

基于遥感影像的大区域植被类型样本快速提取方法研究

——以寒温带针叶林区域为例*

胡 博, 鞠洪波**, 刘 华, 郝 泷, 刘 海

(中国林业科学研究院资源信息研究所, 北京 100091)

摘要: [目的] 利用遥感影像的宏观性, 基于植被分类资料数据, 依据实验区域遥感影像及衍生影像本身特点, 实现大区域样本快速提取。 [方法] 实验参考 1:100 万植被图、WESTDC 中国土地覆盖图, 结合实验区域 2001 年 MODIS 时序 NDVI 影像的非监督分类结果, 利用矢、栅数据的空间特征, 将实验影像非监督分类的类型信息关联为随机样点属性, 依据该属性中包含的非监督分类类型数和各类型的样点比例, 对比类别间样本可分离性指标、标准差变化, 实现样本纯化。 [结果] 纯化后的植被样本与 WESTDC 中植被空间分布基本一致, 主要植被类型空间分布精确程度为 84.82%。将纯化前后的样本输入最大似然分类器, 总体分类精度提高了 32.52%。 [结论] 该采样方法适用于宏观大区域植被样本数据的快速提取。同时, 节省了大区域植被类型调查消耗的人力物力资源, 提高了采样效率。

关键词: 遥感; 大区域; 快速采样; 样点纯化

中图分类号: S771.8

文献标识码: A

文章编号: 1001-1498(2017)01-0111-06

Quick Sampling Method for Large Area Vegetable Types Based on Remote Sensing Images: A Case Study for Cold Temperate Coniferous Forest Region

HU Bo, JU Hong-bo, LIU Hua, HAO Shuang, LIU Hai

(Research Institute of Forestry Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: [Objective] To realize the quick access of the separable and representative training samples on large area remote sensing images. [Method] Based on the macroscopical character of remote sensing images, referring to the 1:1 000 000 vegetation map and China land cover map (WESTDC), and combining the unsupervised classification results of experimental region from the 2001 MODIS time series NDVI images, the spatial features of vector and raster data were used and associated the unsupervised classification results as the random sample properties. Based on the unsupervised classification types and the random point ratio of different classification types, the sample purification was realized. By comparing the difference of the standard deviation and the separability index before and after the sample purification of different vegetable samples, the points which had relatively larger contribution to the supervised classification were saved. [Result] The results showed that the land cover spatial distribution in survey area

收稿日期: 2016-02-04

基金项目: 国家 863 计划课题(2012AA102001)。

作者简介: 胡 博(1985—), 女, 河北人, 博士研究生, 主要从事森林资源监测技术研究

* 致谢: 国家自然科学基金委员会“中国西部环境与生态科学数据中心”提供的 WESTDC 数据

** 通讯作者: 鞠洪波(1956—), 男, 黑龙江人, 研究员, 博士生导师, 林业信息技术

was consistent with the WESTDC map, the overall accuracy of main types was 84.82%. By using the purred samples and applying the maximum likelihood classification method, the overall accuracy was 32.52% higher than that from initial samples. [**Conclusion**] The sampling method is suitable for quick access of remote sensing images of large area and could reduce the time, manpower and material costs.

Keywords: remote sensing; large area; quick sampling; sample purification

通过遥感手段提取大区域(国家级、省级)植被覆盖信息多采用具有高时间分辨率、宽视域、对存储介质要求相对较低、获取方式更便捷的中低分辨率遥感影像,如 AVHRR、MODIS 等。计算机自动提取方法以基于像元的非监督分类方法和监督分类方法为主,非监督分类方法得到的分类结果精度相对较低,并依赖专家经验进行分类后的类型组合,如 GLC2000^[1]、DISCover 全球 1 km 土地覆盖产品等^[2];监督分类方法得到的分类结果精度相对较高,其中选择具有代表性的样本数据比修改分类算法本身对分类结果的影响更为重要。

