基于遥感和 GIS 的巢湖流域人口信息提取

刘建军^{①②} 李春来^③

(①中国科学院地球化学研究所, 贵阳 550002; ②中国科学院研究生院, 北京 100039; ③中国科学院国家天文台, 北京 100012.
E-mail: dhsliujj@hotmail.com)

摘要 利用 GIS 技术和遥感解译的土地利用数据, 采用复合面积内插法, 对巢湖流域人口分布进行了初步研究, 从计算结果看, 此方法能有效地提取小流域的人口信息.

关键词 GIS 小流域 人口信息 巢湖流域

在区域生态环境评价过程中, 由于研究目的不同, 评价单元可以采用行政区单元^{[1]1)}、景观单元^[2]和流域单元^[3]. 以小流域为单元进行区域生态环境评价, 由于小流域是一个独立的地貌单元, 流域内的生态系统具有从上游至下游的生态完整性, 对于生态系统保护与恢复具有重要意义. 因此, 分析小流域的人类活动与生态环境的关系时, 提取流域单元上的人口和社会经济数据就尤为必要. 本文以巢湖流域为例, 结合 GIS 技术和遥感数据, 探讨人口分布信息从行政单元转换到小流域单元上的有效方法.

1 面积内插模型

在很多国家, 人口数据和社会经济数据一般只按照行政单元来统计. 为了便于分析不同评价单元的生态环境与人类活动的关系, 通常需要将这些属性值内插到不同的单元上, 这种将属性数据从源区域向目标区域转换的过程, 就称作面积内插(areal interpolation)^[4-7]. 进行面积内插法时, 必须假设某个区域的属性值分布是均匀的. 目前, 通常采用 3 种假设方法: (1) 源区域属性值分布均匀; (2) 目标区域属性值分布均匀; (3) 控制区域属性值分布均匀^{[8]2)}. 对应的面积内插方法也可以归纳为 3 种:

(i) 面积权重内插法. 假设人口在源数据区域内是均匀分布的, 具体步骤包括: (1) 在源数据区域上叠加目标数据区域; (2) 确定每个源数据区域落在目标区域的比例; (3) 根据面积比例分配人口数. 公式如下:

$$P_t = \sum_s P_s \left(A_{st} / \sum_t A_{st} \right), \quad (1)$$

其中, P_s 是源区域的人口值, P_t 是目标区域的人口值,

A_{st} 是源区域和目标区域叠加后的面积.

这种方法简单, 所需要的数据少, 缺点是误差相对较大, 而且无法对人口分布的细部进行描述.

(ii) 统计内插法. 这种方法把统计或数学模型用于面积内插, 通过样品观测值得到一个覆盖整个区域的通用函数, 然后将这个公式应用到研究区的每个个体上. 通常把土地利用单元作为目标区域, 假设人口在土地利用单元上是均匀分布的. 利用源区域的人口数, 根据(2)式进行回归计算, 获得目标区域的人口密度, 然后再计算目标区域的人口数:

$$P_s = \sum_t C_t A_{st}, \quad (2)$$

其中, P_s 是源区域的人口数, C_t 是目标区域的人口密度, A_{st} 是源区域和目标区域叠加后的面积. 如果考虑误差, 上述回归模型可表述为 $P_s = C_1 A_{s1} + C_2 A_{s2} + \dots + C_t A_{st} + \varepsilon$.

这种算法尽管在回归计算中会丢失部分原始信息, 但可以把原始的人口数据和辅助信息结合起来, 在分析过程中, 用随机的概念处理面积内插的不确定性问题.

(iii) 基于表面模型的面积内插法. 表面模型法(surface modeling)假设高分辨率的规则格网上人口均匀分布, 结合区域内的辅助信息, 采用各种内插方法(如克吕格、傅里叶分析、多项式趋势面分析、反距离权重法、最接近分析、B 样条等)对源数据进行内插, 生成规则栅格, 获得网格上的人口分布后, 将目标区域所有网格人口数进行统计, 实现面积内插^[9-11]. 这种方法在获取人口数据时, 不同的内插方法可能导致结果有较大差别, 但它的优点也很明显: (1) 数据被记录在高分辨率和规则栅格上, 更容易进行任何需要区域的人口汇总; (2) 有利于与其他栅格数据(遥感数据、高程数据)兼容, 更好地进行环境和

1) Organization for Economic Co-operation and Development (OECD). From Landcover to Landscape diversity in the European Union. <http://europa.eu.int/comm/dg06/publi/landscape/index.htm>
2) Deichmann U. A review of spatial population database design and modeling. 1996. http://www.ncgia.ucsb.edu/Publications/Tech_Reports/96/96-3.PDF

社会数据的综合分析^[12]; (3) 能更好地反映区域内人口分布是不均匀性和人口分布变化的重要信息^[13].

