

引用格式:王明明,王卷乐.基于夜间灯光与土地利用数据的山东省乡镇级人口数据空间化[J].地球信息科学学报,2019,21(5):699-709. [Wang M M, Wang J L. Spatialization of township-level population based on nighttime light and land use data in Shandong province[J]. Journal of Geo-information Science, 2019,21(5):699-709.] DOI:10.12082/dqxxkx.2019.180497

基于夜间灯光与土地利用数据的山东省乡镇级人口数据空间化

王明明^{1,2}, 王卷乐^{2,3*}

1. 山东理工大学 建筑工程学院, 淄博 255049; 2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101; 3. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023

Spatialization of Township-level Population based on Nighttime Light and Land Use Data in Shandong Province

WANG Mingming^{1,2}, WANG Juanle^{2,3*}

1. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo, 255049, China; 2. State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

Abstract: Gridded population data can be used to describe the actual spatial distribution of populations and is an effective way to achieve better integration of population data with natural, social, and economic factors. This study analyzed the demand of fine-scale gridded population data. Taking the densely populated Shandong Province in eastern China as an example, the spatialization method of township-level demographic data was investigated using nighttime satellite data and land use data fusion modeling. In this process, EVI was used to reduce saturation of DMSP/OLS nighttime satellite data to increase the difference of population distribution within the urban land. The urban and rural two-level partition method was used to avoid the shortcomings of nighttime data in the low radiance area of rural areas. The demographic values from the rest region were used to evaluate the modeling accuracy, and the results showed that 78% of the administrative units had an absolute relative error of less than 20%. Finally, based on the population data of the township-level which was first published in the fifth census in 2000, the gridded population data SDpop2000 at 100 m-resolution in Shandong Province was generated. The SDpop2000 was compared with the global WorldPop population data product with higher precision. The results showed that the correlation coefficient between SDpop2000 and WorldPop on the 10 km grid scale was as high as 0.93, and the population distribution of SDpop2000 was obviously more accurate than that of WorldPop in the central Shandong, southwestern Taian, southern Jining, southern Linyi,

收稿日期:2018-09-29;修回日期:2019-02-25.

基金项目:中国科学院战略性先导科技专项(A类)(XDA19040501);中国科学院“十三五”信息化专项科学大数据工程项目(XXH13505-07)。[**Foundation items:** the Strategic Priority Research Program of the Chinese Academy of Sciences, No.XDA19040501; the Specific Informatization Scientific Research Science Program of the Chinese Academy of Sciences, No.XXH13505-07.]

作者简介:王明明(1993-),男,河北沧州人,硕士生,主要从事人口数据空间化研究。E-mail: wangmm@lreis.ac.cn

*通讯作者:王卷乐(1976-),男,河南洛阳人,博士,研究员,主要从事科学数据共享、地理信息系统与遥感应用研究。
E-mail: wangjl@igsrr.ac.cn

northern Zaozhuang, and the northern Shandong coastal areas. In addition, the SDpop2000 better described the population distribution trend in Shandong Province, which was denser in western Shandong and the plains of northern Shandong than in the mountainous hilly areas in the central and southern Shandong, the coast of northern Shandong, and the hilly area of Shandong Peninsula. Overall, the spatialization method of township-level population data developed in this study significantly improved spatialization precision and is suitable for township-level population spatialization.

Key words: population; spatialization; nighttime light data; land use data; township-level; Shandong Province

***Corresponding author:** WANG Juanle, E-mail: wangjl@igsnnr.ac.cn

摘要: 格网化人口数据能够刻画实际人口空间分布状况,是实现人口数据更好地与自然、社会、经济等要素融合分析的有效途径。本文面向精细尺度格网人口数据的需求,以中国东部人口稠密的山东省为例,基于乡镇级人口统计数据,研究了结合夜间灯光和土地利用数据的空间化方法。其中以 EVI 修正 DMSP/OLS 夜间灯光数据来增加城镇用地内部人口分布的差异性,以城乡二级分区方法避免夜间灯光数据在农村低辐射亮度区模拟人口的缺点,提高了建模精度。利用其余地区的人口统计值检验建模精度,结果有 78% 的行政单元的相对误差绝对值小于 20%。最终在 2000 年首次公布的乡镇级人口统计数据的基础上,生成了山东省 100 m 格网人口分布数据 SDpop2000。通过与精度较高的全球 WorldPop 人口数据产品对比可见,SDpop2000 和 WorldPop 在 10 km 网格尺度上的相关性系数高达 0.93;SDpop2000 在鲁中部、泰安西南部、济宁南部、临沂南部、枣庄北部和鲁北沿海等地的人口分布明显比 WorldPop 更准确;且 SDpop2000 较好地刻画了山东省在鲁西、鲁北平原区的人口较鲁中南山地丘陵区、鲁北沿海和山东半岛丘陵区的人口更为稠密的人口分布趋势。本文构建的基于 DMSP/OLS 与土地利用的乡镇级人口数据空间化方法明显提高了空间化精度,适用于乡镇尺度的人口精细模拟。

关键词: 人口;空间化;夜间灯光数据;土地利用数据;乡镇级;山东省

1 引言

人口的空间分布是人口地理学研究的核心问题^[1]。研究人口空间分布与其他因素的相关关系,精确估计和掌握详实的人口分布信息,是改善人类生存环境和解决人口问题的重要科学基础。人口统计数据是以最小行政区划为统计单元逐级汇总而得^[2],往往具有以下局限性:①不能在更小的尺度(行政单元内)反映人口空间分布特征^[3];②随着时间的推移,行政单元随着行政系统的改变而变化,使得基于行政单元的长时间序列分析变得困难^[4];③行政边界与自然地物的边界一般不重合,造成地学研究中的“可变元问题”^[5];④人口统计数据的更新周期较长,如中国人口普查一般为 10 年一次,更新周期难以满足科学研究的需要。

人口数据空间化是将一种空间尺度的人口数据合理分配到另一种空间尺度上的过程,通常指将人口统计数据分配到一定尺寸的规则格网上^[6]。产生的格网化人口空间数据可以同自然、地理、生态等数据融合分析,可有效避免上述人口统计数据的局限性,是人口统计数据的拓展和补充,并且现在已经得到广泛应用^[7],如用于支持疾病风险评估、气候变化和人类健康适应性政策和防灾减灾等。当

