

引用格式:Guo Yuchen, Huang Jinchuan, Lin Haoxi. Spatialization of China's Population Data based on Multi-source Data[J]. Remote Sensing Technology and Application, 2020, 35(1): 219-232. [郭雨臣, 黄金川, 林浩曦. 多源数据融合的中国人口数据空间化研究[J]. 遥感技术与应用, 2020, 35(1): 219-232.]
doi:10.11873/j.issn.1004-0323.2020.1.0219

多源数据融合的中国人口数据空间化研究

郭雨臣^{1,2,3}, 黄金川^{1,2,3}, 林浩曦^{1,2,3}

- (1. 中国科学院地理科学与资源研究所, 北京 100101;
2. 中国科学院区域可持续发展分析与模拟重点实验室, 北京 100101;
3. 中国科学院大学资源与环境学院, 北京 100049)

摘要:人口数据通常是以行政区为单元进行统计,在与其他空间数据进行综合分析时往往存在尺度不一等诸多问题。人口数据空间化研究是探究人口空间分布规律的过程,其最终成果可为人口分布与其他数据的空间关联研究奠定基础,也可直接应用于灾害影响评估等领域,具有重要的学术意义和应用价值。基于人口分布和土地覆盖、交通路网数据的空间关系构建人口空间化模型,将人口划分为城镇人口和农村人口,通过相关权重法(Relative Weight)对土地覆盖类型赋权,利用交通路网数据修正,对农村人口进行空间展布;分析不同等级交通路网数据影响方式的不同,对城镇人口进行展布,最后通过规划求解等方法将不同类型数据叠加融合。人口数据空间化结果表明:我国人口分布对海拔地形十分敏感,低海拔平原和台地面积仅占我国国土面积的16.49%,而生活在其上的人口却占总人口的54.88%,相反,我国山地丘陵面积占比为55.78%,但是仅分布22.11%的人口;东部人口集中于黄淮海地区、四川盆地、长江中小游以及东北平原等地;中国城镇人口呈现典型的“点轴”分布特点,沿主要交通线的集中分布比较明显。

关键词:人口数据空间化;土地覆盖;交通路网;相关权重法

中图分类号:TP75 **文献标志码:**A **文章编号:**1004-0323(2020)01-0219-14

1 引言

人口数据空间化,是采用一定的计算方法、算法或基于对人口分布影响因素的分析,对人口统计数据进行离散化处理,探究并运用人口的空间分布规律,最终模拟或再现客观世界的人口地理分布^[1]。当前,人口数据通常是以行政区为单元,通过普查或抽样调查,逐级统计汇总得到,反映某一行政区人口的总和。这样的人口数据最小单位为人口普查小区,普遍以县市为单位发布,其在学术研究中存在以下问题:①统计数据为行政区总体,制图表达时人口数据在行政单元内呈现均匀分布,不能体现人口数据的空间分布特征;②在研究中,研

究区域往往不与行政区划界线一致,以行政区为统计单位的数据难以分割变更统计尺度,增加了数据分析的难度和工作量^[2-3];③不利于与自然或社会空间数据融合,基于行政单元的人口数据无法体现小尺度的空间分异,难以与空间属性明确的栅格或矢量格式自然和社会数据进行融合分析。

区域人口与自然资源环境的关联研究是经济社会研究的趋势之一,鉴于人口统计数据在实际研究和使用中存在着空间单元内均一化、数据结构不一致和空间单元不匹配等问题^[4],要实现人口与资源环境等信息的综合分析,首先要解决的就是人口分布的空间化问题^[2]。人口数据空间化对于学术研

收稿日期:2018-10-03;修订日期:2019-12-15

基金项目:中国科学院战略性先导科技专项(A类)“大数据驱动的美丽中国全景评价与决策支持”课题(XDA19040501)资助。

第一作者:郭雨臣(1994—),男,天津人,硕士研究生,主要从事城市地理与区域规划研究。E-mail: guoyc.17s@igsrr.ac.cn。

通讯作者:黄金川(1973—),男,河南开封人,副研究员,主要从事城市地理与区域规划研究。E-mail: huangjc@igsrr.ac.cn。

究和社会生活都有重要的意义,空间化方法研究是探究和验证人口数据的空间分布规律的过程,空间格网化的人口数据成果可以很好地与土地覆盖、生态环境和气象气候等数据进行空间匹配,以进行交互作用与影响分析^[5-6],在生态环境保护、灾害风险评估与救援、商业决策、区域规划与开发等领域具有广泛的应用价值^[1],可以为行政管理、环境规划等工作提供重要的决策支持^[7]。

人口数据空间化方法研究经历了从格网化到空间化的演变,从单纯的、静态的格网化方法,逐步向自然、经济社会因素综合影响下的空间模型过渡,朝着动态模型方向发展^[2]。早期人口数据格网化只是按一定的数学模型将以行政区为单元的人口统计数据展布到一定尺寸的格网中,在方法上,主要是空间内插法、人口分布模型法。这些方法只是纯数学方法,并没有考虑实际人口分布。随着遥感技术发展,学者开始探究 DEM、NDVI、土地覆盖、夜间灯光和道路路网等空间数据与人口分布的关系,实现在一定空间尺度的栅格单元分配人口数据^[8]。随着网络大数据技术的发展,开始有学者使用手机信令数据^[9]、社交网络登录数据、地图 POI 数据^[10]等新型数据源开展人口空间化研究。这些数据在人口动态分布研究上具有巨大优势,适用于人口职住移动、人口迁移分析。但是在宏观尺度上,这些数据还有些局限:手机信令数据等普遍应用在市域范围研究,全国范围数据量过于庞大;网络大数据为有偏样本,全国范围经济差异巨大,这些数据对经济落后地区估计有较大偏差。

目前在国际上应用较多的人口栅格数据有 GPW 和 LandScan 人口数据。GPW (V1, V2) 简单地将统计人口平均分布于栅格上而未考虑影响人口分布因素^[11];LandScan 根据由道路接近度、坡度、土地覆盖和夜间灯光数据计算的概率系数将省级行政单位的人口分配到 1 km 栅格上^[12]。国内研究,田永中采用“分城乡、分区建模”的原则,对农村人口采用线性加权模型进行模拟,采用逐步回归计算各指标的回归系数,对城市数据采用人口分布模型进行空间展布,构建了 1 km 人口栅格数据^[13]。廖顺宝研究了西藏、青海 2 省区,发现人口密度与海拔高度、土地覆盖、主要道路有较强的相关关系,河流水系对居民点分布的影响也较为明显,以 GIS 软件为工具,通过客观方法赋予各个影响因子人口分布影响权重,运用多源数据融合技术进行了人口统计数

据的空间化^[14]。

借鉴相关研究,本研究主要探究空间要素与人口分布的关系构建影响因素模型,并运用多源数据融合技术,进行人口数据空间化。鉴于城镇人口与农村人口分布特点不同,将人口划分为城镇人口和农村人口,分别进行空间化。假设城镇人口集中分布在城镇用地上,利用交通网络与人口分布具有高度正相关性的特点,分析不同等级道路的影响;假设农村人口分散分布,鉴于相同土地类型在不同地区承载人口能力也不同^[15],所以土地覆盖类型赋予权重应考虑区域差异和并用交通网络数据进行局部修正,最后叠加城镇人口空间化图层和农村人口空间化图层生成全国范围 1 km 人口分布栅格数据。

2 数据来源与技术路线

2.1 数据类型和来源

影响人口分布的因素可以分为自然环境因素和社会因素两大类,前者是人口演变的自然基础,主要有海拔、坡度、植被覆盖等;后者是人类改造自然、生产建设的结果,包括土地覆盖、交通路网等要素^[16-18]。在影响方式上,这些空间要素可以分为承载类和范围影响类,承载类主要是自然地理因素和土地覆盖数据,可以通过技术手段估计其是否分布人口,以及不同类型土地承载人口的相对多寡;范围影响类主要是自然因素中的河流水系和社会因素中的道路路网,这些类型要素会对周边人口分布产生显著影响,并具有距离衰减效应。本研究主要利用的数据包括:

