

中国国土全局空间统计

赵作权, 宋敦江

中国科学院科技政策与管理科学研究所, 北京 100190

摘要 利用全局空间统计方法,估计中国国土在地球球面空间上的整体特征。提出基于空间格局的全局空间随机抽样方法,对中国数字高程模型(DEM, 1:400万)数据进行抽样,分别估计了中国国土的地形曲面和几何轮廓的空间格局参数,其中包括重心、方位、地理范围和空间密集性等。结果表明,中国地形和几何轮廓具有不同的重心、方位、地理范围和空间密集性,揭示了中国国土空间格局的复杂性;中国国土空间格局参数的估计误差很小,反映了全局空间随机抽样方法的有效性。

关键词 全局空间统计;空间抽样;中国国土;地形;几何轮廓

中图分类号 K909

文献标识码 A

文章编号 1000-7857(2009)24-0056-05

Global Spatial Statistics of the Territory of China

ZHAO Zuoquan, SONG Dunjiang

Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China

Abstract In view of describing the territory of China as a whole from a statistical perspective, this paper gives an estimate of the global spatial characteristics of the territory of China using global spatial statistics (or statistics of spatial distribution). A design-and pattern-based method of spatial sampling, or the global spatial random sampling, is proposed, and applied to the China's Digital Elevation Model (DEM, 1:4 000 000). The parameters of the spatial pattern, including the center of gravity, orientation, geographical range and spatial intensity, are estimated for the DEM, which is decomposed into a terrain surface and a geometric outline. The results demonstrate that both the surface and the outline differ remarkably in the center of gravity, orientation, geographical scope and spatial intensity, revealing the complexity of the territory of China on the global surface; the small margins of errors show the efficiency and effectiveness of the proposed pattern- and design-based approach of spatial sampling.

Keywords global spatial statistics; spatial sampling; China's territory; terrain; geometric outline

0 引言

国土作为国家地理标志,对一个国家经济、人口、政治和军事的空间组织具有十分重要的影响^[1]。一个以国土为参照系的社会经济空间分布图通常能够显示人口经济等分布的不均衡性、随机性、集中性和聚集性^[2-3]。不同国家因国土状况不同,常有不一样的社会经济布局。中国人口主要集中在东部平原地区,显示了中国地形包括高山、平原和沿海的控制作用;美国人口经济主要沿美国大陆边界分布,体现了沿海沿湖地区的巨大影响^[4]。近年建立的中国国家基础地理信息系统、国家社会经济统计地理信息系统等大型数据库,均以国土空间为支撑平台^[5]。因此,定量分析中国国土具有重要的认知价值和现实意义。

中国国土是一个在地球球面空间上分布的、具有其独特而复杂几何特征的区域。人们时常用“地域辽阔”、“西高东低”、“屹立于世界东方”等术语概括中国国土的特征。目前已有一些学者对中国地形的起伏度、粗糙度等进行研究、分析^[6-8],但有关中国国土空间的全局或整体特征的统计研究还很少^[2,9]。本研究旨在利用全局空间统计方法^[9],从地球球面空间上的全局特征的角度对中国国土进行估计、研究,首次提出基于空间格局的全局空间抽样方法,对中国数字高程模型数据(DEM, 1:400万)进行抽样,分别估计了中国国土的地形曲面和几何轮廓的空间格局参数,其中包括重心、方位、地理范围和空间密集性。研究结果表明,中国地形和几何轮廓具有不同的重心、方位、地理范围和空间密集性,揭示了中国国土空间格局的复杂性。

学界早就认识到国土的整体特征并通过全局空间统计方法确定这些特征。例如,美国统计局^[10]和 Neft^[12]确定了美国、中国、英国、澳大利亚、日本、巴西、印度 7 个国家国土的地理

收稿日期: 2009-08-11

作者简介: 赵作权(中国科协所属全国学会个人会员登记号:S110006679M),研究员,研究方向为空间统计、聚集创新、空间经济、区域发展政策,电子邮箱:zhao1@casipm.ac.cn

