

引用格式: Wang Bingquan, Ran Youhua. Comparison of Accuracy Assessment Methods of Remote Sensing based Land Cover Products[J]. Remote Sensing Technology and Application, 2022, 37(1): 196-204. [王冰泉, 冉有华. 土地覆被遥感产品真实性检验方法对比[J]. 遥感技术与应用, 2022, 37(1): 196-204.]
DOI: 10.11873/j.issn.1004-0323.2022.1.0196

土地覆被遥感产品真实性检验方法对比

王冰泉^{1,2}, 冉有华^{1,2}

(1. 中国科学院西北生态环境资源研究院, 甘肃 兰州 730000;

2. 中国科学院大学, 北京 100049)

摘要: 真实性检验是保障土地覆被遥感产品生产质量, 支持土地覆被遥感产品应用的重要环节。土地覆被遥感产品真实性检验实践总与理论存在一些差异, 其主要体现在用于检验的参考数据样本数量和空间分布两个方面, 而这种差异对检验结果的影响还不十分清楚。以中国陆地范围为实验区域, GlobeLand30 作为全样本参考数据, ESA CCI-LC 作为检验对象, 通过设计实验, 评估了样本数量、抽样方式和样本单元的不同对检验结果的影响。结果表明: ①样本量对检验结果有重要影响。当样本量减半(约 600 个样本点), 点样本真实性检验的总体精度与在理论样本量下的总体精度接近, 但是, 在样本量较小的情况下, 不同类型的精度对样本量的敏感程度不同, 面积权重小的类别对样本量更敏感, 尤其是在采用分层随机抽样的情况下; ②在理论样本量基础上, 无论是点样本真实性检验还是群样本真实性检验, 抽样方式对总体精度影响不大, 但是对各个类型的精度有重要影响, 尤其是面积权重小的类别; ③群样本检验稳定性比点样本检验差。在理论样本量基础上, 样本单元的增大将增加检验结果的不确定性, 随着样本单元的增加, 虽然检验结果波动性变小, 但是检验结果相较于点样本检验结果仍存在一定偏差。因此, 对于以获取总体精度为目标的真实性检验, 对于点样本, 样本量可以减少到理论样本量的一半, 总体精度的绝对误差在 1% 以内, 标准差约为 2%, 而对于群样本, 则需要增加到理论样本量的 5 倍。对于以获取各类型精度为目标的真实性检验, 建议增加样本量为理论样本量的 2 倍以上, 以确保面积权重小的类型被分配足够的样本点。

关键词: 抽样方式; 样本量; 真实性检验; ESA CCI-LC; GlobeLand30

中图分类号: P237; K903 **文献标志码:** A **文章编号:** 1004-0323(2022)01-0196-09

1 引言

土地覆被是理解人类活动与环境变化及其复杂关系的关键要素^[1], 是生态评估、表层地球系统模拟等的基础数据^[2]。随着遥感数据的积累和分类技术的进步, 各行各业都生产了或正在生产着大量的土地覆被遥感产品^[3], 有力支持了地学研究和国土、农林、水利、环保、气候等业务应用^[4-7], 土地覆被遥感产品已成为应用最为广泛的遥感产品之一。

土地覆被遥感产品由于数据源、分类技术甚至生产者的不同, 可能造成产品精度的较大差异^[8-15]。由于不同的应用目标, 土地覆被遥感产品在空间分辨率、分类系统等方面存在显著差异, 使得对其真实性检验存在复杂性^[16-18]。即使对于同一个产品, 不同的真实性检验方法往往给出不同的真实性检验结果^[19], 这给产品的应用带来一定困扰, 没有经过严格真实性检验的土地覆被遥感产品应用在科学模型中可能导致模型误差难以解释^[20-22], 在管理

收稿日期: 2021-02-18; 修订日期: 2021-12-27

基金项目: 国家自然科学基金项目(42071421)和高分辨率对地观测系统国家重大专项(21-Y20B01-9001-19/22)资助。

作者简介: 王冰泉(1997—), 男, 河南南阳人, 硕士研究生, 主要从事生态遥感研究。E-mail: wangbingquan@nieer.ac.cn

通讯作者: 冉有华(1980—), 男, 甘肃镇原人, 正高级工程师, 主要从事冰冻圈遥感、生态遥感与模型、遥感产品真实性检验和科学数据集研究。E-mail: ranyh@lzb.ac.cn

业务应用中甚至可能带来决策失误^[23-24]。因此,对比不同的土地覆被遥感产品真实性检验方法,有助于理解不同检验方案对检验结果的影响,推进土地覆被遥感产品真实性检验的标准化,促进土地覆被遥感产品质量的提升。

针对土地覆被遥感产品的真实性检验,国内外都做过大量研究,形成了一系列真实性检验理论方法。抽样检验是土地覆被遥感产品真实性检验常用的方法,这些方法的核心是如何确定用于真实性检验的参考数据的样本数量和空间分布^[25-27]。Olofsson等^[28]列举了分层随机抽样和简单随机抽样的样本量估计模型,总结了土地覆被产品精度评估包括抽样设计、响应设计和分析3个主要部分。Stehman^[29]指出分层随机抽样下各个层样本量分配对精度评估指标的影响,为分层随机抽样各个层样本量的分配提供参考。各种概率抽样设计已经普遍用于土地覆被产品的精度评估,最常用的概率抽样方法有简单随机抽样、分层抽样、系统抽样和群抽样^[28]。刘旭拢等^[19]在研究区域评估了简单随机抽样、系统抽样和分层随机抽样下的点样本和群样本验证效果,认为验证方法对精度评估的影响是客观存在的,点样本精度评估结果波动性小于群样本。孟雯等^[30-32]根据空间的数据的相关性,将空间抽样理论应用于土地覆被产品精度评估,大大提高了精度评估的验证效率和准确性。

然而,在真实性检验实践中,受成本和可操作性等方面的影响,实际的样本量和样本分布往往不同于理论计算结果。这样就带来一个问题,样本量及其分布的一定变化会不会对真实性检验结果产生影响?即使相同的样本量或相同的抽样方式,不同的抽样实现其检验结果是否稳定?这种影响到底有多大?在不同土地覆被类型之间的影响有什么差别?

