

中国公里网格人口分布数据集

付晶莹^{1,2}, 江 东¹, 黄耀欢¹

(1. 中国科学院地理科学与资源研究所, 北京 100101;

2. 中国科学院大学, 北京 100049)

摘要: 传统的人口数据来源于全国人口普查数据, 以行政区为基本单元的统计型数据, 空间分辨率低, 无法充分揭示人口数据的空间差异, 同时, 无法与以网格等基础地理单元数据的共享与整合。因此, 统计型人口空间化, 即将以行政区为基本单元的统计人口扩展到一定空间尺度 (本数据集为1公里) 网格上, 构建人口空间分布数据集, 能够与土地利用数据、自然资源及生态环境等自然要素数据进行综合分析, 实现人文社会经济数据与自然要素的整合, 具有重要的理论与现实意义。中国公里网格人口分布数据集的生产, 首先, 综合分析我国人口分布的空间特征及区域差异进行人口区划。在区划后的各子区域中建立多元统计人口空间化模型进行人口空间化, 并利用城市人口密度、交通状况、DEM及总量控制四个因子对模型进行校正, 以确保模型的合理性和准确性。为了对人口数据的精度进行校验, 选择我国东部、西部及中部典型省份中具有完整乡镇数据人口的地区作为检验样本, 以统计数据为真值, 空间化人口数据的相对误差在4.5%到13.6%之间, 而且大部分样本的相对误差小于10%。最终, 形成一套具有统一空间坐标参数、统一数据格式、统一元数据标准的公里网格人口空间分布数据集。

关键词: 中国; 人口; 公里网格; 空间分布

DOI: 10.11821/dlxb2014S006

数据引用格式: 付晶莹, 江东, 黄耀欢. 中国公里网格人口分布数据集. 全球变化科学研究数据出版系统, 2014. DOI: 10.3974/geodb.2014.01.06.v1, <http://www.geodoi.ac.cn/doi.aspx?doi=10.3974/geodb.2014.01.06.v1>

1 前言

中国公里网格人口分布数据集 (PopulationGrid_China) 是以遥感数据获取的土地利用类型数据及人口统计数据为基础, 利用地理信息系统的空间分析功能, 构建人口空间分布模型, 将统计型人口数据空间化, 从而, 生成1 km×1 km的空间人口数据。

2 数据集简介

中国公里网格人口分布数据集 (PopulationGrid_China) 的数据集名称、短名名称、通讯作者、数据作者、地理区域、数据年代、数据空间分辨率、数据出版单位、数据共享网络服务平台、数据集组成、出版及责任编辑等信息一并列于表1。

3 数据研发方法

土地利用状况是影响人口空间分布的重要因素, 土地利用格局与人口密度之间存在着

收稿日期: 2014-03-17; 修订日期: 2014-06-20

基金项目: 中国科学院重点部署项目 (2012) (KZZD-EW-08-03) [Foundation: Key Programs of Chinese Academy of Sciences, No.KZZD-EW-08-03]

作者简介: 付晶莹 (1986-), 女, 博士研究生, 主要从事地理信息科学的应用研究。E-mail: fuyj@lreis.ac.cn

表 1 中国公里网格人口分布数据集 (PopulationGrid_China) 元数据简表

数据库 (集) 名称	中国公里网格人口分布数据集		
数据库 (集) 短名	PopulationGrid_China		
通讯作者	付晶莹 (fujy@lreis.ac.cn)		
数据作者	付晶莹, 中国科学院地理科学与资源研究所, fujy@lreis.ac.cn		
	江东, 中国科学院地理科学与资源研究所, jiangd@igsrr.ac.cn		
	黄耀欢, 中国科学院地理科学与资源研究所, huangyh@lreis.ac.cn		
地理区域	地理区域覆盖中国 31 个省级行政区 (香港、台湾、澳门无数据), 经度范围大致位于 73°33'E~135°05'E, 纬度范围大致位于 3°51'N~53°33'N。		
数据年代	2005 年、2010 年		
数据空间分辨率	1 km	数据时间频率	5 年
数据格式	TIF, ZIP	数据量	TIF 164MB
			ZIP 15MB
数据出版单位	中国科学院地理科学与资源研究所 DOI: 10.3974/		
数据共享服务平台	中国科学院地理科学与资源研究所全球变化科学研究数据出版系统, http://www.geodoi.ac.cn		
	国家地球系统科学数据共享平台, http://www.geodata.cn		
责任编辑	刘闯, 石瑞香, 周翔, 何书金		
数据共享政策	本数据著作权所有者同意依据《全球变化科学研究数据出版系统》管理政策中“全社会公益性共享政策”实施共享。		

