

DOI: 10.13718/j.cnki.xdzk.2018.09.015

# 山地丘陵区社会经济数据空间化模型构建及应用<sup>①</sup>

翁才银<sup>1,2</sup>, 信桂新<sup>3</sup>, 杨庆媛<sup>4</sup>

1. 西南大学 资源环境学院, 重庆 400715; 2. 重庆师范大学 地理与旅游学院, 重庆 401331;
3. 长江师范学院 武陵山区特色资源开发与利用研究中心, 重庆 涪陵 408100;
4. 西南大学 地理科学学院, 重庆 400715

**摘要:** 社会经济数据的空间化, 是地学研究中解决数据结构不一致、空间单元不匹配和数据在空间单元内均一化问题的有效方法. 本文基于多因素回归分析建立了社会经济数据空间化处理办法, 选取重庆市丰都县为研究区, 以人口密度和 GDP 规模为对象进行了回归建模, 为社会经济数据的空间精细化研究提供借鉴. 结果表明, 采取向后剔除法准则的多元线性回归分析数据融合方法, 较好地模拟了研究区社会经济数据的空间分布格局, 拟合方程修正后的可决系数均达到 0.80 以上; 将以乡镇为单元拟合的县域社会经济数据融合模型, 应用于村域尺度社会经济数据的空间精细化表达, 经验证数据吻合度普遍在 80% 以上, 能够较好地实现村域尺度社会经济数据的空间化. 因此, 就研究区而言, 基于多因素回归分析的社会经济数据空间化模型构建, 较好地实现了对社会经济数据的反演, 为创建区域范围内村域尺度连续的社会经济数据表面提供了支撑.

**关键词:** 非空间数据; 精细化; 多元线性回归; 模型; 精度

**中图分类号:** F224

**文献标志码:** A

**文章编号:** 1673-9868(2018)09-0096-08

在地学研究中, 尤其是对土地利用变化的研究, 不管是“解释”还是“评价”, 均趋向于对自然和人文多种影响因子的综合<sup>[1-2]</sup>, 由此对数据科学性、可靠性, 特别是对空间数据和非空间数据精细化对接的要求越来越高. 其中, 空间数据是表示空间实体位置、形状、大小及其分布特征诸方面信息的数据, 具有定位、定性、时间和空间关系等特性; 而非空间数据则是表示空间实体的空间属性以外的其他属性特征信息的数据, 是对空间数据的说明, 又称为属性数据, 主要是面向社会经济数据, 如一个城市的人口、GDP、绿化率等. 建立空间数据与属性数据间的连接, 需要对非空间数据进行空间化处理, 并应尽可能地将各类影响因子的空间分布状态置于同等分辨率下, 以便于精确地分析评估人类活动的自然—人文过程及其效应问题. 由于空间数据的实体性, 当前大量研究均能较为精细地对其表达和利用; 但以大行政区域为空间单元的统计数据却掩盖了区域内部的差异和不均衡<sup>[3]</sup>, 导致社会经济数据在研究和使用中存在数据结构不一致、空间单元不匹配和数据在空间单元内均一化等问题<sup>[4]</sup>. 为反演出社会经济数据在既定时间和既定地理空间中的分布状态, 创建区域范围连续的社会经济数据表面, 当前面向县

① 收稿日期: 2017-05-15

基金项目: 国土资源部土地资源调查评价专项项目(201315106397); 武陵山区特色资源开发与利用研究中心重点项目(WLYJ2017001); 长江师范学院引进人才科研启动项目(2017KYQD95).

作者简介: 翁才银(1973-), 男, 博士研究生, 讲师, 主要从事国土资源与区域规划、旅游规划研究.

通信作者: 杨庆媛, 教授, 博士研究生导师.