(1) 人口数据:2015 年全国地级市人口统计数据主要来自于《中国城市统计年鉴—2016》,少部分是来自相应城市 2015 年国民经济和社会发展统计公报。

(2) 自然环境数据:DEM 数据提供海拔和坡度,来自中国科学院资源与环境数据中心 (<http://www.resdc.cn>) 的中国海拔高度 (DEM) 空间分布数据,源数据是美国奋进号航天飞机的雷达地形测绘 SRTM (Shuttle Radar Topography Mission, SRTM) 数据,于 2000 年 2 月 11 日至 22 日采集。NDVI 数据显示植被情况,来自中国科学院资源与环境数据中心 (<http://www.resdc.cn>) 2015 年中国季度植被指数 (NDVI) 空间分布数据集中夏季数据,基于连续时间序列的 SPOT/VEGETATION NDVI 卫星遥感数据。

(3)土地覆盖数据:土地是人口分布的载体,土地覆盖数据体现了自然基础和人类改造自然的结果。该数据来自中国科学院资源环境科学数据中心(<http://www.resdc.cn>)2015年中国1:10万比例尺土地覆盖现状遥感监测数据库数据集,基于Landsat 8遥感影像,通过人工目视解译生成的1 km栅格数据。

(4)交通路网数据:范围影响要素类型会对周边人口分布产生显著影响,并具有距离衰减效应,如道路、水系。考虑到现代经济建设,道路对人口的影响力更强,所以现纳入道路路网数据。不同等级的道路具有不同的影响,高等级道路包含了铁路、高速公路、国道、省道,这些道路在空间分布上具有高度一致性,所以只选取了国道作为代表;低等级道路选取了城市主干道和县道及其他道路为代表。

2.2 研究思路

模型构建总体思路为,城乡分开建模、自然分

区建模,探索空间要素与人口分布的关系,按照式(1)建立全国1 km栅格人口空间化模型。

$$\frac{\omega_{ij}}{\sum_i \omega_{ij}} = \frac{P_{ij}}{P_{total.i}} \Rightarrow P_{ij} = \omega_{ij} \cdot \frac{P_{total.i}}{\sum_i \omega_{ij}} \quad (1)$$

其中: ω_{ij} 为某种空间要素在第*i*个行政区内第*j*个栅格点对人口分布的影响权重, P_{ij} 为第*i*个行政区内第*j*个栅格点分配到的人口数, $P_{total.i}$ 为第*i*个行政区的人口统计数据。

人口数据空间化过程可具体划分为识别和剔除无人区、城乡人口划分、土地覆盖影响因素分析、交通影响因素分析和人口数据空间分配等步骤,技术路线图详见图1。首先,识别人类活动区及剔除无人区,无人区主要是水域、冰川、荒漠和密林等几乎不存在人类活动的区域。识别和剔除无人区的主要依据是土地覆盖、海拔、坡度和NDVI等数据。分城乡建模根据城乡人口不同分布特征,分别构建人口空间化模型,避免因城乡统一建模的模型和参数误差,提高模型精度^[19]。

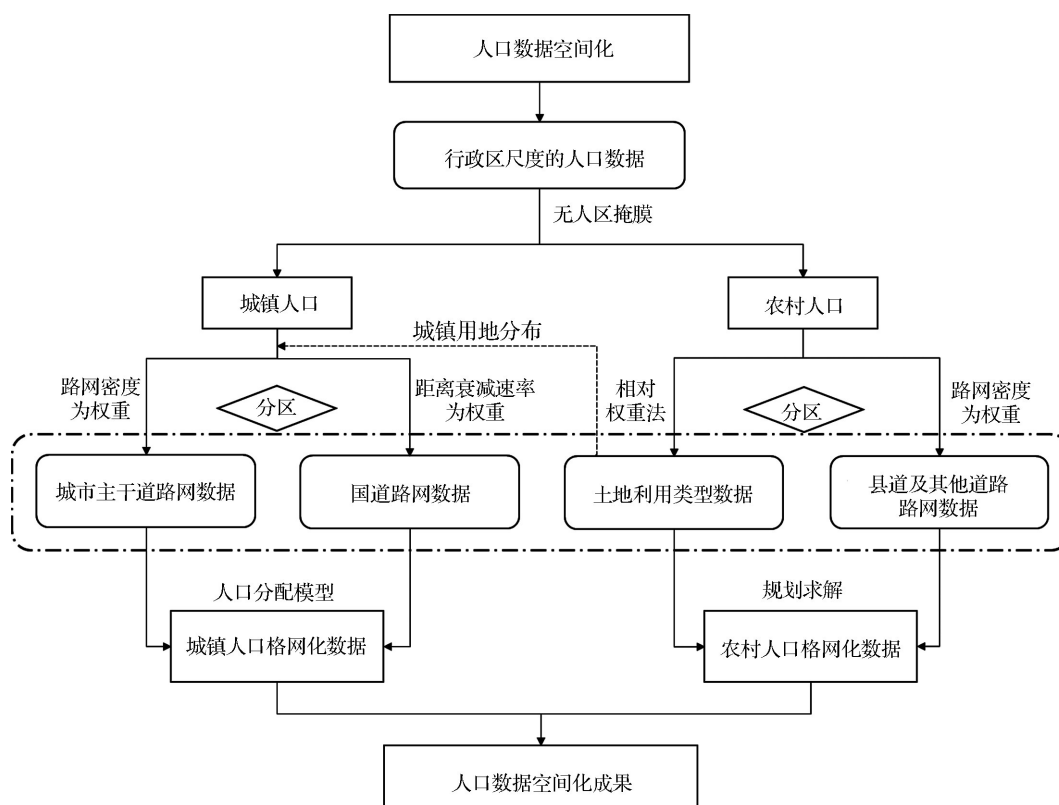


图1 人口数据空间化研究技术路线

Fig.1 Technology route of population data spatialization research

土地覆盖类型作为人与自然相互作用的景观,与人口分布的关系非常密切,基于土地覆盖数据的人口数据空间化方法不仅简单可行、解释性强,而

且人口数据空间化精度也比较高^[20]。利用土地覆盖数据的传统方法主要是根据土地覆盖与人口分布的相关关系,构建多元回归模型^[21],得到不同土

地覆盖类型的人口分布系数以实现人口数据的空间展布^[20],但是土地覆盖数据具有比较强的自相关性,造成多元回归系数无法反映变量独立贡献的问题。本研究采用相关权重法(relative weight)^[22]计算影响因素的权重。具体方法如下:首先,计算所有土地覆盖类型对人口分布的多元回归模型的 R^2 ,作为所有预测变量(土地覆盖类型)对响应变量(人口规模)的整体解释度;然后,对预测变量之间的相关系数矩阵进行奇异值分解,以此对 R^2 进行分配,以每种土地覆盖类型在 R^2 中所占份额作为该种土地覆盖类型对人口空间化的影响权重。

由于土地覆盖数据受限于遥感数据本身的分辨率、制图精度等的条件,忽略了居民点区域之外的一些零星分布的农村人口,所以在此认为所有土地覆盖类型均有农村人口分布,以定量方法测算不同土地类型对人口分布的影响。由于中国地域辽阔,各地自然及社会、经济、历史等条件差异较大,应分区建立土地覆盖影响模型^[13]。

研究假设城镇人口仅分布在城镇用地上,相关研究表明,路网密度与人口密度呈现显著的正相关性,考虑不同等级道路影响方式不同,选取国道和城市主干道分别代表高等级道路和低等级道路,探究其对人口分布的影响。农村人口部分,土地覆盖类型赋权难以反映同一分区同一土地类型内部人口的空间分布差异,也需要利用交通路网进行修正。

在分别得到城镇人口空间化数据和农村人口空间化数据后,将两层叠加可得到最终人口数据空间化成果。

3 人口空间化模型构建

3.1 无人区掩膜模型

人口数据空间化的第一步是要区分无人区和人类活动区,只在人类活动区的格点上赋值分配人口。无人区和人类活动区划分的基础主要是自然要素。本研究的无人区划分是基于土地覆盖数据进行划分,将建设用地(城镇用地、农村居民点、其他建设用地)和耕地(旱地、水田)直接认定为人类活动区,并将这些区域周边一定缓冲区范围也认为是人类活动区,然后其他区域即可定义为无人区。这种方法避免了直接根据自然要素(海拔、坡度、NDVI等)划分时阈值选定问题。采用ArcGIS的邻域分析工具,通过栅格单元邻域范围用地类型搜索,可以识别出初步的无人区,再与海拔、坡度、土

