

一个非均匀密度分布的人口房屋格网数据模型^①

刘 军¹, 孙甲宁¹, 宋立军¹, 李志强², 谭 明¹, 韩贞辉³

(1.新疆维吾尔自治区地震局,新疆 乌鲁木齐 830011; 2.中国地震局地质研究所,北京 100029;

3.河南省地震局,河南 郑州 450016)

摘要:介绍了人口、房屋数据空间化的方法,以伊犁地区为例建立了一个符合非均匀密度分布的公里格网人口和房屋数据模型,并将其进行空间化,最后利用实地调查数据对结果进行了验证。通过数据检验,公里格网的人口密度数据、房屋密度数据客观地反映了伊犁地区统计单元内承灾体的空间分布,具有较高的精度,为地震应急救援和决策提供可靠的依据。

关键词:统计数据;公里格网;现场调查;伊犁地区;震害快速评估

中图分类号: P315

文献标志码: A

文章编号: 1000-0844(2016)增刊 2-0296-06

DOI: 10.3969/j.issn.1000-0844.2016.Supp.2.0296

A Grid Data Model of Population and Building with Heterogeneous Density Distribution

LIU Jun¹, SUN Jia-ning¹, SONG Li-jun¹, LI Zhi-qiang², TAN Ming¹, HAN Zhen-hui³

(1.Earthquake Administration of Xinjiang Uygur Autonomous Region, Urumqi 830011, Xinjiang, China;

2.Institute of Geology, China Earthquake Administrator, Beijing 100029, China;

3.Earthquake Administration of Henan Province, Zhengzhou 450016, Henan, China)

Abstract: This paper introduces the data space method of population and the housing, taking Yili area as an example to establish a consistent with km grid population and housing data model of non-uniform density distribution, and its space, finally the result was verified by the field survey data. Through data inspection, population density data, 1km grid housing density data reasonably reflect the demographic unit and the actual housing distribution, has higher precision, and provide reliable basis for earthquake emergency rescue and decision.

Key words: population and housing statistics; 1km grid; field survey; Yili area; fast assessment of earthquake disaster.

0 引言

人口、房屋建筑是地震灾害脆弱性大小的重要因素,同时也是损失评估计算的核心因子。当前,应用在历次破坏性地震灾害损失评估计算中的人口、房屋等基础数据一般都是采用行政区单元通过统计资料、现场抽样等方式逐级汇总收集,在统计单元内人口、房屋数据是一种均匀密度分布的,不能准确地反映人口、房屋等承载体的分布^[1]。

解决上述问题的主要方法就是基于统计数据,

通过建立数据模型,定量的反演出评估区内的人口、房屋等在线性时间和地理空间中的分布特征,实现基础数据在统计单元内空间化的过程。

自上世纪 90 年代以来,人口统计数据的空间化随着人口数据空间化研究的深入,形成了一系列具有代表性的模型和方法,主要有基于自然、社会经济综合特征的人口密度模型^[2]、基于夜间灯光辐射数据的人口空间分布建模^[3-4]、面积内插法^[5]、基于土

① 收稿日期:2016-03-10

基金项目:中国地震局“三结合”课题(163101);地震科技星火计划项目(XH15044Y)

作者简介:刘 军(1982—),男,硕士,工程师,主要从事地震应急与灾害学研究。E-mail:liujun_eq@sina.com。

地利用关系的人口密度模型^[6-7]等。但人口数据空间化研究仍存在问题有待解决:一方面,在研究方法上,主要注重从宏观上寻求人口分布与影响因子的统计规律,而微观尺度上人口分布的随机性没有受到足够重视,另一方面采用土地利用数据的人口密度模型过度的依赖于研究区域内的土地利用类型,单一影响因子的局限性直接影响了数据质量的可靠性。

目前为止,相比人口统计数据空间化,房屋统计数据的空间化研究还处于探索阶段,特别是应用于防震减灾领域,还没有成熟的方法。本研究以新疆伊犁地区为例,拟在人口统计数据空间化的基础上,利用公里格网实地调查的方法,建立一个服务于防震减灾的非均匀密度分布的房屋数据空间化模型,实现人口与房屋数据的空间化。

