基于不同抽样方法的遥感面积测量方法研究

胡潭高,张锦水,潘耀忠,宋国宝,董燕生,贾 斌
(地表过程与资源生态国家重点实验室(北京师范大学) 北京 100875)

摘要:众多研究结果表明,遥感和抽样技术相结合可以有效地进行地物面积的测量。目前,随机抽样、系统抽样和分层抽样方式在遥感抽样调查技术领域应用比较广泛。本文以遥感图像为基础,从不同角度对随机抽样、系统抽样及分层抽样(包括等样本量、等面积、等丰度抽样)进行了有益探讨,分析发现:对于同一地物,从平均误差百分比、标准差和极差3个角度分析,随机和系统抽样反推得到的总量精度都低于分层抽样精度;对于不同的地物类型,利用3种分层抽样方法反推的结果与地物所占百分比成正相关,地物所占的百分比越大,反推的结果越好;等样本量、等面积、等丰度分层抽样从平均误差百分比、标准差和极差3个角度分析各有优势,跟地物所占的百分比也有密切关系。

关键词:抽样;遥感图像;精度;稳定性
中图分类号:TP 79 **文献标识码**:A **文章编号**:1001-070X(2008)03-0037-05

0 引言

抽样调查和遥感是2种重要的信息获取方式,前者是一种非全面的调查手段,主要通过不同抽样方法选取样本,并根据相应算法估测总体;后者是极其重要的对地观测手段,近些年来发展迅速。二者相互补充,遥感为抽样调查提供详细的抽样框和分层信息,提高抽样调查效率;抽样技术为遥感提供充分的地面数据和验证依据。

抽样技术与遥感技术相结合形成遥感抽样调查技术,在林业、地质、渔业、生态环境保护以及农业统计等领域应用日益广泛。在国外,T. Charles等采用抽样方法对森林资源量变化进行有效估计,对森林资源量的动态变化进行监测[1];美国国家农业统计局(NASS)通过将空间统计抽样方法与遥感监测技术的结合,对全美主要农作物面积进行多样框抽样调查,提高全美农情信息获取速度[2-4];欧盟MARS计划利用遥感技术开发出了能够改善欧洲共同体内部农业统计体系的新方法,以CLC(Corine Land Cover)数据为基础进行土地利用调查,比较随机、系统和分层3种抽样方法的优势,并以分层系统抽样方法选择遥感影像,降低调查成本[5-9]。近年来,国内利用遥感抽样调查技术也取得了许多成果,吴炳方在农作物区划基础上构建面积抽样框架和产量抽样框架,利用样本检验分类精度,获取区划单元内作物种植面积成数,提高了作物面积和产量调查效率[10];李连发等在空间抽样基础上提出空间抽样优化决策模型,并将优化模型用于全国范围内耕地面积调查,降低了抽样成本[11,12];梁进社等利用分层随机抽样方法以北京和重庆市土地利用变化的分类精度为例,为全国大城市建设用地扩展的遥感监测提供了技术支持[13]。

综合国内外的研究发现,在遥感抽样调查技术应用领域,目前主要的抽样方法有3种,即随机抽样、系统抽样和分层抽样,这3种抽样方法具有不同的特点。另外,针对遥感图像,根据不同的分层标志,分层抽样又可以分为等样本量(以样本量作为分层标准,每一层抽选的样本量相等)、等面积(以面积作为分层标准,每一层抽选的样本总面积相等)和等丰度(以丰度作为分层标准,每一层抽选的样本丰度范围相等)3种分层抽样方法。本文分别用等样本量、等面积和等丰度来描述这3种分层方法,并利用其抽样方法分别反推遥感图像中某种典型地物类别的总体面积,比较分析各种抽样方法所带来的精度和多次反推结果间的稳定程度。

1 研究方法

1.1 实验数据

以北京市2006年4月22日QuickBird遥感图像作为实验数据,其分辨率为2.4 m。首先,从中截取一块大小为4 096像元×4 016像元的区域进行预分类,得到包含5类地物的模拟图像(插页彩片

收稿日期:2008-01-31;**修订日期**:2008-04-17
基金项目:国家高技术研究发展计划项目(批准号:2006AA120103;2006AA120101);教育部新世纪人才支持计划资助。

