

集成土地利用数据和夜间灯光数据优化人口空间化模型

陈 晴^{1,2}, 侯西勇^{1*}

(1. 中国科学院烟台海岸带研究所, 烟台 264003 2. 中国科学院大学, 北京 100049)

摘要 :人口统计数据空间化是解决统计数据与自然要素数据融合分析的有效途径。随着RS和GIS技术的发展,人口统计数据空间化方法推陈出新,其中,土地利用数据、夜间灯光数据是人口空间化研究中普遍利用的数据源,但各有优、缺点:土地利用数据中的城镇用地、农村居民点能准确表示人口分布的空间范围,却不能反映其内部的人口密度差异特征;夜间灯光数据的强度信息能体现人口分布的疏密程度,但其像元溢出问题显著夸大人口分布范围,像元过饱和现象也影响着人口数据空间化结果的精度。本研究以中国大陆沿海区域为例,尝试集成土地利用数据和夜间灯光数据优化人口空间化方法,设计了基于精度阈值和动态样本的渐进回归与分区建模的方法,获得了中国沿海2000、2005、2010年1 km分辨率人口空间化数据。结果表明,优化模型显著提高了研究区整体的精度,尤其适用于人口空间结构内部差异较为显著的区域。

关键词 :夜间灯光数据;土地利用数据;人口空间化;中国大陆沿海区域

DOI:10.3724/SP.J.1047.2015.01370

1 引言

人口问题是当今影响可持续发展的重大问题,将人口数据与资源、环境、生态、社会等方面的数据综合集成,已成为国内外多学科研究的基本模式和发展趋势^[1]。我国人口数据一般是以基层行政区为基本单元,通过普查、抽样统计等方式,经逐级汇总形成二维表格,具有权威、系统、规范等特点^[2],但这种统计方式数据空间分辨能力低,在实际应用中存在不足。人口数据空间化是将以行政区域为单元的人口统计数据按照一定的原则,采用某种技术手段合理地分配到统计单元内部一定尺寸的规则格网上的过程,旨在将人口数据与自然生态数据耦合使用,满足多种空间分析、统计和建模的需求^[3]。应用较为普遍的空间化方法有空间插值模型、土地利用/覆被影响模型、多源数据融合模型及遥感反演等^[4]。其中,空间插值模型难以考虑地理环境因素

对人口分布的影响,仅适用于小范围^[5];土地利用/覆被影响模型着重探究人口分布与不同用地类型的关系,对同种用地内部的空间异质性考虑不足^[6];多源数据融合模型综合土地利用、高程、交通网络等多种环境因子,分析各个因子影响人口分布的权重,建立空间化的统计模型,但面临各因子之间存在关联、因子权重难以判定的挑战^[7-9];遥感反演法主要是提取遥感影像中与人口分布有关的信息,进而建立反演模型模拟人口的空间分布^[10],较成熟的方法是利用夜间灯光数据实现人口数据的空间化^[11-12],但夜间灯光数据存在像元过饱和、像元溢出现象,对其应用有不利的影响^[13-15]。鉴于上述各种空间化方法的优劣,集成2种或多种空间化方法建立更精确的人口空间化模型,已成为当前研究的重要趋势之一^[16-18]。

中国的人口空间化研究大多基于全国或地区尺度^[11, 19-20],针对沿海区域的研究相对较少。沿海

收稿日期 2015-01-09;修回日期:2015-03-14.

基金项目 中国科学院战略性先导科技专项“应对气候变化的碳收支认证及相关问题”(XDA05130703);国家自然科学基金项目(31461143032);中国科学院烟台海岸带研究所“一三五”项目(Y254021031)。

作者简介 陈 晴(1989-),女,湖南长沙人,硕士生,主要从事海岸带地理信息系统研究。E-mail: qingchen@yic.ac.cn

*通讯作者 侯西勇(1975-),男,山东泰安人,研究员,博士生导师,主要从事海岸带土地利用变化监测、海岸线变化监测、海岸带脆弱性评价、海岸带综合管理等研究。E-mail: xyhou@yic.ac.cn