(区)域、省域口径统计数据的空间化处理方法主要采用空间插值法、土地利用/覆被影响模型法、多元数据融合分析法、遥感反演验证等<sup>[4]</sup>,揭示县域、省域内部不同空间范围上的差异<sup>[5-6]</sup>。不过,这些方法虽实现了非空间数据与空间数据的对接,但总体上仍处在探索阶段,且适用的数据类型、空间尺度等也不尽相同,其精确性普遍较差,往往难以满足非空间数据在村域尺度精细化空间表达的要求<sup>[7]</sup>。本文依托国土资源部重点区域土地生态状况调查与评估项目之“西南山区生态敏感区土地生态状况调查与评估”,选取“精细化评估”典型区域的重庆市丰都县为研究区,以社会经济数据空间化研究的焦点——人口数据和 GDP 数据为对象,应用多因素回归分析模型探讨社会经济数据空间化处理方法,期望能够为非空间数据的空间精细化研究提供参考和借鉴,为深入推进土地利用变化及其效应的精细化表达提供支撑。

## 1 基于多因素回归分析的社会经济数据空间化模型构建

### 1.1 基本原理

事物的发展变化是在与其他事物相互联系相互影响的过程中进行的,用于描述事物空间特征和属性特征的各类变量是这种相互作用关系的具体表达,可分为函数关系和统计关系两种关系<sup>[8-9]</sup>。函数关系是对一一对应的确定性关系的描述,当一个变量取一个值的时候,另一个变量按照确定的函数关系式也取一个确定的值。而当一个变量的值不能由某一变量唯一确定,或当一个变量取一定值另一变量的取值有几种可能的时候,这种关系就是统计关系。通过对大量数据相互关系的统计分析,可以确定变量间关系的强弱,并可采用相关系数( $R$ )表征,其取值在 $[0, 1]$ 之间。一般认为, $R > 0.8$ 的变量高度相关, $R$ 介于 $0.5 \sim 0.8$ 之间的变量低度相关, $R < 0.3$ 的变量相关性微弱,可视为不相关。由于相关分析与线性回归分析所进行的检验是等价的,因此可据此构建线性回归模型来进一步揭示变量间的相互关系。

非空间数据空间化处理过程中,变量关系复杂,并非一一对应,且一个变量的值难以由某一变量唯一确定,因此对非空间数据空间化的处理适用于统计关系。在进行空间化处理之前,首先拟定非空间数据空间化指标,并识别可能影响其分布的因素,进行统计显著性检验,剔除不相关的因素,进而建立非空间数据空间化的线性回归模型(式 1),最后将非空间数据空间化指标值按土地类型的最小单元分解到对应的空间范围,即可与空间数据构成完整的分析数据体系,开展土地利用变化及其效应评估研究。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \quad i = 1, 2, n \quad (1)$$

式中: $Y$ 为非空间数据空间化指标; $\beta_k$ 为偏回归系数; $X$ 为各类影响因素; $\mu$ 为残差; $k$ 为自变量个数。

### 1.2 因素识别

非空间数据主要对应社会经济数据,即人口、GDP、城市化率、工业产值、固定资产投资等统计指标数据。这些数据反映了统计单元(行政区域)之间的差异,却难以表征行政单元内部的空间差异。在此,本文选取社会经济数据中的人口数据和经济数据作为空间化处理的对象,并拟定人口密度和 GDP 规模两项指标,重点对影响其空间分布的因素加以考察。

参考相关研究<sup>[3, 10-11]</sup>,发现人口分布格局与土地利用/覆被格局具有耦合性,两者关系密切。土地利用/覆被数据,特别是耕地、林地、草地均是人类生产活动在土地利用上的直观反映,已经包含了社会经济数据的绝大部分信息;同时,在农村地区,农村建设用地特别是农村居民点用地是农村人口的直接载体,人口分布越集中,其用地比例越高;人口分布越稀疏,其用地比例越低。由此可估计,人口密度分布与土地利用/覆被指数具有很强的相关关系,科学地确定这种相关关系,可以为人口数据的空间化提供可行路径。此外,人口分布与地形条件也有紧密的联系。山地丘陵区由于地形地貌复杂,人口分布很大程度上还受到地形条件的限制,往往海拔越高,地形条件越差,人口分布相对越少。

而对于 GDP 而言,从收入角度看,等于固定资产折旧、劳动者报酬、生产税净额和营业盈余之和。在农村地区,其空间分布实际上反映了农村地区农、林、牧、渔业及乡镇二三产业的综合情况,最直观的表征就是相应土地利用类型指数的大小。因此,GDP 的空间分布也与土地利用/覆被数据密切相关;同样,类似于人口分布,也会受到地形条件的影响,在地形条件较好的区域,GDP 的面状聚集度往往越高<sup>[12]</sup>。