万方数据

1),然后把此分类结果看作真值,对其进行不同抽样方法的研究分析,其统计信息如表 1 所示。

表 1 模拟图像统计信息

名称	地物类别	*DN* 值	像元数/个	百分比/%
第 1 类	耕地	1	4 750 619	29.215
第 2 类	水体	2	1 281 336	7.880
第 3 类	城镇用地	3	4 497 529	27.659
第 4 类	裸地	4	4 141 188	25.467
第 5 类	草地	5	1 590 112	9.779

1.2 实验步骤

(1)首先,对图像进行重新编码,生成 5 幅不同地物对应的 0 - 1 值图像,如图 1 所示。

(2)根据重编码后的结果,以像元为单位对每幅图像进行格网划分,生成若干个 12 m × 12 m 大小的格网。因为抽样反推总体的效果与抽样样本量有关,样本量太少,必然会影响抽样的精度。在实验中针对的是模拟图像进行的抽样,模拟图像代表的研究区域比较小,所以采用较小的格网进行抽样,产生足够数量的抽样单元来模拟大面积、区域性的实际遥感调查。按照上述方法生成足够数量的格网,然后对这些格网利用不同的抽样方法进行 1% 抽样。

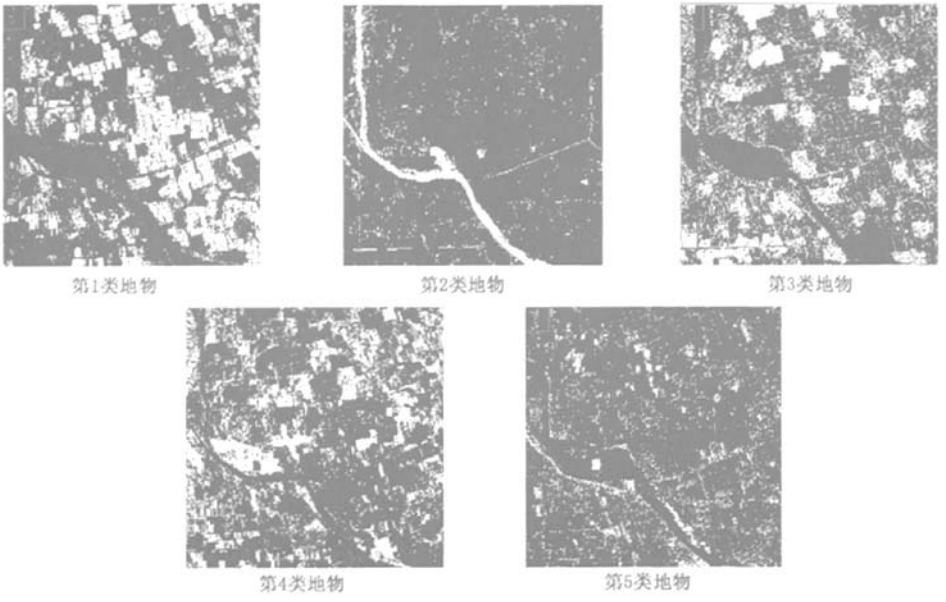


图 1 图像重新编码后,5 种不同地物对应的 0 - 1 值图像

根据抽样结果,利用公式(1)、(2)反推出图像总体结果,即

$$\hat{T}_x = \sum_{i=1}^{N} X_i / a \tag{1}$$

$$N = \frac{N_1 \cdot N_2 \cdot d^2}{D^2} \times a \tag{2}$$

式中,T_x 为抽样反推估算得到的总体;x_i 为第 i 个样本中对应地物的数目;N 为样本大小;N_1 和 N_2 为原始模拟图像的行列数;a 为抽样率,本文中的抽样率为 1%;d 为模拟图像像元的分辨率,此处 d = 2.4 m;D 为抽样单元格网的大小,此处 D = 12 m。

利用公式(3)计算出不同抽样方法的抽样反推总体精度,即

$$\bar{A} = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{T}_{xi}}{T_x} \times 100\% \tag{3}$$