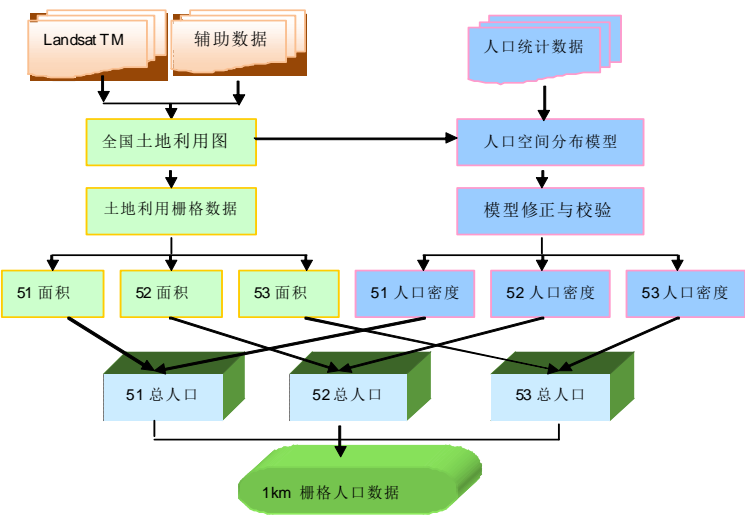


图 1 人口栅格数据生成的流程图

密切的关系。本数据集以遥感反演的土地利用数据及中国统计年鉴人口统计数据为基础，在 GIS 背景数据和空间分析功能的支持下，探询人口密度—土地利用格局的耦合关系，构建空间分布模型，生成 1 km×1 km 人口栅格数据。人口数据空间化实现的具体技术流程引自《黄河流域二元水循环要素反演研究》^[1]（图 1）。

其中：51、52、53 分别代表三种土地利用类型：城市居民点、农村居民点和工交建设用地。

中国公里网格人口分布数据集的生产，首先，综合分析我国人口分布的空间特征及区域差异，在区位理论及空间统计的支持下进行人口区划。在区划后的各子区域中，选择一定具有代表性的县作为建模样本县进行建模，并通过将人口数据与各种土地利用类型进行相关分析，筛选出与人口分布相关性强的具有代表性的土地利用类型，作为建模参数。在

每个子区内,利用建立的多元统计模型进行人口空间化,并利用城市人口密度、交通状况、DEM及总量控制四个因子对模型进行校正,以确保模型的合理性和准确性^[2-4]。最终,形成一套具有统一空间坐标参数、统一数据格式、统一元数据标准的公里网格人口空间分布数据集。其中,分县统计数据来源于[5-6],行政边界数据来源于[7],中国公里网格2005年、2010年人口分布图分别见图2、图3。

4 数据组成

中国公里网格人口分布数据集是TIF格式的中国公里网格人口分布数据集,由2005年和2010年二个数据集组成,分别为中国公里网格2005年人口分布、中国公里网格2010年人口分布。用软件压缩为*.ZIP后,数据量为15MB。

数据可视化图 (ARCGIS TIF格式,单位:人)

5 数据可靠性检验

为了对人口数据的精度进行校验,选择我国东部的北京、上海,中部的吉林、河南以及西部的新疆、重庆六个具有地域代表性的省份,在每个省份中分别选取10个区/县(北京:东城区、朝阳区、房山区、大兴区、延庆县;上海:黄浦区、徐汇区、闸北区、松江区、崇明县;吉林:南关区、永吉县、梨树县、东辽县、通化县、抚松县、通榆县、长岭县、汪清县;河南:中牟县、通许县、汝阳县、郟县、内黄县、淇县、封丘县、温县、台前县;新疆:乌鲁木齐县、独山子区、鄯善县、伊吾县、呼图壁县、温泉县、且末县、温宿县;重庆:万州区、黔江区、涪陵区、渝中区、大渡口区、江北区),这些被选取的区/县具有完整乡镇数据人口作为检验样本,对模型结果进行检验。检验的方法:首先将乡(镇)边界数字化为矢量图,将其转换为统一的投影、坐标系统后,利用GIS的空间分析功能,叠加到人口栅格数据上,统计出各乡(镇)的人口总和,与统计的人口数据进行对比。

