

基于 VGI 的土地覆被遥感产品精度验证

宋宏利¹, 张晓楠²

(1. 河北工程大学 地球科学与工程学院, 河北 邯郸 056038; 2. 河北工程大学 矿业与测绘工程学院, 河北 邯郸 056038)

摘要:以 GEO_WIKI 和 Degree Confluence Project 获取的自发地理信息为参考,通过误差矩阵分析 FROM-GLC-agg 500 m 升尺度数据和 MODIS COLLECTION5 两种相同空间分辨率土地覆被遥感产品的类别精度及类别混淆,采用 Inverse Distance Weighting 模型计算 FROM-GLC-agg 500 m 数据的类别精度空间图谱,结果表明:无论是制图精度还是用户精度,FROM-GLC-agg 500 m 数据与 MODIS 数据相比,并没有表现出显著的优势,两种数据在林地与灌木、草地与裸地、耕地与灌木间存在一定程度的类别错分现象;FROM-GLC-agg 的类别精度图表明在中国的西北部、东北及华北南部等裸地、林地及耕地聚集区,类别精度较高,该区域约占研究区总面积的 36.77%,而在青藏高原区、中国西南部等地表景观异质性显著区域,类别精度相对较低,约占研究区总面积的 20%。

关键词:土地覆被;遥感产品;精度评价;自发地理信息;反距离加权

中图分类号:P237

文献标识码:A

Analysis of land cover category accuracy based on volunteered geographic information

SONG Hongli¹, ZHANG Xiaonan²

(1. School of Earth Science and Engineering, Hebei University of Engineering, Hebei Handan 056038, China;

2. School of Mining and Surveying Engineering, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: The information on land cover at national scales is critical for addressing climate change, biodiversity conservation, ecosystem assessment and environmental modeling. The Volunteered Geographic Information derived from the GEO_WIKI and Degree-Confluence Project was used as referenced data. The category accuracy and the category confusion about FORM-GLC-agg 500m and MODIS COLLECTION5 were compared and the spatial distribution of FORM-GLC-agg's category accuracy was analyzed. The results show that: there is no significant difference on producer and user accuracy between FORM-GLC-agg and MODIS data. They both have a serious confusion between Forest/shrub, cropland/shrub, especially the grassland and bare land. The category accuracy has an uneven distribution, for example, its higher value mainly distributes in northwest and north China, and the lower value more mainly distributes in Tibet Plateau, and southwest China. The results can provide reference for land cover research.

Key words: Land cover products; accuracy assessment; Volunteered Geographic Information; IDW

精确的陆表土地覆被信息,是客观真实地描述全球及区域尺度陆表生态系统结构及其生态过程的关键因子^[1-6],在生物地球化学、环境建模、水土流失、大气环流模型、全球碳循环模型等方面

均得到了广泛应用^[7-10]。但已有研究表明,土地覆被的类别精度具有区域分布特征,全球尺度精度并不能很好的反映国家或区域尺度的类别精度,因此深入了解 FROM-GLC-agg 及其多尺度

产品在国家或区域尺度的类别精度,并比较 FROM - GLC - agg 多尺度产品与目前已有的同级别土地覆被遥感产品的精度是亟待解决的问题。

本文拟以中国为研究区域(不含台湾、南海诸岛统计数据),定量比较 FROM - GLC - agg 500 m 分辨率产品和 MODIS COLLECTION5 产品对应的类别精度;并探求 FROM - GLC - agg 数据类别精度在空间上的分布特征,和构建研究区域类别精度图谱,为中国区域相关领域研究合理选择数据来源提供科学依据。

1 研究数据及预处理

1.1 土地覆被遥感数据

本文拟对两种空间分辨率为 500 m 的土地覆被遥感数据进行精度验证,一种是源于清华大学地球系统研究中心的 FROM - GLC - agg 尺度转换产品^[11];另一种是源于波士顿大学的 MODIS COLLECTION5 数据^[12]。该两种数据均可通过互联网免费获取,鉴于 FROM - GLC - agg 产品的原始 Landsat 遥感影像有 75% 是 2010 年前后获取,为了减小时间差异对于地表土地覆被类别的影响,本文选取 2010 年的 MODIS COLLECTION5 数据作为待验证对象。两种数据的具体参数见表 1。