当前,国内外常采用以下方法对遥感影像监督分类的植被类型样本进行采集:(1)参考资料数据,依据专家经验,借助更高分辨率遥感影像的解译结果或者植被图、土地利用图等先验数据。Darren Pouliot 等^[3]采用 MODIS 数据监测加拿大森林变化,参考了水文、高程、公路网等数据的同时,部分区域借鉴 Google 数据和 TM 等更高分辨率数据采样;Igor Klein 等^[4]采用 MODIS 数据对中亚土地覆盖变化检测,主要依据 TM 影像解译结果进行采样;张永红等^[5]用 MODIS 数据生成全国土地覆盖产品,辅助土地利用图、植被图和局部区域的 TM、SPOT 更高分辨率影像数据获得训练样本;左玉珊等^[6]基于 MODIS 影像的京津冀区域土地覆被分类,辅助京津冀区域土地利用资料、TM 遥感影像、DEM 高程数据;范应龙等^[7]基于 TM 影像通过决策树分类提取热带森林覆盖变化信息,训练样本采用柬埔寨国家林业局野外调查数据;张雨等^[8]基于 MODIS 数据实现的辽宁省土地利用分类,通过分析辽宁省林地更新数据和一类样点数据选取训练样本。(2)结合调查数据的样本采集方法。刘华等^[9]基于四个时相的 TM 影像,根据外业调查信息以及研究区植被图、湿地分布图获取的样点数据,提取湿地信息;陈巧等^[10]基于两期 TM 影像,参照研究区森林资源分布图、土地利用图以及外业调查数据进行训练样区选择,通过监督分类结果监测天保区植被变化;Kun Jia 等^[11]采用 MODIS 时序数据与 TM 影像融合提取了中国北部

5 个县的森林覆盖信息,训练样本的选取借助于样地调查和 Google 数据;竞霞等^[12]采用两期 TM 影像对密云山区植被分类研究,采用野外调查获得的植被类型信息选取训练样本。

从国内外研究可知,遥感影像监督分类的样本选取方式大多凭借专业经验、资料数据或结合调查数据等目视判读采样,有关大区域遥感快速采样方法的类似研究比较少。本研究利用遥感影像的宏观性,基于植被分类资料数据,依据实验区域遥感影像及衍生影像本身特点,实现了大区域样本快速提取。本实验的创新体现在将实验区资料数据与遥感影像相结合,依据统计分析实现植被样本的快速提取。降低遥感监督分类采样过程中的人为干预,提高大区域采样效率,增强样本的客观性。实验结果可以为快速提取实验区植被覆盖信息,提供可靠的样本数据。

1 实验数据与实验区概况

1.1 实验数据

本实验采用 2001 年 23 期时序 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid (简称 MODIS13Q1)数据,每期间隔 16 天,属于陆地专题产品,是 MODIS 的 L3 级科学数据集产品之一。MODIS13Q1 产品原始文件包含多个数据层,实验选取第一层归一化植被指数(Normalized Difference Vegetation Index,简称 NDVI)数据文件^[13]。

本实验采用 2001 年 1:100 万中国植被图矢量数据,可提供实验区植被分布信息;中国植被区划矢量数据提供实验区范围;2000 年中国土地覆盖 1km 栅格数据^[14-15](简称 WESTDC),可提供实验区非植被分布信息。其中,WESTDC 数据来源于国家自然科学基金委员会“中国西部环境与生态科学数据中心”(http://westdc. westgis. ac. cn),该数据借鉴 IGBP 分类系统,在保持了土地利用数据的总体精度的同时,补充了中国植被图中对植被类型及植被季相的信息,更新了中国湿地图,增加了中国冰川图,融合了 MODIS 2001 年土地覆盖产品。

以上矢量、栅格数据均具有空间坐标信息,且经过影像配准,精度限制在一个像元内。

1.2 实验区概况

本研究选取寒温带针叶林区域作为实验区,实验区处于 127°20'E 以西, 49°20'N 以北的大兴安岭北部及其支脉伊勒呼里山地,总面积约为 211 600 km²。植被覆盖率高,地带性植被为兴安落叶松林,有少量灌木植被,是我国重要的木材产地。由于该区域的寒温带气候,形成永久冻土层,水分下渗困难,形成零星水体,地表积水使湿地广泛发育,其中草本沼泽分布在海拔 900 m 以下地带。该区域平坦谷地区域有农业分布但农业不发达。