式中,$\bar{A}$ 为重复抽样估计总体的评价精度;$\hat{T}_{xi}$ 为第 i 次抽样反推结果估算得到的总体;n 为不同抽样方法的抽样次数;T_x 为模拟图像中某类地物的像元真实数目。

(3)在 IDL 4.2 环境下,编程实现不同的抽样方法。分别对 5 类地物利用不同的抽样方法进行抽样,利用抽样结果反推总体,计算实验所需要的各种参数。

2 结果分析

2.1 几种结果分析指标

根据抽样实验步骤,在 IDL 4.2 环境下运行程序,对研究区 5 类不同总量的地物针对每种方法各进行 100 次抽样,利用公式(1)~(3)得到 100 个抽样反推总体的结果,将其与模拟图像的真实结果进行比较,分别得到如下几组参数:

(1)100 次抽样反推结果的平均误差百分比。

$$Average = (1 - \sum_{i=1}^{n} X_i / n) \times 100\% \tag{4}$$

式中,x_i 为每次抽样得到的反推结果精度;n 为抽样次数,此处 n = 100。

万方数据

(2)100 次抽样反推结果的标准差。

$$Stdev = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}} \quad (5)$$

式中,x 为反推结果的平均值 $Average(1,2,\cdots,n)$; n 为样本大小。

(3)100 次抽样反推结果的极差。

$$difference = max(x_{1-n}) - min(x_{1-n}) \quad (6)$$

式中,x_{1-n}表示由 n 次抽样结果精度组成的一维数组,在这里 $n=100$。

2.2 同一地物不同抽样方法的综合分析

针对平均误差百分比指标,对第 1 类地物(在整幅图像中占 29%),分别用上述几种抽样方法,可以得到如图 2 所示的几组曲线。

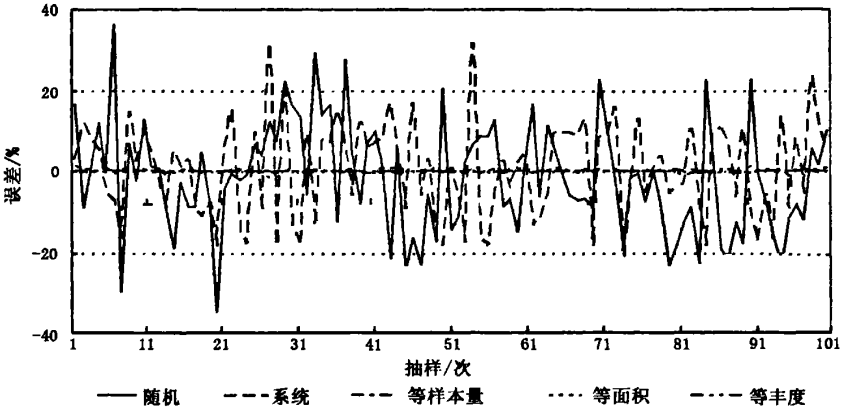


图 2　第 1 类地物不同抽样方法的误差百分比

从图 2 看出,随机抽样与系统抽样总量推算的波动性比较大,单次精度也相对比较低; 3 种分层抽样方法得到的结果稳定性更好,精度较高。

根据以上计算结果,分别以平均误差、标准差和极差 3 种指标为依据,计算不同的抽样方法在 100 次得到的结果平均值,其结果如表 2 所示。

表 2　第 1 类地物不同抽样方法下的 3 种指标

方法	随机	系统	分层		
			等样本量	等面积	等丰度
平均误差/%	-32.5	29.0	-2.3	5.9	1.1
标准差	2.767	2.397	0.173	0.565	0.167
极差	14.213	10.554	0.839	2.855	0.760

从表 2 可以看出:

对于同一地物不同抽样方法的平均误差,随机抽样的平均误差明显大于系统抽样及其它 3 种分层抽样方法的; 3 种分层抽样方法的平均误差比较接近,但是等丰度抽样得到的误差相对较小,即抽样得到的平均精度最高。

对于同一地物不同抽样方法的标准差,随机抽样与系统抽样的标准差明显大于其它 3 种分层抽样的,系统抽样比随机抽样的标准差更小一些; 3 种分层抽样的标准差都比较小,等丰度抽样方法的标准差最小,即抽样得到的反推结果最稳定。