针对这些问题,以CCI-LC作为检验对象,GlobeLand30作为参考对象,设计不同样本量和抽样方式的检验实验,分析抽样方式和样本量的变化对土地覆被产品精度评估的影响,更深入地理解土地覆被遥感产品真实性检验结果不确定性的来源,为土地覆被遥感产品真实性检验实践提供一些参考。

2 数据与方法

2.1 土地覆被遥感数据与预处理

以欧洲空间局全球土地覆被数据集(Climate

Change Initiative-Land Cover, CCI-LC)2010年中国陆地区域的数据作为检验对象,利用2010年GlobeLand 30作为参考数据,开展检验试验。CCI-LC是欧洲空间局研发的1992~2015年逐年全球土地覆被数据集,其空间分辨率为300 m,其采用的是联合国粮食及农业组织研发的LCCS(LC Classification System)分类系统,一共有22个一级类型,36个二级类型,是目前时间序列最长的全球土地覆被产品^[33]。GlobeLand 30是我国研发的首套30 m空间分辨率土地覆被遥感产品,包括2000、2010和2020年3个基准年,总体精度为75.1%,Kappa系数为0.71,分类系统包括耕地、森林和草地等10个一级土地类型^[34]。

为了保证参考数据与检验对象的一致,首先,统一GlobeLand 30数据和CCI-LC数据的坐标系,将GlobeLand 30数据重采样为300 m分辨率,重采样的方法是在300 m栅格内,面积权重最大的类型作为该栅格的类型。其次,建立CCI-LC和GlobeLand30土地覆被产品的公共分类系统,对两种数据进行类型转换,以保持其分类系统的一致性,公共分类体系如表1所示。统一分类系统之后,将CCI-LC和GlobeLand 30数据产品按照统一的分类系统分别进行重分类得到8个类型的土地覆被数据产品。

2.2 实验设计

为了测试不同样本量对检验结果的影响,实验基于样本量估计模型确定理论样本量,在此基础上简单随机抽样、系统抽样和分层随机抽样以100间隔减少和增加样本数量(样本量分别为100,200, ..., 6 000),群抽样以11×11的像元集合增加样本单元的数量(样本单元数量分别为10,11, ..., 50),分析检验结果的变化。为了测试不同抽样方式对检验结果的影响,对于一定数量的样本,对比分析常用的简单随机抽样、系统抽样和分层随机抽样对检验结果的敏感性。为了减少一次检验的随机性,每组实验重复100次,分析100次结果平均的变化和不同实验重复的变异性。以GlobeLand 30全样本检验结果为假设检验结果的真值,评估每组实验的相对变化。另外,本研究还对点样本和群样本检验的差异。

基于概率统计理论的样本量估计模型是通过设置抽样误差和预估分类误差计算研究区域所需的理论样本量,预估分类误差和样本量计算公式

表 1 CCI-LC 和 GlobeLand 30 土地覆被分类产品的公共分类系统

Table 1 The eight aggregated classes for CCI-LC and GlobeLand 30 land cover classification products			
编号	类型	CCI-LC	GlobeLand30
1	耕地	雨养农田(10、11、12);灌溉农田(20);以农田为主(覆盖度大于50%)的农林牧交错区(30);林灌草盖度大于50%的自然植被和农作物交错带(40)	耕地(10)
2	森林	盖度大于15%的常绿阔叶林(50);盖度大于15%的落叶阔叶林(60、61、62);盖度大于15%的常绿针叶林(70、71、72);盖度大于15%的落叶针叶林(80、81、82);阔叶针叶混合林(90);林地盖度大于50%的林草混交带(100);淡水或苦咸水淹没的林地(160);盐水淹没的林地(170)	森林(20)
3	草地和灌木地	以草本植物为主(覆盖度大于50%)的林灌草交错带(110);草地(130);灌木丛(120、121、122)	草地和灌木地(30、40)
4	湿地	灌丛或草本植物覆盖的湿地(180)	湿地(50)
5	人造地表	城市(190)	人造地表(80)
6	裸地和稀疏植被	地衣和苔藓(140);稀疏植被(150、151、152、153);裸地(200、201、202)	苔原(70);裸地(90)
7	水体	水体(210)	水体和海洋(60、255)
8	冰川和常年积雪	冰川和常年积雪(220)	冰川和常年积雪(100)

如下^[35-36]:

$$p=\sum_{i=1}^m p_i\times w_i$$
 (1)

其中: i 表示第 i 个类型; m 表示类型数; p_i 表示第 i 个类型的预估分类误差; w_i 表示第 i 个类型的面积权重。

$$n=\frac{\frac{\mu_{1-\alpha/2}^2(1-p)}{r^2p}}{1+\frac{1}{N}\left(\frac{\mu_{1-\alpha/2}^2(1-p)}{r^2p}-1\right)}$$
 (2)

其中: n 表示所要计算的样本量; $\mu_{1-\alpha/2}$ 表示置信度取 $1-\alpha/2$ 时的标准正态分布临界值; α 取5%; p 表示综合预估分类误差; r 为抽样误差; N 为样本总量,当确定了待检验土地覆被产品各类型的预计分类误差和抽样误差即可获得研究区域所需的样本点数量。

在研究区域进行20次样本量为2 000的简单随机抽样,将20次真实性检验中各个类型的制图精度的平均值作为不同类型的预测精度。通过公式(1)计算8个地物类型的总和预计错误率,再将其代入公式(2)计算中国陆地区域所需要的样本。计算得到的综合预计分类误差 $p=25.2\%$,抽样误差取0.1,因此,中国陆地区域所需样本量 $n=1\,140$ 。

