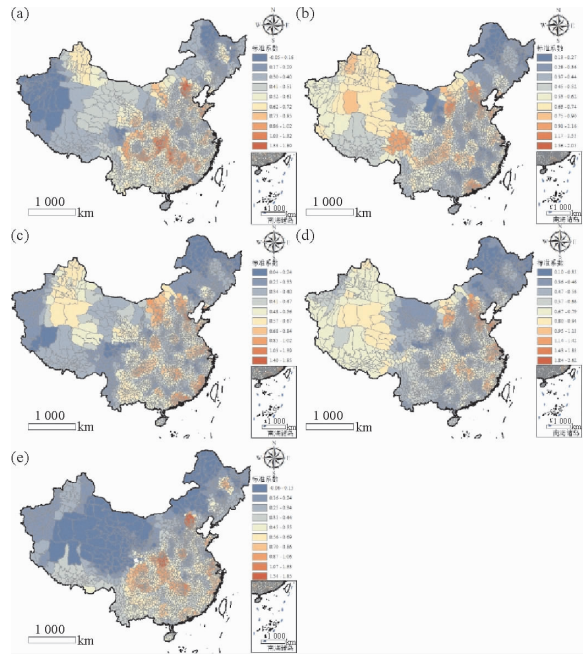
探讨不同建模因子回归系数的空间异质性可以进一步验证其对 GDP 的拟合能力,图 4 展示了单个建模因子的 GWR 标准化回归系数。如图 4 (a) 所示,夜间灯光数据在长江中游城市群、长江三角洲城市群东部、京津冀地区和山东半岛等经济发达地区表现出正向影响,而在西藏、新疆和东北地区则表现出较小的正向影响,这可能是灯光收集时受到下垫面和数据缺失的影响,但也表明夜间灯光能够在一定程度上指代经济发展状况。如图 4(b) ~ 图 4(c) 所示,公司 POI 和政府 POI 在粤港澳大湾区、长江中游城市群、长江三角洲城市群、京津冀城市群和成渝城市群这 5 大国家级城市群均呈现出较其他地区更大的正向影响,代表 POI 对于经济发达城市群的 GDP 预测具有地域普适性。如图 4(d) 所示,较其他数据而言,腾讯位置大数据明显在中国西南部对 GDP 有更大的正向影响,尤其是在新疆、青海和西藏的大部分区县,这其中部分区县 GDP 值甚至不超过 3 亿元,表明人口数量和电子产品普及率能指代相对贫困地区的经济水平。同时,北京市和上海市部分地区回归系数大于 1,说明腾讯位置大数据也能反映出发达地区的经济水平,这也与前人的研究结果一致<sup>[18]</sup>。如图 4(e) 所示,对于城镇建设面积,成渝城市群、京津冀城市群、中原城市群和长江三角洲部分城市获得了较大的回归系数,但在胡焕庸线以西的地区影响却很小,这可能因为这些地区土地集约化利用程度较低。

综上,不同地区的经济水平主导因素不同,POI 数据在较发达地区或城市群有很好的指示作用,但是在西部贫困地区,腾讯位置大数据的指示

作用更大。在中国大陆大范围地域进行 GDP 预测,每种数据之间互相能起到弥补的作用,人类经济活动、人口数量和城镇化率对 GDP 估计都有一定的影响。

### 3.4 验证

仅使用训练数据集当年年份做验证是不够的,使用其他年份的数据做检验能更充分说明建模因子的普适性。本研究也收集了 2018 年和 2019 年 VNP46A4 夜间灯光数据、2018 年土地利用数据(由于缺少 2019 年同源的土地利用数据,因此,将 2018 年土地利用数据作为 2019 年的土地利用方式,POI 数据同理)和 2019 年腾讯位置大数据分别作为 2018 年和 2019 年的自变量验证数据集,将 2020 年 GWR 模型的系数代入到其余年份的自变量数值来估计验证当年地级市和区县的 GDP。如图 5 所示,在两个验证年份中,GDP 统计值和估计值的回归 R 方均在 0.8 以上,回归系数接近 1,说明 GWR 模型的可靠性和建模因子具有普适性、有效性,可以基于这些地理因子快速监测行政单元的 GDP。