重心;Bachi^[11]计算了美国、法国、意大利、以色列、瑞典 5 个国家的地理范围。以往的研究主要关注空间分布在二维平面的整体特征^[12-13],所涉及的数据量不大,似乎没有必要使用空间抽样方法^[29]。与以往研究相比,本研究有 3 点不同:① 提出并使用针对空间格局整体的抽样方法即全局空间抽样方法(global spatial sampling);② 关注空间格局的球面特征,考虑了球面对空间格局描述的影响;③ 不仅关注国土空间的重心和地理范围,还分析它的方位和密集性。

中国国土作为一个几何体具有十分复杂的边界和地形。它不像点状分布那样可以直接进行空间抽样。要准确分析像中国国土空间这样尺度巨大的地理研究对象,需将它剖分成若干形状规则、面积微小的面积单元(如栅格)。这种剖分越精细、分辨率越高,统计计算和估计越准确。精细的空间剖分导致国土空间的样本总量巨大,这是需要空间抽样的根本所在。这里使用的中国 DEM 数据涉及抽样单元高达 950 万,显然需要使用比较有效的空间抽样设计。在此,提出一种针对空间格局整体特征的空间抽样方法即全局空间抽样。这是一种基于空间格局(spatial pattern)的抽样:通过多次(格局)抽样将国土空间转换成空间格局总体并由此空间格局各种空间整体特征指标的概率分布;同时,全局空间抽样是一种基于设计(design-based)的抽样:它确保每一个样本个体或抽样单元有同样的机会被抽到,同时每一个样本格局有同样的机会被抽到,因此这种抽样避免了空间自相关问题^[14-15]。以往的全局空间抽样是基于抽样单元的一次性格局抽样,不能有效估计一个空间格局的各个整体特征^[2],而一般的空间抽样方法关注空间变异而不是空间格局的全局特征^[16-17]。

1 方法与数据

1.1 空间格局统计指标

分别给出在球面坐标系下空间格局的重心、离散性、密集性和方位的测度指标^[18,19]。

图 1 表示一个球面坐标系下一般的空间格局,其中虚线椭圆表示以重心 C 为圆心、以球面圆弧 D 为半径的球面圆,代表球面空间格局 W 的平均展布范围。空间格局的属性总量 W 散布在球面空间若干个;有属性数量 w_i 在区位 $i(i=1,2,$

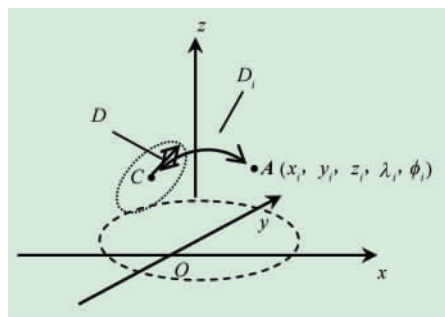


图 1 球面坐标系空间格局统计示意
Fig. 1 Statistics of spatial pattern under global coordinates

$\dots, n$) 分布,对应的球面经纬坐标分别为 λ_i 和 ϕ_i ,对应的三维直角坐标系坐标为 (x_i, y_i, z_i) ; C 为 W 分布的重心点,用坐标 (ϕ, λ) 或 (x, y, z) 表示,而 D_i 是点 A 到点 C 的球面大圆弧距离。

这里面对的是一个属性不均一的空间格局,因此空间格局的 4 个空间特征统计指标以加权形式表达如下。

1.1.1 重心

通过重心的三维坐标与球面坐标变换^[19],重心的球面坐标为

$$\sin \phi = \frac{\sum_i w_i \sin \phi_i}{\sqrt{\left(\sum_i w_i \cos \phi_i \cos \lambda_i\right)^2 + \left(\sum_i w_i \cos \phi_i \sin \lambda_i\right)^2 + \left(\sum_i w_i \sin \phi_i\right)^2}} \quad (1)$$

$$\tan \lambda = \frac{\sum_i w_i \sin \lambda_i \cos \phi_i}{\sum_i w_i \cos \lambda_i \cos \phi_i} \quad (2)$$

1.1.2 空间离散性指标——径向(平均)距离

在球面坐标系下,径向(平均)距离 D 为

$$D = \frac{\sum_i w_i D_i}{\sum_i w_i} \quad (3)$$

这里,

$$D_i = \arccos(\sin \phi_i \sin \phi + \cos \phi_i \cos \phi \cos |\lambda_i - \lambda|)$$

