

文章编号: 1007 4619(2006) 03 0366 07

遥感图像分类精度的点、群样本检验与评估

刘旭拢, 何春阳, 潘耀忠, 杨明川, 张锦水

(北京师范大学 环境演变与自然灾害教育部重点实验室, 北京师范大学 资源学院, 北京 100875)

摘 要: 遥感专题分类结果在使用前, 必须进行客观可靠的精度验证和分析, 以保持遥感分类结果的可靠性。本文利用不同分辨率遥感数据获取的同一地区土地利用 覆盖信息, 进行了简单随机抽样、系统抽样和分层抽样三种不同抽样组织方式下的点样本和群样本检验分析, 评估了不同抽样方式下的点样本和群样本检验效果。研究结果表明: (1) 抽样方式对遥感分类精度评价结果的影响是客观存在的, 不同抽样方式下的点样本和群样本检验结果都存在一定的随机性, 但同一种抽样方式下, 点样本检验精度评估结果的波动幅度小于群样本检验, 稳定性比群样本检验要好; (2) 不同抽样方式下的多次点样本和群样本检验的平均精度检验结果基本上都能够反映分类图像的精度特征, 其中, 点样本检验中, 分层随机抽样点样本检验效果较好; 群样本检验中, 系统抽样群样本检验和分层随机抽样群样本检验的效果优于简单随机抽样群样本检验。

关键词: 遥感分类; 精度评估; 点样本检验; 群样本检验

中图分类号: TP751.1 文献标识码: A

Accuracy Assessment of Thematic Classification Based on Point and Cluster Sample

LIU Xu-long HE Chun-yang PAN Yao-zhong YANG Ming-chuan ZHANG Jin-shui

*(Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education of China, Beijing Normal University,**College of Resources Science & Technology, Beijing Normal University, Beijing 100875 China)*

Abstract In order to assure the application of thematic classification, it is very important and necessary to make a rigorous accuracy assessment. In this paper, we use sampling methods of point and cluster sample to assess the accuracy of the same region's land-use/land-cover thematic maps, which are derived from different resolution remote sensing data. Here, sampling designs are consisted of simple random, systematic and stratified sampling. The results are as following. Firstly, the sampling design has great impact on the accuracy of remote sensing classification. There exists great randomness on the result of points and cluster sample verification on the different sampling design. On the same sampling design, the stability of point sample verification is higher than that of cluster sample verification. Secondly, the average accuracy of different sampling designs of multi-point and multi-cluster sample verification can reflect the accuracy characteristic. During the course of point sample verification, stratified sampling's error is lower than others'. During the course of cluster sample verification, systematic sampling and stratified sampling's accuracy are prior to simple random sampling's.

Key words remote sensing classification; accuracy assessment; point sample; cluster sample

收稿日期: 2005-01-10; 修订日期: 2005-08-02

基金项目: 国家自然科学基金项目(40501001)和国家 863 项目(2003AA131080)支持。

作者简介: 刘旭拢(1978—), 女, 硕士, 现从事土地覆盖遥感监测与精度评估研究。E-mail: klong@ires.cn

1 引言

遥感专题分类结果在使用前, 必须进行客观可靠的精度验证和分析。同一遥感分类结果, 精度评价的方式不同, 评价结果就有可能不同, 因此很有必要对遥感精度评价过程中影响精度评价结果的各种因素进行细致深入的分析。一般地, 目前影响遥感精度评价的因素主要包括抽样方法、参考数据和评估参数三个方面^[1], 其中, 抽样样本的设计和选择尤为关键。