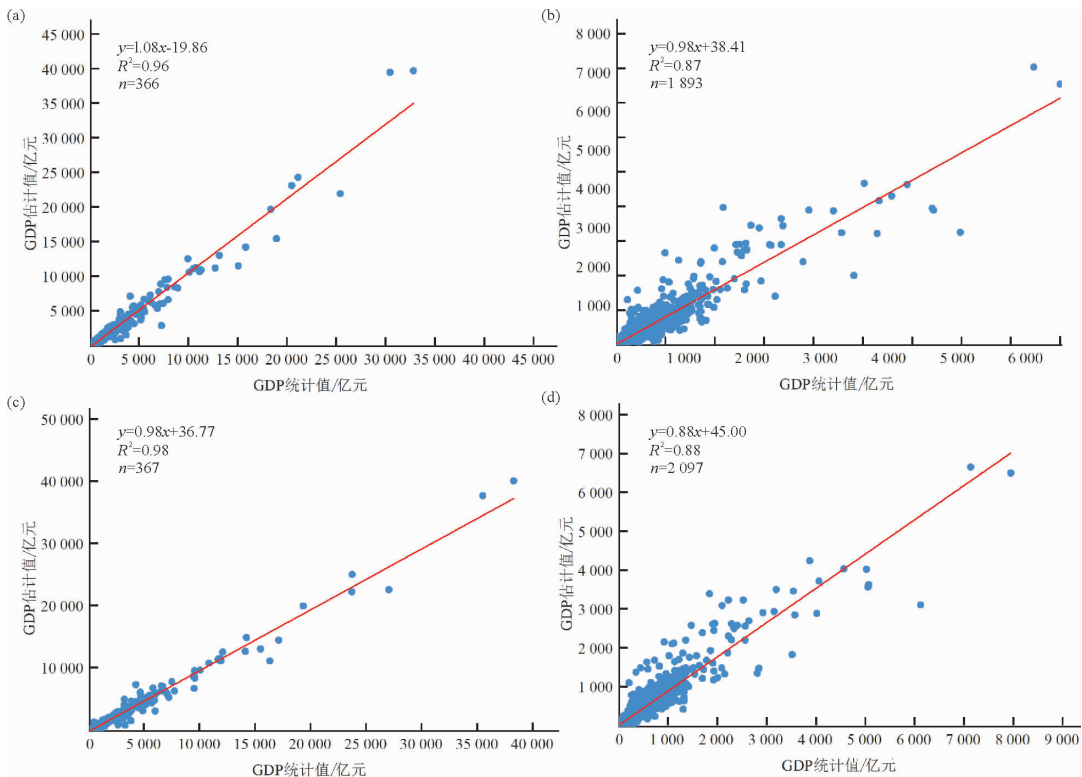


(a) 夜间灯光;(b) 公司 POI;(c) 政府 POI;(d) 腾讯位置大数据;(e) 城镇建设面积

图 4 不同建模因子的 GWR 标准系数图

Fig. 4 The standard coefficient of GWR model using different indicators

注:基于自然资源标注地图服务网站 GS(2020)4619 号标准地图制作,底图边界无修改。



(a) 2018 年地级市验证结果;(b) 2018 年区县验证结果;(c) 2019 年地级市验证结果;(d) 2019 年区县验证结果

图 5 不同年份 GDP 估计值验证图

Fig. 5 Verification of predicted GDP in different years

## 4 结论和讨论

本研究基于夜间灯光遥感数据、POI 数据、腾讯位置大数据和土地利用数据,绘制了中国区县行政单元的 GDP 空间分布图,总体准确率较高。无论是单因子还是多因子建模,GWR 模型都比 OLS 模型的整体预测准确率提高了约 10 个百分点,这也表明基于多源地理大数据比单一地理数据绘制中国 GDP 分布图更可取。同时探讨了不同地理数据源拟合 GDP 的优劣性和适用性,为后续经济研究提供一定的参考依据。结果表明,POI 数据相较于夜间灯光遥感数据和腾讯位置大数据、土地利用数据,在 GDP 空间建模中具有更大的