是人口、城市、经济最密集,人流、物流、资金流、信息流最活跃的地带,同时又是生态环境脆弱带(EC-OTONE)。海陆气的相互耦合作用、全球气候变化及人类活动的影响,导致沿海区域生态环境的脆弱化,使之成为地球表层多种自然灾害频发的地带^[21]。因此,获取空间分辨率更高的沿海人口分布数据,不仅是灾害风险管理的基础,也是提升和促进海岸带综合管理能力的重要前提^[22]。鉴于中国沿海人口空间分布差异显著、不同空间化方法均存在一定不足,本文旨在通过集成土地利用数据和夜间灯光数据,改进人口空间化建模方法,建立中国大陆沿海区域更加准确的1 km栅格人口空间分布数据。

2 研究区背景与数据源

(1)中国海岸线北起鸭绿江口,南抵北仑河口,海岸带跨14个省、市、区。本文选取中国大陆沿海地级市辖区为研究区,某些地级市虽不靠海,但辖区离海域很近,也纳入研究区范围(图1),因此,共包括62个地级市和直辖市的市区、101个县级市和138个县,合计301个统计单元。该区域平均海拔低于500 m,处于中国地势的第三级阶梯,总面积为48.37万km²,占全国国土面积的5%。该区域人口

增长迅速,由2000年的2.72亿增长到2010年的3.18亿,净增加4856万,增长率为16.85%;占全国人口总数的比例由21.01%上升至23.20%。沿海区域人口分布及变化的空间差异显著,大量人口集中在京津冀都市圈、长江三角洲和珠江三角洲地区^[23],且集中的趋势在不断加剧。

(2)研究数据主要包括2000、2005和2010年的人口数据、土地利用与夜间灯光数据等。

① 2000、2005、2010年中国大陆沿海地级市范围人口信息,分别来源于“2000年人口普查分县资料”、“2005年中国全国分县市人口统计资料”,以及“2010年人口普查分县资料”。

② 2000、2005、2010年中国沿海土地利用矢量数据(1:10万),由Landsat TM/ETM+影像解译得到最小图斑覆盖4~8个像元,分类精度达95%^[24],分类系统包含8个一级类型、24个二级类型。

③ 2000、2005、2010年夜间灯光数据,来源于美国国家地理数据中心(<http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>),选用2010年发布的第4版DMSP-OLS夜间稳定灯光值数据,已经滤除闪电、天然气燃烧、火光和渔船等偶然灯光。数据值域是0~63,0为黑暗背景区域,大于0的区域为灯光区域。