目前, 从抽样单元上看, 遥感分类精度评估方式主要可以分为点样本检验和群样本检验两类^[2]。点样本检验以单个像元作为基本的抽样单元, 利用参考数据, 通过对基本抽样像元的分析, 达到对分类结果的总体精度评估。群样本检验则以像元集合(群)作为基本的抽样单元, 利用参考数据, 通过对抽样群的分析达到对分类结果的总体精度评估。点样本检验简单方便, 比较符合抽样调查的随机原则, 是遥感分类评估中常用的检验方式。Janssen等在研究中指出, 以像元为基本抽样单元进行分类结果的精度评估是比较合适的^[2]。Conese^[3]、Knick^[4]、刘正军^[5]、冉有华^[6]等在各自的研究中也广泛采用点样本检验来进行遥感分类结果的精度评价工作。群样本检验的突出特点在于参考数据获取相对集中, 从而降低了野外调查工作量和成本, 比较适合于大区域分类结果的精度评价, 因而近年来逐渐受到了研究者的重视^[7]。Todd等在1980年提出群样本检验方法可以应用于遥感图像精度评估^[8], Congalton在1991年系统分析了群样本检验中抽样单元的采样数量问题^[9], Stehman在1997年讨论了群样本检验的标准差问题^[10], Edwards等在1998年设计了由线性排列像元组成群样本进行大面积土地覆盖制图精度评估的方法^[11]。此外, 史文中等近年来也积极开展遥感分类等空间数据的精度评估工作^[12]。

这些国内外研究者关于点样本检验和群样本检验的理论和实践研究, 大大提高了我们对问题的理解和认识。不过应该看到, 尽管抽样方式是影响点样本和群样本检验效果的一个重要因素, 但在目前的研究中, 对于不同抽样方式对点样本和群样本检验效果影响效果的系统定量分析还比较缺乏。

因此, 本文在利用不同分辨率遥感数据获取同一地区土地利用覆盖信息的基础上, 分别采用简单随机抽样、系统抽样和分层抽样三种不同的抽样组织方式, 对分类结果进行了点样本检验和群样本检

验的实验研究, 进而定量分析不同抽样方式下点样本和群样本检验的检验效果。目的在于更好地认识和理解精度评价过程对遥感分类精度的影响, 以期获得客观可靠的遥感图像分类精度评价结果。

2 方法

本研究使用了覆盖北京市亚运村地区季相相同、质量较好的一景 Landsat ETM+ 多光谱数据(轨道号为 123 /32 获取时间 2001年 05月 19日, 空间分辨率 30m)和一景 IKONOS多光谱数据(获取时间 2001年 04月 26日, 空间分辨率 4m)(图 1)。基本思路是在野外实地调查的基础上, 利用高分辨率遥感数据 IKONOS获取的土地利用覆盖信息, 来对低分辨率 ETM+遥感数据分类结果进行简单随机抽样、系统抽样和分层抽样三种不同抽样组织方式下的点样本和群样本检验分析, 从而评估不同抽样方式下的点样本和群样本检验效果。主要包括研究区土地利用/覆盖信息分类信息获取、分类结果全样本评价和分类信息不同抽样方式下的点样本检验和群样本检验分析三个步骤, 实验依据图 2所示的流程进行。

2.1 研究区土地利用覆盖信息的获取

在几何纠正、图像匹配等相关图像预处理的基础上, 通过研究区的野外实地调查, 首先把研究区划分为裸地、道路、水体、建筑和植被 5种土地利用覆盖类型。然后对 4m 分辨率的 IKONOS数据(图 1(b))进行目视解译、屏幕数字化和矢栅转化, 得到基于 IKONOS数据的 4m 分辨率的研究区土地利用覆盖图(图 3(b))。接着, 在野外实地调查数据和 4m 分辨率 IKONOS数据的帮助下, 通过选择训练样区, 在 ERDAS MAG NE 8.5软件支持下, 采用最大似然分类算法, 对 30m 分辨率 ETM+ 多光谱数据进行分类处理, 得到基于 ETM+ 的 30m 分辨率的研究区土地利用覆盖分类图(图 3(a))。

2.2 分类结果全样本精度评价

由于基于实地调查, 利用 IKONOS数据数字化解译的土地利用覆盖图(图 3(b))空间分辨率是 4m, 远高于基于 ETM+数据的土地利用覆盖分类图(图 3(a)), 相对于图 3(a)而言, 图 3(b)可以基本反映研究区土地利用覆盖的实际状况, 因此能够作为对图 3(a)进行精度评价的基本参考数据。