### 1.3 模型构建

依据非空间数据空间化处理的基本原理,在对人口和 GDP 空间分布影响因素识别的基础上,基于多元数据融合思想和土地利用/覆被的影响<sup>[4]</sup>,即可采用多元线性回归分析构建人口密度、GDP 规模空间分布模型,最终实现人口密度、GDP 数据村域尺度的精细化空间表达。

## 2 实证应用

### 2.1 研究区域概况

丰都县位于东经 107°28′03″—108°12′37″,北纬 29°33′18″—30°16′25″,面积 2 900.86 km<sup>2</sup>,是三峡库区实施生态涵养和生态屏障建设的重要区域。境内地形南高北低,长江北岸以丘陵为主,南岸以山地为主,呈“四山”夹“三槽”的地貌格局,海拔 118.5~2 000.0 m。境内气候温和,四季分明,属亚热带湿润季风气候,年均气温 18.5℃,年均日照时数 1 311.8 h,年平均降雨量 1 091 mm,无霜期 318 d。境内河流均属长江水系。土壤以酸、中、石灰性土壤为主,分别占 48.3%,17.1%,34.6%。全县现辖 30 个乡镇,从经济发展梯度来看,长江沿岸片区乡镇经济实力明显好于北部中山片区和南部高山片区。

### 2.2 数据来源与处理

依据 2015 年丰都县统计年鉴,将分乡镇的人口数据和 GDP 统计数据整理出来。土地利用数据来自丰都县国土资源和房屋管理局,并通过 ArcGIS 软件平台在土地利用现状数据库中提取分镇、分村的土地利用/覆被数据及地形数据。

数据处理方法。借助 ArcGIS 10.1,SPSS 19.0 等软件平台,第一步,将研究区行政区划图和所有土地覆被类型图转换到 Gauss\_Kruger, Xian\_1980\_3\_Degree\_GK\_Zone\_36 同一投影和坐标系下;第二步,在 ArcMap 中将土地利用类型进行调整,剔除实际无人居住或无经济生产功能的水域、滩涂沼泽、自然保留地,筛选并保留与人口分布和经济发展关系密切的建设用地、耕地(园地并入耕地)、林地、草地 4 类土地覆被类型,将研究区行政区划图与处理后的土地覆被图叠加,以“乡镇名称”字段为区域字段,进行分区统计,计算研究区的土地覆被总面积以及各类用地面积;第三步,以研究区 1:10 000 数字高程模型为地图,通过乡镇行政单元提取平均高程、坡度;第四步,将以上统计数据导出到 Excel 表中,用于数据分析和模型建立;第五步,根据第二步提取的研究区所有土地覆被类型的面积,以乡镇为单元,将每种土地覆被类型的面积分别除以各乡镇的行政区划面积,得到各乡镇的土地覆被类型指数;最后,查询统计年鉴得到研究区各乡镇的人口密度及 GDP 数据(表 1)。即基于各乡镇人口密度、GDP 规模数据与土地利用/覆被、地形条件等的相关关系,通过 SPSS 软件对人口密度、GDP 规模数据与相关影响因素数据进行多元线性回归分析,得出表征人口密度、GDP 空间分布的拟合方程,进而推算出村域尺度的人口密度、GDP 数据。

### 2.3 建立并估计社会经济数据多元线性回归模型

在模型构建过程中,自变量虽然较多,但不是每一个自变量的回归关系都有统计学意义,而对回归没有统计学意义的自变量又不能简单地删除;同时,由于自变量之间线性关系的存在,很难抉择自变量的去留。因此,为得到最优回归方程,需要对自变量进行筛选。在此,采取向后剔除法准则(F-to-remove $\geq 0.100$ 的概率),对回归模型中无统计学意义的自变量依次剔除,直到回归模型中自变量都有统计学意义为止。