对于同一地物不同抽样方法的极差,随机抽样与系统抽样极差较大,在 3 种分层抽样方法中,等面积抽样方法极差较大,等丰度抽样方法的极差最小。

综合上述 3 种指标分析可知:

从平均误差角度分析,误差由大至小依次为随机抽样(-32.5%) > 系统抽样(29.0%) > 等面积分层抽样(5.9%) > 等样本量分层抽样(2.3%) > 等丰度分层抽样(1.1%);

从标准差角度分析,由大至小依次为随机抽样(2.77) > 系统抽样(2.40) > 等面积分层抽样(0.56) > 等样本量分层抽样(0.173) > 等丰度分层抽样(0.167);

从极差角度分析,由大至小依次为随机抽样(14.21) > 系统抽样(10.55) > 等面积分层抽样(2.86) > 等样本量分层抽样(0.84) > 等丰度分层抽样(0.76)。

对于第 1 类地物,随机抽样与系统抽样反推结果都不如分层抽样,在 3 种分层抽样中,等丰度分层抽样反推的效果最好。

对于第 2 类地物(在整幅图像中占 7.9%),运用同样的方法,可以得到表 3 和图 3。

表 3　第 2 类地物不同抽样方法下的 3 种指标

方法	随机	系统	分层		
			等样本量	等面积	等丰度
平均误差/%	-12.66	7.16	-1.3	-8.9	6.2
标准差	6.559	6.207	0.499	0.825	0.558
极差	30.230	29.682	2.429	4.233	2.474

万方数据

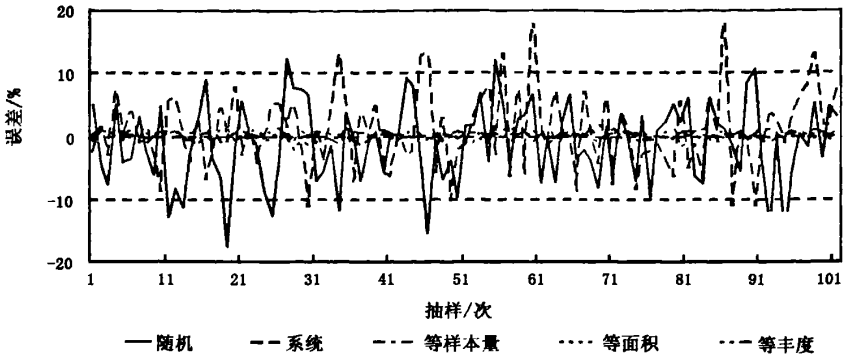


图3 第2类地物不同抽样方法的误差百分比

对第2类地物进行综合分析可知：

从平均误差角度分析，误差由大至小依次为随机抽样(-12.66%)>系统抽样(7.16%)>等面积分层抽样(-8.9%)>等丰度分层抽样(6.2%)>等样本量分层抽样(-1.3%)；

从标准差角度分析，由大至小依次为随机抽样(6.559)>系统抽样(6.207)>等面积分层抽样(0.825)>等丰度分层抽样(0.558)>等样本量分层抽样(0.499)；

从极差角度分析，由大至小依次为随机抽样(30.230)>系统抽样(29.682)>等面积分层抽样(4.233)>等丰度分层抽样(2.474)>等样本量分层抽样(2.429)。

对于第2类地物，随机抽样与系统抽样反推结果都不如分层抽样的，在3种分层抽样中，等样本量分层抽样反推的效果最好。

2.3 不同地物间3种分层抽样的综合分析

3种分层抽样方法随着地物在整幅图像中所占的百分比不同，反推结果也有不同：第1类地物(在整幅图像中占29%)等丰度分层抽样效果最好；第2类地物(在整幅图像中占7.9%)等样本量分层抽样效果最好。为了进一步研究这3者的关系，把所有的5类地物根据平均误差、标准差和极差3种指标进行分析，其结果如表4所示。