为了评价不同抽样方式下点样本检验和群样本

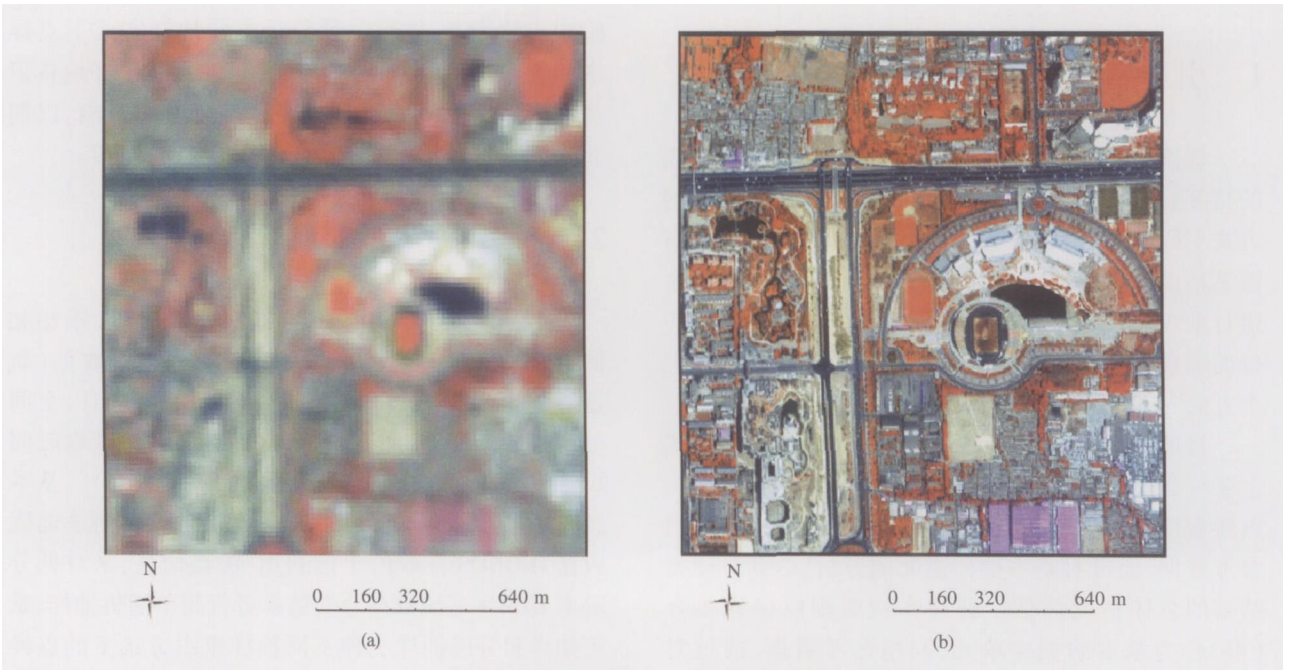


图 1 覆盖研究区 ETM+和 IKONOS 数据

(a)覆盖研究区的 ETM+数据; (b)覆盖研究区的 IKONOS数据

Fig 1 Used remotely sensed data

(a) ETM+ data in study area (b) IKONOS data in study area

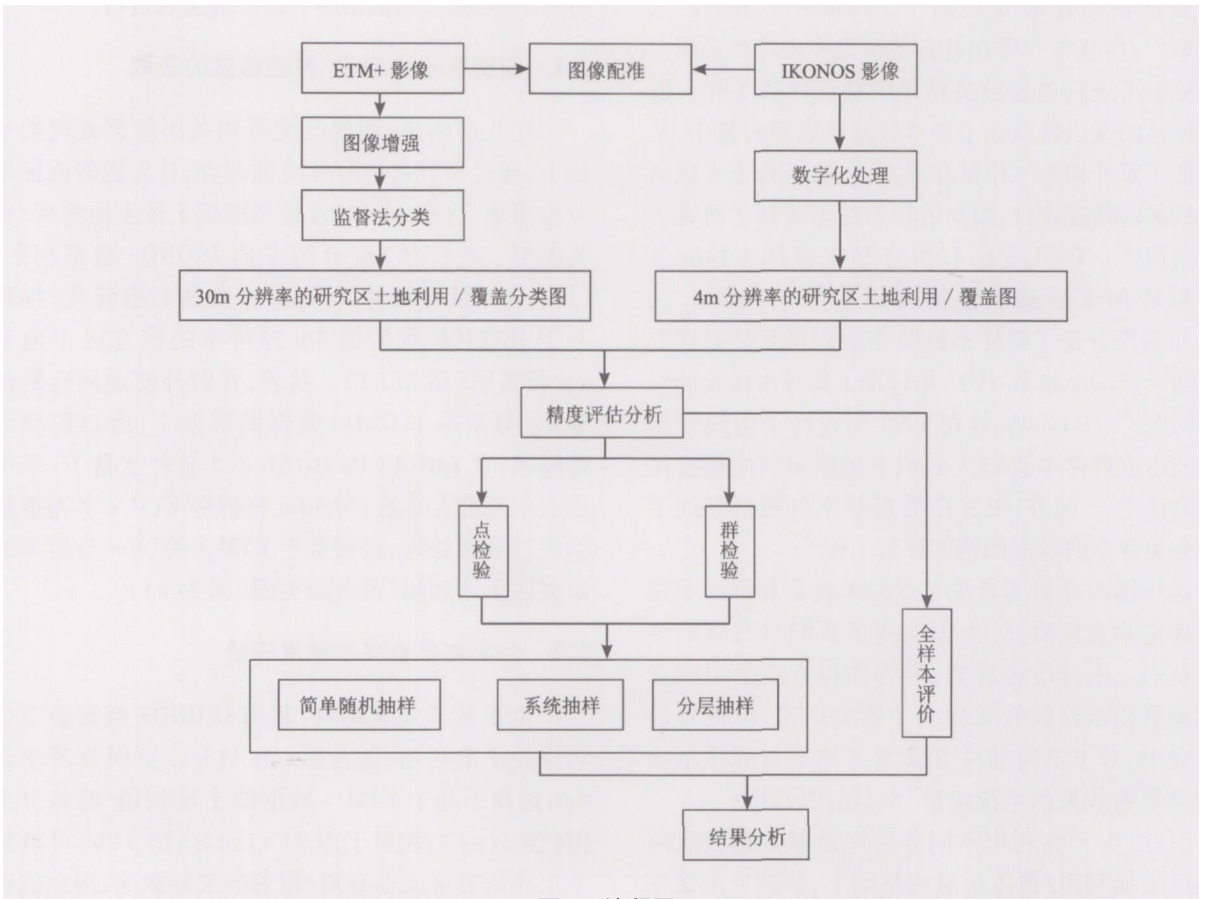


图 2 流程图

Fig 2 Flow chart

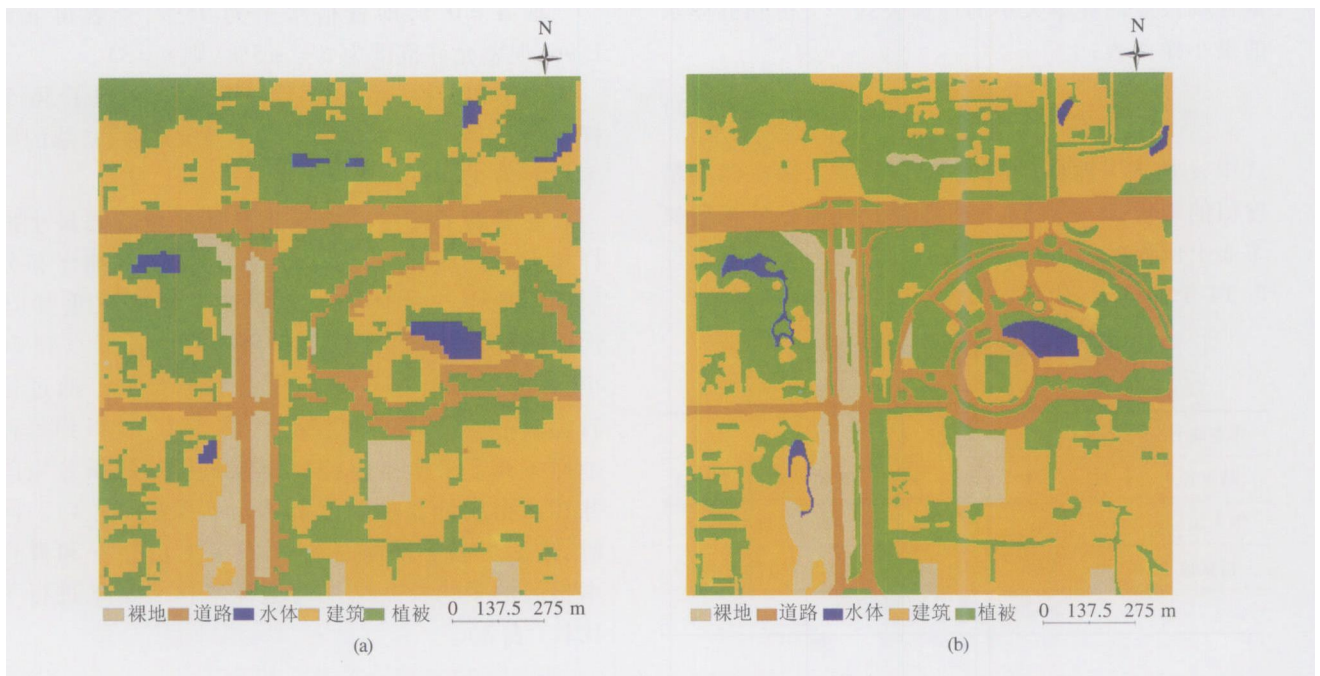


图 3 研究区土地利用 覆盖图

(a) 基于 ETM+ 的 30m 分辨率土地利用 覆盖分类图; (b) 基于 KONOS 的 4m 分辨率土地利用 覆盖图

Fig 3 Land use /cover map in the study area

(a) Land use /cover map with 30m ETM+; (b) Land use /cover map with 4m KONOS