表 1 丰都县 2015 年分乡镇社会经济数据及影响因子数据

乡镇	建设用地	耕地指数	林地指数	草地指数	高程	坡度	人口密度 $Y_1 /$	GDP 规模 $Y_2 /$ 万元
	指数 $X_1 / \%$	$X_2 / \%$	$X_3 / \%$	$X_4 / \%$	$X_5 / \text{m}$	$X_6 / ^\circ$	(人 · $\text{km}^{-2}$ )	
包鸾镇	2.26	19.59	66.63	2.01	883.48	17.25	167	29 201
保合镇	4.34	42.84	34.94	1.15	522.45	10.64	314	22 595
兴龙镇	4.51	37.43	39.66	2.25	423.01	10.17	329	17 530
董家镇	6.58	45.51	30.59	1.44	1 338.46	21.50	455	24 066
都督乡	1.45	14.80	76.22	2.95	697.62	15.24	57	8 083
高家镇	3.14	20.26	56.22	1.99	501.31	11.11	265	92 995
虎威镇	5.36	40.74	32.71	0.52	1 137.94	21.58	327	35 551
暨龙镇	1.64	17.37	69.50	5.12	758.24	16.04	85	18 792
江池镇	3.81	32.41	47.00	2.00	889.69	15.60	249	20 795
栗子乡	3.68	39.83	42.93	1.24	749.72	15.84	289	13 319
龙河镇	5.28	39.69	36.18	2.76	361.11	13.18	369	44 481
龙孔镇	4.69	38.32	25.60	1.66	280.51	12.66	339	27 961
名山街道	9.09	33.33	21.24	1.03	1 057.18	18.31	513	146 204
南天湖镇	1.75	17.08	70.02	3.71	515.94	12.98	120	16 578
青龙乡	4.13	36.59	46.32	1.02	359.46	11.10	263	11 953
仁沙镇	4.25	41.59	36.82	0.83	1 345.08	15.71	337	21 863
仙女湖镇	2.11	8.68	77.50	6.05	421.43	13.92	62	18 973
三合街道	13.65	33.06	26.69	0.55	624.00	23.16	1 600	400 688
三建乡	2.71	24.87	58.86	1.36	388.08	13.57	222	11 916
三元镇	4.15	36.38	42.24	0.59	344.59	7.09	295	21 023
社坛镇	7.09	52.34	19.39	0.68	374.48	11.96	484	39 593
十直镇	4.07	42.19	31.64	0.93	390.89	13.33	325	36 675
树人镇	5.22	44.16	33.12	0.88	418.03	13.16	320	28 594
双龙镇	3.64	37.71	44.77	1.09	659.24	16.81	258	19 139
双路镇	3.96	21.89	59.48	1.41	1 626.45	18.00	184	38 678
太平坝乡	1.18	9.39	74.62	8.20	1 195.03	16.91	50	8 461
武平镇	2.62	19.93	65.17	2.90	315.96	11.10	137	21 332
兴义镇	7.33	26.19	42.29	1.64	457.75	13.80	343	74 993
许明寺镇	4.89	48.04	33.15	0.76	574.98	9.25	349	16 378
湛普镇	5.03	22.11	42.29	0.66	461.03	14.32	254	65 307

以人口密度  $Y_1$  为因变量, 以建设用地指数  $X_1$ 、耕地指数  $X_2$ 、林地指数  $X_3$ 、草地指数  $X_4$ 、高程  $X_5$ 、坡度  $X_6$  为自变量, 利用 SPSS 软件进行多元线性回归分析, 最终得到回归方程如下:

$$Y_1 = -986.956 + 130.517 X_1 + 8.159 X_2 + 9.992 X_3 \quad (2)$$

统计检验显示, 方程拟合优度方面, 相关系数  $R=0.931$ , 可决系数  $R^2=0.867$ , 修正后的可决系数  $\bar{R}^2=0.851$ , 结果表明, 估计的样本回归方程较好地拟合了样本值; 同时, 通过  $F$  检验, 在给定的显著水平  $\alpha=0.05$  下,  $p=0.000$ , 所以, 总体回归方程是有统计学意义的; 另外, 通过  $t$  检验,  $\beta_0, \beta_1, \beta_2, \beta_3$  所对应的  $p$  值均小于  $\alpha=0.05$ , 所以人口密度  $Y_1$  与建设用地指数  $X_1$ 、耕地指数  $X_2$ 、林地指数  $X_3$  之间有统计学意义, 而草地指数  $X_4$ 、高程  $X_5$  和坡度  $X_6$ , 统计结果无统计学意义而被剔除。