地覆盖等数据叠加验证,就可界定无人区范围。研究显示,邻域搜索范围为3 km比较合适,即如果栅格单元周围3 km没有耕地或建设用地,即可定义为无人区,否则就是人类活动区。

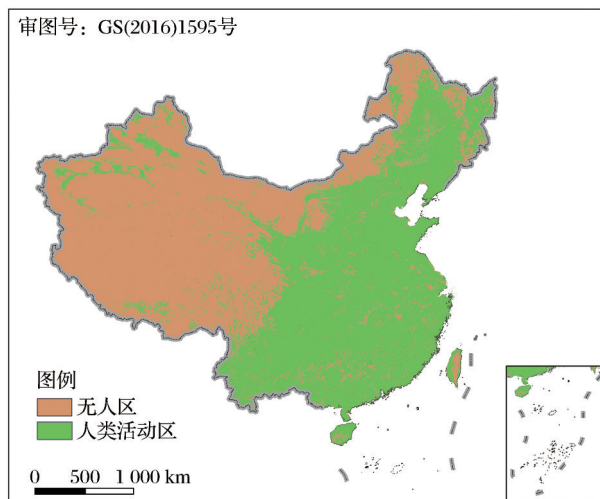


图2 人类活动区与无人区识别结果

Fig.2 Identification between human activity area and no man's area

图2中识别结果显示,我国无人区总面积约515.7 km²,约占全国国土面积的54.2%。“胡焕庸线”以东集中了90%以上的人类活动区,无人区主要是大兴安岭林区、河湖水域区和中高山区。而“胡焕庸线”以西的人类活动区集中在宁夏河套平原、塔里木盆地、天山北坡和西藏一江两河地区和横断山脉河谷地带^[23-24]。

在解释验证无人区划分上选取的自然指标为坡度和NDVI。坡度是影响人类居住建设布局的重要影响因素^[23]。NDVI表示植被覆盖指数,NDVI值过低,代表水域冰川等高反照率地区,NDVI值过高,代表了密林区。通过统计建设用地和耕地分布发现:坡度15%以上,NDVI值在0.1以下或0.89以上,几乎不再存在人类活动。以此为阈值,进行分类可以对无人区类型做出一定的解释。

图3中蓝色区域为河渠、湖泊和水库坑塘等水域和永久性冰川雪地、滩涂、滩地以及沙地、戈壁、盐碱地、沼泽地、裸土地、裸岩石质地等其他指其他未利用土地;深绿色区域坡度大且NDVI值偏高,代表了高山密林区,主要为秦岭、神农架原始森林地区;粉色地区坡度大,但NDVI为正常区间,主要为横断山脉、喜马拉雅山脉、昆仑山脉、天山山脉、南岭山脉以及台湾山脉;棕色区域坡度大且NDVI值偏低,代表了荒芜山地,零散分布在昆仑山脉和喜

马拉雅山脉地区;浅绿色区域NDVI值偏高,但坡度在正常区间,代表了密林区,主要为大小兴安岭林区;红色代表NDVI值偏低,坡度在正常区间,主要在青藏高原北部。灰色为未识别类型的无人区,主要集中在内蒙古北部地区,坡度和NDVI都是正常区间,但是却没有建设用地和耕地等代表人类活动的类型格点。这些区域以牧业为主,可能存在部分游牧人口。

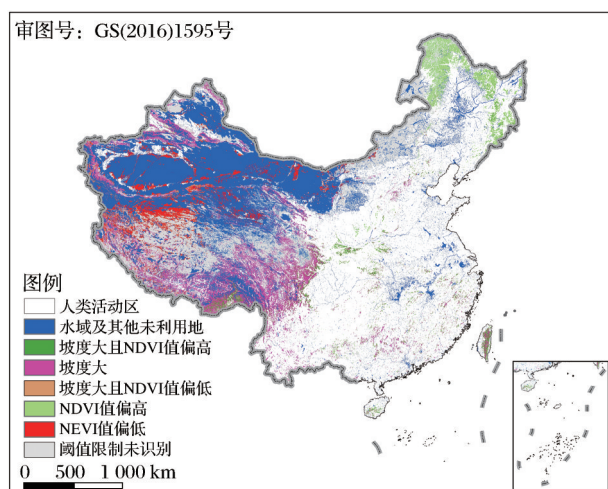


图3 无人区类型划分

Fig.3 Type division of no man's area

3.2 土地覆盖影响权重模型

土地覆盖类型与人口分布具有显著关联,同时不同类型间也有较大差异。就全国范围来说,城镇用地(0.565)、农村居民点(0.492)、其他建设用地(0.393)、耕地(0.374)、水田(0.411)这些因素与人口具有比较明显的正相关关系。城镇用地和农村居民点是人口的主要集聚区,其与人口相关性排在前列,耕地作为人类重要的生态支持系统,主要分布在地势平坦,同样适宜建设,为便于耕种居民点也不会离耕地过远,所以耕地与人口分布的相关性也非常强。有林地指郁闭度 $>30\%$ 的天然林和人工林,相对比较偏远,有林地密集地区,人口较少。灌木林和疏林地相对更接近人类聚集地,作为生态服务支持。在全国范围内分析,草地的相关系数均为负,但是在不同地理分区其与人口分布的关系又可能存在不同。土地覆盖类型之间很多具有较强的相关性,难以应用一般的多元回归模型,因此这里采用相关权重法(relative weight)确定影响权重,避免预测变量自相关性的影响。

各地区由于自然条件、资源禀赋、基础设施、人口发展、历史条件的差异,人口分布区域差异特征

明显。因此,依据土地覆盖类型建立影响权重模型前,应该选取特定指标或方法对研究区进行分区,保证区内的自然、经济条件基本一致,即体现“区间人口分布差异明显、区内人口分布特征相似”的特点,减少因区域条件差异过大而引起的人口空间化误差^[13]。在此综合考虑自然和人类活动因素,采用中国十二大农业生态区划作为分区标准^[25]。

表1 中国12大农业生态区分区

Table 1 Twelve major agro-ecological zones in China

序号	分区	序号	分区
1	东北区	7	黄土高原地区
2	黄淮海区	8	四川盆地
3	长江中下游地区	9	云贵高原区
4	江南区	10	横断山区
5	华南区	11	西北区
6	内蒙古高原及长城沿线区	12	青藏高原区

注:来源为《中国农业资源综合生产能力与人口承载能力》

研究假设城镇人口集中在城镇用地空间,从土地覆盖数据中提取出城镇用地图层,然后探究其他土地覆盖类型对农村人口的影响权重。按照预期,农村居民点是人口集聚区,应该具有最大权重,但是定量计算的结果并不尽然。有研究认为通过统计数据计算的客观权重结果仅仅是一种统计相关关系,它还必须符合一定的地理规则,指出应建立土地类型对人口分布的影响力定性排序原则,修正不符合此原则的系数^[11]。但是,也可以认为农村居民点权重的高低反映了地区居住的集聚程度。农村居民点权重越高,说明农村人口居住越集中,反之,说明很大一部分人口分散到其他土地类型中。同时,将县道及其他道路路网数据与土地覆盖估计结果融合作为最终结果,也可以对某些农村居民点人口偏低的估计做出补偿。因此,本研究没有对各类型权重顺序做出人为规定。表2显示了通过相关权重法(relative weight)计算所得各个分区的土地类型权重,克利夫兰点图直观地反映不同分区土地类型影响排序及差异。

东北区和黄淮海区两大北方平原区的权重分布比较类似,农村居民点和旱地的权重占绝对优势,其余土地类型的权重都比较小,体现北方平原区人口相对集中,有部分人口分散在旱地格点中。长江中下游区包括了长江下游平原、豫南鄂北山地、长江中游平原,旱地的权重最高,农村居民点和水田紧随其后,这表示长江中下游地区人口相对分散,旱地地区地势平坦,适宜居住建设,相当部分人