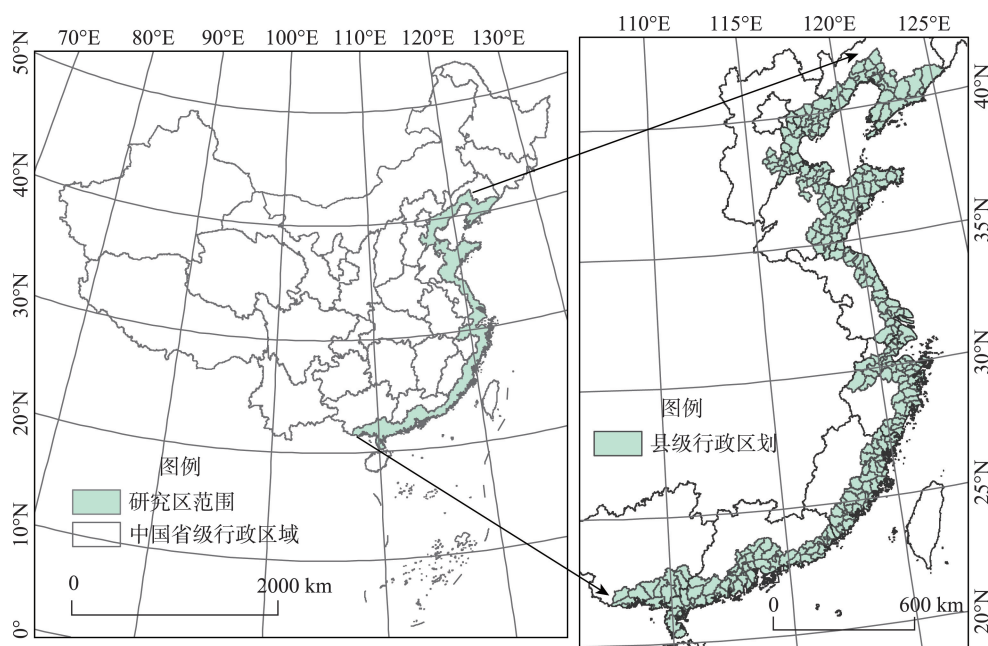


图1 研究区位置与范围

Fig. 1 Location of China's coastal area

3 优化人口空间化方法

3.1 基于土地利用数据提取人口聚居区的夜间灯光指数

基于夜间灯光数据建立人口空间化模型,通常是确定某个阈值来提取夜间灯光数据中的人口聚居区。但是,对于较大范围的空间区域(如本文的中国大陆沿海区域),由于区域内部的空间差异性,统一的阈值难以确定,也并不合理。土地利用的空间格局能客观反映人口的空间分布与活动特征^[25],绝大部分人口聚居在一定类型的用地区域。因此,基于中国沿海土地利用数据,提取城镇用地和农村居民点类型作为人口聚居区,用其分割夜间灯光指数数据,可避免阈值法的不确定性,保证人口聚居区的准确性。具体过程为:建立覆盖研究区的1 km 矢量格网,将土地利用矢量数据与其叠加,统计每一格网单元中城镇用地及农村居民点的分布面积,得到1 km 格网的城镇用地(A_u)和农村居民点(A_r)2类土地利用类型的面积分布数据,计算二者之和(A_u+A_r),得到能表征人口空间分布范围的1 km 格网人口聚居区栅格数据;进而,依据式(1)提取人口聚居区的夜间灯光指数。

$$L'_{\text{night}} = \begin{cases} L_{\text{night}} & (A_u + A_r > 0) \\ 0 & (A_u + A_r = 0) \end{cases} \quad (1)$$

式中, L'_{night} 表示修订后的人口聚居区夜间灯光指数; L_{night} 表示修订前的夜间灯光指数。

3.2 基于修订的夜间灯光数据在县域统计单元回归建模

大量研究表明,人口数量与夜间灯光指数具有较好的相关性^[15, 17, 26]。本研究利用SPSS软件分析沿海区域各县级行政单元人口聚居区的夜间灯光指数值与统计人口的相关性分析表明,一次、二次线性函数及幂函数能较好地体现它们的相关性,但3种建模方法在人口密度较高或较低的区域模拟效果均欠佳。试验表明,通过动态分区,对人口密度较高或较低的区域单独分析人口与灯光数据的关系,线性模型效果更佳。所以,本研究采用线性函数建立空间化模型(式2)。

$$\text{pop}_c = k \times \sum L'_{\text{night}} + b \quad (2)$$

式中, pop_c 是建模区域县级行政单元的人口数; k 是待定的回归系数; $\sum L'_{\text{night}}$ 是对应的县级行政单元内人口聚居区的灯光指数值之和; b 为常数

项,考虑到 k 值非负,而且沿海区域存在无人口分布的聚居区,如空心村^[27]、鬼城,因此,回归过程中将 b 设为0(空心村或鬼城的灯光指数为0,将 b 设为0,则无论 k 为多少,模拟的人口总为0,与现实相符)。

3.3 基于精度阈值和动态样本的渐进回归和分区建模

基于县级行政单元的数据样本,计算出回归系数 k ,将 k 应用于1 km分辨率人口聚居区的夜间灯光数据(式3),得到1 km分辨率的人口空间分布数据。

$$\text{pop}_{1\text{ km}} = k \times L'_{\text{night}} \quad (3)$$

式中, $\text{pop}_{1\text{ km}}$ 表示1 km格网的人口分布模拟数据; L'_{night} 表示人口聚居区的夜间灯光指数。

中国大陆沿海区域人口密度差异较大,例如,2010年长江三角洲、珠江三角洲人口密度均高于1000人/km²,而辽宁沿海、北部湾区域人口密度则低于500人/km²。试验表明,这种差异性显著影响线性回归模型的精度:在人口密集的区域,人口密度被低估,而人口稀疏的区域,人口密度则被高估^[28-29]。因此,本研究设计了基于精度阈值和动态样本的渐进回归和分区建模方法,以提高区域整体的精度:针对初始模拟结果,筛选县域尺度相对误差(式(4))的绝对值较小的县域(阈值设定为 $|\delta| < 20\%$)保留初始模拟结果;而剩余误差较大($|\delta| \geq 20\%$)的县域数量,如超出总样本的10%(本文中约为30),则将这些县域组成新样本再次进行回归分析,得到新的方程,按照同样的阈值进行判断,可实现更进一步的分区;经过多次反复,将获得一组方程,能保证90%以上的县域满足 $|\delta| < 20\%$,而剩余不足10%的县域则保留最后一次回归所得的模型,如此实现了县域层面整个研究区的动态分区建模;将1 km分辨率的夜间灯光数据代入这些分区模型,计算得到1 km尺度人口空间化栅格数据。

$$\delta = \frac{\text{pop}_{\text{mo}} - \text{pop}_{\text{sta}}}{\text{pop}_{\text{sta}}} \times 100\% \quad (4)$$