实验中,不同的检验方法重复进行100次,各种抽样方式的实现分为两步:第一,根据抽样方式的特点创建随机点;第二,获得随机点的待验证地物类型和真实类型。

采用简单随机抽样和分层抽样创建随机点时,运用ArcMap 10.4中的“创建精度评估点”工具,在CCI-LC土地覆被图层上随机创建精度评估点,其中,分层随机抽样是以每个地物类型作为一个层,按面积权重向各个层分配样本点。采用系统抽样

创建随机点时,采用Python 3.7将CCI-LC土地覆被产品的像元按次序排列,随机确定抽样起始点,按照相同的间隔抽取下一个像元直到达到总样本点数量。群抽样则是采用ArcMap 10.4和Python 3.7,在研究区域进行简单随机抽样获得每一个群的中心像元,以该像元为中心抽取窗口大小不同的群,群的大小分别为 3×3 、 5×5 、 7×7 、 9×9 和 11×11 。

获得随机点的待验证地物类型和真实类型采用ArcMap 10.4中的“更新精度评估点”工具。最后,计算混淆矩阵,得到各种检验方法下的总体精度、用户精度和制图精度。

3 结果与分析

3.1 全样本真实性检验

以GlobeLand 30作为参考数据,对CCI-LC土地覆被产品进行全样本检验,以该评价结果作为参考值,讨论不同试验组合的检验结果,2010年CCI-LC数据的全样本检验结果如表2。

分类结果的总体精度为73.50%,Kappa系数为

表 2 全样本真实性检验
Table 2 Accuracy evaluation of full sample

类型	面积权重	用户精度	制图精度	总体精度	Kappa
	/%	/%	/%	/%	系数
耕地	28.48	63.86	83.79	73.50	0.65
森林	18.93	85.02	73.57		
草地和灌木地	28.12	70.82	66.35		
湿地	0.37	44.18	34.55		
人造地表	1.37	81.71	71.89		
水体	1.00	71.70	41.94		
裸地和稀疏植被	21.05	79.92	77.86		
冰川和常年积雪	0.67	70.95	45.75		

65.43%。从用户精度来看,除了湿地和耕地,其他类型的精度均大于70%,森林的用户精度最高,湿地的用户精度最低,从制图精度来看,耕地、森林、人造地表、裸地和稀疏植被精度较高,均大于70%,而湿地、水体、冰川和常年积雪的精度较差,耕地的制图精度最高,湿地的制图精度最低。

3.2 抽样方式对总体精度的影响

基于样本量估计模型计算中国陆地区域所需要的样本量为1 140,分别采用简单随机抽样、系统抽样、分层随机抽样和群抽样4种抽样方式分配样本点,群抽样的样本单元分别采用 3×3 、 5×5 、 7×7 、 9×9 和 11×11 的窗口,得到各种抽样方式下相对应的总体精度和总体精度标准差(表3和图1)。

多次重复实验的结果表明,在理论样本量基础上,虽然不同抽样方式在100次抽样实现中存在一定变异(标准差约为1.20%),但3种抽样方式对总体精度的影响都很小,均能反映检验对象的总体精度,相对于全样本检验结果,偏差不超过0.35%,这与刘旭拢等^[9]的结论是一致的。但如果采用群样本,群样本单元的大小对检验结果有一定的影响。如图1所示,采用不同群样本单元大小对总体精度平均值影响不大,但随着群样本单元的增大,总体精度的标准差随之增大,在群样本大小为 11×11 像元时,标准差超过6.05%,说明在真实性检验实验中,为了提高参考对象数据获取的效率,可以采用群抽样,但在总样本量不变的情况下,会增加检验结果不确定性的风险。

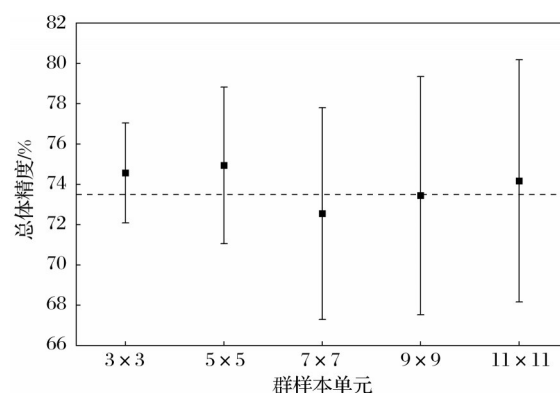
表3 理论样本数量下不同抽样方式对总体精度的影响

Table 3 The influence of different sampling methods on the overall accuracy under the theoretical sample size

	总体精度	标准差	绝对误差
简单随机抽样	73.83%	1.29%	0.33%
系统抽样	73.57%	1.25%	0.07%
分层随机抽样	73.44%	1.20%	0.06%

3.3 样本量对总体精度的影响

样本量以100为间隔从100至6 000分别进行简单随机抽样、系统抽样和分层随机抽样,测试样本量对不同抽样方式下真实性检验结果的影响。如图2(a)所示,总体上,随着样本量的增加,简单随机抽样、系统抽样和分层随机抽样的总体精度在一定的范围内趋于稳定,标准差逐渐减小。当样本量达到理论样本量的一半时(约600个),3种抽样方式真实性检验结果的绝对误差减小到1%以内,但标



(虚线代表全样本检验的总体精度)

图1 不同群抽样下总体精度比较

Fig.1 Comparison of overall accuracy under different cluster sampling units

准差依然较大,超过2%。当样本量达到理论样本量的3倍时(约3 000个),总体精度的标准差接近1%。不仅如此,不同抽样方式对样本量的敏感性不同,特别是在样本量较小时这种差异更加突出,当简单随机抽样下样本量达到300,系统抽样下样本量达到100,分层随机抽样下样本量达到600时,总体精度绝对误差均小于0.50%,敏感程度大小次序是分层随机抽样>简单随机抽样>系统抽样。