表 1 研究数据参数

Tab. 1 Parameters of study datasets

特征	数据名称	
	FROM - GLC - agg 500 m	MODIS COLL ECTION5
数据获取时间	2009. 1 ~ 2010. 12	2010. 1 ~ 2010. 12
空间分辨率	500 m	500 m
投影	Intergrid sinusoidal	Intergrid sinusoidal
类别数	10	17

1.2 参考数据

(1) GEO_WIKI 数据

GEO_WIKI 是以 Google Earth 为平台开发的一个免费的全球尺度土地覆被遥感数据验证数据库,该数据库中数据由世界各国相关领域科研工作者提供,数据库中的每条记录描述了陆表特定经纬度交叉点的土地覆被类别组成,并且在向数据库提交记录时需提供该验证站点的信任度级别(包括绝对确认、非常确认、确认和不确认 4 个等

级)。本文所用的 GEO_WIKI 验证数据库在中国区域验证站点数目为 3 406 个。为了保证验证数据的可靠性,根据数据库提供的关于 GEO_WIKI 验证数据的类别准确度级别,选择验证点准确度级别为 0(绝对准确)和 1(非常准确)的验证点作为验证数据,经此条件过滤,中国区域共有 2 187 条可供使用数据。

(2) 整度验证计划数据

所谓整度验证计划(Degree Confluence Project, DCP)发起于 1996 年,该计划的目标是拜访地球上每一个整数纬度和经度交汇处,并获取每个交汇处及其周围东南西北四个方向的地理标签照片,并且通过网络将这些照片及照片描述发送到指定的网站,以用于各种领域的学术研究。为了保证数据的质量,DCP 团队会对上传的地理照片通过与 OpenStreetMap 及 GoogleEarth 等数据进行对比,从而发现明显的错误和恶意信息。

本文选择研究区内经纬整度交汇点作为验证样点,研究区内共包括 952 个经纬度交汇点,其中完整访问点数为 406 个,不完整访问点数为 35 个,而未访问点个数为 511 个。

为了构建基于整度计划的土地覆被参考数据库,本文通过 4 位有丰富经验的遥感方向专业教师,根据照片纹理内容及照片描述对已访问的 441 个点进行目视解译。对于未访问站点,则通过现有土地覆被遥感产品及 Google Earth 软件平台对相应位置的土地覆被类别进行判读。在解译过程中每个教师须对每张照片的判读记录置信度,置信度分为完全确认、确认、不完全确认和不确认 4 个等级,为了提高验证精度,本文只选择置信度为完全确认和确认的土地覆被类别作为参考数据,经过筛选,共有 854 个站点作为参考数据。

1.3 数据预处理

已有研究表明,土地覆被遥感产品分类体系的差异是阻碍数据间互操作的重要原因。鉴于 FROM - GLC - agg 和 MODIS COLLECTION5 数据采用的分类体系并不相同,同时考虑到与 GEO_WIKI 数据的 9 类别分类体系的兼容性,笔者将陆地表面划分为 8 个类别,并以联合国国际联农组织提供的 LCCS 分类标准为依据,对 FROM - GLC - agg 和 MODIS COLLECTION5 数据的类别进行聚合,具体信息参见表 2。

表 2 土地覆被数据分类体系转换表
Tab. 2 Land cover classification

编号	类别名称	FROM - GLC - agg	MODIS Collection5	GEO_WIKI
1	林地	2	1 - 5, 8, 9	1
2	灌木	4	6, 7	2
3	草地	3	10	3
4	耕地	1	12, 14	4
5	水域	6	11, 15	5, 9
6	建设用地	8	13	6
7	裸地	9	16	8
8	永久性冰雪	10	17	7

