

空间尺度变换对归一化植被指数估计的影响

颜培东, 张蒙蒙, 刘晓英 (西南林业大学, 云南昆明 650224)

摘要 用地理信息系统(GIS)获取、描述和表达栅格地理信息的尺度问题,可抽象为在欧氏空间上进行最小几何图形单元划分,以及从划分的样本集中抽取全部或小样本对总体参量估计的影响问题,以及尺度大小对真实世界几何图形表达的逼近问题,显然后面的问题是尺度越小越好。由于欧氏空间中的几何体划分的复杂性、最小几何图形单元定义的自由性,经典的概率论与数理统计理论还未能完全阐明自由个体形状、尺度、划分方法等对地理信息描述的影响。该研究以归一化植被指数为研究对象,借助地理信息系统软件,利用对常用的5 m×5 m、10 m×10 m、30 m×30 m 栅格尺度表达的已知总体,分别进行了全部自由个体的统计估计和检验,同时也进行了小样本抽样估计和检验。结果表明,地理空间的栅格尺度或地面抽样调查的几何单元大小发生了改变,但对总体的总值和均值的估计或表达没有影响。该研究结果对地理信息系统的研发和建设,对野外空间抽样框的设计,具有一定的指导意义。

关键词 空间尺度;地理信息系统;总值;均值;归一化植被指数

中图分类号 S127;TP79 **文献标识码** A **文章编号** 0517-6611(2015)36-371-03

The Impact of Spatial Scale on the Estimation of NDVI

YAN Pei-dong, ZHANG Meng-meng, LIU Xiao-ying (Southwest Forestry University, Kunming, Yunnan 650224)

Abstract The issue of using geographic information system (GIS) to obtain, describe and express the spatial scale of raster geographic information can be abstracted as the division of minimum geometric figure in the Euclidean space, the impact of all or small sample extracted from the sample set on the estimation of parameters, the approximation of size of spatial scale to the geometric figure in the real world. Obviously, to the last problem, the smaller, the better. Owing to the complexity of geometry division in the Euclidean space and the freedom definition of the minimum geometric graphics unit, classic theories of probability and mathematical statistics have not yet fully illuminated the influence of the shape, spatial scale, partitioning method, etc. of free individual on the description of geographic information. In this paper, with the help of geographic information system software, taking the NDVI as the research object, we separately statistically estimate and test the freedom individual as well as sampling estimate and test the small sample, for commonly used known population expressed by 5m×5m, 10m×10m, 30m×30m. The results show that, the totals and means of the population are not changed. That is to say, the change of grid scale of geographical space or the change of the ground sampling geometry unit size does not affect the estimation and expression of the totals and means of the population. The results of the study have some guiding significance to the geographic information system development and the design of the sampling frame in the wild.

Key words Spatial scale; GIS; Totals; Means; NDVI

GIS 中的空间尺度,一般是指最小空间单元的大小、形状。空间尺度是地学研究的基本问题,大量研究表明,随着空间尺度的改变,对同一地物的表达、描述结果也会有所差别:遥感监测得到的海岸线长度是与遥感图像的空间分辨率密切相关,采用遥感图像的空间分辨率越高,得到的海岸线长度将越长^[1];不同空间尺度下的海岸线理论长度变化量表现出随尺度增加而增大的“边际效应”^[2];河流长度的提取依赖于空间尺度,一般来说,空间尺度越小,河流长度越大,可显示的细节也越多^[3-5];Kustas 等^[6]研究了遥感土地覆盖类型面积估计误差与空间分辨率的关系,发现土地覆盖类型面积随空间分辨率的变化而变化;Arbia^[7]提出了经典的“可塑性面积单元问题(Modifiable areal unit problem MAUP)”对多光谱遥感影像最大似然分类精度的影响,空间分辨率越低,分类误差越大,像元间的空间依赖性可抑制部分误差的增长;在尺度较小并且地物景观比较复杂的城市地区,高分辨率的 QuickBird 影像能够更好地观测到小范围地区的 NDVI 值^[8]。

以上研究都是针对尺度大小对真实世界的几何图形或细节表达的逼近问题,而对于影像的空间尺度效应,从宏观即总体的总值和均值等方面的研究较少。鉴于此,笔者以归一化植被指数(NDVI)为研究对象,从影像本身的总值和均