潜力,尤其是在经济发达地区。但在贫困地区,腾讯位置大数据是反映经济最好的指标。

受限于大范围实时数据的获取和腾讯位置大数据的分辨率,研究只基于 2018 年的 POI 数据和腾讯位置大数据进行回归计算,然而 2020 年我国已经实现了全面脱贫,贫困区县会增加一定数量的 POI 数据,且由于穿戴移动定位设备和电子产品普及率的提高也会导致腾讯位置数据量增多,因此,研究估计 GDP 也会受到一定的影响。目前,GDP 的官方数据只统计到行政单位一级,很难与其他精细尺度的栅格数据匹配从而进行科学研究,因此,未来还将基于研究中的建模因子,利用随机森林模型建立不同产业的模型来绘制 GDP 空间化栅格图。

### 参考文献:

- [1] McCord G C, Rodriguez-Heredia M. Nightlights and subnational economic activity: Estimating departmental GDP in Paraguay[J]. *Remote Sensing*, 2022, 14(5): 1150.
- [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] Ma T, Zhou C, Pei T, et al. Responses of SUOMI-NPP VIIRS-derived nighttime lights to socioeconomic activity in China's cities[J]. *Remote Sensing Letters*, 2014, 5(2): 165-174.
- [4] 柴子为,王帅磊,乔纪纲. 基于夜间灯光数据的珠三角地区镇级 GDP 估算[J]. *热带地理*, 2015, 35(3): 379-385.
- [5] 王正,贾公旭,张清凌,等. COVID-19 疫情背景下 2020 年第一季度广东省二、三产业 GDP 空间分布变化分析[J]. *自然资源遥感*, 2021, 33(3): 184-193.
- [6] Chen Z, Wei Y, Shi K, et al. The potential of nighttime light remote sensing data to evaluate the development of digital economy: A case study of China at the city level[J]. *Computers, Environment and Urban Systems*, 2022, 92: 101749.
- [7] Zhong L, Liu X, Ao J. Spatiotemporal dynamics evaluation of pixel-level gross domestic product, electric power consumption, and carbon emissions in countries along the belt and road[J]. *Energy*, 2022, 239: 121841.
- [8] Putri S R, Wijayanto A W, Sakti A D. Developing relative spatial poverty index using integrated remote sensing and geospatial big data approach: A case study of east JAVA, Indonesia[J]. *ISPRS International Journal of Geo-Information*, 2022, 11(5): 275.
- [9] 刘冰洁,陈佐旗,余柏滢,等. 夜间灯光遥感对城市发展类动能与相似性评估[J]. *遥感学报*, 2021, 25(5): 1187-1200.
- [10] Alahmadi M, Mansour S, Dasgupta N, et al. Using nighttime lights data to assess the resumption of religious and socioeconomic activities post-covid-19[J]. *Remote Sensing*, 2023, 15(4): 1064.
- [11] Shi K, Yu B, Huang Y, et al. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data[J]. *Remote Sensing*, 2014, 6(2): 1705-1724.
- [12] Li X, Xu H, Chen X, et al. Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China[J]. *Remote Sensing*, 2013, 5(6): 3057-3081.
- [13] Wang Z, Román M O, Kalb V L, et al. Quantifying uncertainties in nighttime light retrievals from SUOMI-NPP and NOAA-20 VIIRS day/night band data[J]. *Remote Sensing of Environment*, 2021, 263: 112557.
- [14] Xu Y, Song Y, Cai J, et al. Population mapping in China with Tencent social user and remote sensing data[J]. *Applied Geography*, 2021, 130: 102450.
- [15] Ma X, Yang Z, Wang J, et al. Mapping population on tibetan plateau by fusing VIIRS data and nighttime Tencent location-