式中, pop_{mo} 表示某县模拟的空间化人口数据; pop_{sta} 表示某县统计的人口数据。

上述方法的技术路线如图2所示。

4 人口空间化模拟结果及分析

4.1 人口空间化模型

2000、2005、2010年沿海区域县级单元统计人口及夜间灯光数据建立的初次回归方程的确定性系数均大于0.8,回归模型较为可靠。但是,随着县

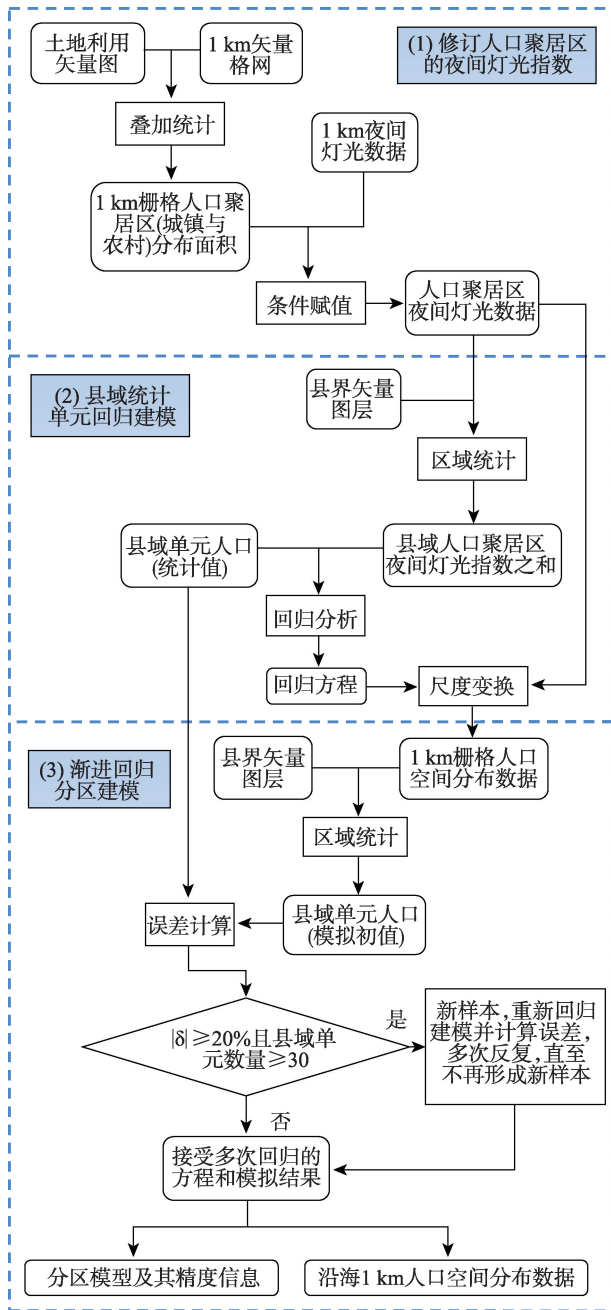


图2 人口空间化技术路线图

Fig. 2 Flowchart of population spatialization

级单元之间人口规模差异的增大,确定性系数逐渐减小。根据初次建模结果县级单元模拟人口数的相对误差(δ),将研究区分成2类区域:保留区域($-20\% < \delta < 20\%$)、需重组样本再次建模区域($\delta \geq 20\%$ 或 $\delta \leq -20\%$),后者区分 $\delta \geq 20\%$ 或 $\delta \leq -20\%$ 分别重新回归建模,随着建模次数的增加,回归方程的确定性系数总体不断提高,趋近于1,表明回归方程拟合精度越来越高(表1)。

中国大陆沿海区域2000、2005、2010年人口空

间化最终结果如图3所示。总体而言,10 a间人口分布的总体格局稳定少变:人口密度较低的区域包括辽宁省的丹东市、盘锦市、葫芦岛市,浙江-福建内陆一侧,粤西及广西沿海区域;人口密度较高的区域包括上海市及紧邻的苏南、浙北地区,浙南的台州市和温州市靠海的区域,闽东南的福州-厦门沿海一带,粤东的潮州市至汕尾市大部分区域,粤中临海城市等。但是,随着经济的快速增长,资源、环境等条件更好的区域对人口产生了较大吸引力,带动了自身及周边地区人口的增长,使得人口密集的区域逐渐蔓延(图4),这种区域尺度较为细微的人口分布变化特征是基于行政区划的“面板”人口统计资料难以体现的。