前全球可开放获得的格网化人口空间分布数据集主要包括中国公里网格人口分布数据集(CnPop)、全球栅格人口数据集(GPW)、全球农村城镇制图计划数据集(GRUMP)和世界人口计划数据集(Worldpop)等。柏中强等^[8]研究发现 Worldpop 在这些数据集中的估计精度最高,准确估计了约 60% 人口的区域人口分布。随着遥感、GIS 和计算机技术的发展,人口数据空间化研究发展迅速,陆续出现了基于理论的方法^[9]、基于遥感和 GIS 的方法^[10-13]、基于多源数据融合的方法^[14]和基于社会感知数据的方法^[7]等不同的人口数据空间化方法。基于遥感和 GIS 的人口数据空间化方法使用最为广泛,其中土地利用和夜间灯光数据是该方法中最为常用的 2 类基础数据^[15-19],但在应用中面临着诸多技术方法挑战。

土地利用类型与人口分布格局有耦合关系,但难以区分土地利用类型内部的人口分布差异。DMSP/OLS 夜间灯光数据是美国国防气象卫星计划(DMSP)卫星的线性扫描业务系统(OLS)探测的某地区夜晚的灯光亮度信息,是监测人类活动的有效数据源。DMSP/OLS 夜间灯光数据已被证实其同人口之间的高相关性^[11,18-19],土地利用数据和夜间灯光数据结合使用,可有效弥补城镇地区土地利用

类型内部的人口分布差异。如梁友嘉等^[20]利用土地利用数据和夜间灯光数据的关系获取更符合真实人口空间分布的灯光辐射数据,基于不同土地利用类型的灯光区面积、非灯光区面积和总灯光辐射强度进行人口空间分布建模,实现对张掖市甘州区乡镇级人口统计数据空间化,模型拟合优度系数为0.88;黄杰等^[21]以土地利用数据和夜间灯光数据为建模参考要素,基于空间滞后回归模型实现江苏省级人口统计数据的空间化,模型拟合优度达到0.93,乡镇级尺度空间化精度检验相关系数为0.90;陈晴等^[6]集成土地利用数据和夜间灯光数据,以县域相对误差的绝对值20%为阈值控制动态样本的逐渐回归和分区建模,实现大陆沿海区域县级人口统计数据的空间化,整体拟合优度系数大于0.8。但由于DMSP卫星传感器本身的缺陷,DMSP/OLS夜间灯光数据在高强度灯光区会存在灯光强度饱和值^[22],在低辐射亮度区域探测能力不足^[23],这大大地影响到城市中心和农村地区的人口估计精度。总体来看,土地利用和夜间灯光数据在人口数据空间化应用中既有优势又有劣势,如何做到扬长避短是人口数据空间化应用中的关键问题之一。

针对这一难题,本文探索基于土地利用和DMSP/OLS夜间灯光的乡镇级人口统计数据的空间化方法。以山东省为例,利用土地利用数据和乡镇级行政单元构建城乡二级分区方法将城镇地区和农村地区分开建模,利用EVI植被指数对DMSP/OLS夜间灯光数据进行修正,尝试解决DMSP/OLS夜间灯光数据在农村低辐射亮度区和城市高辐射亮度区进行人口模拟的不足,最终实现2000年山东省乡镇级人口统计数据的空间化。这不仅为我国当前最精细尺度的乡镇级统计数据空间化提供方法参考,而且因规避了乡镇界限问题而为乡镇级人口空间化数据产品的数据共享创造了条件。

2 研究区概况、数据来源与研究方法

2.1 研究区概况

本文选取中国东部人口稠密的山东省作为研究区(图1)。山东省位于华东地区最北端,东部沿海、西部连接中原,与河北省、河南省、安徽省和江苏省接壤。陆地南北最长约420 km,东西最宽约700 km,陆地总面积15.67万km²。境内包括内陆和半岛两部分,可划分为鲁中南山地丘陵区、山东半



图1 山东省行政区划

Fig. 1 Map of study area (Shandong province)

岛丘陵区 and 鲁西、鲁北平原区3个区,鲁中南山地丘陵区为全省海拔最高、山地面积最广的地区,泰山是最高峰;山东半岛丘陵区位于山东省东部,伸入渤海、黄海间,是典型的丘陵区;鲁西、鲁北平原区是全省海拔最低处,属于华北大平原的一部分。2000年底,山东省境内共有2303个乡镇级行政单元^[24],2000年第五次人口普查结果表明,山东省人口总数为8997.1789万人^[25],约占全国总人口的7%。

2.2 数据源

本文涉及的数据源包括DMSP/OLS夜间灯光数据、EVI植被指数数据、土地利用数据、世界人口计划数据集WorldPop、乡镇级人口统计数据和行政边界数据。数据时间为2000年,数据范围为山东省。

(1)DMSP/OLS夜间灯光数据来源于美国国家海洋和大气管理局国家环境信息中心夜间非辐射定标的平均稳态数据(Stable_Lights.avg_vis Data)^[26],可见像素值范围为0~63,空间分辨率约为850 m。

(2)EVI植被指数数据来源于美国国家航空航天局LAADS DAAC数据中心的3级格网数据产品(MOD13Q1)^[27],空间分辨率约为250 m。

(3)土地利用数据来源于国家科技基础条件平台—国家地球系统科学数据共享平台^[28],空间分辨率为100 m,该数据集包含耕地、林地、草地、水域、建设用地和未利用地在内的一级类和包括有林地、灌木林、疏林地、其他林地和高、中、低覆盖度草地等二级类型数据,共23类属性值。

(4)WorldPop数据来源于WorldPop project官方网站^[29],由英国南安普顿大学地理数据研究所领导的全球人口分布制图计划研制,其空间分辨率为90 m。2000年中国区域WorldPop数据集生产方法

是土地覆被类型法,所用的土地覆被数据为MDA GeoCover数据集,辅助数据包括建筑用地分布数据、OpenStreetMap数据、城镇规划数据等,主要用来校正土地覆被数据上的居民地分布。