表4 5类地物的3种指标结果

平均误差/%	第1类	第2类	第3类	第4类	第5类
等样本量	-2.3	-1.3	-2.8	-0.14	3.8
等面积	5.9	-8.9	-1.4	-3.5	-5.7
等丰度	1.1	6.2	-0.1	-2.1	-4.0
标准差	第1类	第2类	第3类	第4类	第5类
等样本量	0.173	0.499	0.224	0.211	0.492
等面积	0.565	0.825	0.453	0.442	0.563
等丰度	0.167	0.558	0.237	0.247	0.589
极差	第1类	第2类	第3类	第4类	第5类
等样本量	0.839	2.43	1.2	0.938	2.392
等面积	2.855	4.233	2.253	2.066	3.307
等丰度	0.76	2.474	1.106	1.045	2.789

从平均误差角度分析，可以看出，在使用等样本量分层抽样方法进行抽样时，得到的平均误差是第4类地物(在整幅图像中占25.47%)最小，最小值为-0.14%；用等面积分层抽样方法抽样时，得到的结果是第3类地物(在整幅图像中占27.66%)最小，最小值为-1.4%；用等丰度分层抽样方法抽样时，得到的平均误差也是第3类地物最小，最小值为-0.1%。

从标准差角度分析，可以看出，在使用等样本量分层抽样方法进行抽样时，得到的标准差是第1类地物(在整幅图像中占29%)最小，最小值为0.173；用等面积分层抽样方法抽样时，得到的标准差是第4类地物(在整幅图像中占25.47%)最小，最小值为0.442；用等丰度分层抽样方法抽样时，得到的标准差是第1类地物最小，最小值为0.167。

从极差角度分析，可以看出，最小值均出现在第1、3、4类地物中。这是因为第1、3、4类地物在整幅图像中都占20%以上，属于较大面积比例的类型，而第2、5类地物所占比例分别为7.88%和9.78%，在整幅图像中所占比例较小。由此，在利用不同的分层抽样方法进行抽样的时候，地物所占的百分比越大，则抽样反推的效果越好。

3 结论

(1)在地物类型确定的情况下(以第1类地物占29.22%，属于大比例地物为例)，从3种指标出发分析不同的抽样方法，可以看到随机抽样和系统抽样的误差明显大于其它3种分层抽样方法。3种分层抽样方法的平均误差比较接近，相对来说，等丰度抽样得到的误差较小，即抽样得到的平均精度最高。同样，随机抽样与系统抽样的标准差和极差都明显大于其它3种分层抽样，3种分层抽样的标准差都比较小，等丰度抽样方法的标准差和极差均最小，即抽样得到的反推结果最稳定。

(2)以第2类地物(占7.88%，属于小比例地

万方数据

物)为例,从平均误差角度分析,随机抽样与系统抽样反推结果都不如分层抽样。在 3 种分层抽样中,等样本量分层抽样反推的效果最好。由此可以得出,对于不同地物类型,分层抽样方法都优于随机和系统抽样。

(3)在确定分层抽样更优以后,分析了 3 种分层抽样方法受地物百分比的影响,地物所占的百分比越大,则抽样得到的反推效果越好。

本研究是在固定了抽样百分比的情况下(1% 抽样)对不同抽样方法的优劣性进行研究的。在以后的研究中,还需要进一步探讨抽样百分比对不同抽样方法反推总体结果的影响。另外,为了方便比较分析,以及生成足够数量的抽样单元来模拟大面积、区域性的实际遥感调查,对于抽样框的格网大小我们也进行了固定(12 m × 12 m),没有进一步分析格网大小对于抽样方法的影响。

参考文献:

[1] Charles T Scott. Sampling Method for Estimating Change in Forest Resources[J]. Ecological Application, 1998, 8(2):228 - 233.

[2] Cotter J, Nealon J. Area Frame Design for Agricultural Surveys [R]. U. S. Department of Agriculture: National Agricultural Statistics Service, 1987.

[3] Doraiswamy P C, Hatfield J L, Jackson T J, et al. Crop Condition and Yield Simulations Using Landsat and MODIS[J]. Remote Sensing of Environment, 2004, 92: 548 - 559.

[4] Hartley H O. Discussion: Sampling Frames for Collection of Agricultural Statistics[J]. Journal of Farm Economics, 1965, 47(5): 1542 - 1544.

[5] Ryerson R A, Dobbins R N, Thibault C H. Timely Crop Area Estimates from Landsat[J]. Photogrammetric Engineering and Remote Sensing, 1985, 51: 1735 - 1743.