值出发来研究影像的空间尺度效应。

1 数据来源与研究方法

1.1 数据来源 该研究采用的数据是2006年12月3日成像的云南省昆明市宜良县 QuickBird 数据,其多光谱波段空间分辨率为2.44 m,全色波段空间分辨率为0.61 m。

1.2 原理和方法 假定一个二维欧氏空间上的永远栅格数据模型表达的已知总体,利用重采样技术进行几种尺度不同的栅格自由个体划分,以及野外抽样模拟试验,利用统计检验方法,研究了规则格网几何单元尺度、随机抽样技术对总体总值和均值估测的影响。

1.2.1 归一化植被指数(NDVI)。归一化植被指数(NDVI)被定义为近红外波段与可见光波段数值之差和这2个波段数据之和的比值^[9]。即:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (1)$$

其中,在 QuickBird 影像中第4波段(0.76~0.90 μm)对应于近红外波段 NIR,第3波段(0.63~0.69 μm)对应于红波段 R。

1.2.2 空间数据的重采样。重采样方法是获取不同空间分辨率遥感影像的主要手段之一。对遥感影像进行重采样的主要方法有:最邻近法(Nearest Neighbor)、双线性内插法(Bilinear Interpolation)、三次卷积法(Cubic Convolution)。该研究采用最邻近法,因最邻近法直接将某像元位置最邻近的像元值作为该像元的新值。该方法具有简单、处理速度快

且不会改变原始栅格值的优点,适用于离散型数据(如土地利用数据、土壤类型数据等)的尺度转换^[10]。

1.2.3 自由个体划分方法。欧氏空间中自由个体的几何形状、面积,个体间的重叠、分离、相切都是人为确定的。如图1所示,圆形为个体,点为个体颗粒—离散化的现实世界的点。个体形状和大小的自由裁量和选定,是地理空间抽样中的特殊性问题,它直接影响调查的精度和成本,影响调查的有效性和可靠性,但相关广泛性和共性的理论较少,特别是结合GIS手段,在欧氏空间模拟、逼近和管理的定量化研究较少。鉴于此,该研究借助GIS软件对NDVI进行不同的自由个体划分,并研究其对总体总值和均值的影响。

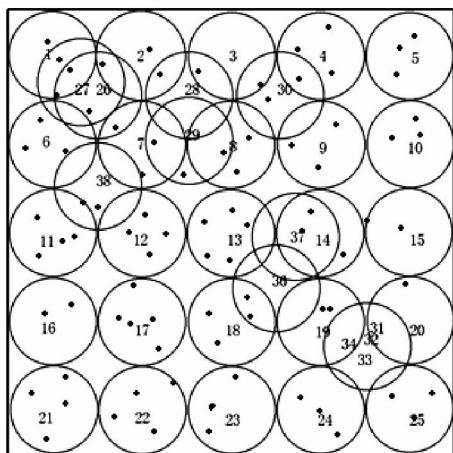


图1 欧氏空间中圆形个体概念

以空间分辨率5 m的NDVI为例,借助GIS软件,把分析窗口设为(303 439.257 839 7,2 745 637.453 32,315 539.257 839 7,2 758 837.453 32)通过以下公式:

$$\begin{cases} X_{\min} = 303\ 439.257\ 84 + [\text{random } 0\ 5] \\ Y_{\min} = 2\ 745\ 637.453\ 32 + [\text{random } 0\ 5] \end{cases} \quad (2)$$

式中, X_{\min} 为起始点坐标移动后的 X 值; Y_{\min} 为起始点坐标移动后的 Y 值。

对起始坐标点进行移动, X_{\max} 、 Y_{\max} 相应地加上 X_{\min} 、 Y_{\min} 移动的距离,在此基础上采用最邻近重采样法对原始空间分辨率为2.44 m的NDVI数据进行重采样。由于起始坐标点移动,使自由个体产生新的划分。

1.2.4 统计假设检验。统计假设检验是先对总体的分布规律做出某种假说,然后根据样本提供的数据进行统计运算,根据运算结果,对假说做出肯定或否定的决策^[11]。因该研究对NDVI的描述都集中于总值和均值,故主要检验试验组和对照组的NDVI值(μ_1 和 μ_2)的差异性,其步骤为:

(1)建立虚无假设,即先认为两者没有差异,用“ $H_0: \mu_1 = \mu_2$ ”表示。

(2)通过统计运算,确定假设 H_0 成立的概率 P 。

(3)根据 P 的大小,判断假设 H_0 是否成立。 P 值与 H_0 的关系见表1。

常用的假设检验方法有 Z 检验(u -检验)、 T 检验、卡方检验、 F -检验法、秩和检验等。该研究主要采用 Z 检验法对

全部总体进行统计假设检验,用 T 检验法对抽样数据进行统计假设检验。

表1 P 值与 H_0 的关系

P 值	H_0 成立概率大小	差异显著程度
$P \leq 0.01$	H_0 成立概率极小	差异非常显著
$P \leq 0.05$	H_0 成立概率较小	差异显著
$P > 0.05$	H_0 成立概率较大	差异不显著

2 结果与分析

2.1 全个体统计检验方法 Z 检验属于统计假设检验的一种,它适用于大样本(样本容量大于30)的2个平均数之间的差异显著性检验。它是通过计算2个平均数之间差的 Z 分数与规定的理论 Z 值的对比分析结果,判定2个平均数的差异是否显著的一种差异显著性检验方法。

表2 P 值与 $|Z|$ 值的关系

$ Z $	P 值	差异显著程度
$ Z \geq 2.58$	$P \leq 0.01$	差异非常显著
$ Z \geq 1.96$	$P \leq 0.05$	差异显著
$ Z < 1.96$	$P > 0.05$	差异不显著

该研究采用的 Z 检验方法是:通过检验来自2组样本平均数的差异性,判断它们各自代表的总体的差异是否显著。其 Z 值计算公式为:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (3)$$

式中, \bar{X}_1 、 \bar{X}_2 为样本1、样本2的平均数; S_1 、 S_2 为样本1、样本2的标准差; n_1 、 n_2 为样本1,样本2的容量。

该研究 Z 检验的样本来自12幅NDVI影像,其中,空间分辨率为5 m、10 m、30 m各4幅。以空间分辨率为5 m的NDVI为例,通过不同自由个体划分技术和重采样技术相结合的方法,得到4幅影像。其中第一幅是原始NDVI即起始坐标点不动重采样得到的,其余3幅则是用起始点坐标加上random函数产生的随机值即起始坐标点改变后再进行重采样而得到的。借助GIS工具,通过describe函数可得到总体的均值,该研究所采取的总值的计算公式为:

$$\text{总值} = \text{均值} \times \text{行数} \times \text{列数} \times \text{单元大小} \quad (4)$$

试验结果见表3~5。

表3 5 m分辨率的NDVI起始坐标点及总值和均值

序号	起始点坐标		总值	均值
	X	Y		
1	303 439.257 839 7	2 745 637.453 320 0	47 650 705.08	0.298 339
2	303 440.257 839 7	2 745 641.453 320 0	47 648 628.72	0.298 326
3	303 444.257 839 7	2 745 642.453 320 0	47 653 899.48	0.298 359
4	303 443.257 839 7	2 745 641.453 320 0	47 648 309.28	0.298 324

先对5 m分辨率的NDVI数据进行统计检验,根据公式(3)对起始点坐标不动的总体和起始点坐标移动后的总体进行 Z 检验,结果表明:起始点坐标为(303 440.257 839 7, 2 745 641.453 32)、(303 444.257 839 7, 2 745 642.453 32)和

(303 443. 257 839 7, 2 745 641. 453 32) 的总体与起始坐标点为(303 439. 257 839 7, 2 745 637. 453 32) 的总体 Z 检验得到的 $|Z|$ 值分别为 0. 029 552、0. 045 462、0. 034 098, 明显低于 1. 96, 表明它们各自代表的总体的差异不显著。由此可以得出起始坐标点的移动即不同的自由个体划分方法对总体无影响。

对起始坐标点为(303 439. 257 839 7, 2 745 637. 453 32) 的空间分辨率为 5 m、10 m、30 m 的 NDVI 数据进行 Z 检验, 结果见表 6。