以 11×11 为样本单元,分别抽取10,11,...,50个群样本单元(对应的点样本量为1 210,1 331,...,6 050),如图2(b)所示,总体上,随着样本单元数量的增加,总体精度的绝对误差和标准差都在逐渐下降。与点样本相比,在相似样本量情况下,群样本检验方式的稳定性要差的多,当样本量达到理论样本量的5倍时(约50个样本单元,6 050个点样本),总体精度的标准差才接近于点样本理论样本量时的标准差。

3.4 不同的验证方法在各类型之间的差异

以耕地、人造地表和湿地为研究对象,分别代表面积权重为大、中、小的类型,耕地、人造地表和湿地的用户精度和制图精度随样本量的变化情况见图3~图6。总体上,随着样本量的增加,点样本检验下耕地、湿地和人造地表的精度和制图精度在一定的范围内趋于稳定,标准差逐渐减小。在相同样本量下,3个类型的用户精度和制图精度的标准差大小次序是:湿地>人造地表>耕地(图3~图5)。简单随机抽样和系统抽样下检验结果相似,当样本量达到理论样本量的一半时(约600个),耕地用户精度和制图精度的绝对误差在1%以内,但标准差依然较大,约为3.50%,湿地用户精度和制图

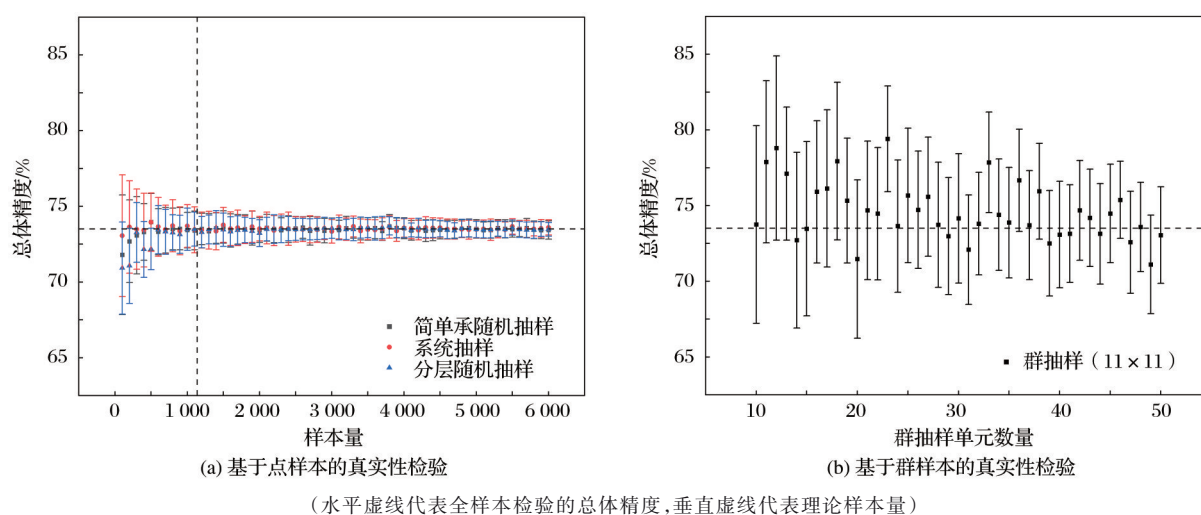


图 2 总体精度随样本量的变化

Fig.2 The variation of overall accuracy with sample size

精度的绝对误差减小到5%以内,标准差超过30%,人造地表用户精度和制图精度的绝对误差在5%以内,标准差小于15%。当样本量达到理论样本量的3倍时(约3000个),耕地用户精度和制图精度的绝对误差在0.5%以内,标准差接近1.5%,湿地用户精度和制图精度的绝对误差小于3%,标准差接近15%,人造地表用户精度和制图精度的绝对误差小

于0.6%,标准差接近6%。此外,不同抽样方式下各类型精度指标对样本量的敏感性也不同,尤其是分层随机抽样下,各类型的制图精度对样本量的敏感性与类型的面积权重有关,面积权重越小,该类型对样本量的敏感性就越大。当样本量分别达到600、1300和2300时,耕地、人造地表和湿地的制图精度才开始稳定,此时的制图精度绝对误差均小于4%。