2 研究思路

根据数据格网化研究的需求、空间模拟尺度及区域特点,在已有研究的基础上对人口数据空间化方法做了以下改进:(1)在研究尺度方面,城区以街道办,农村以乡、镇及兵团团场为数据建模单元,减小了数据模型空间尺度转换的跨度;(2)在土地利用数据方法的数据空间化模型基础上,用道路、居民点和坡度作为土地利用数据的信息补充因子;(3)采用多元逐步回归分析建模,避免因子间的多重共线性;(4)采用基于公里格网的实地人口调查数据对空间化结果进行检验。

房屋统计数据的空间化是在人口数据空间化的基础上,利用六普数据结合实地调查资料,确定不同人口密度条件下人均住房面积及住房类型的比例关系,最终实现房屋数据的空间化。

3 人口统计数据空间化

3.1 模型建立

3.1.1 城区人口密度权重模型

城区人口空间分布宏观上与城镇建成区的区位密切相关,以场位论的城市人口密度理论模型中最常用的是 Clark 的距离衰减模型^[8]:

$$\rho(r) = \rho_0 \times \exp(-r/r_0)^\sigma \quad (1)$$

其中 $\rho(r)$ 表示与城市中心距为 r 处的人口密度; ρ_0 表示特征半径内的人口密度; r 为人口密度单元与特征中心的距离; r_0 为城市特征半径; σ 为反映城区人口分布的空间变化信息熵的约束常数,城市处在发展阶段一般取值为 1.2。

然而,与实际相比,仅以离城市中心距离作为影响人口分布的因素,得到的仅是一种近似结果,不能表达出城市人口分布的细节信息,需要寻求一种合理的方法从微观角度反映城区人口分布的非均匀性和随机性的因子辅助建立模型。道路分布与房屋占地面积是人口分布的重要指示因子。房屋占地面积对人口的分布,具有明确的指示作用,能够准确的反映出人口的具体分布位置,是比较理想的数据源。基于以上分析,本研究构建了伊犁地区城区人口密度权重模型:

$$w = \alpha_1 \times w_r + \alpha_2 \times w_i + \alpha_3 \times w_h \quad (2)$$

式中: w 表示城区单元区划内的人口密度的权重; w_r 表示为城区范围内距离衰减模型中单元区划人口密度的权; w_i 、表示城区范围内道路密度因子的单元区划人口密度的权; w_h 基于房屋占地面积因子的权; $\alpha_1, \alpha_2, \alpha_3$ 是权重的配比系数,其中 $\alpha_1 + \alpha_2 + \alpha_3 = 1$,且都大于 0,各参数按以下 3 个步骤进行计算:

(1) 基于距离衰减的人口密度权重 w_r

$$w_r = w(r) = \exp(-r/r_0)^\sigma \quad (3)$$

城市特征半径 r_0 的确定利用遥感影像对伊犁地区的城区及近郊进行量测估计,伊犁地区的城市特征半径取 4 km。

(2) 基于道路密度的人口密度权重 w_i :

$$w_i = w(d_r) = \begin{cases} 1.0 & d_r \geq d_h \\ 0.4 + 0.6 * (d_r - d_l) / (d_h - d_l) & d_l < d_r < d_h \\ 0.4 * (d_r / d_l) & d_r \leq d_l \end{cases} \quad (4)$$

式中: w_i 为基于道路密度的人口密度权重; d_r 为区域内道路密度; d_l 表示区域单元内道路稀疏上限阈值; d_h 表示区域单元内道路密集下限阈值。在考察伊犁城区及周边地区的道路密度后,确定 $d_l = 0.25, d_h = 0.75$ 。

(3) 权重配比系数

基于距离衰减的人口密度因子能较好的反映城市区位控制下人口空间分布的宏观特征,而基于区域道路密度的人口密度因子^[8]、房屋占地面积因子的空间分布则反映了微观尺度下人口分布的随机特征,三因子的影响比重暂定为是等量的,赋予相同权重比例系数,即: $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$ 。