表 2 分区土地覆盖权重矩阵
Table 2 Matrix of partitioned land use's weights

	东北区	黄淮 海区	长江中下 游地区	江南区	华南区	内蒙古高原及 长城沿线区	黄土高 原地区	四川盆 地区	云贵高 原区	横断山区	西北区	青藏高 原区
农村居民点	39.9	49.0	14.4	4.5	20.7	11.1	55.0	2.0	0.4	0.5	22.5	7.7
其他建设用地	0.5	2.1	1.0	8.1	1.9	8.7	3.3	14.6	0.5	0.2	12.5	3.8
旱地	37.2	36.1	29.2	18.5	18.4	32.0	18.1	22.4	51.5	17.2	25.2	26.1
水田	11.6	2.9	10.7	31.6	8.3	0.6	0.5	21.3	0.7	13.9 (22.6)	3.2	8.5
有林地	1.1	0.5	5.1	13.3	9.3	5.4	6.2	3.2	4.7	8.1 (13.2)	1.9	9.7
灌木林	0.5	0.8	3.9	1.9	4.6	24.1	0.7	8.1	15.7	11.0	2.1	2.5
疏林地	0.7	0.6	4.3	13.3	13.0	4.9	1.7	6.7	2.5	3.0 (5.0)	6.4	3.2
其他林地	0.8	1.1	5.5	3.5	10.9	4.2	4.2	15.4	1.9	9.8	14.3	10.9
高覆盖草地	2.8	2.6	9.6	2.7	6.5	4.0	2.0	1.3	1.8	4.3 (7.0)	3.8	11.0
中覆盖草地	4.0	2.4	9.1	1.1	3.9	2.6	3.7	4.1	17.6	16.0 (26.1)	5.5	8.8
低覆盖草地	1.0	1.8	7.2	1.4	2.5	2.3	4.8	0.9	2.7	16.0 (26.2)	2.5	7.7

口会分散在旱地格点。江南区包括东部山地丘陵、中部山地丘陵、西部山地丘陵、南岭山地,农村居民点的权重更小,体现这里受到地形的限制,人口更为分散,耕地作为地势平坦区域,与人口相关性更大,同时,江南区旱地较少,所以水田的权重最高。华南区与北方平原区类似,也是符合一般认知的农村居民点和旱地占前两位,这一地区由闽粤丘陵平原、粤桂丘陵、雷州半岛和海南丘陵和台湾山地组成,权重分布表明该地区农村人口主要在丘陵平原地带聚集。蒙古高原及长城沿线区农村居民点的权重不高,这里地广人稀,人口也比较分散,旱地一般分布在水热条件好的地区,这些地区同样是适宜居住区,所以旱地的权重最高。黄土高原区千山万壑,地形破碎,人口更为集中,所以农村居民点权重最高,地势平坦的旱地其次。四川盆地区有两部分,分别是盆周山地和盆地丘陵平原,农村居民点权重较低主要是周边山地范围的地级市样本造成的,这些地区受地形限制,人口较为分散,水田和旱地代表了地势平坦区,散落分布着大量农村人口。云贵高原区农村人口更为分散,这一地区的第一产业兼有农业和牧业,二者发展条件好的地区为人口主要分布区。横断山区的样本量严重不足,因此,根据各土地类型格点数目和它们之间的相关性,剔除了部分土地类型,计算后又参考土地类型之间的相关性和相邻区域对剔除的土地类型进行了人为

权重赋值。表中括号中的数值为保留全部土地类型通过相关权重法(relative weight)直接计算的数值,横断山区与云贵高原类似,也是旱地和中低覆盖草地最高。西北区主要为绿洲农业,人口相对分散,人口分布具有粮食生产指向性,即具有粮食生产能力的地区更适合居住,所以旱地权重最高,农村居民点其次。青藏地区环境比较恶劣,农村居民点权重较低,表示人口比较分散,旱地地势平坦,分布了大部分人口,同时在这一地区,高覆盖草地代表气候较好、水草丰美地区^[23],所以高覆盖草地权重高居第二位。

3.3 交通路网影响权重模型

从近代开始,交通网络对人口分布产生了的重大影响^[26]。石家庄市和郑州市都因交通干线崛起,超越了历史悠久的名城,成为区域中心城市。交通网络一方面与现有人口分布格局有很强的相关性,另一方面交通网络可能会打破原有的人口分布格局,促成新的集聚流向。在分析中应注意不同等级道路的作用方式、影响衰减程度。

3.3.1 低等级道路

低等级道路对人口空间分布具有近距离影响作用,可以通过ArcGIS的线密度分析工具生成城市道路和县乡村道路的路网密度栅格数据,分别作为城市人口和农村人口数据空间化的权重依据。从图5(a)可以看出,县道和其他道路的路网密度最