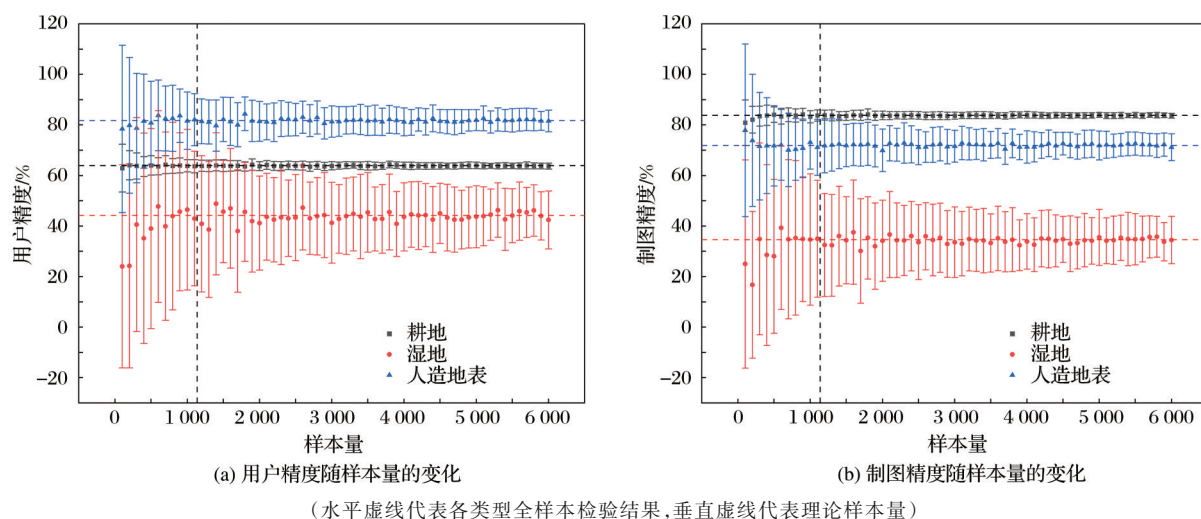


图 3 简单随机抽样下精度指标随样本量的变化

Fig.3 Changes of accuracy index with sample size under simple random sampling

以 11×11 为样本单元,分别抽取10,11,...,50个群样本单元(对应的点样本量为1210,1331,...,6050),如图6所示,群抽样下的各个类型真实性检验结果波动性大,耕地验证效果优于湿地和人造地表,湿地和人造地表的真实性检验结果低估全样本真实性检验结果。当样本量达到理论样本量的3倍时(约25个样本单元,3025个点样本),耕地用户精

度和制图精度绝对误差小于5%,标准差大于5%,湿地和人造地表的精度和制图精度绝对误差大于10%,标准差大于24%,即使样本量达到理论样本量的5倍,湿地和人造地表用户精度和制图精度的绝对误差依然较大。这说明对于面积权重较小(小于2%)的类型,需要专门设计检验方案,利用更多的样本开展真实性检验。

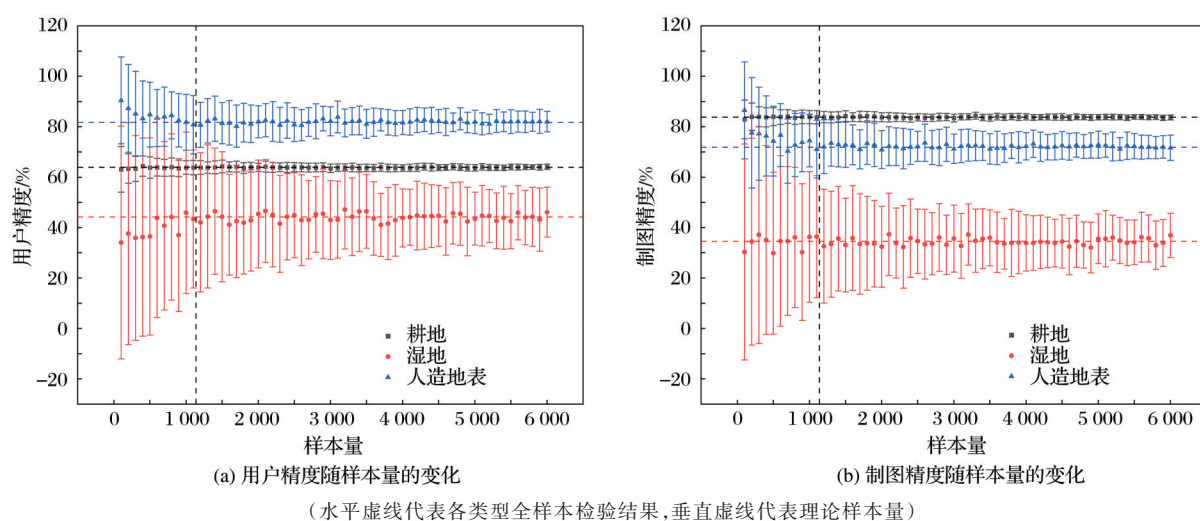


图 4 系统抽样下精度指标随样本量的变化

Fig.4 Changes of accuracy index with sample size under systematic sampling

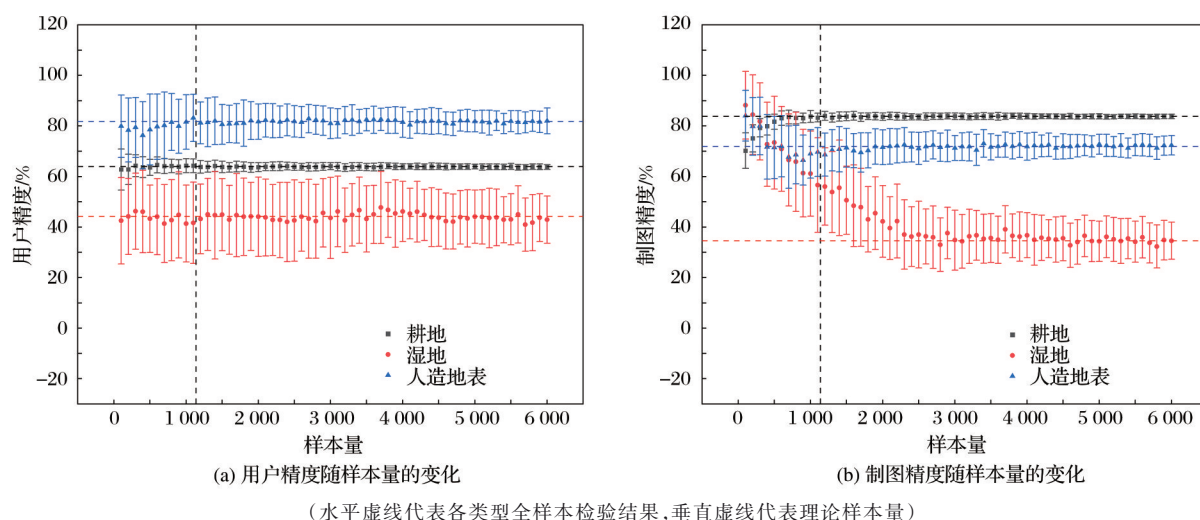


图 5 分层随机抽样下精度指标随样本量的变化

Fig.5 Changes of accuracy index with sample size under stratified random sampling