4.2 精度分析

按照动态样本分区建模,90%以上的县级单元模拟人口的 $|\delta|$ 应低于20%。统计表明(表2),2000、2005、2010年模拟人口的 $|\delta|$ 小于20%的县级单元的数量比例分别高达94%、92%和93%;并且 $|\delta|$ 的标准差较小,表明其分布较集中,模拟的整体精度较高。

另外,对比初次建模及最终结果的模拟误差(图5),可显示出动态样本分区建模在精度改善方面的效果:初次建模的相对误差总体偏大且较为离散,而通过动态样本分区和渐进回归则显著提高了模拟的精度。

5 结论

本研究针对中国大陆沿海区域,集成土地利用数据和夜间灯光数据,采取精度阈值和动态样本控制的渐进回归与分区建模方法,显著提高了人口空间化模型的整体精度,获得了中国大陆沿海2000、2005和2010年高质量的1 km栅格人口空间分布数据。研究表明:

(1) 夜间灯光数据能反映居民点、交通道路、产业结构等多种信息,是进行人口数据空间化建模的理想数据源之一。但是,夜间灯光数据具有像元溢出的特点,多数学者采取先设定阈值的方法,以便提取人口分布的空间范围,即人口聚居区,进而选择聚居区的夜间灯光数据,与人口统计数据做相关分析,缓和像元溢出对空间化模型带来的影响,并最终建立人口空间化模型^[18]。本研究则直接利用高精度的土地利用数据提取城镇用地、农村居民点2类用地作为人口分布的聚居区,可更有效地解决

表1 人口空间化线性回归模型及确定性系数
Tab. 1 Linear models and R² for population spatialization

年份	建模区域	确定性系数(R ²)	回归系数(k)	模型精度及动态分区情况		
				$ \delta < 20\%$	$\delta \leq -20\%$	$\delta \geq 20\%$
2000	所有县级单元	0.906	97.753	保留结果	A1	C1
	A1	0.920	119.871	保留结果	A2	不存在
	A2	0.914	210.124	保留结果	A3	不存在
	A3	0.916	341.578	保留结果	A4	不存在
	A4	0.932	480.864	保留结果	A5	不存在
	A5	0.934	675.961	保留结果	样本少,保留结果	样本少,保留结果
	C1	0.975	69.462	保留结果	不存在	C2
	C2	0.983	53.194	保留结果	不存在	样本少,保留结果
2005	所有县级单元	0.847	85.079	保留结果	A1	C1
	A1	0.916	120.028	保留结果	A11	A12
	A11	0.910	543.291	保留结果	A13	不存在
	A12	0.953	343.717	保留结果	不存在	不存在
	A13	0.928	771.784	保留结果	A14	不存在
	A14	0.967	1112.758	保留结果	样本少,保留结果	不存在
	C1	0.941	61.178	保留结果	不存在	C2
	C2	0.975	42.487	保留结果	不存在	样本少,保留结果
2010	所有县级单元	0.829	75.130	保留结果	A1	C1
	A1	0.954	107.403	保留结果	A2	不存在
	A2	0.968	145.739	保留结果	A3	不存在
	A3	0.939	195.424	保留结果	A4	不存在
	A4	0.943	265.406	保留结果	样本少,保留结果	不存在
	C1	0.917	50.343	保留结果	不存在	C2
	C2	0.973	37.562	保留结果	不存在	C3
	C3	0.990	33.162	保留结果	不存在	样本少,保留结果