检验表明本次工作获得了很好的结果:以统计数据为真值,空间化人口数据的相对误差在4.5%到13.6%之间,而且大部分样本的相对误差小于10%。

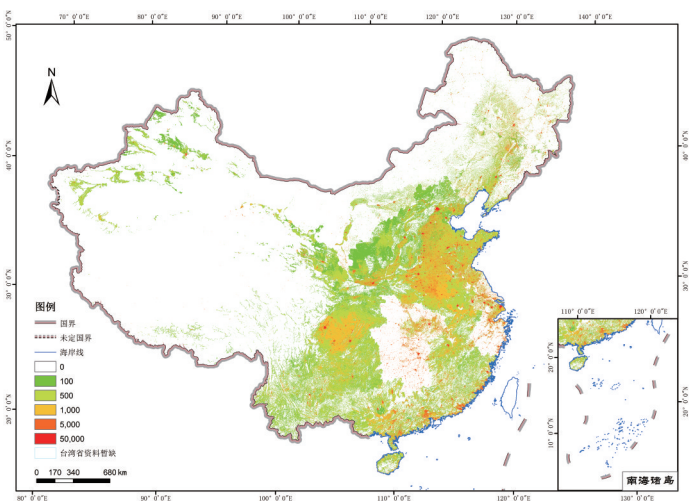


图2 中国公里网格2005年人口分布图
(ARCGIS TIF数据格式)

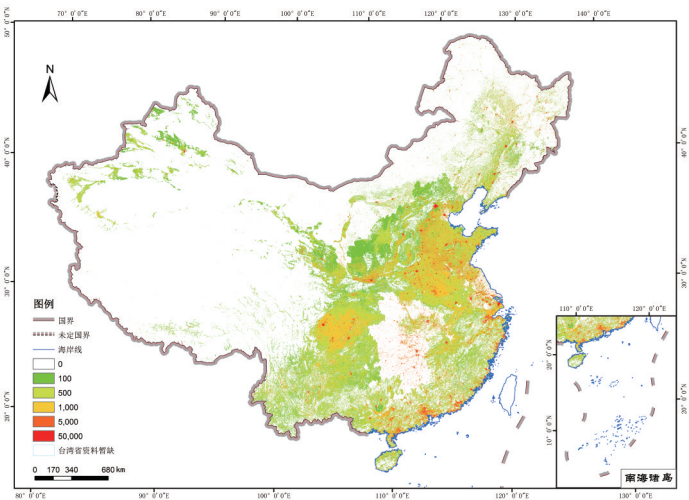


图3 中国公里网格2010年人口分布图
(ARCGIS TIF数据格式)

6 结 论

该数据是以 1 km 空间分辨率为基本单元的中国人口空间分布数据。人口数据空间化解决了人口统计数据与所属区域的空间特征脱节, 以及人口数据与环境数据、自然资源数据、经济统计数据等的集成等问题。人口统计数据空间化, 尤其是栅格化的人口数据对跨学科的研究与应用具有重要价值。

致谢: 本数据集中用到的中国 1:10 万土地利用数据来自中国科学院资源环境科学数据中心, 数据提供者刘纪远研究员等, 在此表示感谢。

参考文献

- [1] 王建华, 江东. 黄河流域二元水循环要素反演研究. 北京: 科学出版社, 2006.
- [2] 江东, 杨小唤, 王乃斌 等. 基于 RS、GIS 的人口空间分布研究. 地理科学进展, 2002, 17(5): 734-738.
- [3] 江东. 人文要素空间化研究进展. 甘肃科学学报, 2007, 19(2): 91-94.
- [4] 黄耀欢, 杨小唤, 刘业森. 人口区划及其在人口空间化中的 GIS 分析应用: 以山东省为例. 地球信息科学, 2007, 9(2): 49-54.
- [5] 国家统计局. 2005 中国统计年鉴. 北京: 中国统计出版社, 2005.
- [6] 国家统计局. 第 6 次人口普查数据. 北京, 2010.
- [7] 国家基础地理信息中心. 中华人民共和国国界地理信息系统数据 (1:100 万比例尺). 2008.