(5)乡镇级行政边界数据来源于国家科技基础条件平台—国家地球系统科学数据共享平台^[30],比例尺为1:25万。

(6)乡镇级人口统计数据来源于《中国乡、镇、街道人口资料》^[25]。

2.3 数据预处理

在乡镇级人口数据空间化之前,需先整编构建乡镇级人口空间数据库、修正DMSP/OLS夜间灯光数据和对土地利用数据进行重分类。所有数据源均统一为China Geodetic Coordinate System 2000地理坐标系统和Albers投影。

(1)乡镇级人口空间数据。利用政府官网、行政区划网、百度地图等查阅相关乡镇(街道)的区划调整、更名或缺失等信息,根据该信息将部分乡镇(街道)的行政边界数据或人口统计数据进行合并或更名,以使两者能够匹配。利用ArcGIS的“拓扑关系构建”功能处理行政边界之间的重叠和空隙等拓扑错误,完成所有行政单元的精确匹配。

(2)夜间灯光数据。借鉴卓莉等^[22]提出的方法对DMSP/OLS夜间灯光数据进行修正,得到基于EVI指数修正的增强型植被指数的夜间灯光调整指数(Enhanced Vegetation Index Adjusted Nighttime Light Index, EANTLI):①利用ArcGIS的栅格计算器取7、8月份4景EVI影像的均值,并做归一化处理,剔除值小于0.01的像元,得到 EVI_a 影像的值为 $[0.010025, 0.8642]$;②将DMSP/OLS夜间灯光数据做归一化处理,并利用栅格计算器根据式(1)对EANTLI影像进行栅格计算;③根据以往研究的做法^[17,20],将EANTLI影像重采样为100 m分辨率。

$$EANTLI = \frac{1 + (NTL_n - EVI_a)}{1 - (NTL_n - EVI_a)} \times NTL \quad (1)$$

式中: NTL_n 为归一化后的DMSP/OLS夜间灯光指数; EVI_a 为归一化的剔除值小于0.01的EVI均值; NTL 为未归一化的DMSP/OLS夜间灯光指数。

(3)土地利用数据。将土地利用数据中的水田和旱地合并为耕地,有林地、灌木林、疏林地和其他林地合并为林地,高覆盖度草地、中覆盖度草地和低覆盖度草地合并为草地,河渠、湖泊、水库坑

塘、滩涂、滩地和沼泽地合并为水系用地,沙地、盐碱地、裸土地、裸岩石质地和其他合并为未利用地,保留城镇用地、农村居民点、其他建设用地共8个属性。

2.4 城乡分区回归模型构建及检验

2.4.1 城乡二级分区方法

城乡二级分区方法如图2所示。一级分区:将所有乡镇级行政单元按照其内部居民地的类型分为3类:城镇地区、农村地区和其余地区。只包含城镇用地的行政单元属于城镇地区;只包含农村居民点的行政单元属于农村地区;同时包含城镇用地和农村居民点两种居民地类型的行政单元属于其余地区。二级分区:在一级分区的基础上,将农村地区和其余地区再各自分为3个分区,其中农村地区按照农村居民点人口密度(式(2))大小,将农村地区的行政单元分为低人口密度区 A_2 、中人口密度区 A_3 和高人口密度区 A_4 ,具体分区比例根据模型检验结果调整,当建模精度达到最优,则最终确定分区比例。其余地区用来检验建模精度,其分区方法依赖于模型检验过程,具体在模型检验中叙述。

$$density_v = \frac{pop_v}{area_v} \quad (2)$$

式中: $density_v$ 为农村居民点人口密度; pop_v 为农村地区行政单元总人口数; $area_v$ 为农村居民点面积。

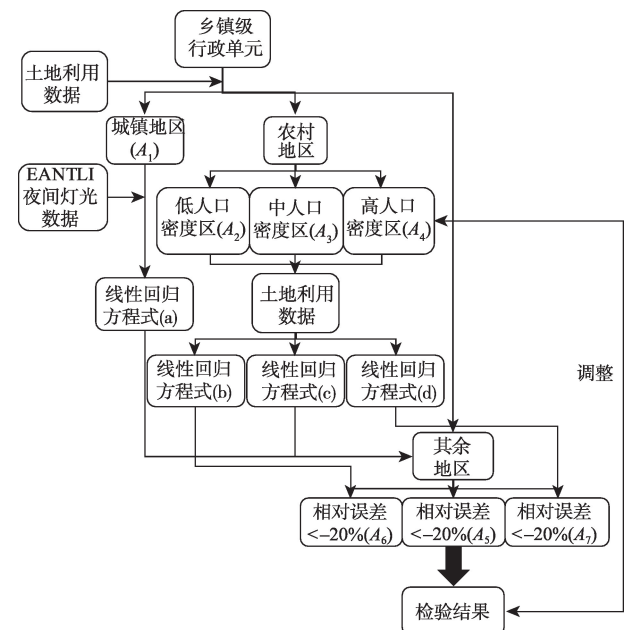


图2 城乡二级分区建模和检验技术流程

Fig. 2 Flowchart of modeling and inspection based on urban and rural two-level partition

2.4.2 城乡分区回归模型构建

城镇地区作为 A_1 分区, 在乡镇级行政单元尺度上, 建立 EANTLI 夜间灯光总值和人口总值的回归模型, 回归方程如式(3)所示, 建模结果对应图2中的方程式(a)。

$$pop_c = a \times EANTLI + b \quad (3)$$

式中: pop_c 为城镇地区行政单元总人口数量; EANTLI 为城镇地区行政单元中城镇用地范围内的灯光总强度; a 为待定回归系数; b 为常数项, 代表城镇地区行政单元人口数量的修正值。

在农村地区的低人口密度区、中人口密度区和高人口密度区分别建立人口总数同农村居民点面积的回归模型, 回归方程如式(4)所示, 建模结果分别对应图2中的方程式(b)、方程式(c)和方程式(d)。

$$pop_{vi} = a_i \times area_{vi} + b_i \quad (4)$$

式中: pop_{vi} 为农村地区第 i 个分区的行政单元人口总数; $area_{vi}$ 为农村地区第 i 个分区的行政单元农村居民点面积; a_i 为待定回归系数, 代表农村居民点的人口密度; b_i 为常数项, 代表农村地区行政单元人口数量的修正值; i 的取值为 2、3、4, 分别代表分区 A_2 、 A_3 和 A_4 。