[6] Meyer - roux J, Vossen P. The First Phase of the MARS Project, 1988 - 1993: Overview, Methods and Results[A]. In Conference on the MARS Project[C]. Luxembourg: Office for Publications of the EC, 1994: 33 - 81.

[7] Tsiligirides T A. Remote Sensing as a Tool for Agricultural Statistics: a Case Study of Area Frame Sampling Methodology in Hellas[J]. Computers and Electronics in Agriculture, 1998, 20: 45 - 77.

[8] Plouroe Lucie, Congalton Russell G. Sampling Method and Sample Placement: PLOURDE Lucie, CONGALTON Russell G. Sampling Method and Sample Placement: How do They Affect the Accuracy of Remotely Sensed Maps[J]. Photogrammetric Engineering and Remote Sensing, 2003, 69(3): 289 - 297.

[9] 刘海启. 欧盟 MARS 计划简介与我国农业遥感应用思路[J]. 中国农业资源与区划,1999,20(3):55 - 57.

[10] 吴炳方. 全国农情监测与估产的运行化遥感方法[J]. 地理学报,2000,55(1):24 - 35.

[11] 李连发,王劲峰,刘纪远. 国土遥感调查的空间抽样优化决策[J]. 中国科学(D 辑,地球科学),2004,34(10):975 - 982.

[12] Wang J, Liu J, Zhuang D. Spatial Sampling Design for Monitoring the Area of Cultivated Land[J]. International Remote Sensing, 2002, 23: 263 - 284.

[13] 梁进社,张 华. 土地利用变化遥感监测精度评价系统——以随机抽样为基础的方法[J]. 地理研究,2004,23(1):29 - 37.

RESEARCHES ON REMOTE SENSING AREA MEASUREMENT BASED ON DIFFERENT SAMPLING METHODS

HU Tan - gao, ZHANG Jin - shui, PAN Yao - zhong, SONG Guo - bao, DONG yan - sheng, JIA Bin

(*State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing* 100875, *China*)

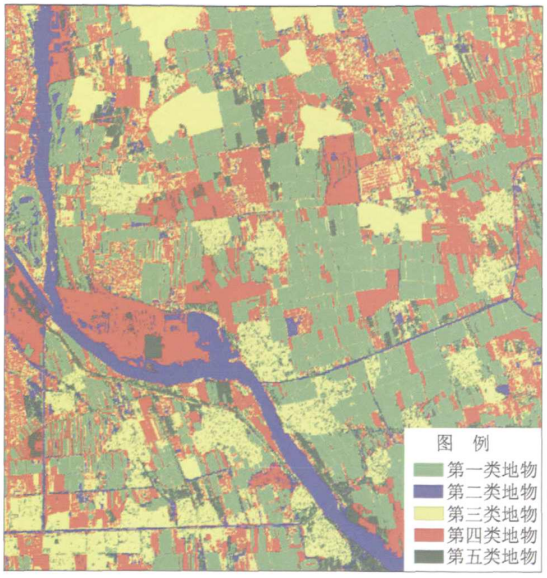
Abstract: It is held that the combination of remote sensing and sampling technology can be effectively used in area measurement. At present, the methods of random sampling, systematic sampling and stratified sampling are widely used in remote sensing sampling. Based on remote sensing images, this paper has discussed from different angles the random sampling, systematic sampling, and stratified sampling that includes equal - count, equal - area and equal - dist. It is found that: (1) for the same features, from three indicators (the average error percentage, standard deviation and difference), the precision of stratified sampling is better than that of random sampling and systematic sampling, (2) for different features, the use of three stratified sampling method and the results obtained are proportional to the percentage of features, i. e., the greater the percentage of features, the better the results, and (3) from three indicators (the average error percentage, standard deviation and difference), the three kinds of stratified sampling methods (equal - count, equal - area and equal - dist) have their respective own merits, and the precision is also closely related to the percentage of features.