2 巢湖流域人口提取方法

2.1 研究区

巢湖流域位于安徽省中部, 东经 116° 24' 30"~118° 00' 00", 北纬 30° 58' 00"~32° 06' 00", 位于长江、淮河两大河流之间, 巢湖流域汇水面积为 9131 km²^[14]. 巢湖是我国著名的五大淡水湖之一, 因流域内人口急剧增长, 人类活动加强, 大量营养盐类流入巢湖, 导致湖泊水体的严重富营养化, 已被国家列为重点治理的“三河三湖”工程之一和“中国跨世纪绿色工程计划”之列, 进行整个流域生态环境评价、保护和管理已迫在眉睫.

2.2 数据源

根据 1:25 万数字高程模型(DEM)提取的巢湖流域 80 个小流域分布图, 2000 年美国陆地卫星专题制图仪(Landsat TM)数据人工解译的 1:10 万土地利用分布图¹⁾, 2000 年的巢湖流域 9 个县市人口统计数据^[15].

2.3 方法

由于原始人口数据分布点少, 在进行人口信息提取过程中, 常常需要用到辅助信息, 例如, 环境信息、地形、土地利用等等^[16,17]. 巢湖流域内的地理环境特征相似, 绝大多数区域地势起伏不大, 人口分布受地形的影响较小; 另外, 人口分布与土地利用分布呈很好的相关性. 考虑到上述因素, 本文以遥感解译的土地利用分布图为辅助信息, 结合统计模型和表面模型的优点, 采用复合面积内插获取巢湖流域的人口分布.

(i) 统计模型建立. 将巢湖流域内 9 个县市的人口数(P_s)和每个县的土地利用面积(水田面积 A_{s1} , 旱地面积 A_{s2} , 建设用地面积 A_{s3} , 林地面积 A_{s4} , 草地面积 A_{s5}), 按照下式进行回归运算:

$$P_s = C_1A_{s1} + C_2A_{s2} + C_3A_{s3} + C_4A_{s4} + C_5A_{s5} + \varepsilon, \quad (3)$$
其中, C_1, C_2, C_3, C_4, C_5 为回归系数, ε 为误差.

(ii) 人口潜力计算. 本文的人口潜力不同于承载力研究中的人口承载力, 它是计算区域内各种要素(如土地利用)的位置和大小对某点的影响或通达性^[18,19]. 计算网格人口潜力时, 假设它与土地利用的面积成正比, 与土地利用中心点的距离成反比²⁾. 采用 GIS 中的反距离权重内插法(IDW), 根据各种土地类型

的面积, 以它们多边形的中心点进行内插, 获得每个网格的人口潜力矩阵 V_{ij} , 网格大小定义为 100 m × 100 m.

$$V_{ij} = \left[\frac{\sum_{k=1}^n \frac{A_{k1}}{d_{(ij)k}^b}}{\sum_{k=1}^n \frac{1}{d_{(ij)k}^b}}, \frac{\sum_{k=1}^n \frac{A_{k2}}{d_{(ij)k}^b}}{\sum_{k=1}^n \frac{1}{d_{(ij)k}^b}}, \dots, \frac{\sum_{k=1}^n \frac{A_{k5}}{d_{(ij)k}^b}}{\sum_{k=1}^n \frac{1}{d_{(ij)k}^b}} \right], \quad (4)$$

其中, $A_{k1}, A_{k2}, \dots, A_{k5}$ 分别为第 k 个水田、旱地、建设用地、林地和草地多边形的面积, $d_{(ij)k}$ 为网格 ij 到第 k 个中心点的距离, b 为适当的距离指数. 根据这个公式计算出巢湖流域的网格人口潜力分布.

(iii) 人口信息获取. 获得巢湖流域基于网格的人口潜力分布后, 将每个网格的潜力矩阵值代入(3)式, 获得每个网格的人口数, 对流域内所有网格进行统计获得流域总人口数.

3 结果

3.1 人口分布

根据(3)式的人口回归计算, 得到回归方程如下:

$$P_s = -0.78A_{s1} + 6.93A_{s2} + 38.28A_{s3} - 2.26A_{s4} - 3.29A_{s5} + \varepsilon, \quad (5)$$

其中, 相关系数 $R = 0.986$, 变量解释率: 97.23%(图 1). 利用(5)式和网格的人口潜力, 得到人口分布的网格数据(图 2).