2.4.3 模型检验

本文利用其余地区的乡镇级人口统计值对建模精度进行检验, 即利用其余地区的人口统计值和模拟值之间的相对误差(式(5))评估建模精度。其中其余地区的人口模拟值利用图2中方程式(a)~(d)计算所得: ① 利用方程式(a)、(c)对其余地区的人口做初次模拟, 并通过式(5)计算人口统计值和模拟值的相对误差 δ 。② 统计 $-20 < \delta < 20$ 的行政单元作为 A_5 分区, $\delta < -20$ 的行政单元作为 A_6 分区, $\delta > 20$ 的行政单元作为 A_7 分区。③ 保留 A_5 分区的人口模拟值不变, A_6 和 A_7 分区的人口分别利用方程式(a)、(b)和(a)、(d)进行重新模拟。

$$\delta = \frac{pop_i^{sta} - pop_i^{mo}}{pop_i^{sta}} \times 100\% \quad (5)$$

式中: pop_i^{sta} 为其余地区第 i 个行政单元的统计人口数; pop_i^{mo} 为其余地区第 i 个行政单元的模拟人口数。

2.5 人口数据空间化及结果对比

2.5.1 人口数据空间化

基于上文建立的各个分区, 利用图2中的方程式(a)~方程式(d)去掉公式的常数项——行政单元尺度人口数量的修正值)模拟每个分区网格单元的初

始人口数。其中 A_1 、 A_2 、 A_3 和 A_4 分区分别利用方程式(a)~(d)模拟, A_5 分区利用方程式(a)和(c)模拟, A_6 分区利用方程式(a)和(b)模拟, A_7 分区利用方程式(a)和(d)模拟。然后统计每个乡镇级行政单元内的初始人口模拟总数 pop_i^0 , 因受模型公式误差的影响, 每个行政单元的初始人口模拟值 pop_i^0 和实际人口统计值 pop_i^{sta} 不一致, 因此建立各行政单元的修正公式(式(6)), 对各网格的初始模拟结果进行调整, 使得各行政单元模拟的人口总数与实际统计人口数完全吻合。最后将调整后的网格人口数转换成 100 m 栅格数据。

$$p_{ij}^{mo} = \frac{p_{ij}^0}{pop_i^0} \times pop_i^{sta} \quad (6)$$

式中: p_{ij}^{mo} 表示第 i 个乡镇级行政单元第 j 个格网单元修正后的人口模拟值; p_{ij}^0 表示第 i 个行政单元第 j 个格网单元初始人口模拟值; pop_i^0 表示第 i 个行政单元的初始人口模拟值; pop_i^{sta} 表示第 i 个行政单元人口统计值。

2.5.2 空间化结果对比

对比 WorldPop 和 SDpop2000 的制图效果、10 km 网格内人口数的相关性、10 km 网格内人口数的相对误差和相对误差的空间分布。其中相对误差计算公式如下:

$$\delta = \frac{p_s - p_w}{p_w} \times 100\% \quad (7)$$

式中: p_s 代表 SDpop2000 在 10 km 网格内的人口数; p_w 代表 WorldPop 在 10 km 网格内的人口数, δ 代表相对误差。

3 结果及分析

3.1 分区建模及模型检验

经过多次调整农村地区二级分区比例, 最终确定 A_2 、 A_3 和 A_4 分别取农村地区行政单元数量的 1/6、2/3 和 1/6 时模型精度达到最高。经过检验, 有占比 78% 的行政单元的相对误差绝对值小于 20%, 且其余地区的模拟人口同统计人口的相关系数为 0.931, 建模精度较高。建模及分区结果如表1所示。

3.2 行政单元人口统计值与建模数据的相关性分析

在各建模区域, 检验行政单元内人口统计值同各类土地利用类型面积及 EANTLI 夜间灯光数据的相关性, 如表2所示。其中, 城镇地区行政单元人

表1 各地区分区和建模结果
Tab. 1 Results of partitioning and modeling

分区		行政单元数/个	建模结果	方程拟合优度系数
城镇地区	A_1	31	$pop_c = 0.165 \times EANTLI + 8399.931$	0.992
	A_2	181	$pop_{v2} = 0.0025 \times area_{v2} + 3210.143$	0.778
农村地区	A_3	724	$pop_{v3} = 0.004 \times area_{v3} + 6403.842$	0.763
	A_4	181	$pop_{v4} = 0.006 \times area_{v4} + 8873.137$	0.838
	A_5	322	—	—
其余地区	A_6	177	—	—
	A_7	210	—	—

表2 各建模区域行政单元人口统计值同各要素之间的相关性
Tab. 2 Correlation coefficients between demographic values of administrative units in each partition and various modeling elements

分区		耕地	林地	草地	水系用地	未利用地	城镇用地	农村居民点	其他建设用地	EANTLI
城镇地区	A_1	0.294	0.093	0.092	0.204	0	0.819**	0	-0.095	0.996**
	A_2	0.524**	0.022	-0.133	-0.160*	-0.184*	0	0.883**	-0.216**	0.228**
农村地区	A_3	0.635**	0.015	0.09*	0.025	-0.027	0	0.873**	0.034	0.279**
	A_4	0.595**	-0.227	-0.068	0.053	0.045	0	0.915**	0.112	0.143

注:*表示在0.05水平(双侧)上显著相关,**表示在0.01水平(双侧)上显著相关。

口统计值与城镇用地面积、夜间灯光辐射值有较高的相关性;农村地区行政单元人口统计值与农村居民点面积有较高的相关性,与耕地面积有较弱相关性。说明本文的建模数据选择合理,其中农村地区的人口统计值与耕地面积有较弱的相关性,主要是因为耕地面积在一定程度上可以表征一个乡镇行政单元内的人口数量,但并不表征人口分布在耕地上的数量多少。

3.3 空间化结果与对比

3.3.1 乡镇级人口数据空间化结果

山东省2000年乡镇级人口统计数据空间化结果如图3所示,分辨率为100 m。从格网人口分布数量来看,山东全省人口密度的值域在0~335之间,人口密度超过100人的地区基本分布在地级市。从100 m格网人口空间分布趋势来看,城镇地区的人口密度明显高于农村地区,且高人口密度主要集中在济南、青岛、潍坊、淄博、烟台、德州、威海和济宁等地级城市驻地,济南和青岛两地的人口密度和数量最高;鲁西、鲁北平原区的人口分布较鲁中南山地丘陵区、鲁北沿海和山东半岛丘陵区的更为稠密。