同样, 以 GDP 规模  $Y_2$  为因变量, 以建设用地指数  $X_1$ 、耕地指数  $X_2$ 、林地指数  $X_3$ 、草地指数  $X_4$ 、高

程  $X_5$ 、坡度  $X_6$  为自变量, 利用 SPSS 软件进行多元线性回归分析, 最终得到回归方程如下:

$$Y_2 = 3\,905.035 + 30\,710.186 X_1 - 3\,035.647 X_2 \quad (3)$$

统计检验显示, 方程拟合优度方面, 相关系数  $R=0.922$ , 可决系数  $R^2=0.850$ , 修正后的可决系数  $\bar{R}^2=0.838$ , 结果表明, 估计的样本回归方程较好的拟合了样本值; 同时, 通过  $F$  检验, 在给定的显著水平  $\alpha=0.05$  下,  $p=0.000$ , 所以, 总体回归方程是有统计学意义的; 另外, 通过  $t$  检验,  $\beta_0, \beta_1, \beta_2$  所对应的  $p$  值均小于  $\alpha=0.05$ , 所以 GDP 规模  $Y_2$  与建设用地指数  $X_1$ 、耕地指数  $X_2$  之间有统计学意义, 而林地指数  $X_3$ 、草地指数  $X_4$ 、高程  $X_5$  和坡度  $X_6$ , 统计结果无统计学意义而被剔除。

## 2.4 社会经济数据多元线性回归预测结果检验

虽然构建了较为显著的社会经济数据多元线性回归模型, 但模型预测结果的精度是否达到空间化分析的要求, 还有待验证。为此, 对丰都县 330 个村(居委会)按 5% 的比例随机抽取约 15 个村(居委会), 检验人口密度  $Y_1$  和 GDP 规模  $Y_2$  的预测结果与实地调查结果的差异。其中, 将各村建设用地指数  $X_1$ 、耕地指数  $X_2$ 、林地指数  $X_3$  代入式 2(表 2), 计算各村人口密度预测值; 将各村建设用地指数  $X_1$ 、耕地指数  $X_2$  代入式 3(表 2), 计算各村 GDP 规模。而实地调查数据来自对 15 个村经济社会概况的摸底, 并参考各村 2015 年农村经济统计报表, 以及征求村镇干部意见综合予以确定。预测吻合度的检验, 就是将人口密度和 GDP 的预测结果与实地调查的各村人口密度和 GDP 进行对比分析。以人口密度为例, 用人口密度预测值与实地调查的人口密度值之差取绝对值, 除以实地调查的人口密度值, 再用 1 减去所得比值后乘以 100%。吻合度计算公式为:

$$P_i = \left(1 - \frac{|D_i - D_0|}{D_0}\right) \times 100\% \quad (4)$$

式中:  $D_i$  为非空间数据多元线性回归预测得到的  $i$  村人口密度值;  $D_0$  为实地调查的  $i$  村人口密度值;  $P_i$  为该村吻合度。

表 2 丰都县 2015 年抽样村非空间数据及影响因子数据

乡镇	抽样村	建设用地	耕地指数	林地指数	草地指数	高程	坡度
		指数 $X_1/\%$	$X_2/\%$	$X_3/\%$	$X_4/\%$	$X_5/\text{m}$	$X_6/^\circ$
包鸾镇	白果元村	0.94	10.06	82.95	1.33	1 069.95	21.09
保合镇	范家沟村	4.48	46.08	24.52	0.05	548.72	8.55
董家镇	关圣场村	5.32	54.22	17.20	5.28	370.06	12.49
都督乡	都督居委	1.22	12.87	79.64	1.87	1 254.17	24.47
虎威镇	大池居委	5.14	51.51	30.54	0.11	499.21	9.80
江池镇	大安村	4.19	42.80	36.58	1.61	776.89	16.37
栗子乡	建龙村	4.75	48.66	29.63	0.95	816.95	17.20
龙孔镇	大坝村	3.43	35.34	37.49	0.53	351.08	13.11
南天湖	厂天坝村	1.45	15.57	70.89	4.86	1 057.18	18.31
青龙乡	黄岭村	3.45	35.88	47.49	1.78	515.14	13.27
仁沙镇	打谷坝村	4.54	46.63	25.87	0.48	420.27	9.59
三合街道	刀溪村	2.02	19.37	64.61	1.34	426.88	14.09
三建乡	蔡森坝村	2.62	27.34	63.28	2.57	726.48	21.76
三元镇	大城寨村	4.55	46.76	34.52	0.76	534.99	16.64
社坛镇	蔡家庙村	5.14	52.74	25.99	0.90	234.91	9.39