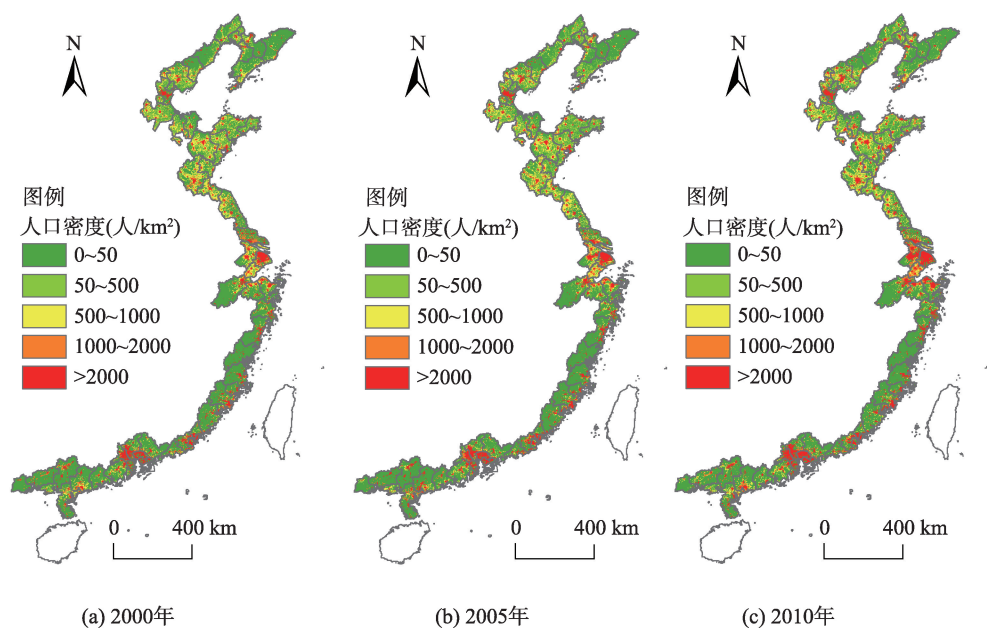


图3 中国大陆沿海人口空间化分布图
Fig. 3 Population spatialization map in China's coastal area

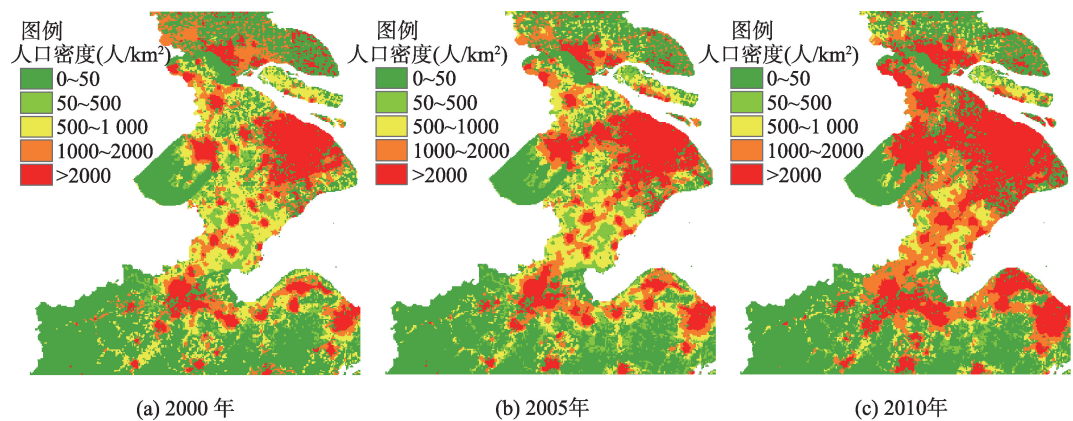


图4 2000、2005、2010年长江三角洲人口空间分布对比图

Fig. 4 Comparison of population spatialization maps for Yangtze River Delta in 2000, 2005 and 2010

表2 模拟人口的 $|\delta|$ 分布区间统计表

Tab. 2 Statistics for interval distribution of $|\delta|$

年份	不同 $ \delta $ 分布区间对应的县级行政单元(个数)				$ \delta $ 的标准差
	0-0.1	0.1-0.2	0.2-0.4	> 0.4	
2000	160	122	11	8	0.17
2005	165	111	16	9	0.15
2010	156	125	11	9	0.17

像元溢出带来的影响,排除了非人口聚居区的夜间灯光指数值对模型的误差。

(2) 夜间灯光数据的像元过饱和特征显著影响人口密度较高区域的建模精度,例如,不同地区夜间灯光数据的最大值相差无几,都对应当地人口最密集的区域,但不同地区之间相比,人口密度的最大值相差悬殊,夜间灯光数据并不能反映这种实际差异;此外,中国大陆沿海区域人口分布的空间异质性极为显著,包括南北差异、海陆梯度及城乡差异等,单次建模的结果存在大量被显著高估或低估的区域,尤其是人口密度处于“两极”的县域,误差非常大,即便进一步基于县域单元总人口进行栅格数据像元尺度的线性调整,也不能进一步改善人口空间展布的特征。因此,本研究依据精度阈值确定动态样本和分区建模,将前次模拟误差较大的县域重组样本再次建模,实质是对人口密度格局差别较大的空间区域进行分区和单独建模,将人口密度特征相近的区域划分在一起,成为研究区的子区域,各个子区域采用同一个空间化模型,但不同的子区域空间化模型不同。这种建模方法的过程相对简单,并且很好地解决了上述空间异质性的问题。但在农村区域,尤其是山区,夜间灯