其中, D 的单位是球面大圆弧角度,可以转变成球面大圆弧距离。

1.1.3 空间密集性指标

在球面空间上,空间密集指数 I 需要考虑球面曲率对计算距离和面积的影响

$$I = \frac{\sum_i w_i}{2\pi R^2(1 - \cos D)} \quad (4)$$

其中, D 为圆弧 D 在球面上与球心所对应的角度, R 为完美球体地球的半径, $R=6371$ km。式(4)表明空间密集指数 I 与 W 呈正比,与 D 呈反比。

式(4)可以用来区分两种不同地理尺度的聚集:如果一个空间格局的 D 明显小于它所在的封闭地理空间的 D ,则该空间格局处于局域聚集状态;如果该空间格局的 D 明显大于它所在的封闭地理空间的 D ,则该空间格局处于全局聚集状态(如果该空间格局的 D 等于它所在的封闭地理空间的 D ,则该空间格局很可能没有任何程度的聚集状态)。

1.1.4 方位

在球面坐标系下,设点 i 到经过分布重心 C 的任意球面大圆 G 的距离为 d_i ,则全空间格局到该大圆的加权平均距离的最小值为

$$d = \min \left\{ \frac{\sum_i w_i d_i}{\sum_i w_i} \right\} \quad (5)$$

其中,与球面大圆 G 相垂直,并同时经过重心 C 与球心的球面大圆 P 代表该空间格局的方位,用 P 和同时经过重心和球心的经度线所对应的球面大圆的夹角表示。

1.2 数据与全局空间抽样

中国国土(包括大陆、台湾和海南岛)数据由中国数字地形模型(DEM,1:400 万)、中国边界两部分组成。它包含 9 498 198 个尺度为 1 km×1 km 的正方形网格(grid),其中东西向最多有 4 845 个栅格,南北向最多有 4 045 栅格。中国国土体积为 17 374 922 064 km³。中国 DEM 数据和边界数据来源于中国科学院资源环境科学数据中心提供的中国基础空间数据。

图 2、图 3 反映了中国国土在地球球面空间上围绕其重心点的径向统计分布。图 2 显示中国地形为单峰陡峭的分布,展现了中国高原地形的局域性和高度密集性;图 3 显示中国几何轮廓为单峰舒缓倾斜的分布,展现了中国国土的地域广阔性和相对密集性。这两个曲线通过以中国地形重心(格尔木附近)、国土几何重心(兰州附近)为中心一系列等面

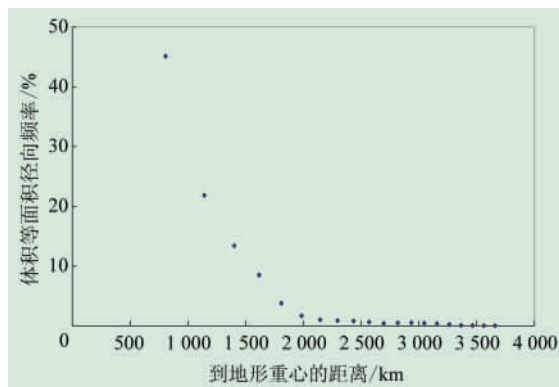


图 2 中国国土地形体积等面积径向频率分布
Fig. 2 Equal-area-based radical frequency distribution of the terrain of China by volume

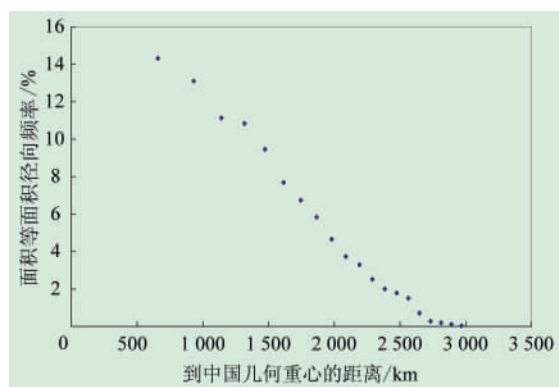


图 3 中国国土面积等面积径向频率分布
Fig. 3 Equal-area-based radical frequency distribution of the outline of China by area