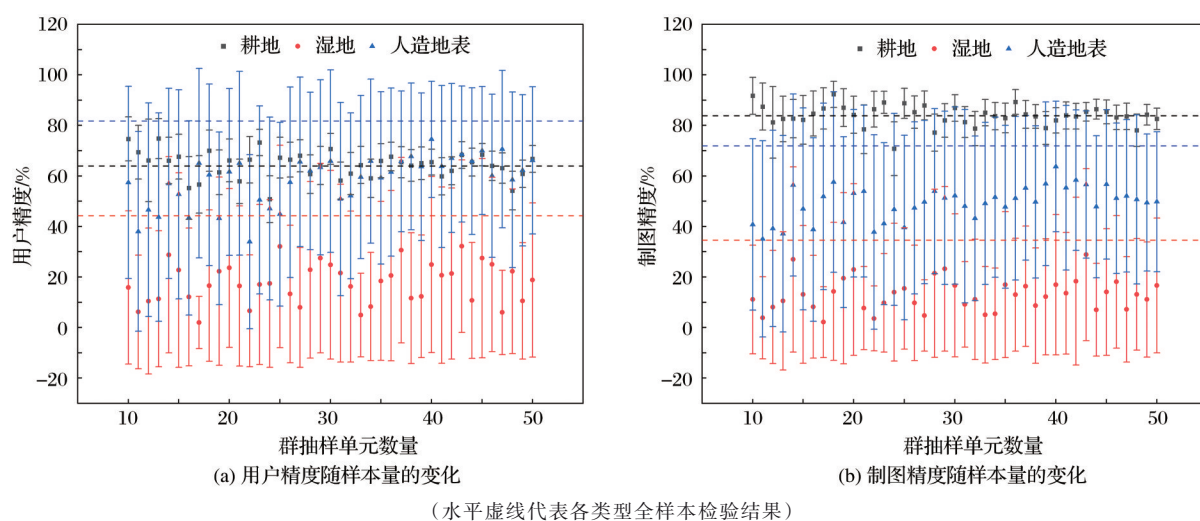


图 6 群抽样(11×11)下精度指标随样本量的变化

Fig.6 Changes of accuracy index with sample size under cluster sampling (11×11)

4 结 语

以2010年GlobeLand30为参考数据,在中国陆地范围对ESA CCI-LC土地覆被产品开展检验试验,测试不同样本量、不同抽样方式和不同样本单元大小对检验结果的影响。结果表明:

(1)样本量对类型水平上的检验结果有重要影响。当样本量达到理论样本量的一半时(约600个)或理论样本量的3倍时(约3 000个),真实性检验总体精度的绝对误差接近1%,但理论样本量一半下的总体精度的标准差较大,约为2%。土地覆被类型水平上的用户精度和制图精度对样本量较为敏感,面积权重小的类型对样本量更为敏感。

(2)抽样方式对总体精度的影响不大,但对样本量的敏感性有所不同。在理论样本量情况下,利用简单随机抽样、系统抽样和分层随机抽样检验的总体精度绝对误差都小于1%。但不同抽样方式对样本量变化的敏感性不同,特别是在样本量较小时(小于理论样本量的一半)这种差异更加突出。

(3)不同土地覆被类型检验结果对样本量的敏感性不同。对于点样本真实性检验,面积权重小的类型对样本量更为敏感,尤其是在分层随机抽样下,当样本量达到理论样本量的2倍时(约2 300个点样本),湿地的制图精度的绝对误差才接近于简单随机抽样和系统抽样在理论样本量时的绝对误差。对于群样本真实性检验,各个类型真实性检验结果比点样本的波动性大,即使样本量达到理论样本量的5倍(约50个群样本单元,6 050个点样本),湿地和人造地表用户精度和制图精度的绝对误差依然较大。

上述研究发现可为真实性检验实践中,抽样方式、样本量和抽样单元的确定提供参考。在实际土地覆被遥感产品真实性检验过程中,抽样方式、样本量和样本单元大小都对检验结果有一定影响。对于以获取总体精度为目标的真实性检验,样本量和抽样方式对检验结果的影响不大,但是,点样本真实性检验结果的稳定性要比群样本检验结果稳定性好。对于点样本,样本量可以减少到理论样本量的一半,总体精度的绝对误差在1%以内,标准差约为2%。而对于群样本,则存在增加总体精度不确定性的风险,即使当样本量达到理论样本量的5倍时,总体精度的标准差才接近点样本的标准差。对于以获取各个类型精度为目标的真实性检验,样本量、抽样方式和样本单元对检验结果有重要影

响。对于点样本,除了分层随机抽样,其他的抽样方式可能无法保证面积权重小的类型分配有足够的样本,但是,即使是分层随机抽样,对于面积权重小的类型在理论样本量时也存在较大的不确定性,建议通过增加样本量为理论样本量的2倍以上,以减小这方面的不确定性,此时各个类型的用户精度和制图精度绝对误差约为1%,标准差减小了5%。相较于点样本,群样本虽然可以节约参考数据的调查成本,但是,可能会增加面积权重小的类型检验结果的不确定性,可以增大样本单元的数量或者适当减小样本单元的大小以减小检验结果的不确定性。

本研究以2010年GlobeLand30作为参考数据,以全样本检验结果为参考精度,存在一定的不确定性。未来随着地面验证数据的积累,可利用大量的地面验证数据开展检验实验,更精确地理解不同检验方案对检验结果的影响。从理论角度,传统的样本量估算模型没有显式考虑检验对象的空间异质性,可能会导致样本量的冗余或欠缺^[37],即样本的代表性出现偏差。同时,传统的样本量估算模型中也没有考虑检验范围的大小,导致不同区域大小的样本量一样,造成检验结果的空间代表性差异。因此,未来可以通过进一步发展考虑检验对象空间异质性和空间范围的样本量估算模型,以提高样本的代表性和检验结果的可比性。

参考文献(References):