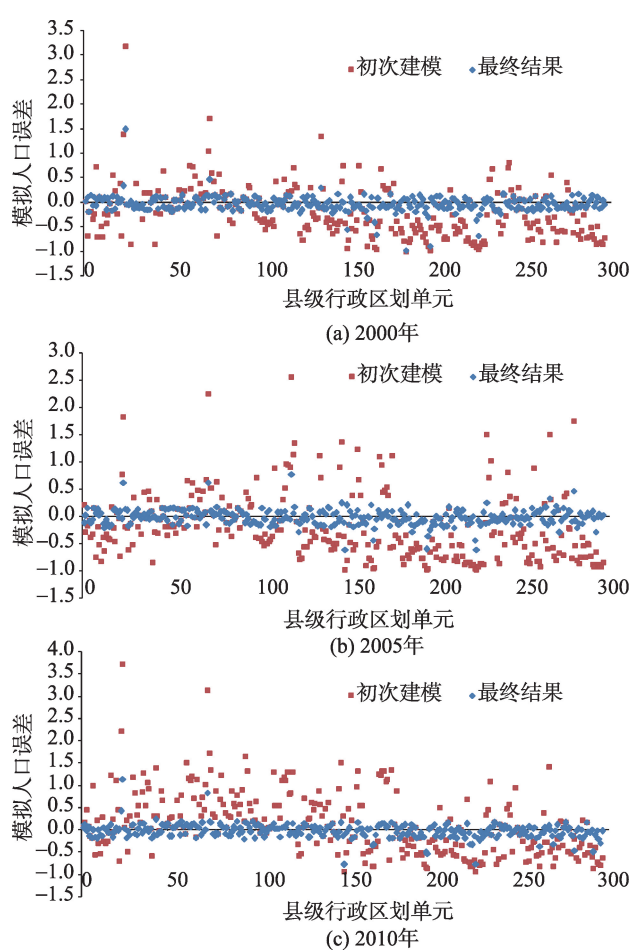


图5 模拟人口相对误差的散点图

Fig. 5 Scatter plots of relative errors of modeled population in the research area

光暗淡,本文的动态分区建模方法能在一定程度上缓解灯光不足带来的模拟误差,但是夜间灯光数据本身难以较好地反映山区人口的分布,这将对模拟精度有一定的影响。

参考文献:

- [1] 林丽洁,林广发,颜小霞,等.人口统计数据空间化模型综述[J].亚热带资源与环境学报,2010,5(4):10-16.
- [2] 胡云锋,王倩倩,刘越,等.国家尺度社会经济数据格网化原理和方法[J].地球信息科学学报,2011,13(5):573-578.
- [3] 闫庆武,卞正富.基于GIS的社会统计数据空间化处理方法[J].云南地理环境研究,2007,19(2):92-97.
- [4] 李飞,张树文,杨久春,等.社会经济数据空间化研究进展[J].地理与地理信息科学,2014,30(4):102-107.
- [5] 丁文秀,赵伟,左德霖,等.基于土地利用分类模型和重力模型耦合的人口分布模拟——以武汉市人口数据为例[J].大地测量与地球动力学,2011,31(增刊):127-131.
- [6] 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.
- [7] 刘梦鑫,邬群勇,鲁毅敏.加权人口密度连续分布模拟模型[J].地球信息科学学报,2014,16(2):199-206.
- [8] 杨妮,吴良林,郑士科.基于GIS的县域人口统计数据空间化方法[J].地理空间信息,2013,11(5):74-77.
- [9] Yang X H, Huang Y H, Dong P L, *et al.* An updating system for the gridded population database of China based on remote sensing, GIS and spatial database technologies[J]. *Sensors (Basel)*, 2009,9(2):1128-1140.
- [10] Azar D, Graesser J, Engstrom R, *et al.* Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti[J]. *International Journal of Remote Sensing*, 2010,31(21):5635-5655.
- [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.
- [12] 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.
- [13] 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[J].人文地理,2014,29(5):94-100.
- [14] 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.
- [15] Townsend A C, Bruce D A. The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution[J]. *International Journal of Remote Sensing*, 2010,31(16):4459-4480.
- [16] 张建辰,王艳慧.基于土地利用类型的村级人口空间分布模拟——以湖北鹤峰县为例[J].地球信息科学学报,2014,16(3):435-442.
- [17] 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.
- [18] 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[J].遥感信息,2009(1):83-87.
- [19] 凡宸,秦建桥,高洋,等.基于GIS数据的县域人口密度空间降尺度模拟——以广东省惠东县为例[J].中国人口·资源与环境,2014,24(3):26-230.
- [20] 叶靖,杨小唤,江东.人口数据空间化的格网尺度效应分析[J].地球信息科学学报,2010,12(1):1-8.
- [21] 朱晓东,李杨帆,桂峰.我国海岸带灾害成因分析及减灾对策[J].自然灾害学报,2001,10(4):26-29.
- [22] 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.
- [23] 牛海燕.中国沿海台风灾害风险评估研究[D].上海:华东师范大学,2012.
- [24] Di X H, Hou X Y, Wang Y D, *et al.* Spatial-temporal characteristics of land use intensity of coastal zone in China during 2000-2010[J]. *Chinese Geographical Science*, doi: 10.1007/s11769-014-0707-0.
- [25] 江东,杨小唤,王乃斌,等.基于RS、GIS的人口空间分布研究[J].地理科学进展,2002,17(5):734-738.
- [26] Small C, Elvidge C D. Night on Earth: Mapping decadal changes of anthropogenic night light in Asia[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2013,22: 40-52.
- [27] 刘彦随,刘玉.中国农村空心化问题研究的进展与展望[J].地理研究,2010,29(1):35-41.
- [28] Harvey J T. Estimating census district populations from satellite imagery: Some approaches and limitations[J]. *International Journal of Remote Sensing*, 2010,23(10):2071-2095.
- [29] Liu Q, Sutton P C, Elvidge C D. Relationships between nighttime imagery and population density for Hong Kong[J]. *Proceedings of the Asia-Pacific Advanced Network*, 2011,31:79-90.

An Improved Population Spatialization Model by Combining Land Use Data and DMSP/OLS Data

CHEN Qing^{1,2} and HOU Xiyong^{1*}

(1. *Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, 264003, China;*

2. *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Population spatialization could effectively remove the obstacles for data fusion between census data and geographic data. With the rapid development of RS and GIS techniques, the research methods of population spatialization have been updated greatly, and among them, both land use data and DMSP/OLS data are the most widely used data sources for population spatialization. However, both of them have advantages and disadvantages, in specific, the patches of cities and rural settlements in land use data indicate the area of population distribution accurately, but land use data lacks detailed features of population distribution, especially for that existing in the same type of land use; DMSP/OLS data manifests the spatial variations of population density, however it often overestimates the distribution area of population due to its 'overglow' effect, and at the same time, the problem of pixel saturation in DMSP/OLS data also impairs the fitness of this data. In this paper, land use data and DMSP/OLS data are combined together to distinguish the night light data value of populated area, and the methodologies of population spatialization are improved by introducing the precision threshold and dynamic regionalization method. Census data in 2000, 2005 and 2010 for China's coastal area are taken as examples to test the outcomes of the improved methodologies for population spatialization. The results show that: (1) due to the prominent spatial heterogeneity of population distribution in China's coastal area, unitary model based on DMSP/OLS data for the whole study area exhibits very poor precision, therefore, land use maps are utilized to distinguish the populated and non-populated area based on DMSP/OLS images. Compared with the threshold method, land use maps more effectively removes the 'overglow' effect of DMSP/OLS data; (2) precision threshold of the regression model is adopted to dynamically divide the whole study area into several sub-regions, in specific, only counties that meet the testing accuracy defined by the precision threshold could retain in the regression model, otherwise, they should be regrouped into a new set of samples and be fitted by a new regression model. It is named as 'dynamic regionalization' method in this paper. The results show that this method further improves the overall accuracies of population spatialization data.

Key words: DMSP/OLS data; land use data; population spatialization; China's coastal area

*Corresponding author: HOU Xiyong, E-mail: xyhou@yic.ac.cn