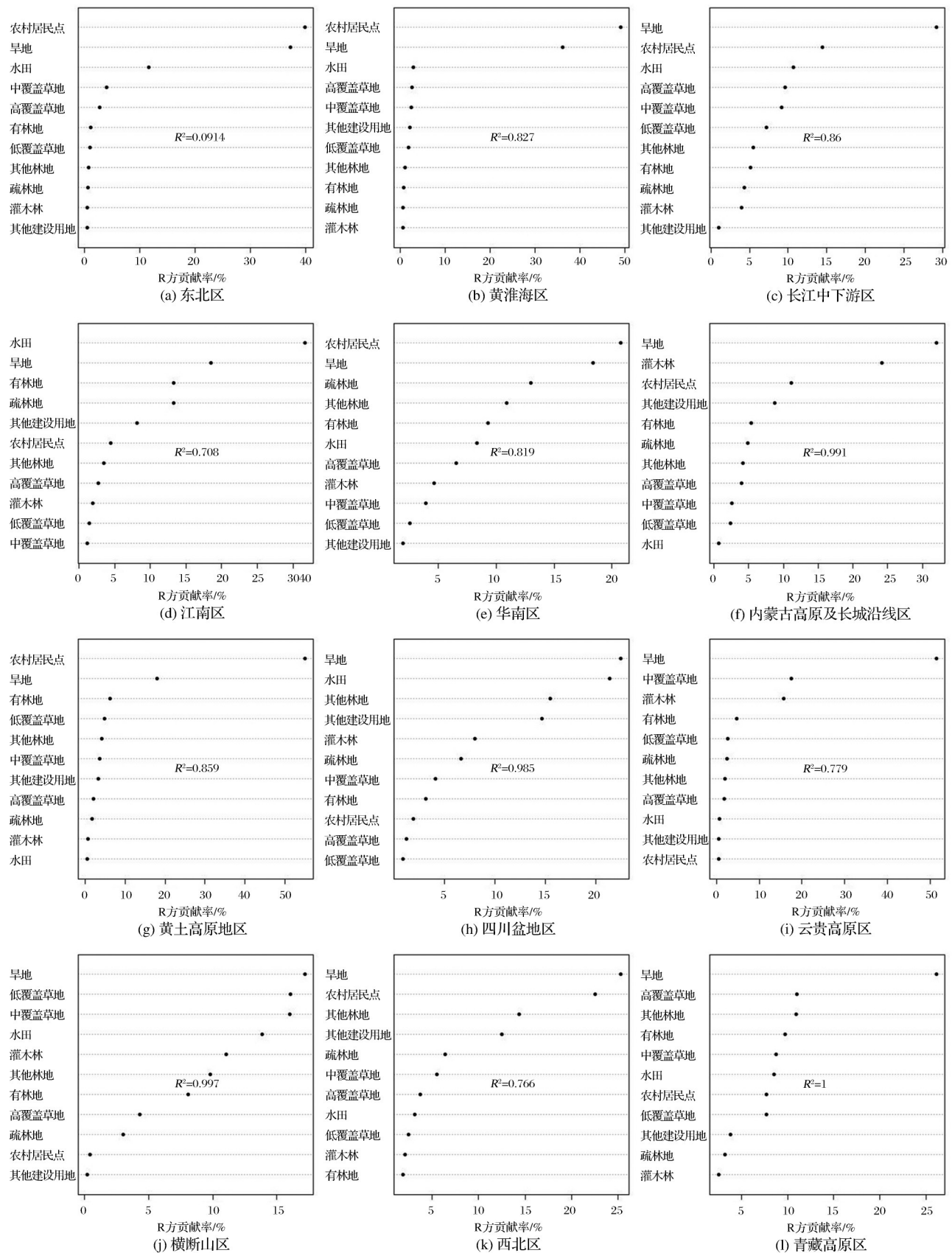


图 4 分区土地覆盖类型权重克利夫兰点图

Fig.4 Cleveland point map of land use type's weight

大的是京津周边、长江下游平原、成都平原等区域，与这些区域高密度的农村人口分布吻合。从图 5 (b)可以看出，直辖市和重要省会城市的城市干道

路网密度最大，京津冀、长三角、珠三角三大城市群最为突出，而华北平原和长江中下游平原则凸显了很多中小城市点。

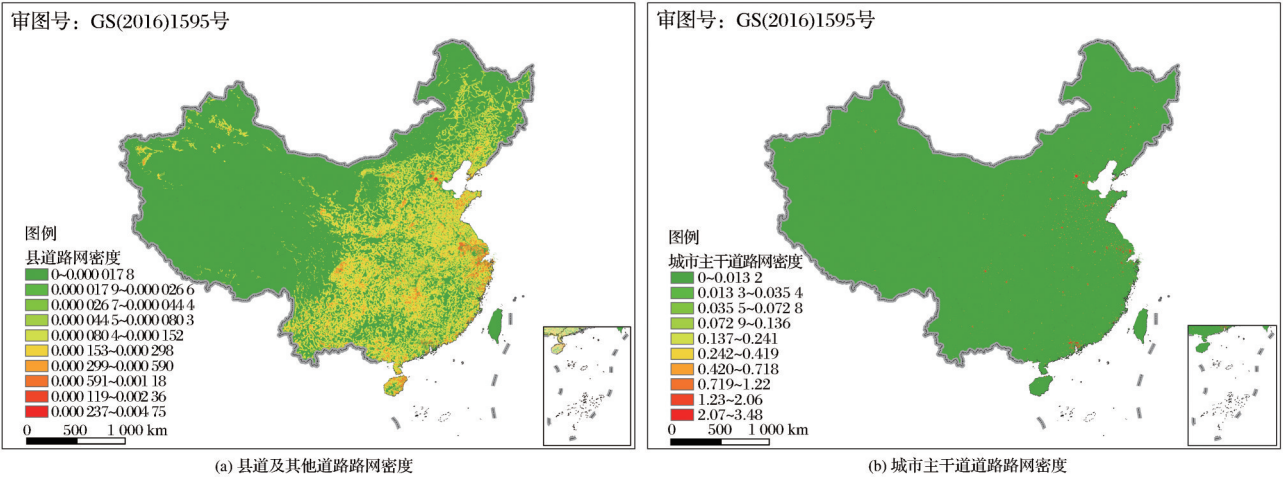


图 5 路网密度

Fig.5 Network density

3.3.2 高等级道路

国道作为高等级道路，与城市主干道和县道有所不同，其在分布上呈现稀疏网状，路网密度区分性不强。这里考虑其距离衰减效应，对其做缓冲区分析，统计与国道不同距离范围的城镇用地面积（城镇用地格点），建立回归模型，以距离衰减速率为城镇用地格点赋权。

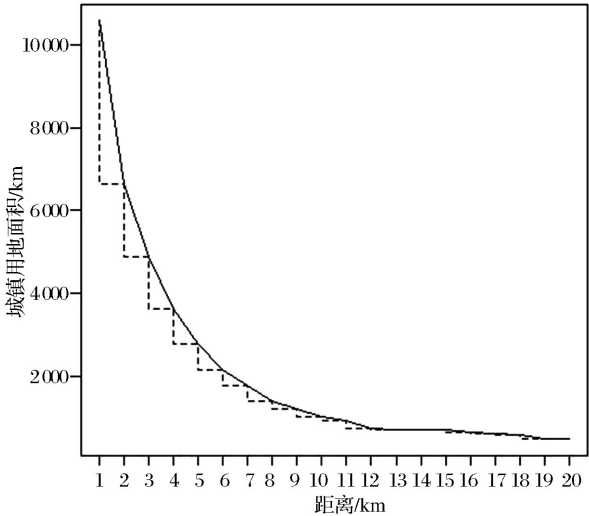


图 6 城镇用地面积与距国道距离曲线图

Fig.6 Urban land area and distance from national highway

图 6 反映了距离国道路网一定距离的城镇用地面积（城镇用地格点数，1 个栅格像元表示 1 km²）。由图可知，随着与国道距离增加，城镇用地面积迅速减小，当大于 10 km 后，基本不再变化。这表明国

道对城市作用的影响范围约为 10 km。对城镇用地面积取对数，与距离区间的中值进行回归分析，得到的结果 R^2 为 0.899，显示拟合程度比较好，特别地，在确定了国道的影响范围为 10 km 后，只以 10 km 以内的样本进行回归， R^2 高达 0.978，回归曲线高度拟合了城镇用地面积随距离的变化情况。具体结果如表 3 所示，y 表示城镇用地面积，x 表示距离区间中值。

表 3 \ln (城镇用地面积)与距离区间中值回归结果

Table 3 Results of \ln (town land area) and distance interval median regression

距离	回归公式	R Square	P-value
<20 km	$\ln(y) = -0.1471x + 8.6643$	0.899	2.2008E-10***
<10 km	$\ln(y) = -0.2518x + 9.1703$	0.978	6.39126E-08***

注:Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

将回归表达式变为原变量形式，对其求导可得城镇用地面积随与国道距离变化的速率。

$$y = e^{-0.2518x + 9.1703} \tag{2}$$

$$\frac{dy}{dx} = -0.2518e^{-0.2518x + 9.1703} \tag{3}$$

代入相应区间中值，以区间中值对应变化速率为这一距离区间国道对人口分布影响的权重，结果如表 4 所示。

3.4 多元数据融合模型

根据以上各模型得到的不同类型人口分配权重，对省级行政区统计的人口数据进行空间格网分

表 4 不同距离区间权重值		
Table 4 Weights in different distance intervals		
距离区间	距离区间中值 x_i	$ \frac{dy}{dx}(x=x_i) $
0 ~ 1 km	0.5	2133.0
1 ~ 2 km	1.5	1658.2
2 ~ 3 km	2.5	1289.1
3 ~ 4 km	3.5	1002.1
4 ~ 5 km	4.5	779.1
5 ~ 6 km	5.5	605.6
6 ~ 7 km	6.5	470.8
7 ~ 8 km	7.5	366.0
8 ~ 9 km	8.5	284.5
9 ~ 10 km	9.5	221.2

配,再以地级市为单位统计估计的人口,以此为样本,分析多源数据的融合模型。

3.4.1 农村人口数据融合

首先分析不同类型数据对农村人口的估计情况,考虑道路对人口的影响会与地方经济发展水平

有关,首先根据城市化发展水平对地级市进行了分组。参照城市化曲线,按照各地级市的人口城市化率,以 25%、50%、75% 进行分组,并将其分别命名为城市化起步区,城市化提速区,城市化发展区,城市化成熟区。统计各地级市通过土地覆盖权重估计的人口,发现各组平均值差异不大;统计各地级市通过县道及其他道路路网密度权重估计的人口,发现各组平均值差异比较大,明显多估了城市化高的地区,少估了城市化地的地区。主要原因是城市化程度高度地区经济发达,基础设施完善,路网密度比较大,但是其农村人口却相对较少。以城市发展阶段进行分组,对土地覆盖估计差值和县道估计差值进行单因素方差分析(ANOVA),结果显示土地覆盖估计差值不显著,表明各组均值没有显著性差异;而县道估计差值在小于 0.001 的水平上显著,表明各组均值具有显著性差异,证实了县道及其他道路估计情况在城市发展阶段上确实存在系统性误差。

表 5 农村人口估计误差单因素方差分析						
Table 5 One-way analysis of variance of rural population estimation error						
		自由度	总方差和	平均方差和	F value	Pr(>F)
土地覆盖	发展阶段分组	3	15 995	5 332	0.728	0.536
	残差	339	2 481 619	7 320		
县道	发展阶段分组	3	370 548	123 516	14.53	6.30E-09***
	残差	339.0	2 881 140	8 499		

注:Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

根据土地覆盖权重和县道及其他道路路网密度权重估计农村人口的评价分析结果,在融合两种数据应该将地级市按照城市发展水平分组,分别确定融合比重。由于土地覆盖数据和县道及其他道路路网数据为不同类型的数据,可赋予一定比例进行直接组合。

$$P_{country} = a \cdot P_{landuse} + b \cdot P_{county-road} \tag{4}$$

其中: $P_{country}$ 为某一行政区农村人口, $P_{landuse}$ 为土地覆盖权重估计人口, $P_{county-road}$ 县道及其他道路路网数据估计人口。

a、b 的取值应该是使残差(观测值与模型提供的拟合值之间的差距)平方总和最小化。同时为了保证地级市范围人口总和等于实际统计值,应控制 $a+b=1$ 。

$$\sum (P_{real} - P_{country})^2 \rightarrow min \tag{5}$$

$$\sum (P_{real} - a \cdot P_{landuse} - b \cdot P_{county-road})^2 \rightarrow min \quad (a+b=1) \tag{6}$$

式中:a、b 的值可以在 Excel 中利用规划求解功能得到。按照之前的分析,在城市化成熟区,县道及其

他道路路网显著地高估了人口,具有较大误差,所以规划求解结果中县道系数值很小,即在城市化成熟区融合两种数据几乎没有纳入县道及其他道路的估计结果。在城市化发展区和提速区,土地覆盖权重与县道及其他道路权重几乎是等比例融合。在城市化起步区,县道及其他道路的权重比较高,约是土地覆盖权重的 3 倍,体现这些城市人口普遍集中在基础设施相对完善的区域。

3.4.2 城镇人口数据融合

与农村人口统计分析类似,根据城市主干道路网密度和国道影响权重对省级行政区统计的人口

表 6 农村人口融合比例分析			
Table 6 Analysis of the proportion of rural population integration			
发展阶段	观测值	土地覆盖系数	道路网络系数
城市化成熟区	31	0.959	0.041
城市化发展区	143	0.569	0.431
城市化提速区	150	0.584	0.416
城市化起步区	19	0.238	0.762

进行分配,在地级市尺度加总,评价估计情况。统计各地级市通过城市主干道路网密度估计的人口,发现各组平均值差异比较大,明显多估了城市化高的地区,少估了城市化低的地区;统计各地级市通过国道影响权重估计的人口,发现各组平均值差异也比较大,但是差异的趋势性不是太明晰。城市化程度高度的地区基础设施完善,城市主干道更密集,造成了估计值偏大。

通过国道影响情况估计,城市化高的地区国道覆盖范围更大,以其为权重造成分配了比实际多的人口;城市化提速区国道道路建设速度低于人口集

中速度,可能是城市化提速区依据国道影响权重估计值偏小的原因;城市化起步区可能由于国家大量投资修建国道,所以用其估计人口时比实际值偏大。

以城市发展阶段进行分组,对城市主干道估计差值和国道估计差值进行单因素方差分析(ANOVA),城市主干道估计差值在小于0.001显著,表明各组均值具有显著性差异;而国道估计差值在小于0.01的水平上显著,表明各组均值也具有显著性差异,证实了二者估计情况确实都在城市发展阶段上存在系统性误差。

表 7 城镇人口估计误差单因素方差分析

Table 7 One-way analysis of variance of urban population estimation error

		自由度	总方差和	平均方差和	F value	Pr(>F)
城市主干道	发展阶段分组	3	4 240 560	1 413 520	16.02	9.20E-10***
	残差	339	29 910 800	88 232		
国道	发展阶段分组	3	290 854	96 951	5.54	0.001 01**
	残差	339	5 932 187	17 499		

注:Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

以城市主干道和国道估计的城镇人口融合的情况比较复杂:①城市主干道和国道影响区(周围10 km)都没有覆盖全部的城镇用地,所以有部分城镇用地没有被赋予权重,将二者简单组合则有部分城镇用地没有分配人口;②城市主干道和国道影响区对人口分配权重是多个地级市宏观范围计算结果,一部分城市处在国道影响区内或是主干道路网覆盖的城镇用地比例较小,赋予同样的权重分配人口,造成了局部的异常高值点。

因此,在考虑城市主干道和国道估计融合模型时,总体上需要考虑的是二者能够解释城市人口分布原因的多少,其不能解释的部分平均处理,即均匀分配给所有城镇用地格点。在局部上需要考虑的是,部分城市被赋值的城镇用地格点占该市城镇用地比重较小,造成异常高值点,因此,需要针对每个城市引入面积调整因子修正。最终模型如下:

$$P_{road} = P_{total} \cdot \omega_{explained} \cdot \omega_{area}^k \tag{7}$$

其中: P_{road} 为按某一类型道路分配的人口, P_{total} 为城市总人口, $\omega_{explained}$ 为总体统计上道路的解释力权重, ω_{area} 为具有道路权重赋值的城镇用地面积占城镇用地总面积的比重,k为路网集聚能力系数。

模型中k为路网集聚系数,代表道路影响对人口的集聚影响程度。理论上,k的取值为 $(-\infty, +\infty)$,因为 $\omega_{area} \leq 1$,所以k值越大表示路网对人口集聚的

影响程度越弱,k值越小表示路网对人口集聚的影响程度越强。经过多次验证,k取2比较符合本次模型构建。回归分析中的 R^2 可以表征预测变量对响应变量的解释程度,但是由于城市主干道路网密度和国道影响权重自身具有一定的相关性,二者单独与人口进行回归的 R^2 不能完全代表自身对人口的解释程度,所以这里再次利用相关权重法(relative weight)分解出自身的独立解释份额。

表 8 两类道路解释力权重

Table 8 Interpretation power weight of the two types of road

	R Square	城市主干道 独立解释份 额	国道独立解 释份额
城市化成熟区	0.673	44.19	55.81
城市化发展区	0.746	31.39	68.61
城市化提速区	0.658	32.47	67.53
城市化起步区	0.715	46.16	53.84

将城市主干道分配人口和国道分配人口数据按照各自的权重赋予城镇用地每个格点,有部分城镇用地格点城市主干道权重和国道权重均为0值,此时没有分配到人口数据。需要在每个地级市总人口中减去以上按道路权重分配的人口数据,把这部分人口数据均匀分配给相应地级市每个城镇用地栅格,最终叠加3个图层可得到最终城镇人口空

间分布数据。

4 人口空间格网化成果分析

将城镇人口空间化数据和农村人口空间化数据图层叠加,即可得到最终的全国人口空间化数据产品(图7)。从人口空间化数据产品看,我国“胡焕庸线”以东的人口聚集区主要在东北平原、华北平原、长江中下游平原、四川盆地、雷州半岛周边等地,“胡焕庸线”以西的人口主要分布在河西走廊、宁夏平原、塔里木盆地周围、天山北坡、西藏的一江两河地区、横断山区的河谷地带。城镇人口呈现明显的“点轴式”分布,哈大线、京广线、京沪线、陇海线是明显的城市聚集带。选取了哈大线、京广线、京沪线、陇海线和沪昆线等5个主要交通通道,发现主要交通通道为中心的狭长条带集中了大量人口,以20 km做缓冲区分析,以这5个主要交通通道为中轴,宽为40 km的条带面积仅占我国国土面积的

4.38%,而集中的人口却达到了20.22%。

人口空间化数据与其他空间要素数据可以很好地进行融合分析。将人口空间化数据与中国1:100万地貌类型空间分布数据叠加分析发现:中国人口分布对海拔比较敏感,低海拔平原和台地面积仅占我国国土面积的16.49%,而生活在其上的人口却占总人口的54.88%,相反,我国山地丘陵面积占比为55.78%,但是仅分布有22.11%的人口(表10)。