积同心圆环(concentric rings)获得。

中国 DEM 由近 950 万个形状和面积相同或非常接近的栅格构成。DEM 单位经度的长度在北纬 26°、30°、35°、41°分别为 6 093、6 088、6 091、6 081 个栅格。这种划分基本克服了单位经度的长度随纬度下降的影响,能够保证对中国国土进行空间随机抽样。

全局空间随机抽样:使用随机二维布尔矩阵方法,从含有约 950 万个样本个体(栅格)中随机抽取 950 个栅格,形成一个抽样格局,使得每一个栅格都有相同的概率被抽中;再从余下的栅格中随机抽取 950 个栅格,构成一个新的抽样格局,同样使每一个栅格都有相同的概率被抽中;由此重复,共获得 20 个抽样格局,其中包括 19 000 抽样栅格,占中国 DEM 全部栅格的 2/1000。图 4 显示其中的一个抽样格局。它从统计意义讲代表着中国国土的整体特征。

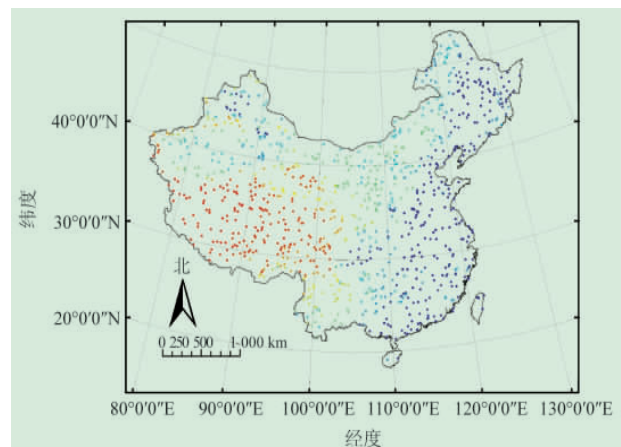


图 4 中国国土空间随机抽样分布示意(从深蓝色到红色的颜色系列代表地形高程的逐渐增加)

Fig. 4 Distribution of a sampling pattern for the territory of China (the color sequence, from blue to red, represent the gradual increase in elevation)

2 结果

图 5 显示出中国地形和几何轮廓抽样格局的重心、径向距离的分布。这两个图表明,中国国土的空间格局参数(这里是重心和径向距离)具有相当高的分布集中性。针对中国地形来讲,其抽样格局的重心分布于青藏高原的中心地带(格尔木南),其抽样格局的径向距离围绕着青藏高原的主体部分;针对中国几何轮廓来讲,其抽样格局的重心分布于中国国土的中心地带(兰州南),其抽样格局的径向距离包围着中国国土从北京到西藏的半壁江山。

根据 20 个抽样格局的空间格局指标(重心、径向距离、方位和空间密集指数)估计中国国土空间格局参数,发现中国国土地形和几何轮廓(边界)具有不同的重心、径向距离、方位和空间密集性(图 6 和表 1)。中国地形的球面重心在(94.629 0°E,34.475 8°N),位于格尔木南 217.0 km 处,通天河南侧的俄麻生多村附近,球面径向距离为 983.83 km,方位为