表 4 10 m 分辨率的 NDVI 起始坐标点及总值和均值

序号	起始点坐标		总值	均值
	X	Y		
1	303 439. 257 839 7	2 745 637. 453 320 0	47 599 594. 68	0. 298 019
2	303 442. 257 839 7	2 745 647. 453 320 0	47 608 059. 84	0. 298 072
3	303 441. 257 839 7	2 745 642. 453 320 0	47 603 108. 52	0. 298 041
4	303 448. 257 839 7	2 745 641. 453 320 0	47 607 420. 96	0. 298 068

表 5 30 m 分辨率的 NDVI 起始坐标点及总值和均值

序号	起始点坐标		总值	均值
	X	Y		
1	303 439. 257 839 7	2 745 637. 453 320 0	47 547 808. 31	0. 297 941
2	303 440. 257 839 7	2 745 664. 453 320 0	47 536 317. 97	0. 297 869
3	303 453. 257 839 7	2 745 646. 453 320 0	47 636 220. 06	0. 298 495
4	303 459. 257 839 7	2 745 657. 453 320 0	47 564 565. 05	0. 298 046

表 6 不同空间分辨率的 NDVI 数据的 Z 检验结果

分辨率	5 m	10 m	30 m
5 m	-	0. 92	0. 42
10 m	0. 92	-	0. 08
30 m	0. 42	0. 08	-

由表 6 可知, 起始坐标点相同、空间分辨率不同的 NDVI 之间的 $|Z|$ 值明显低于 1. 96, 表明它们各自代表的总体的差异不显著。由此可以得出空间分辨率的不同对总体无影响。

表 7 空间分辨率相同、起始坐标点不同的两独立样本的 T 检验结果

NDVI	方差齐性检验		平均值 t 检验				偏差的 95% 置信区间		
	F 值	Sig. 值	t 值	df 值	Sig. 双边检验	平均差	标准误差	最低值	最高值
假定方差相等	0. 311	0. 579	0. 537	58. 000	0. 594	0. 022 741	0. 042 388	-0. 062	0. 107 591
假定方差不等			0. 537	57. 447	0. 594	0. 022 741	0. 042 388	-0. 062	0. 107 609

由表 7 可知, 对于 NDVI 来说, F 值为 0. 311, 相伴概率为 (P 值) 0. 579, 大于显著性水平 0. 05, 不能拒绝方差相等的假设, 认为空间分辨率为 10 m、起始点坐标为(303 442. 257 839 7, 2 745 647. 453 32) 的 NDVI 数据和空间分辨率为 10 m、起始点坐标为(303 439. 257 839 7, 2 745 637. 453 32) 的 NDVI 数据方差无显著差异; 方差相等时 T 检验的结果中 t 统计量的相伴概率为 0. 594, 大于显著性水平 0. 05, 不能拒绝 T 检验的零假设。也就是说, 被检验的两个总体的平均值不存在显著差异。此外从样本均值差的 95% 置信区间看, 区间跨 0, 也说明被检验的 2 个总体的平均值无显著差异。

在此基础上对不同空间分辨率不同自由个体划分的 NDVI 数据进行 Z 检验, 其结果都低于 1. 96。

综上, NDVI 数据总体的总值和均值不受空间分辨率的影响, 同样也不受不同自由个体划分方法的影响。

2.2 随机小样本估计方法 T 检验是适用于小样本(样本容量小于 30 时) 2 个平均值差异程度的检验方法。它是用 T 分布理论来推断差异发生的概率, 判定 2 个平均数的差异是否显著。 T 检验分为 3 类: 单样本 T 检验、两独立样本 T 检验以及两配对样本 T 检验。该研究采用两独立样本 T 检验, 以检验来自 2 个总体样本平均数的差异性, 并以此判断它们各自代表的总体的差异是否显著。

借助 GIS 软件, 采用放回随机抽样^[12] 的方法对不同空间分辨率不同自由个体划分的 NDVI 数据进行随机抽样。以空间分辨率为 5 m、起点坐标为(30 3439. 257 839 7, 2 745 637. 453 32) 的 NDVI 数据为例, 研究区共有 2 640 行, 2 420 列, 用 random 函数随机产生 30 对行列值, 用以下公式计算出随机抽取的第 m 行第 n 列的左下角坐标:

$$\begin{cases} X_{\min} = 303\ 439.257\ 8397 + (m-1) \times 5 \\ Y_{\min} = 2\ 745\ 637.453\ 32 + (n-1) \times 5 \end{cases} \quad (5)$$