检验的效果, 首先把图 3(a)重采样为与图 3(b)一致的 4m 分辨率, 然后以图 3(b)为标准, 对图 3(a)进行像元对像元的全样本评价, 即对分类结果的每一个像元都进行比较评价, 并假定这种评价结果是对遥感分类结果图 3(a)的最真实评价结果(表 1)。

表 1 全样本精度评价结果

Tab 1 Accuracy assessment based on all the pixels

	分类 图像					总和
	裸地	道路	水域	建筑	植被	
裸地	8394	997	195	649	792	11027
道路	205	13544	3	3170	4474	21396
水域	32	92	1686	248	434	2492
建筑	665	1230	301	52156	9515	63867
植被	891	2508	538	8408	46666	59011
总和	10187	18371	2723	64631	61881	157793
总体精度 = 77.6% Kappa = 0.661						

2.3 不同抽样方式下的点样本检验和群样本检验

简单随机抽样、系统抽样和分层随机抽样是目前遥感精度检验中最常用的三种抽样方式^[13], 为了评估这三种抽样方式下点样本和群样本检验的效

果, 首先需要确定点样本和群样本检验的最小抽样样本数据。

2.3.1 点检验最小抽样点数的确定

在总体样本量较大的情况下, 点检验可以按照下面的公式计算所需的最少的随机抽样点的个数^[11]。