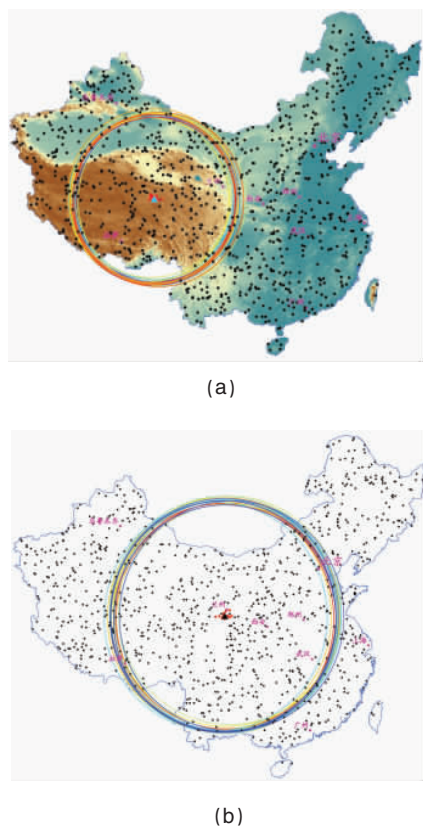


图 5 中国地形(a)和几何轮廓(b)抽样格局(20 个)的重心和径向距离分布示意

Fig. 5 Distribution of the center of gravity and radical distance of the 20 sampling patterns (20) for the terrain (a) and the outline (b) of China

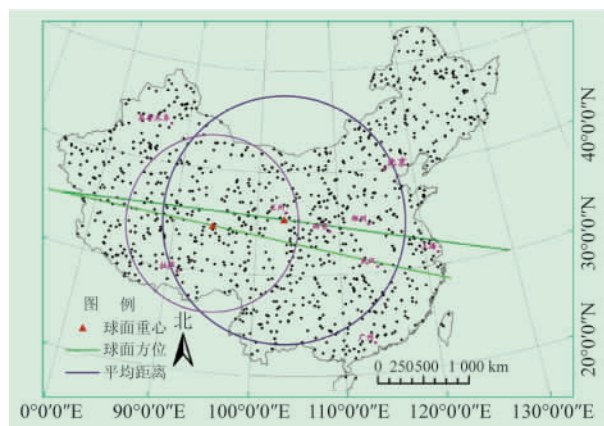


图 6 中国国土地形和几何轮廓的空间格局参数估计值示意(重心、径向距离和方位分别由红色三角形、紫色圆环和绿色直线代表)

Fig. 6 Distribution of the estimated parameters of the spatial pattern of the terrain and outline of China (the center of gravity, radical distance, and orientation are represented, respectively, by red triangles, purple circles and green straight lines)

表 1 中国国土地形和几何轮廓的空间格局参数估计

Table 1 Estimated parameters of the spatial pattern of the terrain and outline of China

空间格局参数	中国地形		中国几何轮廓	
	均值	95%边际误差	均值	95%边际误差
重心经度	94.629 0	±0.182 1	103.320 3	±0.218 8
重心纬度	34.475 8	±0.082 1	35.460 2	±0.101 8
方位(北东)	91.72	±2.50	97.29	±3.08
径向距离/km	983.83	±8.89	1 344.48	±8.15
空间密集度	5.731	±0.102	1.676	±0.020

北东 91.72° ,密集度为 $5.731 \text{ km}^3/\text{km}^2$ 。中国几何轮廓的球面重心在 $(103.320 \text{ }^{\circ}\text{E}, 35.460 \text{ }^{\circ}\text{N})$,位于兰州西南方向 78.0 km 处,和政县城附近,球面径向距离为 $1\,344.48 \text{ km}$,方位为北东 97.29° ,密集度为 $1.676 \text{ km}^2/\text{km}^2$ 。比较中国地形和几何轮廓,其重心相距 799.2 km ,反映中国地势西高东低;径向距离相差 360.7 km ,反映中国高地形主要分布在西部;空间方位基本一致,地形展布方向比几何轮廓展方向布偏北 5.57° 。这里,兰州市的地理位置为 $(103.750 \text{ }^{\circ}\text{E}, 36.068 \text{ }^{\circ}\text{N})$,格尔木为 $(94.883 \text{ }^{\circ}\text{E}, 36.416 \text{ }^{\circ}\text{N})$ 。