- [1] Chen Jun, Chen Jin, Gong Peng, *et al.* High resolution global land cover mapping[J]. *Geomatics World*, 2011, 9(2): 12-14. [陈军,陈晋,宫鹏,等. 全球地表覆盖高分辨率遥感制图[J]. *地理信息世界*, 2011, 9(2): 12-14.]
- [2] Li Xiaobing, Chen Yunhao, Yu Feng. Global and regional cover mapping from remote sensing data: status quo, strategies and trends[J]. *Advance in Earth Sciences*, 2004, 19(1): 71-80. [李晓兵,陈云浩,喻锋. 基于遥感数据的全球及区域土地覆盖制图——现状、战略和趋势[J]. *地球科学进展*, 2004, 19(1): 71-80.]
- [3] Grekousis G, Mountrakis G, Kavouras M. An overview of 21 global and 43 regional land-cover mapping products[J]. *International Journal of Remote Sensing*, 2015, 36(21): 5309-5335. DOI: 10.1080/01431161.2015.1093195.
- [4] Chen Jun, Chen Jin, Liao Anping, *et al.* Concepts and key techniques for 30 m global land cover mapping[J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(6): 551-557. [陈军,陈晋,廖安平,等. 全球30 m地表覆盖遥感制图的总体技术[J]. *测绘学报*, 2014, 43(6): 551-557.]
- [5] Foley J A, DeFries R, Asner G P, *et al.* Global consequences of land use[J]. *Science*, 2005, 309(5734): 570-574. DOI:

- 10.1126/science.1111772.
- [6] Zell E, Huff A K, Carpenter A T, *et al.* A user-driven approach to determining critical earth observation priorities for societal benefit[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2012, 5(6): 1594-1602. DOI: 10.1109/JSTARS.2012.2199467.
- [7] Sterling S M, Ducharme A, Polcher J. The impact of global land-cover change on the terrestrial water cycle[J]. *Nature Climate Change*, 2013, 3(4): 385-390. DOI: 10.1038/nclimate1690.
- [8] Coppin P, Lambin E, Jonckheere I, *et al.* Digital change detection methods in natural ecosystem monitoring: a review[J]. *Analysis of Multi-temporal Remote Sensing Images*, 2002: 3-36. DOI: 10.1142/9789812777249_0001.
- [9] Croke B F W, Merritt W S, Jakeman A J. A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments[J]. *Journal of Hydrology*, 2004, 291(1-2): 115-131. DOI: 10.1016/j.jhydrol.2003.12.012.
- [10] He C, Okada N, Zhang Q, *et al.* Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China[J]. *Applied Geography*, 2006, 26(3-4): 323-345. DOI: 10.1016/j.apgeog.2006.09.006.
- [11] Herold M, Mayaux P, Woodcock C E, *et al.* Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets[J]. *Remote Sensing of Environment*, 2008, 112(5): 2538-2556. DOI: 10.1016/j.rse.2007.11.013.
- [12] Giri C, Zhu Z, Reed B. A comparative analysis of the global land cover 2000 and MODIS land cover data sets[J]. *Remote Sensing of Environment*, 2005, 94(1): 123-132. DOI: 10.1016/j.rse.2004.09.005.
- [13] Verburg P H, Neumann K, Nol L. Challenges in using land use and land cover data for global change studies[J]. *Global Change Biology*, 2011, 17(2): 974-989. DOI: 10.1111/j.1365-2486.2010.02307.x.
- [14] Gong P, Wang J, Yu L, *et al.* Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data[J]. *International Journal of Remote Sensing*, 2013, 34(7): 2607-2654. DOI: 10.1080/01431161.2012.748992.
- [15] Congalton R G, Gu J, Yadav K, *et al.* Global land cover mapping: a review and uncertainty analysis[J]. *Remote Sensing*, 2014, 6(12): 12070-12093. DOI: 10.3390/rs61212070.
- [16] Wu Wenbin, Yang Peng, Zhang Li, *et al.* Accuracy assessment of four global land cover datasets in China[J]. *Transactions of the CSAE*, 2009, 25(12): 167-173.[吴文斌, 鹏杨, 张莉, 等. 四类全球土地覆盖数据在中国区域的精度评价[J]. *农业工程学报*, 2009, 25(12): 167-173.]
- [17] Ran Youhua, Li Xin, Lu Ling. Accuracy evaluation of the four remote sensing based land cover productions over China[J]. *Journal of Glaciology and Geocryology*, 2009, 31(3): 490-500.[冉有华, 李新, 卢玲. 四种常用的全球 1 km 土地覆盖数据中国区域的精度评价[J]. *冰川冻土*, 2009, 31(3): 490-500.]
- [18] Liu Qionghuan, Zhang Yili, Liu Linshan, *et al.* Accuracy evaluation of the seven land cover data in Qiangtang Plateau[J]. *Geographical Research*, 2017, 36(11): 2061-2074.[刘琼欢, 张懿锂, 刘林山, 等. 七套土地覆被数据在羌塘高原的精度评价[J]. *地理研究*, 2017, 36(11): 2061-2074.]
- [19] Liu Xulong, He Chunyang, Pan Yaozhong, *et al.* Accuracy assessment of thematic classification based on point and cluster sample[J]. *Journal of Remote Sensing*, 2006, 10(3): 366-372.[刘旭珑, 何春阳, 潘耀忠, 等. 遥感图像分类精度的点、群样本检验与评估[J]. *遥感学报*, 2006, 10(3): 366-372.]
- [20] Foody G M. Status of land cover classification accuracy assessment[J]. *Remote Sensing of Environment*, 2002, 80(1): 185-201. DOI: 10.1016/S0034-4257(01)00295-4.
- [21] Wulder M A, Coops N C. Satellites: make earth observations open access[J]. *Nature News*, 2014, 513(7516): 30. DOI: 10.1038/513030a.
- [22] Moristette J T, Privette J L, Christopher O, *et al.* A framework for the validation of MODIS land cover products[J]. *Remote Sensing of Environment*, 2002, 83(1/2): 77-96. DOI: 10.1016/S0034-4257(02)00088-3.
- [23] Foody G M. Harshness in image classification accuracy assessment[J]. *International Journal of Remote Sensing*, 2008, 29(11): 3137-3158. DOI: 10.1080/01431160701442120.
- [24] Tsensdbazar N E, De Bruin S, Herold M. Assessing global land cover reference datasets for different user communities[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 103: 93-114. DOI: 10.1016/j.isprsjprs.2014.02.008.
- [25] McCallum I, Obersteiner M, Nilsson S, *et al.* A spatial comparison of four satellite derived 1 km global land cover datasets[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2006, 8(4): 246-255. DOI: 10.1016/j.jag.2005.12.002.
- [26] Fritz S, See L, McCallum I, *et al.* Highlighting continued uncertainty in global land cover maps for the user community[J]. *Environmental Research Letters*, 2011, 6(4): 044005. DOI: 10.1088/1748-9326/6/4/044005.
- [27] Tchuenté A T K, Roujean J L, De Jong S M. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2011, 13(2): 207-219. DOI: 10.1016/j.jag.2010.11.005.
- [28] Olofsson P, Foody G M, Herold M, *et al.* Good practices for estimating area and assessing accuracy of land change[J]. *Remote Sensing of Environment*, 2014, 148: 42-57. DOI: 10.1016/j.rse.2014.02.015.
- [29] Stehman S V. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change[J]. *Remote Sensing Letters*, 2012, 3(2): 111-120. DOI: 10.1080/01431161.2010.541950.
- [30] Meng Wen, Tong Xiaotong, Xie Huan, *et al.* Accuracy assessment for regional land cover remote sensing mapping product based on spatial sampling: a case study of Shaanxi Province, China[J]. *Journal of Geo-information Science*, 2015, 17(6): 742-749.[孟雯, 童小华, 谢欢, 等. 基于空间抽样的区域地表覆盖遥感制图产品精度评估——以中国陕西省为例[J]. *地球信息科学学报*, 2015, 17(6): 742-749.]