式中, X_{\min} 为随机抽取的第 m 行第 n 列所对应像元的左下角 X 坐标值; Y_{\min} 为随机抽取的第 m 行第 n 列所对应像元的左下角 Y 坐标值。

用 sample 函数得到随机抽取的 30 对行列号对应像元的像元值。对不同空间分辨率、不同自由个体划分的 NDVI 数据分别进行随机抽样。

用 SPSS 20.0 软件进行独立样本的 T 检验。以 10 m 分辨率的 NDVI 数据为例, 在 SPSS 20.0 软件中对起始点坐标为(303 442. 257 839 7, 2 745 647. 453 32) 和起始坐标点为(303 439. 257 839 7, 2 745 637. 453 32) 的采样数据进行独立样本的 T 检验, 结果见表 7。

由以上分析可知, 被检验的 2 个总体无显著差异, 即对于空间分辨率相同的 NDVI 数据来说, 其总体不受自由个体划分方法的影响。

按照以上分析方法对起始点坐标为(303 441. 257 839 7, 2 745 642. 453 32) 和(303 448. 257 839 7, 2 745 641. 453 32) 样本数据与起始点坐标为(303 439. 257 839 7, 2 745 637. 453 32) 的样本数据进行独立样本的 T 检验, 其相伴概率分别为 0. 446、0. 266, 大于显著性水平 0. 05, 表明起始坐标点移动对总体无影响。

对起始坐标点为(303 439. 257 839 7, 2 745 637. 453 32)

(下转第 378 页)

参考文献

- [1] 马颂德,张正友. 计算机视觉:计算理论与算法基础[M]. 北京:科学出版社,1998.
- [2] 崔岩. 基于双目立体视觉的距离测量[D]. 长春:长春理工大学,2007.
- [3] 田国会,李晓磊,赵守鹏,等. 家庭服务机器人智能空间技术研究与进展[J]. 山东大学学报(工学版),2007,37(5):53-59.
- [4] 徐冠华,田国良,王超,等. 遥感信息科学的进展和展望[J]. 地理学报,

- 1996,51(5):385-397.
- [5] 马玉真,胡亮,方志强,等. 计算机视觉检测技术的发展及应用研究[J]. 济南大学学报(自然科学版),2004,18(3):222-227.
- [6] 乔毅. 移动机器人双目视觉的三维重构[D]. 哈尔滨:哈尔滨工程大学,2007.
- [7] 徐杰. 机器视觉中摄像机标定 Tsai 两步法的分析与改进[J]. 计算机工程与科学,2010,32(4):45-48.

(上接第373页)

的空间分辨率为5 m、10 m的NDVI数据进行两独立样本的

T检验,结果见表8。

表8 空间分辨率不同、起始坐标点相同的两独立样本的T检验结果

NDVI	方差齐性检验		平均值t检验				偏差的95%置信区间		
	F值	Sig.值	t值	df值	Sig. 双边检验	平均差	标准误差	最低值	最高值
假定方差相等	0.755	0.388	0.807	58.00	0.423	0.031 9	0.039 51	-0.047	0.110 96
假定方差不等			0.807	57.92	0.423	0.031 9	0.039 51	-0.047	0.110 97

按照分析表7同样的方法分析可知,表8中被检验的2个总体没有显著差异。说明起始坐标点不变或分析窗口不变,仅改变空间分辨率对总体无影响。

在以上研究的基础上,对起始坐标点为(303 439.257 839 7,2 745 637.453 32)的空间分辨率为5 m、10 m、30 m的NDVI数据两组合并进行独立样本的T检验,结果表明5 m与10 m、10 m与30 m分辨率的总体的相伴概率分别为0.113、0.578,其相伴概率明显高于显著性水平0.05,表明了起始坐标点相同或分析窗口相同的NDVI总体不受空间分辨率的影响。