3.3.2 结果对比

本文从3个方面对比SDpop2000和WorldPop。

(1)2000年WorldPop在山东省的成图效果如

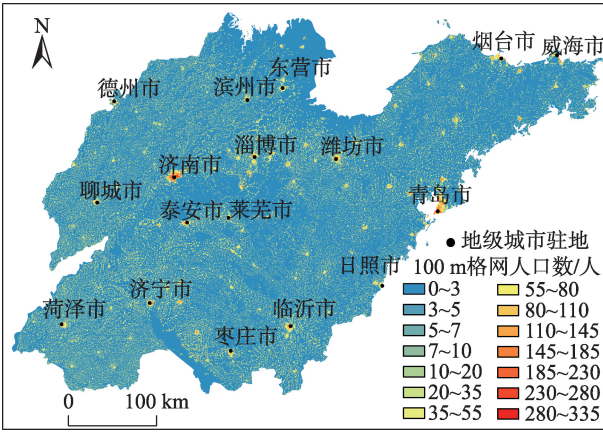


图3 山东省2000年100 m分辨率人口分布
Fig. 3 Spatial distribution of population in Shandong Province in year 2000 with 100 m×100 m grids

图4所示,SDpop2000和WorldPop的局部对比如图5所示。可以发现SDpop2000和WorldPop有相似的人口分布趋势,但SDpop2000在城市地区较WorldPop有更详细的人口分布信息,且局部区域的人口密度比WorldPop高,如图5中的济南市和青岛市;WorldPop在鲁中部、泰安西南部、济宁南部、临沂南部和枣庄北部存在大面积连续区域人口密度相同的错误分布现象,而SDpop2000在这些区域的人口分布主要集中在农村居民点处,人口分布符合实际情况,如图5展示的SDpop2000和WorldPop在

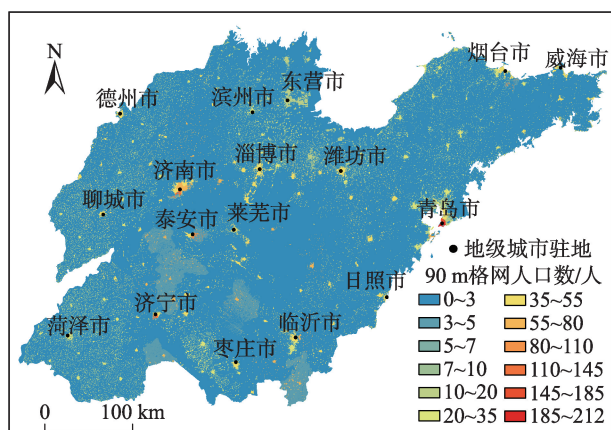


图4 山东省2000年WorldPop成果

Fig. 4 WorldPop of Shandong Province area in 2000

泰安西南部存在的人口分布差异。

(2) 本文分别统计 SDpop2000 和 WorldPop 在 10 km 网格内的人口数, 对两者进行相关性分析。如图 6 所示, SDpop2000 和 WorldPop 具有显著的线性关系, R^2 达到 0.866, 且两者之间的相关系数为 0.93, 达到较高的相关性。

(3) 通过式 (7) 计算 SDpop2000 和 WorldPop 的相对误差, 相对误差值域分布如图 7 所示, 其中相对误差绝对值在 50% 以内的 10 km 网格数占总网格数的 88%; 将相对误差分布到空间 (图 8), 发现 SD-

pop2000 的人口估计较 WorldPop 严重偏低的蓝色区域主要分布在鲁北沿海和济宁市以南, 严重偏高的红色区域分布较为分散, 其中滨州市、东营市、聊城市和烟台市等地级市驻地处的人口估计较 WorldPop 偏高, 其余地级市驻地处的人口估计同 WorldPop 相当。

4 结论与讨论

本文结合夜间灯光数据和土地利用数据在人口数据空间化应用中的优势, 采用城乡二级分区建模的方法, 利用 EVI 植被指数对 DMSP/OLS 夜间灯光数据进行了修正, 对山东省 2000 年乡镇级人口统计数据进行 100 m 格网的空间化。结果显示本文所建模型拟合度较好, 空间化结果在较细尺度下能反映研究区 2000 年的人口分布状况, SDpop2000 较 WorldPop 的人口分布更加精细, 空间差异性更加显著。具体分析如下:

(1) 乡镇级二级分区建模及检验。一级分区将城镇用地和农村居民点用地区分开, 实现了不同居民地类型采用不同模型的建模方法, 有效地避免了 DMSP/OLS 夜间灯光数据在低辐射亮度区域探测能力不足的缺点, 其中乡镇级行政单元的使用是实