3.1.2 农村人口密度权重模型

农村地区人口的空间分布主要受自然环境、居住条件以及农业生产资料等因素的空间分布影响,

且表现出很鲜明的地域特征。综合前人的研究成果,本文以地貌类型分区为基础,选取路网密度指数、土地利用指数、平均坡度和居民点指数,通过 15 个指标建立衡量影响农村人口分布的居住、自然环境条件和生产资料的空间分布模型,采用多元逐步回归分析法,建立农村人口密度权重模型。

(1) 建模因子

土地利用的空间格局是影响人口空间分布的主要因素^[9],基于土地利用的人口数据空间化已经成为大尺度人口数据空间化的主要手段。然而,利用遥感影像对人口集聚区域、地貌类型及不同尺度等土地利用数据的信息提取过程中,均存在不同程度的信息遗漏^[4],这在人口数据空间化过程中是不可忽视的。同时相关研究也证明,人口分布与道路、居民点和坡度密切相关,可用于补偿土地利用数据的误差和信息损失。

因此,研究选取土地利用、道路分布、居民点分布以及坡度 4 类数据,共 15 个建模因子作为人口分布影响因素,来衡量影响人口分布的自然环境条件和社会经济状况。

(2) 多元回归模型的建立

基于地貌类型分区,各分区选取若干乡、镇、街道及团场作为样本,通过人口密度或人口密度对数与各建模因子进行多元线性回归分析试验。各分区多元线性拟合优度值表明:人口密度与建模因子存在一定的对数相关和线性相关关系。为避免人口密度模拟值为负值的不合理现象,本文采用对数相关关系建立人口密度权重模型,模型数学表达式为:

$$Y = \exp(\beta_0 + \beta_1 \times \chi_1 + \beta_2 \times \chi_2 + \dots + \beta_n \times \chi_n) \quad (5)$$

式中: Y 为建模样本单元平均人口密度; $\chi_1, \chi_2, \dots, \chi_n$ 为逐步回归筛选出的 n 个建模因子; $\beta_1, \beta_2, \dots, \beta_n$ 为建模因子对应的回归系数, β_0 为常数项。

(3) 模型系数修正

回归拟合的结果是基于一种统计的相关关系,仍需要符合特定的地理规则,如不同土地利用类型的食物生产力对人口分布的影响力存在一定的排序原则^[7]。本研究的建模因子中不仅包含了土地利用类型,还考虑了道路密度、居民点密度以及坡度。作为信息补偿因子,道路密度和居民点密度应与城镇用地和农村居民点与人口密度分布存在较强的相关性。

3.2 人口密度图生成

在 GIS 软件的支持下,对伊犁地区的城镇行政

单元采用城镇人口密度权重模型,计算其对应公里格网的人口密度权重值;对广大的农村地区采用农村人口密度权重模型,计算其对应公里格网的人口密度权重。以乡、镇、街道单元人口统计数据为总量计算生成行政单元内公里格网人口密度。通过式(2),假设空间任意一个格网内人口密度权值表示为 w_i , pop 表示格网所在乡镇级别的人口统计总数, n 表示研究区内该乡镇的格网统计数目,则该网格内人口密度 $D_{\text{pop}}(i)$ 为:

$$D_{\text{pop}}(i) = w_i / \sum_i w_i \times \text{pop} \quad (6)$$

最后,将城区和农村对应的公里格网矢量数据图层合并,并将矢量图层导出为像元尺寸为 $1 \text{ km} \times 1 \text{ km}$ 的人口密度栅格图,如下图 1 所示。