Key words: Sampling; Remote sensing images; Accuracy; Stability

第一作者简介: 胡潭高(1983 -),男,硕士,主要从事遥感图像处理及信息提取方面的研究。

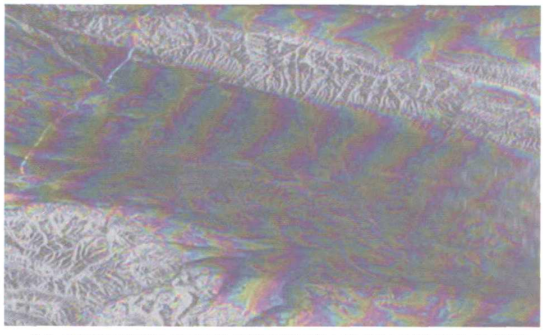
(责任编辑: 肖继春)

万方数据

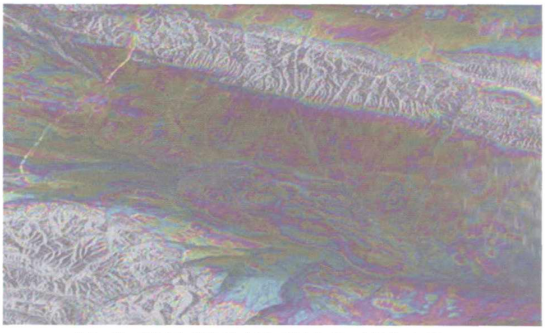


彩片1 分类模拟图像示意图

彩片2 研究区2004年12月8日CBERS 4.3.2波段合成图像

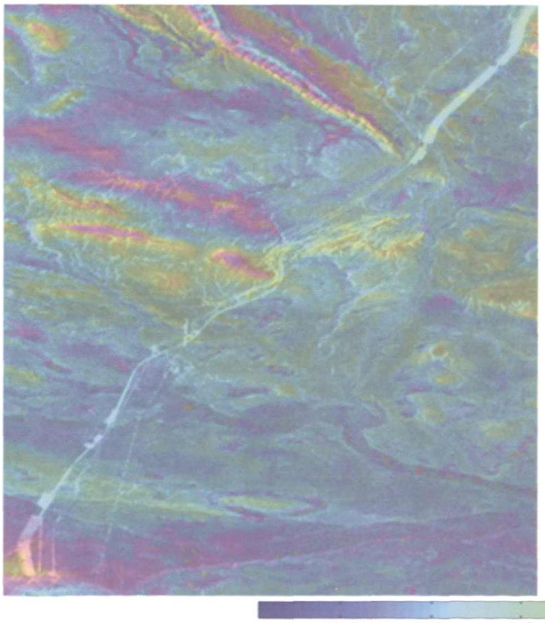


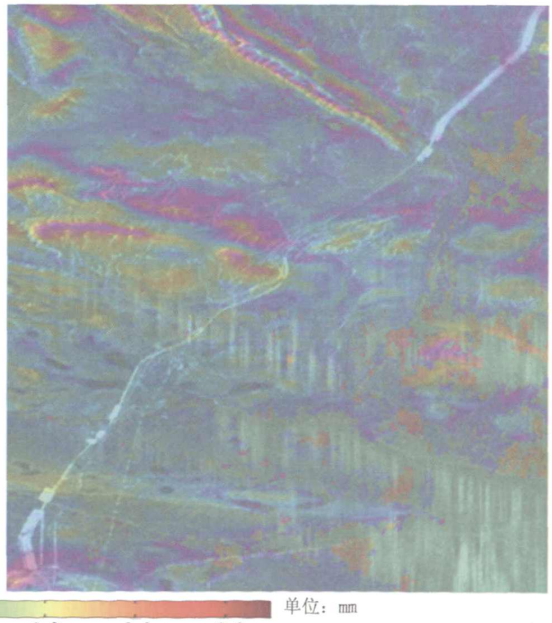
(a) 利用不精确PALSAR轨道数据计算去除平地相位后的条纹图



(b) 修正轨道后计算去除平地相位后的条纹图

彩片3 残余平地相位去除





(a) 干涉对2006-06-06/2006-09-06形变信息　(b) 干涉对2006-06-23/2006-08-08形变信息线性变化结果

彩片4 利用PALSAR干涉对计算出来的研究区形变结果

(彩片2-4见谢酬一文)

万方数据