图 1 预测值与实际值的回归图

3.2 人口检验

由于流域内合肥市和舒城县所有面积都位于流域内, 本文对两个市县的模拟人口数和实际统计人口数进行对比. 另外, 根据 1995 年统计资料^[20]和

1) 人工解译的 1:10 万土地利用数据由中国环境监测总站遥感和 GIS 实验室王文杰主任提供
2) 见 1835 页脚注 2) <https://engine.scichina.com/doi/10.1360/csb2002-47-23-1835>
3) 中国环境状况公报'99/水环境 <http://sd-ep.cei.gov.cn/soechina99/waterc/watermain.htm>

图 2 巢湖流域人口网格分布

1995 年的土地利用图模拟得到的流域人口进行了对比(表 1)。从检验结果看,由模型获得的人口数与实际空间的人口数据符合程度很好。

表 1 模拟人口的精度

	合肥市	舒城县	巢湖流域
实际统计人口数/人($\times 10^4$)	130.28	97.77	611.8
模拟人口数/人($\times 10^4$)	112.94	102.16	574.1
精度/%	86.7	95.7	93.8

4 结语

本文以 GIS 为空间分析平台,利用遥感数据为人口分布的辅助信息,采用复合面积内插法,提取了巢湖流域的人口分布。GIS 和遥感技术的引入为流域人口信息获取提供了有效的方法,有利于实现人口数据与生态数据进行相关分析和人口分布的动态分析。下一步的研究将增强对复杂条件(地形、气候等因素)下人口分布的模拟能力,利用辅助信息(如地形、公路)对模型进行调整。

致谢 欧阳自远院士对本文提出了修改意见,特此致谢。本工作受中国科学院知识创新工程项目(ZCX2-115 和 KCX10G-D00-08)资助。

参 考 文 献

1 叶亚平, 刘鲁君. 中国省域生态环境质量评价指标体系研究, 环境科学研究. 2000, 13(3): 33~36

2 王文杰, 潘英姿, 李雪. 区域生态质量评价指标选择基础框架

及其实现. 中国环境监测, 2001, 17(5): 17~20

3 Jones K B, Riitters K H. A Landscape Atlas in Mid-Atlantic Region, An Ecological Assessment of the United States. 1997, EPA/600/R-97/130

4 Flowerdew R, Green M, Kehris E. Using areal interpolation methods in geographic information systems. Papers in Regional Science, 1991, 70: 303~315

5 Goodchild M F, Lam N S N. Areal Interpolation: A variant of the traditional spatial problem. Geo-Processing, 1980, (1): 297~312

6 Fisher P F, Langford M. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. Environment and Planning A, 1995, 27: 211~224

7 吕安民, 李成名, 林宗坚. 面积内插算法初探. 测绘通报, 2002, (1): 44~46

8 Goodchild M F, Anselin L, Deichmann U. A framework for the areal interpolation of socioeconomic data. Environment and Planning A, 1993, 25: 383~397

9 Martin D. An assessment of surface and zonal models of population. Int J Geographical Information Systems, 1996, 10(8): 973~989

10 Willmott C J, Matsuura K. Smart interpolation of annually averaged air temperature in the United States. Journal of Applied Meteorology, 1995, 34(12): 2577~2586

11 Langford M, Unwin D J. Generating and mapping population density surfaces within a geographical information system. The Cartographic Journal, 1994, 21(June): 21~26

12 Clarke J I, Rhind D W. Population data and global environmental change. Paris: IISC/UNESCO, 1992

13 Bracken I, Martin D. The generation of spatial population distributions from census centroid data. Environment and Planning A, 1989, 21: 537~543

14 屠清瑛, 顾丁锡, 尹澄清, 等. 巢湖——富营养化研究. 合肥: 中国科学技术大学出版社, 1990. 1~2

15 安徽统计局. 安徽统计年鉴 2000. 北京: 中国统计出版社, 2001

16 Moxey A, Allanson P. Areal interpolation of spatially extensive variables. A comparison of alternative techniques. International Journal of Geographic Information Systems, 1994, 8(5): 479~487

17 Langford M, Maguire D J, Unwin D J. The areal interpolation problem: estimating population using remote sensing in a GIS framework. In: Masser E, Blakemore M, eds. Handling Geographical Information, Methodology and Potential Applications. Harlow: Longman, 1991. 55~77

18 Stewart J Q, Warntz W. The Physics of population distributions. In: Berry B J L, Marble D F, eds. Spatial Analysis: A Reader in Statistical Geography. New Jersey: Prentice Hall, 1986. 130~146

19 Plane D A, Rogerson P A. The geographical analysis of population. New York: Wiley, 1994

20 李如忠. 巢湖流域生态环境质量评价初步研究. 合肥工业大学学报(自然科学版), 2001, 2(5): 987~990

(2002-06-24 收稿, 2002-08-30 收修改稿)