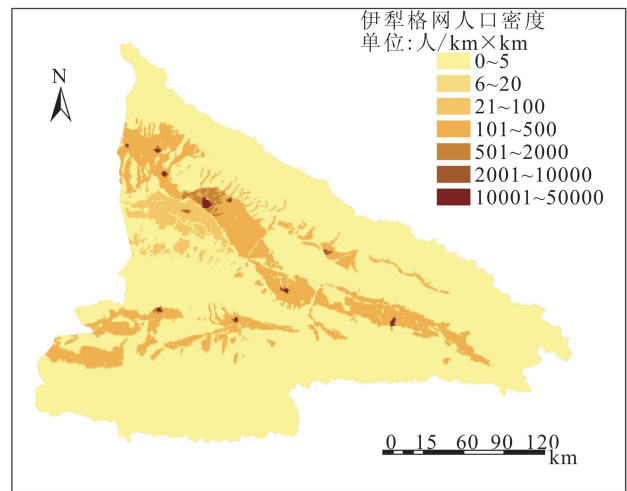


图 1 伊犁地区公里格网人口密度图

4 房屋数据的空间化

本研究利用公里格网实地调查的方法在研究区域获得足够的样本,利用公里格网人口密度分级的方法,分别获取不同格网人口密度条件下人均住房面积和不同住房类型的比例关系。

4.1 房屋数据空间化模型

房屋比例关系是房屋数据空间化过程中的难点,结合历次灾害评估报告及地震灾害快速评估的实际需要。本研究结合伊犁地区实际情况,将房屋结构类型分为土木房屋结构、砖木房屋结构、砖混房屋结构、框架房屋结构及富民安居房屋结构 5 种类型。通过现场调研,伊犁农村地区无框架结构房屋,故将农村地区房屋结构划分为土木房屋结构、砖木房屋结构、砖混房屋结构及富民安居房屋结构四种结构类型,将城区房屋结构划分为土木房屋结构、砖