$$n = \frac{u_{1-\alpha/2}^2}{d^2} \times p \times (1 - p) \quad (1)$$

式中, n 为最少的抽样点的个数, p 为分类正确的百分比, u 是对应于置信水平从正态分布的概率表上所查的值, d 为误差允许范围。

p 值可以通过事先抽取少量样本, 做精度评估得出。经试探性样本实验, 取 $p=0.74$ 取 $\alpha=0.1$, 即置信水平为 95%, 查表得 $u=1.96$ 误差允许范围为 $d=\pm 5\%$, 则 $n=296$ 。

实际抽样时, 为了和下面做的群样本检验样本数具有可比性, 总的抽样点数目确定为 324 个。

2.3.2 群样本检验最小样本数的确定

群抽样的抽样单元也存在一个最小样本数的问题, 但是群中的二级样本点都限定在所选的基础样本内, 空间距离接近, 不适合用公式(2)计算最小样本量。通常是把群中的每一个像元群看成整体, 作为一个分析单元, 这样每一个像元群都可以得到一个精度值, 所有像元群的精度值是连续的。然后根

据连续尺度的样本大小的计算公式^[13],得到群样本的最小样本数:

$$n = \left[\frac{s \times u_{1-\alpha/2}}{d \times \bar{x}} \right]^2 \quad (2)$$