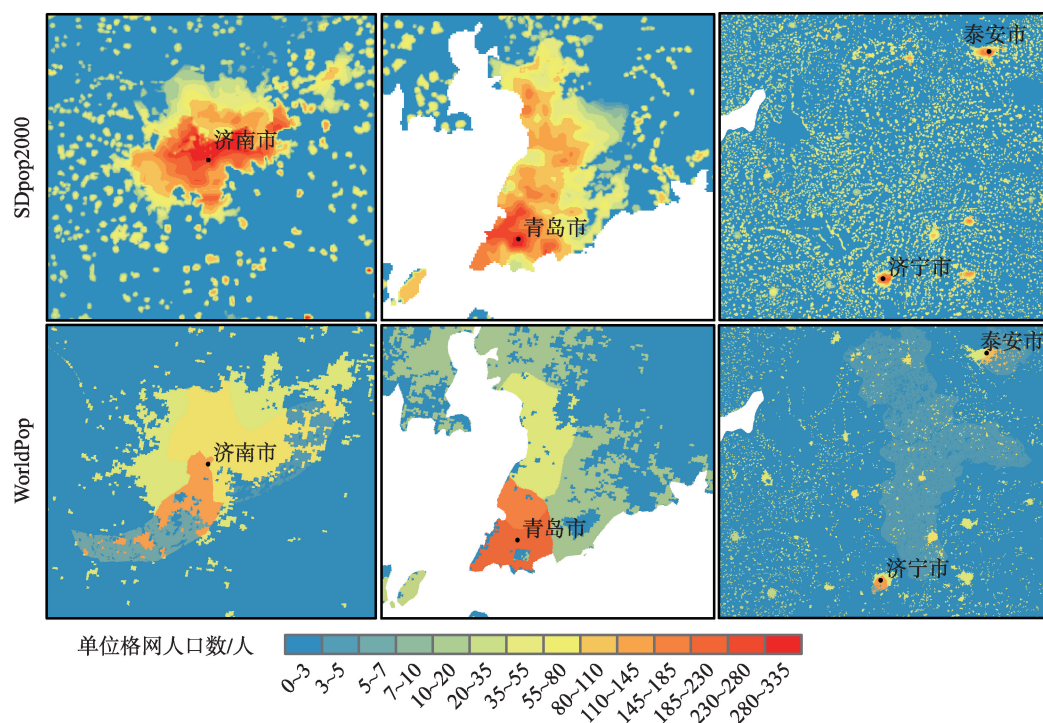


图5 SDpop2000和WorldPop局部区域对比

Fig. 5 Comparison of SDpop2000 and WorldPop in local areas

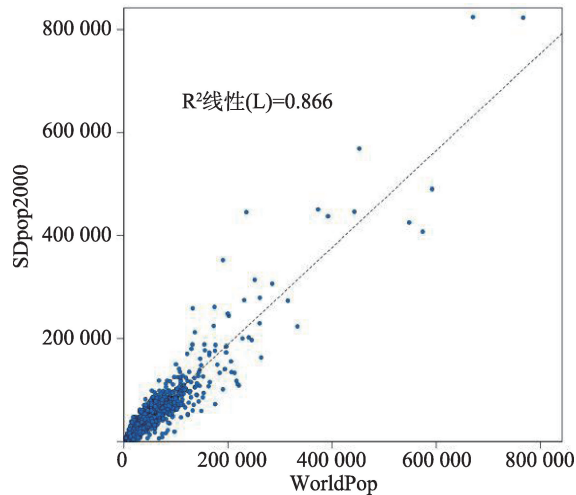


图6 10 km 网格内SDpop2000与WorldPop人口数散点图

Fig. 6 Scatter plots of population of SDpop2000 and WorldPop with 10 km×10 km grids

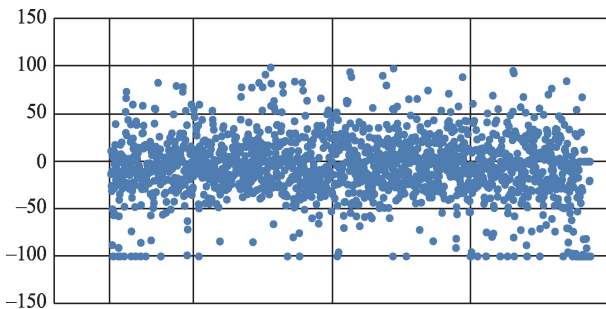


图7 10 km 格网内SDpop2000与WorldPop的人口相对误差散点图

Fig. 7 Scatter plots of population relative errors between SDpop2000 and WorldPop with 10 km×10 km grids

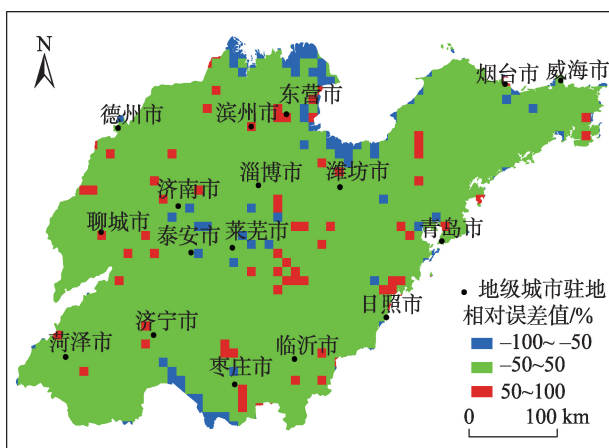


图8 10 km 格网内SDpop2000与WorldPop人口相对误差的空间分布

Fig. 8 Spatial distribution of population relative error between SDpop2000 and WorldPop with 10 km×10 km grids

现城乡分区的有利条件。二级分区将农村地区和其他地区进行了更细致的分区,通过调整农村地区二级分区比例,使得建模精度达到最优。本方法的模型精度利用其余地区的人口统计值进行检验,结果显示,有占比78%的行政单元的相对误差绝对值小于20%,且模拟人口同统计人口的相关系数达到0.931,说明本建模方法估计结果可靠。

(2)基于EVI指数修正的EANTLI夜间灯光数据处理。此方法由卓莉等^[22]提出,从城市内部灯光强度差异的可区分性、同辐射定标数据RCNTL的相似性、估算电能消耗等方面证实了EANTLI较植被调节型城市灯光指数(VANUI)更具有优越性。因此本文选择此方法对DMSP/OLS夜间灯光数据进行修正。本文在EVI数据的选择上有所不同:使用7、8月份EVI的均值而非年均值,是因为山东省的冬季基本没有绿色植被,年均值会降低植被覆盖程度的有效信息,从而影响到EANTLI的精度,而7、8月份是植被生长最茂盛的时候,可较好体现植被覆盖度的程度。经上文检验,城镇地区行政单元人口数与EANTLI夜间灯光辐射值之间有较高的相关性,且从SDpop2000和WorldPop的对比结果来看,SDpop2000在城镇地区比WorldPop有更详细的人口分布信息,且在城市中心SDpop2000的人口数量要比WorldPop高,说明EANTLI夜间灯光数据的使用,明显提高了城市中心人口分布的实际差异性。

(3)城镇用地和农村居民点用地处理。城镇用地指城市及县镇以上建成区用地,农村居民点指独立于城镇用地以外的农村居民点,本文将人口统计数据空间化到这两种土地利用类型。人口在农村地区的分布,除居民地上的人口具有一定的稳定状态外,其余土地利用类型上的人口均有一定的流动性,处于不稳定状态,很难通过分析行政单元尺度上的人口同土地利用类型之间的相关关系,得到网格尺度上这类土地利用类型上的人口分布概率密度。因此本文只考虑人口分布于居民点处的稳定状态,有效避免了SDpop2000在耕地、林地和草地等土地利用类型上出现大面积相同人口密度区域的偏差错误分布现象,提高了人口数据空间化的精度,如SDpop2000和WorldPop在鲁中部、泰安西南部、济宁南部、临沂南部和枣庄北部所表现出的人口分布差异。