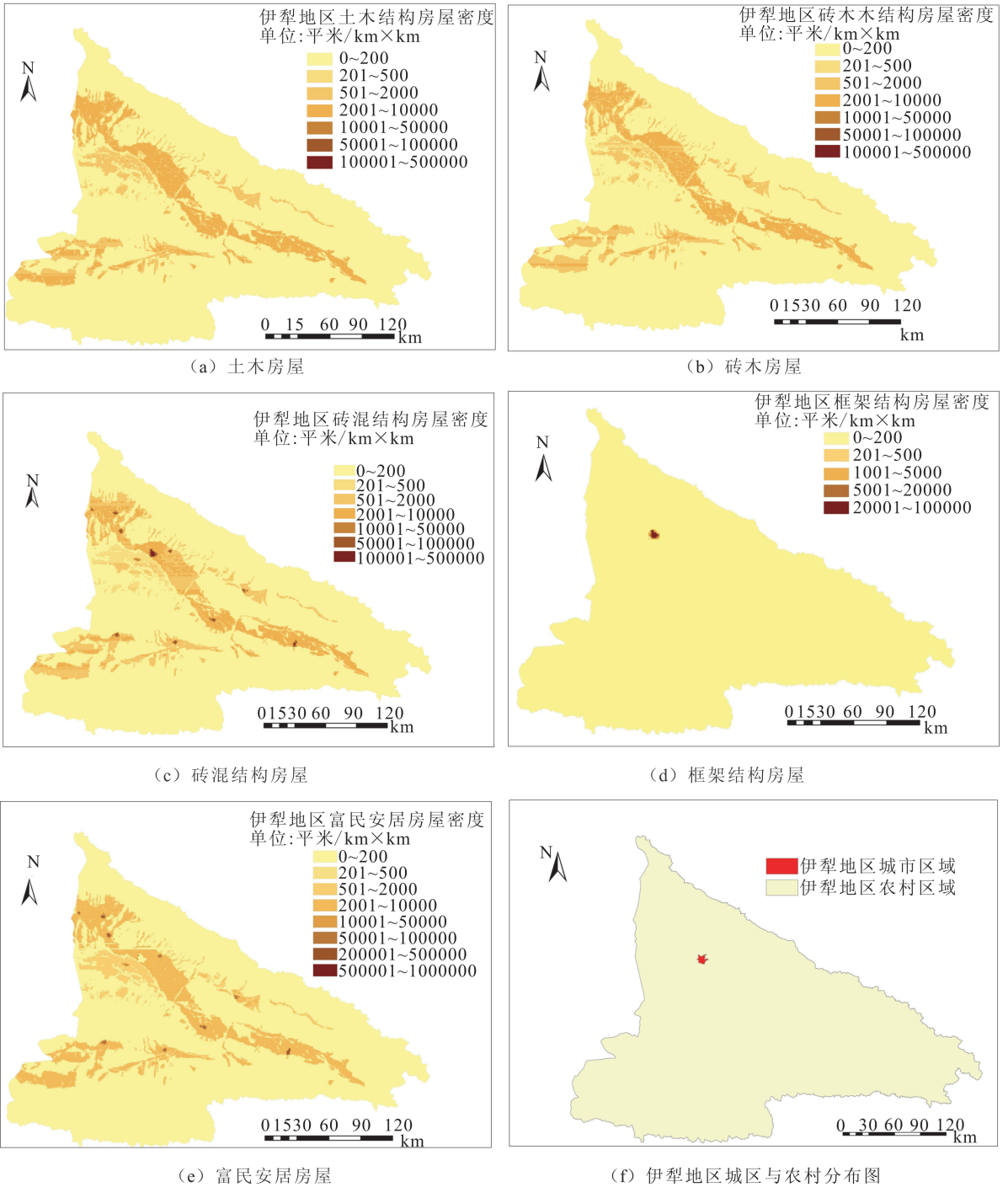


图 2 伊犁地区公里格网房屋密度图

木房屋结构、砖混房屋结构及框架结构四种结构类型。

4.1.1.1 农村房屋计算模型

行政单元内人均土木房屋面积、砖木房屋面积以六普人均土木房屋面积、砖木房屋面积数据为基础,结合相应年度房屋面积减少率,人均土木房屋面

积、人均砖木房屋面积如下式所示:

$$A_i = A_{i2010} * (1 - p\%)^n \tag{8}$$

行政单元内人均砖混房屋、富民安居房屋面积以六普人均砖混房屋、富民安居房屋面积数据为基础,结合相应年度房屋面积增长率,人均砖混房屋、富民安居房屋面积及人均框架结构房屋面积如下式所示:

$$A_j = A_{j2010} * (1 + p\%)^n \quad (9)$$

A_i 、 A_j 表示土木房屋面积、人均砖木房屋面积和人均砖混房屋、富民安居房屋面积及人均框架结构房屋面积, A_{i2000} 、 A_{j2000} 分别表示各类房屋在 2000 年的基数。

4.1.2 城市房屋计算模型

伊犁地区的城区主要由 8 个街道办构成, 房屋结构为土木房屋、砖木房屋、砖混结构、框架结构。伊犁地区城市房屋计算以街道办为单元, 采用实地抽样调查的方法, 获得抽样区内的相对于 2010 年六普房屋数据结构类型的年度变化比例, 由此计算各街道办各类房屋面积的比例; 2013 年度人均住房面积采用上式 6 中计算方式, 利用伊犁地区统计年鉴, 以 2004 年为基数计算 2013 年度城区的人均房屋面积。

4.2 房屋数据空间化的实现

在 GIS 软件的支持下, 基于房屋数据格网化模型, 将伊犁地区的 $1 \text{ km} \times 1 \text{ km}$ 人口密度数据分别乘以相应的人均住房面积就可获得公里格网房屋数据。通过合并运算, 就可以获得研究区域的房屋密度图。具体的做法如下:

$$D_{\text{bui}} = D_{\text{pop}} \times A \quad (10)$$

式中: D_{bui} 指的是空间任意一个公里格网内房屋密度, D_{pop} 指的是对应的该公里格网内人口密度, A 指的是对应该公里格网内人均住房面积。

$$D_{\text{bui}}(i) = D_{\text{bui}} \times R_i \quad (11)$$

式 11 中, $D_{\text{bui}}(i)$ 指公里格网内 i 结构类型的房屋密度。 R_i 表示该公里格网内 i 结构类型的房屋在不同

行政区划条件下占总房屋面积的比例, 在不同行政区划中所对应值不同, 其中农村地区以村为统计单元, 城区以街道办为统计单元。

通过式 10, 可以计算出空间任意一个公里格网内的房屋密度, 通过式 11, 可以计算空间任意一个公里格网内不同住房类型的房屋密度。然后在 GIS 软件中, 把伊犁地区所对应的房屋密度图进行合并计算, 可以得到该公里格网的房屋密度图及不同住房类型的房屋密度图。

从上图可以看出, 在伊犁农村地区土木结构、砖木结构等简易房屋仍然占有一定比例, 富民安居房屋具有较广的覆盖, 特别是随人口密度加大安居房密度也越大; 在城市地区框架结构房屋在伊犁城市中心位置密度较大, 市郊明显减少, 在农村地区完全无框架架构房屋, 符合事实。