由表 3 可以看出, 随机抽取的 15 个村中, 通过回归分析得到的空间化结果吻合度普遍在 80% 以上, 说明研究采用的非空间数据多元线性回归模型能够较真实地反映研究区实际情况, 能够满足研究区村域尺度社会经济数据精细化空间表达要求。

表 3 抽样村空间化精度检验表

乡镇	抽样村	统计数据		预测结果		吻合度/%	
		人口密度/ (人·km <sup>-2</sup> )	GDP 规模/ 万元	人口密度/ (人·km <sup>-2</sup> )	GDP 规模/ 万元	人口密度	GDP 总量
包鸾镇	白果元村	47	1 911	46	2 195.22	98.89	85.16
保合镇	范家沟村	264	1 483	219	1 635.24	82.89	89.72
董家镇	关圣场村	363	2 495	322	2 690.44	88.60	92.15
都督乡	都督居委	66	1 874	73	2 167.12	90.14	84.34
虎威镇	大池居委	498	4 754	409	5 389.21	82.20	86.64
江池镇	大安村	339	2 276	275	2 655.02	80.95	83.36
栗子乡	建龙村	389	1 806	326	2 063.84	83.84	85.70
龙孔乡	大坝村	143	1 633	123	1 826.64	86.09	88.17
南天湖	厂天坝村	41	1 330	38	1 169.78	91.78	87.93
青龙乡	黄岭村	220	955	231	1 061.61	94.96	88.89
仁沙乡	打谷坝村	291	1 843	245	1 899.23	84.23	96.96
三合街道	刀溪村	95	8 310	81	7 230.07	84.92	87.01
三建乡	蔡森坝村	239	1 684	210	1 371.13	88.04	81.42
三元镇	大城寨村	323	1 595	334	1 825.14	96.63	85.56
社坛镇	蔡家庙村	448	1 447	374	1 655.37	83.45	85.57

### 3 结论与讨论

基于多元线性回归分析的数据融合方法,能够较好地模拟山地丘陵区县域社会经济数据的空间分布格局.目前,社会经济数据空间化处理的方法主要包括空间内插法、土地利用/覆被影响模型、多源数据融合法和遥感反演法.面积权重内插法是空间内插法中最为简单常用的一种社会经济数据空间化处理方法<sup>[13]</sup>,其假设前提是区域的属性值是均匀分布的,因此当区域内存在湖泊、河流、沼泽等无人居住的斑块时,就需要对该方法进行改进.由此可见,该方法在样点密集的小区域应用较为合适;在尺度较大、采样稀疏或地理环境变化剧烈的区域,则很难保证其模拟结果与实际情况的吻合度.而基于高时空分辨率遥感影像提取的土地利用/覆被等空间信息,建立的社会经济数据与相应土地利用/覆被类型所占面积之间的模型关系,可以较好地实现社会经济数据的空间化<sup>[14]</sup>;但其最大问题在于,遥感影像解译过程中信息的丢失和损坏,使得提取结果存在一定误差.当然,基于合适的空间分布单元,采用相关性分析,确定影响社会经济数据空间分布的主要因子,由此建立社会经济数据与各影响因子间回归模型的多元数据融合方法,也能够得到社会经济数据的空间分布模拟图<sup>[15]</sup>;不过,其主要问题是,受指标选取影响,往往造成信息冗余和同类信息影响程度被夸大,以致增加了模拟结果的不确定性.此外,还有利用遥感数据反演社会经济特征的遥感反演法<sup>[16]</sup>,遥感数据本身的像元过饱和、像元溢出等问题以及遥感影像判读时的人为误差,均限制了其对社会经济数据模拟的精度.鉴于上述方法的优缺点,本研究中,针对西南山地丘陵区地理环境复杂多变的特征,结合人口密度和 GDP 规模的影响因子识别,以土地利用现状数据库为基础提取数据,一方面最大程度地避免了遥感影像提取导致的信息遗失和破坏;另一方面,以相关性分析判定变量间关系的强弱,以多元线性回归分析进行数据融合,并依据向后剔除法准则剔除统计结果无统计学意义的因子,既建立了社会经济数据与影响因子(土地利用/覆被类型)间的关系,又降低了信息冗余度,拟合方程修正后的可决系数  $\bar{R}^2$  均达到 0.80 以上,可以说较好地模拟了研究区社会经济数据的空间分布格局.