- based services data[J]. *Ecological Indicators*, 2022, 139: 108893.
- [16] 李慧敏, 罗大伟, 窦世卿. 利用腾讯位置大数据进行多尺度人口空间化估算[J]. *测绘通报*, 2022(6): 93-97.
- [17] 吴中元, 许捍卫, 胡钟敏. 基于腾讯位置大数据的精细尺度人口空间化——以南京市江宁区秣陵街道为例[J]. *地理与地理信息科学*, 2019, 35(6): 61-65.
- [18] Huang Z, Li S, Gao F, et al. Evaluating the performance of LBSM data to estimate the gross domestic product of China at multiple scales: A comparison with NPP-VIIRS nighttime light data[J]. *Journal of Cleaner Production*, 2021, 328: 129558.
- [19] Huang Z, Li S, Peng Y, et al. Spatial non-stationarity of influencing factors of China's county economic development base on a multiscale geographically weighted regression model[J]. *ISPRS International Journal of Geo-Information*, 2023, 12(3): 109.
- [20] Ma T. Multi-level relationships between satellite-derived nighttime lighting signals and social media-derived human population dynamics[J]. *Remote Sensing*, 2018, 10(7): 1128.
- [21] Zhao N, Cao G, Zhang W, et al. Tweets or nighttime lights: Comparison for preeminence in estimating socioeconomic factors[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 146: 1-10.
- [22] 池娇, 焦利民, 董婷, 等. 基于 POI 数据的城市功能区定量识别及其可视化[J]. *测绘地理信息*, 2016, 41(2): 68-73.
- [23] 陈子龙, 王芳, 李少英, 等. 基于多源数据的县域主导功能类型划分及其空间结构模式识别[J]. *地球信息科学学报*, 2021, 23(12): 2215-2231.
- [24] Chen Q, Ye T, Zhao N, et al. Mapping China's regional economic activity by integrating points-of-interest and remote sensing data with random forest[J]. *Environment and Planning B: Urban Analytics and City Science*, 2021, 48(7): 1876-1894.
- [25] 刘杨, 李宏伟, 杨斌程, 等. 基于遥感数据和 POI 数据的 GDP 空间化研究——以北京市为例[J]. *地域研究与开发*, 2021, 40(2): 27-32,39.
- [26] 杨妮, 邓树林. 基于 POI 大数据与土地利用数据协同的小尺度区域 GDP 空间化方法[J]. *测绘地理信息*, 2022, 47(2): 144-147.
- [27] 张爱华, 潘耀忠, 明艳芳, 等. 多源信息耦合的 GDP 空间化研究——以北京市为例[J]. *遥感技术与应用*, 2021, 36(2): 463-472.
- [28] 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(2): 120-123.
- [29] 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, 27(3): 396-405.
- [30] 卢秀, 李佳, 段平, 等. 基于夜间灯光和土地利用数据的 GDP 空间化及预测——以云南沿边地区为例[J]. *地域研究与开发*, 2020, 39(2): 36-39,81.
- [31] 张怡哲, 杨续超, 胡可嘉, 等. 基于多源遥感信息和土地利用数据的中国海岸带 GDP 空间化模拟[J]. *长江流域资源与环境*, 2018, 27(2): 235-242.
- [32] Chen Y, Wu G, Ge Y, et al. Mapping gridded gross domestic product distribution of China using deep learning with multiple geospatial big data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 1791-1802.
- [33] Xu Z, Wang Y, Sun G, et al. Generating gridded gross domestic product data for China using geographically weighted ensemble learning[J]. *ISPRS International Journal of Geo-Information*, 2023, 12(3): 123.
- [34] Li X, Ma R, Zhang Q, et al. Anisotropic characteristic of artificial light at night-systematic investigation with VIIRS dnb multi-temporal observations[J]. *Remote Sensing of Environment*, 2019, 233: 111357.
- [35] 原国家质检总局、国家标准委. 国民经济行业分类(GB/T 4754—2017)[S]. 北京: 中国统计出版社, 2018.
- [36] 张海平, 周星星, 汤国安, 等. 基于 GIS 场模型的城市餐饮服务热点探测及空间格局分析[J]. *地理研究*, 2020, 39(2): 354-369.
- [37] Chen M, Xian Y, Huang Y, et al. Open fine-scale population spatialization data descriptor data of China in 2018 based on real location-based big data[J]. *Scientific Data*, 2022, 9: 624.