4.3 数据检验

按城区、乡、镇及兵团 4 个类别, 共在伊犁地区布置了 117 个调查点进行数据检验, 获得了 117 个公里格网的实际房屋密度值。通过对公里格网房屋密度数据与现场调查格网内房屋密度数值进行误差分析, 对模拟结果的数据精度进行检验。

格网内房屋密度误差的计算方法采用式(12)进行:

$$E_{\text{bui}} = (\text{模拟数据} - \text{调查数据}) / \text{调查数据} \times 100\% \quad (12)$$

通过下表 3 的误差统计结果表明伊犁地区的房屋密度数据具有较高的精度。

表 1 伊犁地区房屋数据误差统计表

误差范围/%	土木结构		砖木结构		砖混结构		框架结构		富民安居房	
	调查点个数	所占比例/%	调查点个数	所占比例/%	调查点个数	所占比例/%	调查点个数	所占比例/%	调查点个数	所占比例/%
<-30	3	10.71	2	9.09	3	10.71	1	5.88	3	9.68
-30~-20	3	10.71	1	4.55	2	7.14	2	11.76	3	9.68
-20~-10	4	14.29	4	18.18	4	14.29	2	11.76	5	16.13
-10~0	4	14.29	5	22.73	6	21.43	3	17.65	7	22.58
0~10	6	21.43	5	22.73	5	17.86	4	23.53	5	16.13
10~20	3	10.71	2	9.09	4	14.29	2	11.76	4	12.90
20~30	3	10.71	2	9.09	2	7.14	1	5.88	2	6.45

5 结 语

本文通过对伊犁地区人口、房屋数据的格网化研究, 获得高质量的格网人口和格网房屋数据的方法, 并对模拟的公里格网人口数据进行检验。通过将公里格网人口密度数据与研究区内的遥感影像、居民分布、道路及河流数据叠加分析, 发现建筑物分布与所生成的人口密度存在较好的空间耦合关系。研究首

次采用实地调查数据进行了人口密度结果精度检验以及结合六普数据及实地调查数据确定房屋类型的比例关系, 生成的房屋密度数据质量具有较高的精度, 达到了预期目标。同时在建模过程中, 作为信息补偿因子, 道路密度和居民点分布较好的弥补了传统土地利用数据带来的信息损失, 取得了较好的效果。

基于公里格网人口和房屋数据的震害评估中将明显优于传统的基于平均密度的方法, 能够满足实际

应用要求,可以为防震减灾工作特别是地震灾害损失评估中提供可靠的依据。

参考文献

- [1] 陈振拓.服务于地震应急的人口数据格网化方法研究[D].中国地震局地质研究所,2012.
- [2] David Martin.An Assessment of Surface and Zonal Models of Population[J].International Journal of Geographical Information Systems,1996,10(8):973-989.
- [3] Sutton P.Modeling Population Density with Night-time Satellite Imagery and GIS[J].Computers,Environment and Urban Systems,1997,45(2):95-118.
- [4] Lo C P.Modeling the Population of China using DMSP Operational Linescan System Nighttime Data[J].Photogrammetric Engineering and Remote Sensing,2001,22(7):333-339.
- [5] 吕安民,刘海启,李成名,等.人口密度的面积内插算法研究[J].中国农业资源与区划,2002,23(1):35-40.
- [6] 廖顺宝,李泽辉.基于人口分布与土地利用关系的人口数据空间化研究[J].自然资源学报,2003,23(01):117-122.
- [7] 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.
- [8] Clark, Colin.Urban Population Densities[J].Journal of the Royal Statistical Society.Series A,General,1951.
- [9] 江东,杨小唤,王乃斌,等.基于RS、GIS的人口空间分布研究[J].地球科学进展,2002,17(5):734-738.
- [10] Lo C P.Modeling the Population of China Using DMSP Operational Linescan System Nighttime Data[J].Annals of the Association of American Geographers,2002,92(2):224-240.
- [11] Newling B E.The Spatial Variation of Urban Population Densities[J].Geographical Review,1969.
- [12] Tobler W,Deichmann U,Gottsegen J.World Population in a Grid of Spherical Quadrilaterals[J].International Journal of Population Geography,1997,36(5):363-366.