- [31] Feng J, Bai L, Liu S, *et al.* Optimization of decision-making for spatial sampling in the North China Plain, based on remote-sensing a priori knowledge[J]. ISPRS-International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, 2012. DOI: 10.5194/isprsarchives-XXX-IX-B8-103-2012.
- [32] Wang J, Liu J, Zhuan D, *et al.* Spatial sampling design for monitoring the area of cultivated land[J]. International Journal of Remote Sensing, 2002, 23(2): 263-284. DOI: 10.1080/01431160010025998.
- [33] Bontemps S, Boettcher M, Brockmann C, *et al.* Multi-year global land cover mapping at 300 m and characterization for climate modelling: achievements of the land cover component of the ESA climate change initiative[C]// 2015 36th International Symposium on Remote Sensing of Environment. International Society for Photogrammetry and Remote Sensing, 2015: 323-328. DOI:10.5194/isprsarchives-XL-7-W3-323-2015.
- [34] Chen J, Chen J, Liao A, *et al.* Global land cover mapping at 30 m resolution: a POK-based operational approach[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 103: 7-27. DOI: 10.1016/j.isprsjprs.2014.09.002.
- [35] Wang Zhenghua, Tong Xiahua. Sampling inspect schemes for continuous lot spatial data[J]. Journal of Tongji University (Nature Science Edition), 2010, 38(5): 749-752. [王振华, 童小华. 连续大批量空间数据质量抽样检验方案[J]. 同济大学学报:自然科学版, 2010, 38(5): 749-752.]
- [36] Tong X, Wang Z, Xie H, *et al.* Designing a two-rank acceptance sampling plan for quality inspection of geospatial data products[J]. Computers & Geosciences, 2011, 37(10): 1570-1583. DOI: 10.1016/j.cageo.2011.02.006.
- [37] Ren Huiqun. Research on sample size estimation for accuracy assessment of land cover products[D]. Beijing: Beijing University of Civil Engineering and Architecture, 2020. [任慧群. 地表覆盖产品精度评定样本量估算方法研究[D]. 北京: 北京建筑大学, 2020.]

Comparison of Accuracy Assessment Methods of Remote Sensing based Land Cover Products

Wang Bingquan^{1,2}, Ran Youhua^{1,2}

(1. Northwest Institution of Eco-Environment and Resources, Chinese Academy of Sciences,
Lanzhou 730000, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Accuracy assessment is important for the application of land cover products. In practice, there are many differences for sample size and spatial distribution of the reference data used for accuracy assessment, and the impact of this difference on accuracy assessment results is not very clear. This paper validated the ESA CCI-LC using GlobeLand30 as the reference data in China. We test the impact of sample size, sampling model, and sample unit on the overall accuracy and various types of accuracy. The results are as following. Firstly, the sample size makes a difference to the accuracy assessment results. The overall accuracy based on point sample is close to that of the theoretical sample size while the sample size in half (about 600 sample points), however, the sensitive degree of different classes of precision on the sample size is different, classes with low area weight are more sensitive to sample size, especially stratified random sampling used in the case of small sample size. Secondly, based on the theoretical sample size, the sampling model has little effect on the overall accuracy whether it is the accuracy assessment of point sample or cluster sample, but it has an impact on the accuracy of each class, especially the class with a small area weight. Finally, the stability of accuracy assessment of simple random sampling based on cluster sample is worse than the accuracy assessment based on point sample. The increase of the sample unit will increase the uncertainty of the accuracy assessment result on the basis of the theoretical sample size, with the increase of the sample unit, although the volatility of the accuracy assessment result becomes smaller, the result is still large compared to the point sample. Therefore, in order to reduce the uncertainty of the accuracy assessment results of land cover remote sensing products, the sample size can be increased for point sample, and the number of sample units can be increased or the size of sample units can be appropriately reduced for cluster sample.

Key words: Sampling model; Sample size; Sample unit; ESA CCI-LC; GlobeLand 30