式中, s 是试探性样本的标准差, \bar{x} 是试探性样本精度值的均值, u 是对应于置信水平从正态分布的概率表上所查的值, d 为误差允许范围。实验中试探取 14个群样本,计算 n 的大小,结果见表 2

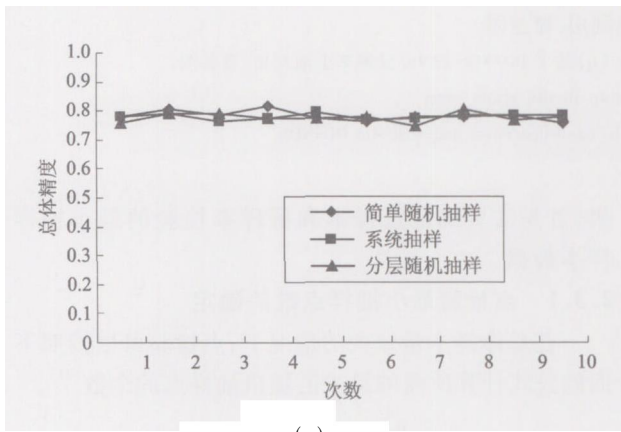
表 2 试探性样本的精度值
Table 2 Precision of trial sample

样本编号	1	2	3	4	5	6	7
精度值	0.71	0.56	0.61	0.67	0.78	0.81	0.83
样本编号	8	9	10	11	12	13	14
精度值	0.89	0.78	0.85	0.78	0.82	0.56	0.89
$s^2 = 0.012 \quad \bar{x} = 0.753$							

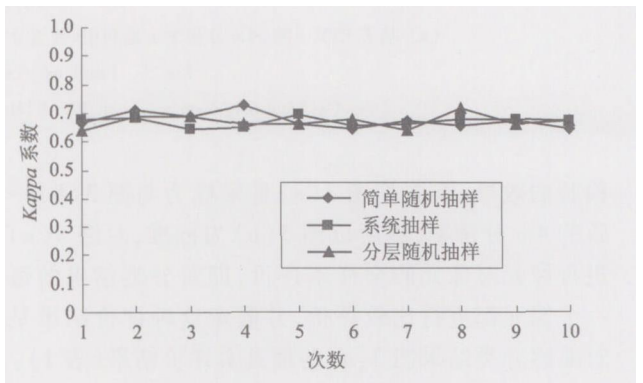
取 $\alpha = 0.1$ 即置信水平为 95%,查表得 $u = 1.96$ 误差允许范围为 $d = \pm 5\%$,则 $n = 33$

为了和点检验的样本数保持一致,抽取了 36个样本单元,每个样本单元设置为 3×3 的群,总的样本点数为 324个。

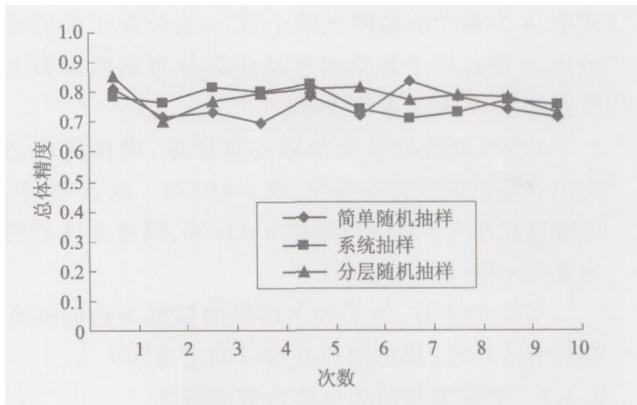
在得到了点样本检验和群样本检验的最小抽样单元后,分别采用简单随机抽样、系统抽样和分层随机抽样三种抽样方式,每种抽样方式重复 10次,在分类图上每次分别采用 324个抽样点和 36个 3×3 的抽样群进行点检验和群检验。通过比较每次抽样结果与参考结果(图 3(b)),得到每种抽样方式下多次点检验和群检验得到的误差矩阵并计算相应的总体精度和 $Kappa$ 系数(图 4)。同时,进一步将不同抽样方式下多次点样本和群样本检验的平均检验结果与全样本检验结果进行了比较(表 3)。