2 研究方法

首先,通过 ARCGIS 软件在寒温带针叶林区域内随机布点,生成的样点具有中国植被区划文件的 空间信息,每个样点在实验影像上体现为一个像素,代表实际地物 250 m × 250 m 的样地范围。参考 1:100 万中国植被图和中国土地覆盖数据,分别赋予植被类样点的类型属性,包括:乔木、草地、湿地、灌木、农田;非植被类样点的类型属性,包括:建筑物和水体。并对各类样本的初始可分离性进行计算。

然后,对实验区 2001 年 23 期 MODIS 时序 NDVI 影像数据产品,依据不同植被类型的 NDVI 特征向量不同,采用非监督分类迭代算法^[15]得到时序 NDVI 数据波段的 聚类聚类信息。结合该非监督分类结果,依据矢、栅数据的空间特征,将非监督分类

的类型信息作为各类样本中的样点属性进行关联,统计各类样本包含的非监督分类类型和样点个数 (j, N_{Ri}),并将对应类型的样点比例作为权重值 (P_{Ri})。权重越大则越能够代表该样本类型,权重越小则越可能是该类样本中的非纯样点,按照权重大小降序排列样点 ($\text{Descend}(P_{Ri})$) 并计算累积百分比 (P_{Racc})。其中, j 为样点个数、 R 为植被类型、 i 为非监督分类类型。

大区域植被样本快速提取总体思路,如图 1 所示:

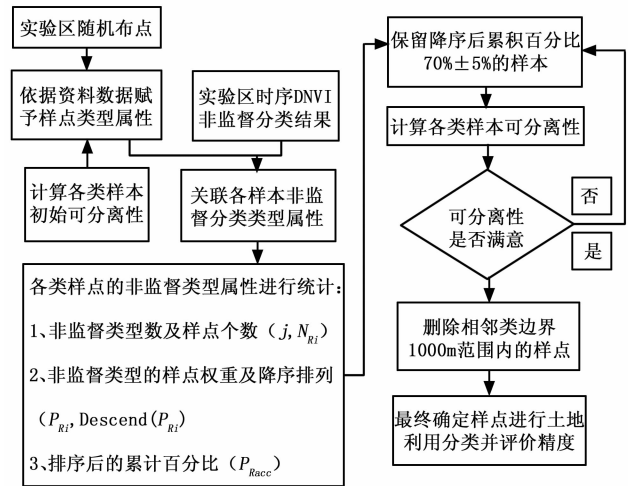


图 1 大区域植被类型样本快速提取方法流程

Fig. 1 Methodological framework of quick sampling in large area

将初始 7 类样本对应的非监督分类的结果,依据每一类样本中包含的非监督分类类型及样点个数,绘制分布直方图,如图 2 所示:

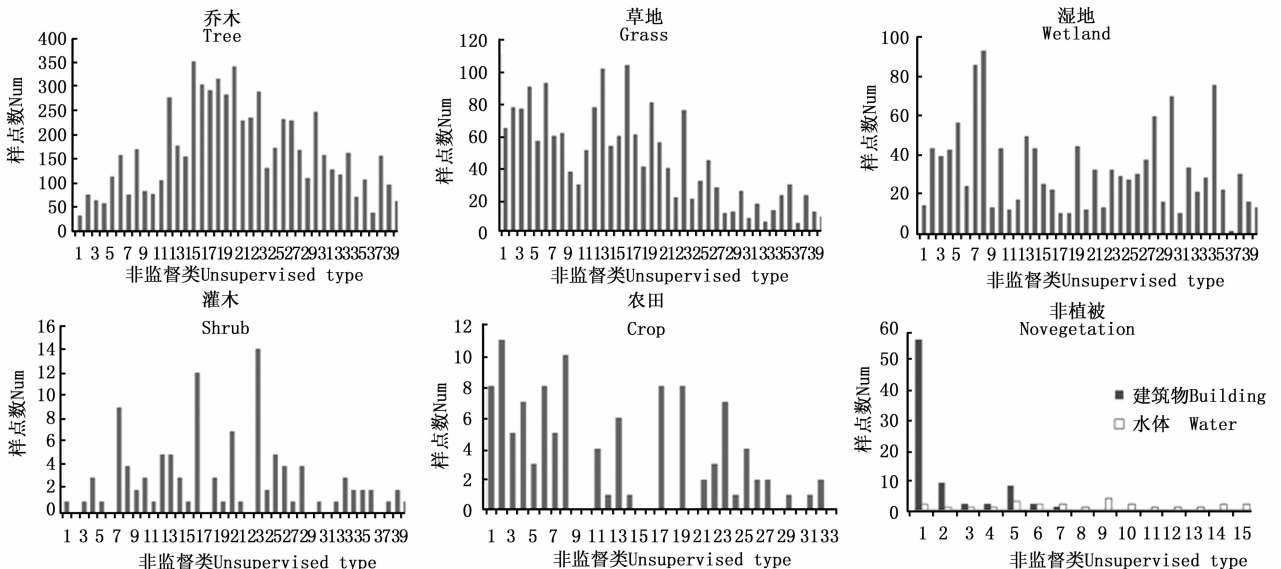


图 2 各类样本非监督分类类型属性特点直方图

Fig. 2 Histogram of samplings with unsupervised feature

统计结果表明,研究区域内植被以乔木、草地、湿地类为主,灌木、农田较少。该区域建筑物在非监督分类中特征明显,与其他地物很好的区分。水体零星分布,样点较少。本研究以植被类样本纯化提取为主,建筑物和水体为非植被类型样本不予讨论。

依据像元类型权重信息和空间信息剔除各类样本中不合要求的样点,进行样本纯化^[16]。保留权重更大的更有代表性的样点,剔除掉权重比例小于2.0%的样点,剩余的样点累积百分比为70%±5%,在此样点比例范围内再结合可分离性指标进行样点的取舍。

对保留样点的非监督分类类型、权重、累积百分比进行统计,如表1所示:

最后,从空间信息上判断,地类边界过渡区域的像元往往都是混合的非纯像元,因此剔除相邻类型边界1 000 m范围内过渡区域内的样点。

表1 纯化后样本中非监督分类类型及保留样点的权重和累积百分比

Table 1 Unsupervised types with weight factor and accumulative percent of purified points %

类型 Type	乔木 Tree	草地 Grass	湿地 Wetland	灌木 Shrub	农田 Crop
1	5.31	5.83	7.00	12.96	10.00
2	5.15	5.72	6.48	11.11	9.09
3	4.77	5.21	5.72	8.33	7.27
4	4.58	5.09	5.27	6.48	7.27
5	4.40	4.53	4.52	4.63	7.27
6	4.35	4.36	4.29	4.63	7.27
7	4.26	4.36	3.77	4.63	6.36
8	4.19	4.30	3.39	3.70	6.36
9	3.72	4.24	3.31	3.70	5.45
10	3.55	3.62	3.31	3.70	
11	3.50	3.45	3.31	2.78	
12	3.46	3.40	3.24	2.78	
13	3.46	3.34	3.01		
14	2.66	3.34	2.86		
15	2.60	3.17	2.56		
16	2.54	3.11	2.48		
17	2.53		2.48		
18	2.44				
19	2.38				
20	2.36				
总计 Total	72.21	67.06	67.00	69.44	66.36

3 结果与分析

3.1 纯化前后样点可分离性比较

采用 Jeffries-Matusita 距离(以下简称 J-M 距离)和转换分离度(Transformed Divergence,以下简称 T-

D)作为测度指标来检测所选取的任意两个类别训练样本之间的分离程度。可分离性指标(Separability)取两者中较小值,即 $Separability = \min(J-M, T-D)$ 。可分离性指标值为0~2.0,大于1.9说明样本之间可分离性好,1.9~1.8属于合格样本,小于1.8需要重新选择样本,小于1考虑将两类样本合成一类样本。

如表2所示,对于可分离性指标小于1.8的样本,认为样本中混有其他类型样点,因此,删除此类样本中权重比例小的干扰样点,再计算可分离性指标。反复操作以上过程,并比较分析可分离性指标,决定样点取舍。最终得到具有代表性的植被类型样本。