对不同空间分辨率,不同起始点坐标的NDVI数据分别进行独立样本的T检验,其相伴概率都要高于显著性水平,表明空间分辨率及起始坐标点的不同对总体无影响。因总体不变,故可得出空间尺度的差异对总体总值、均值估计无影响。

3 结论与讨论

3.1 结论 该研究借助GIS软件,以5 m×5 m、10 m×10 m、30 m×30 m为空间栅格单元,以不同的坐标原点、不同坐标系剖分出多种不同栅格数据集表达同一已知总体;对不同数据模型下形成的总体分别进行了全部自由个体的统计检验,抽取部分样品进行假设检验,结果表明,这些样本的总值、均值都来自同一总体,即地理空间的栅格尺度、建立栅格数据的坐标原点、或野外调查的样地大小发生了改变,对总体的总值和均值的估计没有影响。

显然,利用不同空间尺度、不同坐标系建立地图的栅格数据表达模型,对其表达的地理指标的均值、总值估计没有影响,但前提是指标与栅格形状、栅格的尺度无关,这对于描述与几何形状或尺度无关的地理指标、或指导野外抽样调查是有意义的。

由于空间尺度即栅格大小的改变,影响了地理指标表达的图像的精细程度,影响了其在空间中的连续分布格局的现实逼近及地图的可视化表达和理解应用,但这与地理指标的统计估计不是同一问题。换言之,该研究建立的几个数据模

型是为了估计地理指标,而在不需要强调地面格局精确化表达时,不见得需要栅格分辨率越细越好。

3.2 讨论

(1)该研究中因起始坐标点移动而使研究区有小部分变动,对空间分辨率为5 m、10 m、30 m的NDVI数据而言,因random函数产生的在随机值分别为0~10、0~30、0~100,故由于起始坐标的移动而产生的误差在1个像元之内。

(2)数据重采样产生的误差。最邻近重采样法最大会产生0.5个像元大小的位移。因NDVI是植被生长状态及植被覆盖度的最佳指示因子,且在植被覆盖的情况下,NDVI为正值,并随植被覆盖度的增大而增大;对于云、雪、水等而言,因其在可见光波段比近红外波段有较高的反射作用,故其NDVI为负值;岩石、裸土在两波段有相似的反射作用,因而其NDVI值接近于0^[9]。故从研究意义上讲,该研究在随机采样的过程中只取了正值,而忽略了0值和负值。

参考文献

- [1] 张华国,黄韦良. 基于分形的海岸线遥感信息空间尺度研究[J]. 遥感学报,2006,10(4):463-468.
- [2] 刘勇,黄海军,严立文. 不同空间尺度下石臼陀岛海岸线提取的遥感应用研究[J]. 遥感技术与应用,2013,28(1):144-149.
- [3] 李静静,陈健,朱金玲. 基于DEM的河流长度尺度转换与不确定性分析[J]. 人民长江,2010,40(8):55-58.
- [4] 沈涛,苏山舞. 基于DEM的河流多尺度显示研究[J]. 测绘科学,2003,28(2):20-22.
- [5] 张宏才. 不同尺度数字高程模型提取水系的尺度效应[D]. 西安:西北大学,2003.
- [6] KUSTAS W P, HUMES K S. Sensible heat flux from remotely-sensed data at different resolutions[M]//STEWART J B, ENGMAN E T, FEDDES R A, et al. Scaling up in hydrology using remote sensing. New York: John Wiley & Sons, 1996:127-145.
- [7] ARBIA G, ESPA G. Effects of the MAUP on image classification[J]. Geographical systems, 1996, 3:123-141.
- [8] 杨培玉,陈圣波,吴琼,等. 城市地区ETM和QUICKBIRD影像NDVI值比较研究[J]. 遥感技术与应用,2008,23(5):69-72.
- [9] 赵英时. 遥感应用分析原理与方法[M]. 北京:科学出版社,2003.
- [10] 胡云锋,徐芝英,刘越,等. 地理空间数据的尺度转换[J]. 地球科学进展,2013,28(3):297-304.
- [11] 茆诗松,程依明,濮晓龙,等. 概率论与数理统计教程[M]. 北京:高等教育出版社,2010.
- [12] W·G·科克伦. 抽样技术[M]. 北京:中国统计出版社,1985.