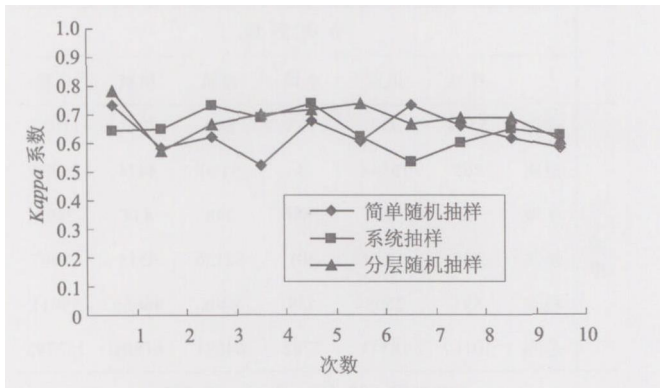
(a)



(b)



(c)



(d)

图 4 基于点样本和群样本检验的精度评价

(a)、(b)为点样本检验结果; (c)、(d)为群样本检验结果

Fig 4 Accuracy assessment based on point sample and cluster sample

(a) Overall accuracy based on point sample; (b) Kappa coefficient based on point sample

(c) Overall accuracy based on cluster sample; (d) Kappa coefficient based on cluster sample

表 3 点样本检验和群样本检验多次平均精度评价结果比较

Table 3 Average accuracy comparison between point sample and cluster sample

		总体精度		<i>Kappa</i> 系数	
		评估参数	绝对误差	评估参数	绝对误差
点样本检验	简单随机点样本检验	0.782	0.006	0.674	0.013
	系统点样本检验	0.780	0.004	0.669	0.008
	分层点样本检验	0.777	0.001	0.671	0.010
群样本检验	简单随机群样本检验	0.756	0.020	0.633	0.028
	系统抽样群样本检验	0.771	0.005	0.648	0.012
	分层抽样群样本检验	0.783	0.007	0.683	0.022

3 结果分析

对 ETM+数据土地利用覆盖分类结果(图 3(b))的全样本检验表明,分类结果的总精度是 77.6%, *Kappa* 系数 0.667(表 1)。不同抽样方式下的多次点样本检验和群样本检验结果(图 4 表 3)表明:

(1)无论是点样本检验还是群样本检验,不同抽样方式下的检验结果都存在一定的波动,这说明抽样方式对最终精度评价结果的影响是存在的。同时,无论是点样本检验还是群样本检验,即使是在同一种抽样方式下,每次的精度评价结果也不完全相同,表明这两种检验方式的评价结果都存在一定的随机性。

(2)多次重复检验的平均结果表明,不同抽样方式下的点样本和群样本检验都能够大致反映出分类图像的精度特征。其中,分层随机抽样点样本检验的总体精度与全样本评价结果最为接近,二者差值的绝对值为 0.001。系统抽样点样本检验的 *Kappa* 系数与全样本评价结果最为接近,二者差值的绝对值为 0.008。系统抽样点样本检验和分层随机抽样点样本检验的检验结果均优于简单随机抽样点样本检验结果。而系统抽样群样本检验和分层随机抽样群样本检验的精度值,尤其是系统抽样群样本检验结果与全样本检验结果比较接近,效果较好,但简单随机抽样群样本检验结果和全样本评价结果之间则相差相对较大。

(3)同一种抽样方式下,点样本检验精度评估结果的波动幅度小于群样本检验,这表明点样本检验的稳定性比群样本的稳定要好。但从具体操作角度上看,尤其是大面积土地利用覆盖分类精度验证

时,在精度评价要求不是太高的情况下,群样本检验比点样本检验具有节约野外检验数据获取成本的优势。另外,简单随机抽样和系统抽样两种抽样方式下的点样本检验和群样本检验,都可能出现个别空间上分布稀少的土地利用覆盖类型不能检验到的情况,这是在采用这两种精度评价方式时应该注意的问题。