(4)SDpop2000与WorldPop的对比分析。2000

年 WorldPop 数据被柏中强等^[9]利用乡镇级人口统计数据验证,结果显示其人口分布是目前可开放获取的人口分布产品中最为精确的,因此本文利用 WorldPop 和 SDpop2000 做对比分析。对比结果显示 SDpop2000 与 WorldPop 有相似的人口分布趋势,在异于行政区划尺度的 10 km 网格内,两者的人口数具有较高的相关性和较低的相对误差,表明 SDpop2000 同样有着较为准确的人口分布,而 SDpop2000 在城市、泰安西南部、济宁南部、临沂南部、枣庄北部、鲁中部和鲁北沿海等地的人口分布与 WorldPop 存在差异。在城市,SDpop2000 的城市中心和城市周边的人口密度差异性较大,SDpop2000 较 WorldPop 反映出的人口分布信息更丰富,更为准确地反映了中国的实际人口分布,如图 5 中的济南市和青岛市。图 4 中,WorldPop 在泰安西南部、济宁南部、临沂南部、枣庄北部和鲁中部等地存在着大面积相同人口密度区域,不符合实际人口分布,而 SDpop2000 在这些地方展现出更具差异性的人口空间分布信息(图 3)。图 8 显示 SDpop2000 和 WorldPop 在济宁南部、鲁北沿海和鲁中三地存在较大面积的人口分布差异,济宁南部和鲁北沿海的蓝色区域表示 SDpop2000 的人口分布密度比 WorldPop 低,鲁中莱芜市东南部的红色区域表示 SDpop2000 的人口分布密度比 WorldPop 高。其中济宁南部是山东省微山湖和独山湖等水域所在地,人口分布较为稀少;鲁北沿海蓝色区域靠近海洋,多为沙滩、生态旅游区、农田、渔港等,人口分布同样较为稀少;从图 4 可以看出,worldPop 在鲁中莱芜市东南部为大面积连续较低人口分布区域,由此可见 SDpop2000 在济宁南部、鲁北沿海和鲁中三地的人口分布较 WorldPop 更加准确。

(5) 本文基于 2000 年乡镇级人口统计数据进行空间化,主要有以下 3 点突破意义:① 乡镇级人口统计数据是我国目前公开的最精细的统计数据,但对于乡镇级人口统计数据的空间化研究较少。因为,这其中的人口统计数据和行政边界数据的空间匹配等需要使用乡镇界限,并且要做大量的空间数据处理工作。本文的研究工作丰富了乡镇级人口统计数据的空间化方法;② 2000 年是我国首次公布乡镇级统计数据的年份,以该年份开展研究可以为乡镇级人口数据空间化提供一个基准;③ 乡镇级的人口空间数据是不能直接发布的,因为乡镇界线具有保密限制。面对数据共享的需要,可以将该数

据按实际人口分布进行重新展布,形成空间化格网产品。这既提高产品的实际地理学含义,也解决了共享问题。

本文方法的不足与展望:① 本文的建模思路之一是将城镇地区和农村地区分开建模,由于缺少能够精确区分城镇地区和农村地区的矢量边界数据,以及与之对应的农村人口和城镇人口数据,本文暂利用土地利用类型区分城镇地区和农村地区,这样缩减了建模的样本数量,对建模精度有所影响。在今后的研究中,可以进一步思考和获取精确的城乡矢量边界,辅以统计数据中的农业人口和非农人口,对空间化模型进行改进;② 针对本文产出 100 m 格网化人口分布数据,比较理想的结果评估方法是利用村级人口实际统计数据对其精度进行评价。但目前还没有公开可获得的村级人口统计数据和村级行政边界数据,希望今后可以获取这一微观样本数据提高模型精度,并对 SDpop2000 做更加准确的精度评价。

参考文献(References):

- [1] 胡焕庸.论中国人口之分布[M].北京:科学出版社,1983. [Hu H Y. Distribution of China's population[M]. Beijing: Science Press, 1983.]
- [2] 廖顺宝,孙九林.基于 GIS 的青藏高原人口统计数据空间化[J].地理学报,2003,58(1):25-33. [Liao S B, Sun J L. GIS based spatialization of population census data in Qinghai-Tibet Plateau[J]. Acta Geographica Sinica, 2003,58 (1):25-33.]
- [3] Briggs D J, Gulliver J, Fecht D, et al. Dasymetric modeling of small-area population distribution using land cover and light emissions data[J]. Remote sensing of Environment, 2007,108(4):451-466.
- [4] Zeng C Q, Zhou Y, Wang S X, et al. Population spatialization in China based on night-time imagery and land use data[J]. International journal of remote sensing, 2011,32 (24):9599-9620.
- [5] 柏中强,王卷乐,杨飞.人口数据空间化研究综述[J].地理科学进展,2013,32(11):1692-1702. [Bai Z Q, Wang J L, Yang F. Research progress in spatialization of population data[J]. Progress in Geography, 2013,32(11):1692-1702.]
- [6] 陈晴,侯西勇.集成土地利用数据和夜间灯光数据优化人口空间化模型[J].地球信息科学学报,2015,17(11):1370-1377. [Chen Q, Hou X Y. An improved population spatialization model by combining land use data and DMSP/OLS data[J]. Journal of Geo-information Science, 2015,