多元线性回归模型能够较好地实现村域尺度社会经济数据的空间精细化表达.村是农村最小、最基本的社区,是一个系统完备的社会单元,具有社会结构的基本构成——经济、政治、文化<sup>[17]</sup>.典型区域的案例研究是认识全球变化的重要途径,但离不开微观尺度的土地利用研究,而村域尺度的土地利用行为能够为此提供依据.随着对自然和人文多种影响因子综合的研究趋向加强,对村域尺度社会经济数据空间精细化

表达的要求也愈益迫切。然而,目前宏观—中观尺度的社会经济数据空间化是研究报道的重点,如全国尺度的GDP密度分布图<sup>[18]</sup>、人口密度空间化<sup>[19]</sup>,省域尺度以县为单元的人口分布<sup>[20]</sup>、GDP空间分布<sup>[21]</sup>等,关于村域尺度社会经济数据精细化表达的研究尚不多见。本研究通过多元线性回归分析,将以乡镇为单元拟合的县域社会经济数据融合模型,应用于村域尺度社会经济数据的空间精细化表达,并随机抽取5%的村,参考各村2015年农村经济统计报表,以及征求村镇干部意见综合确定各村人口密度和GDP规模,对照多元线性回归模型预测结果进行验证,结果显示吻合度普遍在80%以上,说明该模型能够较好地用于村域尺度社会经济数据的空间化,对开展土地利用变化相关的典型区域案例研究具有支撑作用。

村域尺度社会经济数据空间化的实现有助于深化对典型区域土地利用变化原因和土地生态安全状况的研究。以往在数据处理过程中,获得和使用的统计数据都是通过自下而上逐级汇总获得的,所以基本上都是各级行政单位为数据的收集和统计单位,但这种数据的获取和表达方式并不能揭示这些数据内在的空间差异特征。很多实际研究中,这些数据难以与典型区域自然因素数据进行叠加分析。如在土地生态安全评估中,生态压力、生态建设保护与发展协调指标涉及的人口密度、GDP规模数据,若要与气象数据、土壤数据、土地利用覆被数据等进行叠加分析,而不细化社会经济数据的统计单位(村),这些叠加分析将难以进行。同时,若仅仅采用简单的将统计数据平均分配到各个栅格的方法,不仅无法体现评估的科学性、精细化,更会致使评估结果同质化,也不能体现土地生态安全状况在评价单元(村)间的连续性和差异性。因此,村域尺度社会经济数据的空间化,实质上是对传统社会经济统计数据的二次挖掘,赋予了其新的内涵。本研究中,虽然模型的普适性还有待进一步检验,但就研究区而言,较好地实现了对社会经济数据的反演,为创建区域范围内村域尺度连续的社会经济数据表面提供了支撑;同时,研究也发现,在山地丘陵区社会经济数据空间化处理过程中,建设用地、耕地、林地具有重要的社会经济指示意义,是进行社会经济数据空间化处理的关键变量,而表征地形地貌的高程、坡度指标在村域微观尺度下被剔除,没有预期的重要,从某种意义上看,这为该模型向同类区域或平原区的推广提供了依据,值得进一步深入探讨。