表 1 分别给出了中国国土地形和几何轮廓重心、径向距离、方位和空间密集指数的 95% 边际误差。这表明在 100 个抽样格局中,中国国土地形和几何轮廓的空间格局参数真值会落在其中 95 个抽样格局的置信区间之中。这里重心的边际误差可以从球面坐标对转变成到重心的球面距离;其中中国地形重心的 95% 边际误差是一个半径为 21.60 km 的球面圆,中国几何轮廓重心的 95% 边际误差是一个半径为 26.00 km 的球面圆。这里把中国地形和几何轮廓数据转变成空间格局总体,其空间特征包括重心、径向距离、方位以及密集指数。根据中心极限定理,不管这个空间格局总体是否服从正态分布,它是样本量为 1 万的大样本,因此空间格局样本的重心均值、径向距离均值、方位均值以及密集指数均值的抽样分布都将随着样本量的增大而趋向正态分布^[20]。这里使用样本量为 20 的小样本,空间格局样本的各个均值经过标准化以后的随机变量服从自由度为 19 的 t 分布,即一种比正态分布更平坦和分散的对称分布^[20]。因此,本研究采用 t 分布建立空间格局总体各个均值的置信区间。这里边际误差的计算公式为 $t_{0.025} \cdot s/n^{1/2}$,其中 $t_{0.025}=2.131, n=20$ (样本容量)。

3 讨论

3.1 空间统计误差的来源

空间统计误差并不像一般统计误差那样完全来源于抽样^[14]。它不仅与抽样方法和样本量有关,也与 DEM 等空间(分布)数据特点密切相关。首先,空间统计误差与边界效应有

关。中国国土几何轮廓是复杂的连续空间分布,不可能与各式各样的空间格网结构完全叠合,常有信息丢失;其次,空间统计误差高程分布的与空间切割(或分区)效应有关。空间数据转换(如分辨率变低、空间参照系转变、空间划分方法变化、矢量数据栅格化)常常导致空间信息扭曲或丢失;再次,空间统计误差与球面效应有关。球面空间分布不容易转变成由相同形状和面积栅格构成的均一结构,从而影响空间随机抽样的有效性^[21]。特别是这三种效应常常相互关联,和空间抽样方法一起共同制约着全局空间统计包括本研究的误差估计。

3.2 空间格局数据的双层结构

全局空间统计分析空间格局数据的双层结构,它并不像一般统计那样侧重数据属性值的大小。本研究中,使用统计方法探讨高程的空间分布而不是它的频度分布;高程的空间分布远比高程的径向等面积分布复杂,如图2高程的径向分布只是反映了地理范围和密集性,并没有反映高程分布的方位和重心。这里,把高程的空间分布和它的足迹(footprint)分布(几何轮廓)区分开,后者是纯几何分布;同时,确定二者的空间整体特征即它们在球面空间的重心、地理范围、密集性和方位,其中重心和方位表达这两个分布与地球球面空间的关系(位置和方向),地理范围和密集性表达这两个分布自身的空间属性。特别是几何轮廓分布成为识别高程分布的聚集性、不均衡性和随机性提高了参照标准。例如,中国地形的径向距离远远小于中国几何轮廓的径向距离,表明中国地形分布的空间聚集性;中国地形的重心远离中国几何轮廓的重心,表明中国地形分布的空间不均衡性;只有当中国地形(或其他自然、社会、经济要素)和中国几何轮廓具有相同的重心、径向距离和方位,中国地形才可能是完全空间随机的。

4 结论

利用全局空间统计方法估计中国国土在地球球面空间上的整体特征。提出了基于空间格局的全局空间随机抽样方法,对中国数字高程模型数据进行抽样,分别估计了中国国土地形曲面和几何轮廓两者的空间格局参数(包括重心、方位、地理范围和空间密集性)及其边际误差。研究表明,中国地形和几何轮廓具有不同的重心、方位、径向距离和空间密集性,揭示了中国国土空间格局的复杂性。同时,中国空间格局参数的边际误差很小,反映了全局空间随机抽样方法的有效性。

针对中国国土的全局空间统计有待更深入、更全面的研究。下一步工作可以探讨其他抽样方法如空间分层抽样的适用性和有效性,本研究只是拓展了简单随机抽样方法;其次,分析中国国土的形状特征,因为形状和重心、地理范围、密集程度和方位等一样是国土等空间格局重要的整体特征;再次,将本研究方法用于其他尺度中国DEM数据以及其他国家如美国的DEM数据,进而比较不同尺度条件下国土空间格局参数的差异和变化规律;最后,把本研究方法用于非连续空间分布如点状、线状或网络状分布。当然,本研究方法也可以用于研究全球变化、土地利用等各种涉及大地理尺度球