- 17(11):1370-1377.]
- [7] Patel N N, Stevens F R, Huang Z, et al. Improving large area population mapping using Geotweet densities[J]. Transactions in Gis, 2017,21(2):317-331.
- [8] Bai Z Q, Wang J L, Wang M M, et al. Accuracy assessment of multi-source gridded population distribution datasets in China[J]. Sustainability, 2018,10.
- [9] 江东,杨小唤,王乃斌,等.基于RS、GIS的人口空间分布研究[J].地球科学进展,2002,17(5):734-738. [Jiang D, Yang X H, Wang N B, et al. Study on spatial distribution of population based on remote sensing and GIS[J]. Advance in Earth Sciences, 2002,17(5):734-738.]
- [10] Harvey J T. Population estimation models based on individual TM pixels[J]. Photogrammetric Engineering & Remote Sensing, 2002,68(11):1181-1192.
- [11] 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[J].地理学报,2005,60(2):266-276. [Zhuo L, Chen J, Shi P J, et al. Modeling population density of China in 1998 based on DMSP/OLS nighttime light image [J]. Acta Geographica Sinica, 2005,60(2):266-276.]
- [12] Azar D, Engstrom R, Graesser J, et al. Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data[J]. Remote Sensing of Environment, 2013,130:219-232.
- [13] 田永中,陈述彭,岳天祥,等.基于土地利用的中国人口密度模拟[J].地理学报,2004,59(2):283-292. [Tian Y Z, Chen S P, Yue T X, et al. Simulation of Chinese population density based on land use[J]. Acta Geographica Sinica, 2004,59(2):283-292.]
- [14] 柏中强,王卷乐,姜浩,等.基于多源信息的人口分布格网化方法研究[J].地球信息科学学报,2015,17(6):653-660. [Bai Z Q, Wang J L, Jiang H, et al. The gridding approach to redistribute population based on multi-source data[J]. Journal of Geo-information Science, 2015,17(6):653-660.]
- [15] 符海月,李满春,赵军,等.人口数据格网化模型研究进展综述[J].人文地理,2006,21(3):115-119. [Fu H Y, Li M C, Zhao J, et al. Summary of grid transformation models of population data[J]. Human Geography, 2006,21(3):115-119.]
- [16] Eicher C L, Brewer C A. Dasymetric mapping and areal interpolation: Implementation and evaluation[J]. American Cartographer, 2001,28(2):125-138.
- [17] 李翔,陈振杰,吴洁璇,等.基于夜间灯光数据和空间回归模型的城市常住人口格网化方法研究[J].地球信息科学学报,2017,19(10):1298-1305. [Li X, Chen Z J, Wu J X, et al. Gridding methods of city permanent population based on night light data and spatial regression models[J]. Journal of Geo-information Science, 2017,19(10):1298-1305.]
- [18] Elvidge C D, Baugh K E, Dietz J B, et al. Radiance calibration of DMSP-OLS low-light imaging data of human settlements[J]. Remote Sensing & Environment, 1999,68(1):77-88.
- [19] 高义,王辉,王培涛,等.基于人口普查与多源夜间灯光数据的海岸带人口空间化分析[J].资源科学,2013,35(12):2517-2523. [Gao Y, Wang H, Wang P T, et al. Analysis of population spatialization for Chinese coastal zones based on census and multiple night light data[J]. Resource Science, 2013,35(12):2518-2525.]
- [20] 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[J].冰川冻土,2012,34(4):999-1006. [Liang Y J, Xu Z M. Modeling the spatial distribution of population based on night light radiation and LUCC: A case study in Ganzhou District, Zhangye Municipality[J]. Journal of Glaciology and Geocryology, 2012,34(4):999-1006.]
- [21] 黄杰,闫庆武,刘永伟,等.基于DMSP/OLS与土地利用的江苏省人口数据空间化研究[J].长江流域资源与环境,2015,24(5):735-741. [Huang J, Yan Q W, Liu Y W, et al. Modeling the population density of Jiangsu Province based on DMSP/OLS satellite imagery and land use data [J]. Resources and Environment in the Yangtze Basin, 2015,24(5):735-741.]
- [22] 卓莉,张晓帆,郑璟,等.基于EVI指数的DMSP/OLS夜间灯光数据去饱和方法[J].地理学报,2015,70(8):1339-1350. [Zhuo L, Zhang X F, Zheng J, et al. An EVI-based method to reduce saturation of DMSP/OLS nighttime light data[J]. Acta Geographica Sinica, 2015,70(8):1339-1350.]
- [23] Elvidge C D, Baugh K E, Zhizhin M, et al. Why VIIRS data are superior to DMSP for mapping nighttime lights [J]. Proceedings of the Asia-Pacific Advanced Network, 2013,35:62-69.
- [24] 中华人民共和国民政部.中华人民共和国行政区划简册[M].北京:中国地图出版社,2001. [Ministry of Civil Affairs of the People's Republic of China. The administrative divisions of the People's Republic of China[M]. Beijing: China Cartographic Publishing House, 2001.]
- [25] 国家统计局人口和社会科技统计司.中国乡、镇、街道人口资料[M].北京:中国统计出版社,2002. [Department of Population Social Science and Technology Statistics National Bureau of Statistics of China. China population by township[M]. Beijing: China Statistics Press, 2002.]
- [26] National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information.

- Version 4 DMSP-OLS nighttime lights time series [DB/OL]. <https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>, 2018-08-12.
- [27] National Aeronautics and Space Administration (NASA) Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC). MOD13Q1 - MODIS/Terra vegetation indices 16-Day L3 global 250 m SIN Grid [DB/OL]. <https://ladsweb.modaps.eosdis.nasa.gov/api/v1/productPage/product=MOD13Q1>, 2018-08-12.
- [28] 国家科技基础条件平台—国家地球系统科学数据共享服务平台. 山东省 100 m 土地利用数据 (2000 年) [DB/OL]. <http://www.geodata.cn/data/datadetails.html?dataguid=276366180064167&docid=15360>, 2018-08-12. [National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China. Land use data of Shandong Province in 100 m (2000) [DB/OL]. <http://www.geodata.cn/data/datadetails.html?dataguid=276366180064167&docid=15360>, 2018-08-12.]
- [29] WorldPop. China population 2000 [DB/OL]. <https://www.worldpop.org/geodata/summary?id=1552>, 2018-08-12.
- [30] 国家科技基础条件平台—国家地球系统科学数据共享服务平台. 中国 1:25 万乡镇界限数据 (2000 年) [DB/OL]. <http://www.geodata.cn/data/datadetails.html?dataguid=23491717434844&docId=10968>, 2018-08-12. [National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China. China's 1:25 million township boundary data (2000) [DB/OL]. <http://www.geodata.cn/data/datadetails.html?dataguid=23491717434844&docId=10968>, 2018-08-12.]