## 参考文献:

- [1] 解利剑,周素红,闫小培.近年来中国城市化与全球环境变化研究述评[J].地理科学进展,2010,29(8):952-960.
- [2] 李承桢,信桂新,杨朝现,等.传统农区土地利用与覆被变化(LUCC)及其生态环境效应[J].西南大学学报(自然科学版),2016,38(5):139-145.
- [3] 李军,胡云峰.我国经济社会统计数据空间均匀程度分析[J].地域研究与开发,2009,28(3):130-134.
- [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] 李云波,沈岐平.房地产空间数据与非空间数据的关联——门牌号的研究[J].哈尔滨建筑大学学报,2000,33(6):117-120.
- [6] 康相武,吴绍洪,戴尔阜,等.区域社会经济财产数据空间化方法研究[J].中国软科学,2006(8):104-108.
- [7] 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.
- [8] 汪海春,王闯,姜润东.股票市场与人民生活相关的分析[J].时代报告(学术版),2014(10):355.
- [9] 曾子芳.测量数据处理中基于SPSS的回归分析[J].工程技术(文摘版),2016(5):295-296.
- [10] 廖一兰,王劲峰,孟斌,等.人口统计数据空间化的一种方法[J].地理学报,2007,62(10):1110-1119.
- [11] 任斐鹏,江源,董满宇,等.基于遥感和GIS的流域社会经济数据空间化方法研究[J].长江科学院院报,2015,32(3):112-116.
- [12] 赵军,杨东辉,潘竟虎.基于空间化技术和土地利用的兰州市GDP空间格局研究[J].西北师范大学学报(自然科学版),2010,46(5):92-96.
- [13] 闫庆武,卞正富,赵华.人口密度空间化的一种方法[J].地理与地理信息科学,2005,21(5):45-48.
- [14] 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.
- [15] 熊俊楠,韦方强,苏鹏程,等.基于多源数据的四川省GDP公里格网化研究[J].应用基础与工程科学学报,2013,

21(2): 317—327.

- [16] 王雪梅, 李 新, 马明国. 基于遥感和 GIS 的人口数据空间化研究进展及案例分析 [J]. 遥感技术与应用, 2004, 19(5): 320—327.
- [17] 丁耀琪. 村级社区的特性及其档案工作 [J]. 档案与建设, 1999(1): 45—46.
- [18] 韩向娣, 周 艺, 王世新, 等. 基于夜间灯光和土地利用数据的 GDP 空间化 [J]. 遥感技术与应用, 2012, 27(3): 396—405.
- [19] 王英安. 中国人口空间分布数字模拟的曲面建模方法 [D]. 北京: 中国科学院地理科学与资源研究所, 2006.
- [20] 王 磊, 蔡运龙. 人口密度的空间降尺度分析与模拟——以贵州猫跳河流域为例 [J]. 地理科学进展, 2011, 30(5): 635—640.
- [21] 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.

## Building of a Spatialization Model of Socioeconomic Data in Mountainous and Hilly Regions and Its Application

WENG Cai-yin<sup>1,2</sup>, XIN Gui-xin<sup>3</sup>, YANG Qing-yuan<sup>4</sup>

1. School of Resources and Environment, Southwest University, Chongqing 400715, China;
2. School of Geography and Tourism, Chongqing Normal University, Chongqing 401331, China;
3. Research Center for Development and Utility of Unique Resources in the Wulingshan Region, Yangtze Normal University, Fuling Chongqing 408100, China;
4. School of Geographical Sciences, Southwest University, Chongqing 400715, China

**Abstract:** In geoscientific research, spatialization of socioeconomic data is an effective way to solve such problems as inconformity of data structure, mismatching of space units and data homogenization in the space units. In a case study, this paper develops a method based on multiple regression analysis and makes regression modeling with population density and GDP of Fengdu County of Chongqing as research objects, so as to provide a reference for spatio-detailed research of socioeconomic data. The results show that in the study area the data fusion method based on multiple regression analysis with the criterion of “backward regression” can well simulate the spatial distribution pattern of the socioeconomic data, and the revised coefficients of determination of all the fitted equations are more than 0.80. When the data fusion model fitted with town as the unit are applied to spatio-detailed presentation of socioeconomic data under the village-scale domain, spatialization of village’s socioeconomic data is satisfactorily realized, the goodness of fit of data between the predicted value and the actual value being generally more than 80%. Therefore, for this study area, the spatialization model of socioeconomic data based on multi-factor regression analysis has a relatively precise inversion to socioeconomic data, and supplies a new way to build regional continuous socioeconomic data surface under the village-scale.

**Key words:** non-spatial data; detailed; multiple linear regression; model; precision