面空间现象的演化规律。

参考文献 (References)

- [1] Davis D R, Weinstein D E. Bones, bombs, and break points: The geography of economic activity[J]. *American Economic Review*, 2002, 92: 1269-1289.
- [2] Neft D S. Statistical Analysis for Areal Distributions [M]. Philadelphia: Regional Science Research Institute, 1966.
- [3] Zhao Z, Stough R R, Li N. Note on the measurement of spatial imbalance [J]. *Geographical Analysis*, 2003, 35: 170-176.
- [4] Rappaport J, Sachs J D. The United States as a coastal nation[J]. *Journal of Economic Growth*, 2003(8): 5-46.
- [5] 杨宽宽, 宋雪清, 安凯. 国家社会经济统计地理信息系统的设计与实现[J]. *地理信息世界*, 2005(3): 32-34.
Yang Kuankuan, Song Xueqing, An Kai. *Geomatics World*, 2005(3): 32-34.
- [6] 封志明, 唐焰, 杨艳昭, 等. 中国地形起伏度及其与人口分布的相关性[J]. *地理学报*, 2007, 62(10): 1073-1082.
Feng Zhiming, Tang Yan, Yang Yanzhao, et al. *Acta Geographica Sinica*, 2007, 62(10): 1073-1082.
- [7] 李军, 游松财, 黄敬峰. 基于GIS的中国陆地表面粗糙度长度的空间分布[J]. *上海交通大学学报: 农业科学版*, 2006, 24(2): 185-189.
Li Jun, You Songcai, Huang Jingfeng. *Journal of Shanghai Jiaotong University: Agricultural Science Edition*, 2006, 24(2): 185-189.
- [8] 涂汉明, 刘振东. 中国地形起伏度研究[J]. *测绘学报*, 1991, 20(4): 311-319.
Tu Hanming, Liu Zhendong. *Acta Geodaetica et Cartographica Sinica*, 1991, 20(4): 311-319.
- [9] 赵作权. 地理空间分布整体统计研究进展 [J]. *地理科学进展*, 2009(1): 1-8.
Zhao Zuoquan. *Progress in Geography*, 2009(1): 1-8.
- [10] US Census Bureau. Centers of Population Computation for 1950, 1960, 1970, 1980, 1990 and 2000[R]. Washington, DC, 2001.
- [11] Bachi R. Standard distance measures and related methods for spatial analysis[J]. *Papers-Regional Science Association*, 1962, 10: 83-133.
- [12] Kellerman A. Centographic measures in geography [J]. *Concepts and Techniques in Modern Geography*, 1981: 32.
- [13] Tobler W. Global spatial analysis [J]. *Computers, Environment and Urban Systems*, 2002, 26: 493-500.
- [14] de Grujter J. Spatial sampling schemes for remote sensing [C]//Stein A, et al. ed. *Spatial Statistics for Remote Sensing*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1999: 211-242..
- [15] de Grujter J, ter Braak C J F. Model-free estimation from spatial samples: A reappraisal of classical sampling theory [J]. *Mathematical Geology*, 1990, 22(4): 407-415.
- [16] de Grujter J, Brus D, Bierkens M, et al. Sampling for natural resource monitoring[M]. Springer, 2006.
- [17] Stehman S V, Overton W C. Spatial sampling [C]//Arlinghaus S L ed. *Practical Handbook of Spatial Statistics*. Boca Raton, FL: CRC Press, 1996.
- [18] Zhao Z. Spatial distribution and economic growth: The U.S. economy 1969-2000[M]. Germany: VDM Verlag, 2008.
- [19] Aboufagel E, Austin D. A new method for computing the mean center of population of the United States [J]. *Professional Geographer*, 2006, 58: 65-69.
- [20] 贾俊平. 统计学[M]. 2版. 北京: 中国人民大学出版社, 2006.
Jia Junping. *Statistics* [M]. 2nd ed. Beijing: China's Renmin University Press, 2006.
- [21] White D, Kimerling A J, Overton W S. Cartographic and geometric components of a global sampling design for environmental monitoring[J]. *Cartography and Geographic Information Systems*, 1992, 19(1): 5-22.

(责任编辑 陈广仁)