表2 纯化前后植被类样本可分离性指标对比

Table 2 Separabilities of purified and unpurified samples

地类 Types	纯化 Purify	乔木 Tree	灌木 Shrub	草地 Grass	湿地 Wetland
灌木 Shrub	前 Before	1.631			
	后 After	1.992			
草地 Grass	前 Before	0.901	1.488		
	后 After	1.970	1.997		
湿地 Wetland	前 Before	1.049	1.607	0.860	
	后 After	1.933	1.998	1.780	
农田 Crop	前 Before	1.915	1.959	1.741	1.887
	后 After	1.998	2.000	1.938	1.995

对比纯化前后的结果可以看出纯化后各类样本的可分离性指标提高到1.9以上,可分离性良好。只有湿地和草地类型可分离性指标值为1.78,认为湿地区域覆盖的植被和草地区域覆盖的植被多为草本植物,受实验中特征影像制约辨识度不高。

3.2 初始样点与纯化后样点的平均值、标准差比较

对比2001年MODIS全年数据的第33天、第145天、第257天、第353天的影像,纯化前后的各植被样本的平均值和标准差,如表3所示。

通过对比4个时相影像的统计结果可知,纯化前后平均值的变化无显著规律,纯化后的样本标准差大多缩小。纯化后的样本数据的分布曲线变窄,重叠变小,可分性增大,更易于类型间的区分。

3.3 纯化后的植被样本结果分析

3.3.1 纯化后的植被样本精度分析 中国植被图表明实验区植被覆盖以乔木、草地和湿地为主。植被覆盖率为乔木67.40%、草地18.04%和湿地12.48%。灌木0.97%,农田1.08%两类植被在实验区域较少分布。

为验证纯化后样本的空间分布状况,统计各植

被样本在中国土地覆盖图上分布的混淆矩阵,如表 3、4 所示。

表 4 中乔木、草地纯化后样本与中国土地覆盖类型中相应植被类型的空间分布基本一致,精确程

度达到 87.53% 和 90.46%。湿地与草地均以草本植被覆盖为主,两种植被类型在实验影像上不易区分,因此样本代表性相对较差。乔木、草地、湿地三者空间分布的总体精确程度为 84.82%。

表 3 纯化前后样本的平均值和标准差

Table 3 Mean and Stdev. values of purified and unpurified samples

地类 Types	纯化 Purify	灰度平均值 Mean gray value				灰度标准差 Gray Stdev.				像元数 Pixel amount	保留百分比 Retain percent/%
		B33d	B145d	B257d	B353d	B33d	B145d	B257d	B353d		
乔木 Tree	前 Before	1 693.5	7 509.6	5 776.4	2 294.4	1 957.3	1 038.6	852.6	1 832.5	6 685	
	后 After	1 424.1	7 677.8	5 879.9	2 229.0	1 640.6	769.1	812.4	1 706.8	3 489	52.19
灌木 Shrub	前 Before	1 406.1	7 299.2	5 879.9	2 229.0	1 510.1	1 128.7	687.8	1 694.7	108	
	后 After	861.2	7 364.6	4 877.4	1 939.8	753.8	1 080.3	627.8	1 193.0	31	28.70
草地 Grass	前 Before	960.0	6 709.1	5 257.3	1 896.2	1 391.8	1 459.9	928.3	1 653.1	1 767	
	后 After	337.5	5 375.8	4 692.7	1 207.7	468.5	1 340.4	754.2	1 106.7	409	23.15
湿地 Wetland	前 Before	1 732.6	6 788.4	5 441.0	2 348.7	1 845.6	1 410.5	854.9	1 881.0	1 328	
	后 After	1 004.6	5 865.5	5 119.6	1 842.5	1 063.5	1 306.3	804.9	1 574.2	212	15.96
农田 Crop	前 Before	590.1	6 300.5	5 369.6	1 694.0	1 167.8	1 659.6	1 197.2	1 639.9	110	
	后 After	268.3	5 508.8	5 225.2	1 190.6	479.3	1 374.9	946.0	968.3	31	28.18

表 4 高覆盖率植被样本与 WESTDC 中植被类匹配的混淆矩阵

Table 4 Compared the distribution of high coverage vegetation samples with WESTDC by confusion matrix