4 结 论

本文利用高分辨率遥感数据 IKONOS 获取的土地利用覆盖信息,来对低分辨率 ETM+遥感数据分类结果进行简单随机抽样、系统抽样和分层随机抽样三种不同抽样组织方式下的点样本检验和群样本检验分析,从而评估不同抽样方式下的点样本检验和群样本检验效果。研究结果表明:(1)抽样方式对遥感分类精度评价结果的影响是客观存在的,不同抽样方式下的点样本检验和群样本检验结果都存在一定的随机性,但同一种抽样方式下,点样本检验精度评估结果的波动幅度比群样本检验的小,稳定性比群样本检验要好;(2)不同抽样方式下的多次点样本检验和群样本检验的平均精度检验结果,基本上都能够反映分类图像的精度特征,其中点样本检验中,分层随机抽样点样本检验效果较好,群样本检验中,系统抽样群样本检验和分层随机抽样群样本检验的效果优于简单随机抽样群样本检验。

但需要指出的是,本文的研究结论主要还是建立在基于研究区数据的多次实验分析基础上的,如何从内部统计规律出发,定量推导和阐述不同抽样方式对点样本检验和群样本检验的影响,进而更好地认识和理解遥感分类精度评价过程和机理,将是

目前遥感分类精度评价中需要进一步关注的问题。

参 考 文 献 (References)

- [1] Stehman S V, Czaplewski R L. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles [J]. *Remote Sensing of Environment* 1998, **64**: 331—334
- [2] Janssen L L F, Van Der Weide F J M. Accuracy Assessment of Satellite Derived Land cover Data: A Review [J]. *Photogrammetric Engineering & Remote Sensing* 1994, **60**: 419—426
- [3] Conese G, Maselli F. Use of Error Matrices to Improve Area Estimates with Maximum Likelihood Classification Procedures [J]. *Remote Sensing of Environment* 1992, **40**: 113—124
- [4] Knick S T, Rotenberry J T, Zarniello T J. Supervised Classification of Landsat TM Imagery in A Semi arid Rangeland by Nonparametric Discriminant Analysis [J]. *Photogrammetric Engineering & Remote Sensing* 1997, **63**: 79—86
- [5] Liu Z J, Wang C Y, Yan H, *et al*. High Resolution Land Cover Image Classification and Evaluation Based on Fuzzy Armap Neural Network [J]. *Journal of Image and Graphics* 2003, **8**(2): 151—154 [刘正军, 王长耀, 延昊等. 基于 Fuzzy ARMAP神经网络的高分辨率图像土地覆盖分类及其评价 [J]. *中国图象图形学报* 2003, **8**(2): 151—154.]
- [6] Ran Y H, Li W J, Chen X Z. Verification and Assessment of Land Use Classification by Using TM Image taking Dingxi County as an Example [J]. *Remote Sensing Technology and Application* 2003, **18**(2): 81—86 [冉有华, 李文君, 陈贤章. TM 图像土地利用分类精度验证与评估——以定西县为例 [J]. *遥感技术与应用*, 2003, **18**(2): 81—86.]
- [7] Alfred S F, Van D M, Ben G. Spatial Statistics for Remote Sensing [M]. Kluwer Academic Publishers, 1999.
- [8] Todd W J, Gehring D G. Landsat Wildland Mapping Accuracy [J]. *Photogrammetric Engineering & Remote Sensing* 1980, **46**(4): 509—520
- [9] Congalton R G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data [J]. *Remote Sensing of Environment* 1991, **37**: 35—46.
- [10] Stehman S V. Estimating Standard Errors of Accuracy Assessment Statistics under Cluster Sampling [J]. *Remote Sensing of Environment* 1997, **60**: 258—269.
- [11] Edwards T G, Moisen G G, Culter D R. Assessing Map Accuracy in a Remotely sensed Ecoregion scale Covermap [J]. *Remote Sensing of Environment* 1998, **63**: 73—83.
- [12] Shi W Z. Theory and Methods for Handling Errors in Spatial Data [M]. Science Publishers, 1998. [史文中. 空间数据误差处理的理论与方法 [M]. 科学出版社, 1998.]
- [13] Zhao Y S, *et al*. Principle and Method of Remote Sensing Application [M]. Science Publishers, 2003 [赵英时等. 遥感应用分析原理与方法 [M]. 科学出版社, 2003.]