表9 主要交通通道两侧人口分布情况

Table 9 Population distribution on both sides of the main traffic corridor

交通通道两侧距离/km	面积占比/%	人口占比/%
5	1.08	7.07
10	2.17	12.95
15	3.27	17.12
20	4.38	20.22

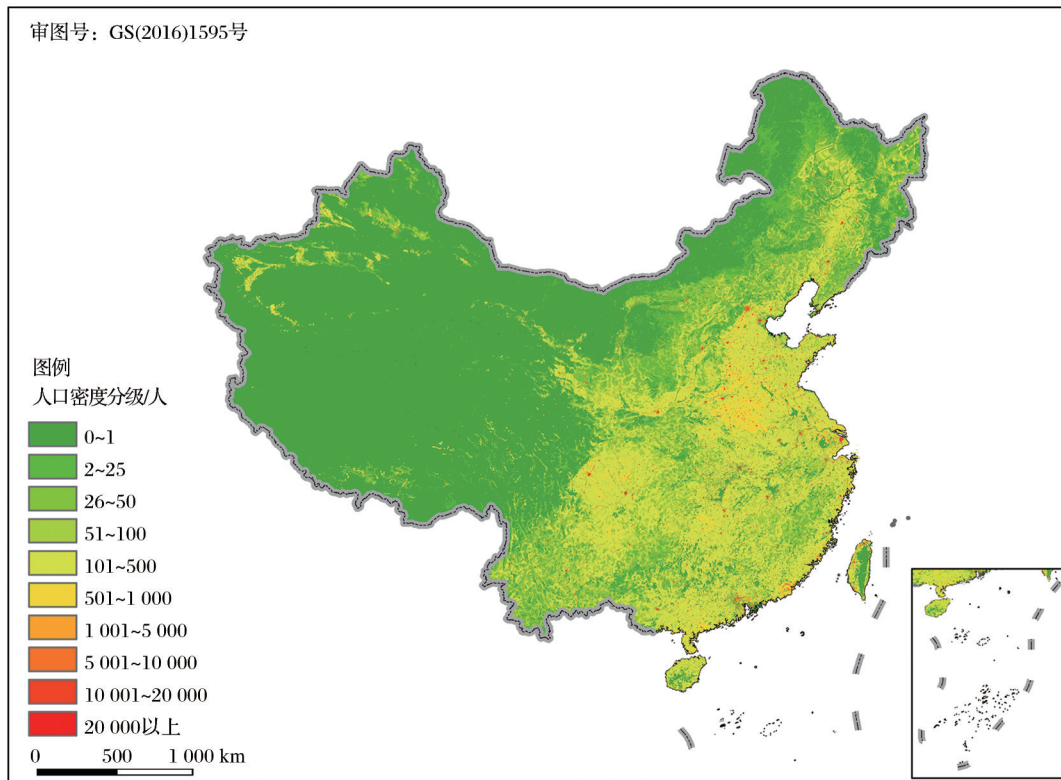


图7 全国公里格网人口空间化数据

Fig.7 Spatial population data of national kilometer grid

5 结 语

本研究通过探究不同类型要素对人口分布的影响模型,并运用多源数据融合的方法,最终得到了全国1 km人口栅格数据。主要结论如下:

(1)中国地域辽阔,不同地区自然条件,社会历史背景差距巨大,在研究全国范围自然要素分布影响时,首要做的是以一定标准分区。以12个农业生态区划分,计算土地覆盖影响权重,各区差异非常大,农村人口呈现出不同的分布状态,北方更加集

表 10 基于人口空间化数据对全国不同地形的人口统计结果

Table 10 Demographic results of different terrains across the country based on population spatialization data

地形	面积占比/%	人口占比/%
低海拔平原	11.48	43.56
中海拔平原	6.82	4.90
高海拔平原	2.99	0.80
极高海拔平原	0.94	1.26
低海拔台地	5.11	11.33
中海拔台地	2.45	1.50
高海拔台地	1.20	0.35
极高海拔台地	0.71	0.79
低海拔丘陵	6.67	6.31
中海拔丘陵	8.79	1.50
高海拔丘陵	1.23	0.17
极高海拔丘陵	1.05	0.45
山地	38.03	13.67

聚,南方相对分散,平原区更加集聚,山区和丘陵相对分散。

(2)在使用交通网络数据时,要注意不同等级道路在作用方式、影响范围上的不同。低等级道路可以路网密度体现区域内部差异性,高等级道路主要通过缓冲区分析来确定其影响范围。但是以交通网络为权重对人口分布进行拟合时,不同发展水平的城市间往往会存在系统误差。经济发达的城市人均道路拥有量高于经济落后的城市,统一权重估计时会显著高估经济发达的城市,所以应该按城市发展水平分组进行权重赋予。

(3)在设计模型时要主要总体因子和局部个体因子结合,若仅考虑全体城市样本或某一类型城市样本,简单将其用在所有城市上,会造成某些异常值。城镇用地数据在融合时,综合考虑了总体上的交通网络解释力,又考虑了局域因子—某一类型道路赋权用地面积比例,有效消除了异常值。

鉴于数据和篇幅所限,本研究尚有如下不足,需要在后续研究中完善:

(1)本研究主要集中在总体人口数量与自然环境空间要素的关系,没有反映人口特征诸如年龄、性别、受教育状况及流动特征等的空间结构性。在构建模型时也主要考虑的是客观环境的适宜性,没有考虑人类行为对于分布的影响,没有体现时间分辨率特征。

(2)本研究最终成果是将地级市尺度统计的人口在空间上分配,在后续研究中可改为县级行政区的统计人口,提高精准度。

(3)后续研究可考虑纳入其他与人口分布相关性更高的空间要素,例如全国建筑物密度栅格数据^[27]、电子地图 POI 数据。

参考文献 (References):

- [1] Bai Zhongqiang, Wang Juanle, Yang Fei. Research Progress in Spatialization of Population Data[J]. Progress in Geography, 2013, 32(11): 1692-1702.[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.]
- [2] Fu Haiyue, Li Manchun, Zhao Jun, *et al.* Summary of Grid Transformation Models of Population Data[J]. Human Geography, 2006, 21(3): 115-119.[符海月,李满春,赵军,等.人口数据网格化模型研究进展综述[J].人文地理,2006,21(3): 115-119.]
- [3] Xu Ying. The Relations of Tourism Region and Administrative Region: on the Case of Huangshan City[J]. Economic Geography, 2003, 23(6): 786-790.[薛莹.论旅游区域与行政区域的关系—以黄山市为例[J].经济地理,2003,23(6): 786-790.]
- [4] Li Fei, Zhang Shuwen, Yang Jiuchun, *et al.* A Review on Research about Spatialization of Socioeconomic Data[J]. Geography and Geo-Information Science, 2014, 30(4): 102-107.[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] Wu Jiansheng, Xu Duo, Xie Wudan, *et al.* Spatialization of Demographic Data at Medium Scale based on Remote Sensing Images: Regarding Beijing-Tianjin-Hebei as an Example[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2015, 51(4): 707-717.[吴健生,许多,谢舞丹,等.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[J].北京大学学报(自然科学版),2015,51(4): 707-717.]
- [6] Liu Jinsong. A Review of Population Geography Research in China[J]. Acta Geographica Sinica, 2014, 69(8): 1177-1189.[刘劲松.中国人口地理研究进展[J].地理学报,2014,69(8): 1177-1189.]
- [7] Ye Yu, Liu Gaohuan, Feng Xianfeng. Presentation of Spatial Distribution of Population and Its Application[J]. Journal of Geo-Information Science, 2006, 8(2): 59-65.[叶宇,刘高焕,冯险峰.人口数据空间化表达与应用[J].地球信息科学学报,2006,8(2): 59-65.]
- [8] Hu Yunfeng, Wang Qianqian, Liu Yue, *et al.* Index System and Transferring Methods to Build the National Society and Economy Grid Database[J]. Journal of Geo-Information Science, 2011, 13(5): 573-578.[胡云锋,王倩倩,刘越,等.国家尺度社会经济数据网格化原理和方法[J].地球信息科学学报,2011,13(5): 573-578.]
- [9] Zhong Weijing, Wang De, Xie Dongcan, *et al.* Dynamic Characteristics of Shanghai's Population Distribution Using Cell Phone Signaling Data[J]. Geographical Research, 2017, 36(5): 972-984.[钟炜菁,王德,谢栋灿,等.上海市人口分