地类 Types	乔木 Tree		草地 Grass		湿地 Wetland		样点总数 Total
	样点数 Num.	百分比 Percent/%	样点数 Num.	百分比 Percent/%	样点数 Num.	百分比 Percent/%	
乔木 Tree	3 054	87.53	16	3.91	16	7.55	3 086
草地 Grass	377	10.81	370	90.46	130	61.32	877
湿地 Wetland	51	1.46	20	4.89	62	29.25	133
其他 Other	7	0.20	3	0.73	3	1.42	13
总计 Total	3 489	100.00	409	100.00	212	100.00	4 110

表 5 低覆盖率植被样本与 WESTDC 中植被类型匹配的混淆矩阵

Table 5 Compared the distribution of less coverage vegetation samples with WESTDC by confusion matrix

地类 Types	农田 Crop		灌木 Shrub		样点总数 Total
	样点数 Num.	百分比 Percent/%	样点数 Num.	百分比 Percent/%	
农田 Crop	28	90.32	1	3.23	29
灌木 Shrub	0	0.00	2	6.45	2
其他 Other	3	9.68	28	90.32	31
总计 Total	31	100.00	31	100.00	62

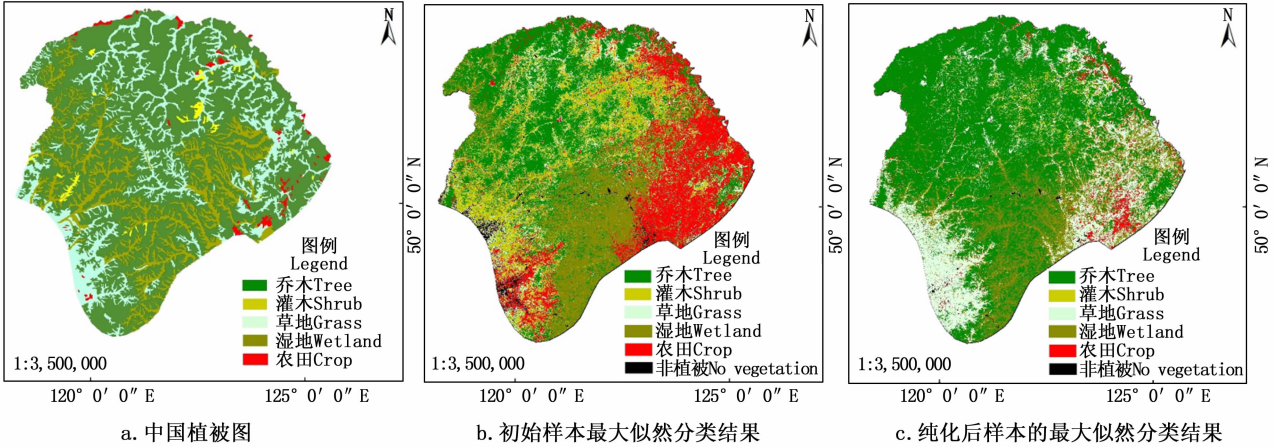
表 5 中农田样本与中国土地覆盖类型中相应植被类型的空间分布基本一致,精确程度达到 90.32%,由于实验区灌木分布极少,覆盖率仅为 0.97%,同时中国土地覆盖图与中国植被图中灌木类空间分布不一致,导致灌木样本以 WESTDC 为参考的空间分布状况精确程度较低。但灌木样本可分离性指标高于 1.9,不影响与其他类型的区分,故认为灌木样本基本不会影响到各类植被覆盖信息的表达。

实验样本中的植被空间分布与中国土地覆盖图中植被的空间分布相似度超过 80%,因此认为纯化后的样本在该区域具有代表性。

3.3.2 纯化后的植被样本分类结果分析 将纯化前后的样本,分别输入最大似然分类器,得到实验区域的分类结果,对比中国植被图图 3(a)中的植被分布,如图 3 所示。

验证样点在中国植被图和中国土地覆盖图中植被类型相同的区域随机选取。纯化后样本的最大似然分类结果图 3(c),总体精度为 76.78%;相比未经过纯化的初始样本最大似然分类结果图 3(b),总体精度 44.26%,提高了 32.52%。

由于特征影像、分类器选择会对分类精度造成影响,验证数据本身也与真实地物存在一定的精度局限性等,以上因素都可能降低实验影像最终分类结果的精度。因此,在三者相同的情况下,纯化后的样本分类结果精度较大提高,纯化后样本的分类结果与中国植被图中植被分布的总体趋势能够达到基本一致,认为采样方法是可行的。



a. Test area in vegetation atlas of China; b. Maximum likelihood classification result with initial samples; c. Maximum likelihood classification result with purified samples

图3 样本纯化前后分类结果与中国植被图中植被分布的比较分析

Fig. 3 Comparisons of classification results with (a) (b) (c)

4 结论

(1)以遥感监督分类方法快速精准的提取大区域植被覆盖信息为研究背景,本研究结合专家经验和实验数据同质特征向量的聚类信息,采用统计分析的方法,实现了针对宏观大区域遥感影像快速提取植被样本的方法。

(2)实验方法增强了样本的可分离性、缩小了样本标准差,经验证纯化后样本对实验区域的植被覆盖信息具有代表性,采样方法可行。所得样本具有实验影像时效特征,用于遥感监督分类可以缩短大区域植被变化监测时间间隔。

(3)实验节省了大区域植被类型调查耗费的人力物力资源和时间,避免了引入多源数据造成的数据冗余和误差影响,提高了采样效率。

参考文献:

[1] Chandra Giri, Zhiliang Zhu, Bradley Reed. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets[J]. *Remote Sensing of Environment*, 2005, 94: 123–132.

[2] M. C. Hansen, B. Reed. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products [J]. *Remote Sensing*, 2000, 21(6&7): 1365–1373.

[3] Darren Pouliot, Rasim Latifovic, Natalie Zabcic, *et al.* Development and assessment of a 250 m spatial resolution MODIS annual land cover time series(2000–2011) for the forest region of Canada derived from change-based updating [J]. *Remote Sensing of Environment*, 2014, 140: 731–743.

[4] Igro Klein, Ursula Gessner, Claudia Kuenzer. Regional land cover mapping and change detection in Central Asia using MODIS time-se-

ries[J]. *Applied Geography*, 2012, 35: 219–234.

[5] 张永红,张继贤,郭建,等.全球地图第一版中国土地覆盖产品的生成[J]. *武汉大学学报:信息科学版*, 2009, 34(8): 914–925.

[6] 左玉珊,王卫,郝彦莉,等.基于MODIS影像的土地覆被分类研究——以京津冀地区为例[J]. *地理科学进展*, 2014, 33(11): 1556–1565.

[7] 范应龙,谭炳香.两种热带森林覆盖变化检测方法比较研究[J]. *林业科学研究*, 2015, 28(3): 325–331.

[8] 张雨,林辉,臧卓,等.基于MODIS数据的辽宁省土地利用分类[J]. *西南林业大学学报*, 2014, 34(1): 52–57.

[9] 刘华,鞠洪波,邹文涛,等.长江源典型区湿地对区域气候变化的响应[J]. *林业科学研究*, 2013, 26(4): 406–413.

[10] 陈巧,陈永富,鞠洪波.基于3S技术的天保区植被变化监测方法研究[J]. *林业科学研究*, 2013, 26(6): 736–743.

[11] Kun Jia, Shunlin Liang, Lei Zhang, *et al.* Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 33: 32–38.

[12] 竞霞,王锦地,王纪华,等.基于分区和多时相遥感数据的山区植被分类研究[J]. *遥感技术与应用*, 2008, 23(4): 394–398.

[13] 廖靖,覃先林,周汝良.利用2种植被指数监测中国6种典型森林生长期的比较研究[J]. *西南林业大学学报*, 2014, 34(3): 57–61.

[14] 冉有华,李新,卢玲.基于多源数据融合方法的中国1 km土地覆盖分类制图[J]. *地球科学进展*, 2009, 24(2): 192–203.

[15] 李俊祥,达良俊,王玉洁,等.基于NOAA-AVHRR数据的中国东部地区植被遥感分类研究[J]. *植物生态学报*, 2005, 29(3): 436–443.

[16] 吴健平,杨卫星.遥感数据监督分类中训练样本的纯化[J]. *国土资源遥感*, 1996, (27): 36–41.

(责任编辑:彭南轩)