- 布与空间活动的动态特征研究——基于手机信令数据的探索[J]. 地理研究, 2017, 36(5): 972-984.]
- [10] Xu Zening, Gao Xiaolu. A Novel Method for Identifying the Boundary of Urban Built-up Areas with POI Data[J]. *Acta Geographica Sinica*, 2016, 71(6): 928-939.[许泽宁, 高晓路. 基于电子地图兴趣点的城市建成区边界识别方法[J]. 地理学报, 2016, 71(6): 928-939.]
- [11] Tobler W, Deichmann U, Gottsegen J, *et al.* World Population in a Grid of Spherical Quadrilaterals[J]. *International Journal of Population Geography*, 1997, 3(3): 203-225.
- [12] Dobson J E, Bright E A, Coleman P R, *et al.* LandScan: A Global Population Database for Estimating Populations at Risk[J]. *Photogrammetric Engineering & Remote Sensing*, 2000, 66(7): 849-858.
- [13] Tian Yongzhong, Chen Shupeng, Yue Tianxiang, *et al.* Simulation of Chinese Population Density based on Land Use[J]. *Acta Geographica Sinica*, 2004, 59(2): 283-292.[田永中, 陈述彭, 岳天祥, 等. 基于土地覆盖的中国人口密度模拟[J]. 地理学报, 2004, 59(2): 283-292.]
- [14] Liao Shunbao, Sun Jiulin. GIS based Spatialization of Population Census Data in Qinghai-Tibet Plateau[J]. *Acta Geographica Sinica*, 2003, 58(1): 25-33.[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.]
- [15] Chen Qing, Hou Xiyong. 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.[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.]
- [16] Zhao Tongtong, Song Bangguo, Chen Yuansheng, *et al.* Analysis of Population Distribution and Its Spatial Relationship with Terrain Elements in the Yarlung Zangbo River, Nyangqu River and Lhasa River Region, Tibet[J]. *Journal of Geo-information Science*, 2017(2): 225-237.[赵彤彤, 宋邦国, 陈远生, 等. 西藏一江两河地区人口分布与地形要素关系分析[J]. 地球信息科学学报, 2017(2): 225-237.]
- [17] Bai Zhongqiang, Wang Juanle, Yang Yaping, *et al.* Characterizing Spatial Patterns of Population Distribution at Township Level Across the 25 Provinces in China[J]. *Acta Geographica Sinica*, 2015, 70(8): 1229-1242.[柏中强, 王卷乐, 杨雅萍, 等. 基于乡镇尺度的中国25省区人口分布特征及影响因素[J]. 地理学报, 2015, 70(8): 1229-1242.]
- [18] Wang Juanle, Cheng Kai, Bian Lingling, *et al.* Integration Framework and Key Technology of Big Earth Data for SDGS and Beautiful China Evaluation[J]. *Remote Sensing Technology and Application*, 2018, 33(5): 775-783.[王卷乐, 程凯, 边玲玲, 等. 面向SDGS和美丽中国评价的地球大数据集成框架与关键技术[J]. 遥感技术与应用, 2018, 33(5): 775-783.]
- [19] Dong Nan, Yang Xiaohuan, Cai Hongyan. Research Progress and Perspective on the Spatialization of Population Data[J]. *Journal of Geo-Information Science*, 2016, 18(10): 1295-1304.[董南, 杨小唤, 蔡红艳. 人口数据空间化研究进展[J]. 地球信息科学学报, 2016, 18(10): 1295-1304.]
- [20] Wang Kejing, Cai Hongyan, Yang Xiaohuan, *et al.* Spatialization Method for Census Data based on Reclassifying Residential Land Use in Urban Areas: A Case Study in the Middle Reaches of the Yangtze River Watershed[J]. *Remote Sensing Technology and Application*, 2015, 30(5): 987-995.[王珂靖, 蔡红艳, 杨小唤, 等. 基于城镇居民用地再分类的人口数据空间化方法研究——以长江中游4省为例[J]. 遥感技术与应用, 2015, 30(5): 987-995.]
- [21] Wang Kejing, Cai Hongyan, Yang Xiaohuan. Multiple Scale Spatialization of Demographic Data with Multi-factor Linear Regression and Geographically Weighted Regression Models[J]. *Progress in Geography*, 2016, 35(12): 1494-1505.[王珂靖, 蔡红艳, 杨小唤. 多元统计回归及地理加权回归方法在多尺度人口空间化研究中的应用[J]. 地理科学进展, 2016, 35(12): 1494-1505.]
- [22] Jeff W. Johnson. A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression[J]. *Multivariate Behavioral Research*, 2000, 35(1): 1-19.
- [23] Wang Chao, Kan Aike, Zeng Yelong, *et al.* Population Distribution Pattern and Influencing Factors in Tibet based on Random Forest Model[J]. *Acta Geographica Sinica*, 2019, 74(04): 664-680.[王超, 阚媛珂, 曾业隆, 等. 基于随机森林模型的西藏人口分布格局及影响因素[J]. 地理学报, 2019, 74(04): 664-680.]
- [24] Li Jiaming, Lu Dadao, Xu Chengdong, *et al.* Spatial Heterogeneity and Its Changes of Population on the Two Sides of Hu Line[J]. *Acta Geographica Sinica*, 2017, 72(01): 148-160.[李佳铭, 陆大道, 徐成东, 等. 胡焕庸线两侧人口的空间分异性及其变化[J]. 地理学报, 2017, 72(01): 148-160.]
- [25] Chen Baiming. China's Comprehensive Agricultural Resources Production Capacity and Population Carrying Capacity[M]. Beijing: Meteorological Press, 2001.[陈百明. 中国农业资源综合生产能力与人口承载能力[M]. 北京: 气象出版社, 2001.]
- [26] Zhang Haixia, Niu Shuwen, Qi Jinghui, *et al.* Geological Statistics Analysis of Population Distribution at Township Level in Henan Province[J]. *Geographical Research*, 2016, 35(2): 325-336.[张海霞, 牛叔文, 齐敬辉, 等. 基于乡镇尺度的河南省人口分布的地统计学分析[J]. 地理研究, 2016, 35(2): 325-336.]
- [27] Feng Tiantian, Gong Jianya. Investigation on Small-area Population Estimation based on Building Extraction[J]. *Remote Sensing Technology and Application*, 2010, 25(3): 323-327.[冯甜甜, 龚健雅. 基于建筑物提取的精细尺度人口估算研究[J]. 遥感技术与应用, 2010, 25(3): 323-327.]

Spatialization of China's Population Data based on Multi-source Data

Guo Yuchen^{1,2,3}, Huang Jinchuan^{1,2,3}, Lin Haoxi^{1,2,3}

(1.*Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences,*

Beijing 100101, China;

2.*Key Laboratory of Regional Sustainable Development Analysis and Simulation, Chinese Academy of*

Sciences, Beijing 100101, China;

3.*College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China)*

Abstract: The population data is usually calculated by administrative districts. When comprehensive analysis is carried out with other spatial data, there are often many problems such as different scales. The spatial data of population data is the process of exploring the spatial distribution of population. The final result can be other spatial correlation research. It lays the foundation and can also be directly applied to disaster impact assessment and other fields, which has important academic significance and application value. The population spatialization model is based on the spatial relationship between population and land use and traffic road network data. The population is divided into urban population and rural population. The land use type is empowered by the relative weight method, and the traffic network data is used to correct it. Spatial distribution of rural population; analysis of different ways of affecting different levels of traffic road network data, comprehensive distribution of urban population, and finally through the planning and other methods to superimpose and integrate different types of data. The results of population data spatialization results show that China's population is very sensitive to altitude and topography. Low-altitude plains and terraces only account for 16.49% of China's land area, while the population living on it accounts for 54.88% of the total population. On the contrary, China's mountains The proportion of hills is 55.78%, but only 22.11% of the population is distributed. The eastern population is concentrated in the Huanghuaihai area, the Sichuan Basin, the Yangtze River small and medium-sized tourist, and the North-east Plain. The urban population in China presents a typical "point-axis" distribution. The distribution along the main traffic lines is more prominent.

Key words: Population data spatialization; Land cover; Traffic network; Related weights