2 研究方法

2.1 基于误差矩阵的类别精度分析

误差矩阵已成为遥感领域制图精度比较的常用方法。为了得到 FROM - GLC - agg 和 MODIS COLLECTION5 数据与参考数据间的误差矩阵,本文以 GEO_WIKI 和 Degree Confluence Project 两种自发地理信息为参考点数据,在 GIS 软件中,利用点属性提取函数分别获取参考点和待验证数据指定位置土地覆被类别,通过统计比较参考点类别与遥感数据类别差异,按规则填充误差矩阵中各单元数值。根据误差矩阵,计算出用户精度(UA)、制图精度(PA)、总体精度(OA)和 Kappa 系数(KA)等参数。

2.2 类别精度空间分布特征分析

误差矩阵尽管可以揭示土地覆被遥感数据类别精度的总体特征,但无法表达类别精度在空间上的分异规律。为此,本文提出了在误差矩阵指示转换基础上,采用反距离加权模型(Inverse Distance Weighting, IDW)计算 FROM - GLC - agg 500 m分辨率土地覆被产品在中国区域的类别精度空间特征,其计算思路为:首先,将研究区域内的参考数据与 FROM - GLC - agg 数据在对应位置一一进行比较,如果参考数据的类别与验证数据的类别相同,则该点被赋予 1,否则,该点被赋予 0;在进行指示转换完成后,采用 IDW 计算模型,对研究区域进行插值预测处理,其计算公式为:

$$z_p = \sum_{i=1}^n \lambda_i z_i \tag{1}$$

式中: z_p 为 p 点的预测值; λ_i 为 p 点周围第 i 个已知点的权重,其计算公式为 $\lambda_i = \frac{1/d_i^r}{\sum_{i=1}^n 1/d_i^r}$, d_i 为第 i 个已知点距离预测点 p 的距离, r 为权函数系数,

本文 $r=2$ 。IDW 模型在计算过程中需确定搜索半径及搜索区域内已知点个数,鉴于参考数据的空间分布并不均一,本文并未设定固定搜索半径,而是将搜索半径设定为自适应带宽,邻域内已知点数量设为 50。

3 结果分析

3.1 类别精度分析

图 1 揭示了以 GEO_WIKI 和 Degree Confluence Project 为参考数据, FORM 和 MODIS 两种 500 m 分辨率土地覆被遥感数据的类别精度。从图 1 中可以看出, FROM - GLC - agg 数据的林地、耕地、建设用地及永久性冰雪的制图精度均高于 MODIS 数据,但灌木、草地、水域和裸地的制图精度均低于 MODIS 数据。其主要原因是有两点:一是本文所采用的研究数据为 FROM - GLC - agg 的空间聚合产品,该数据是在原始 30 m 分辨率数据基础上,采用比例聚合方法进行升尺度转换后的 500 m 土地覆被数据,由于灌木、水域等类别在我国分布较少,且成零星分布特征,因此在聚合过程中会导致被其优势分布土地覆被类别取代,进而降低了该两种类别的制图精度。二是在参考数据的类别比例中,灌木、草地、水域和裸地的比例相对较低,其中,灌木仅为 41 个点,占参考数据总数的 1.3%;水域为 30 个点,占参考数据总数的 0.96%;草地和裸地均为 120 点,约占参考数据总数的 3.84%,由于验证点数量的稀少,一定程度上无法真实表达出土地覆被产品的类别精度,最终导致 FROM 数据的制图精度低于 MODIS 数据。

对于用户精度, FROM - GLC - agg 数据的水域、裸地和永久性冰雪的精度均高于 MODIS 数据,但林地、灌木的精度均低于 MODIS 数据,而两者草地和耕地的精度几乎相同。这说明,空间分辨率在提取陆表无物候变化特征地物上具有一定

优势,但当地物存在典型的物候变化特征时,其传感器的空间分辨率优势即被时间和光谱分辨率优势取代。FROM - GLC - agg 数据是以 Landsat TM/ETM + 为源数据,该数据具有 30 m 的空间分辨率,但其时间分辨率为 16 天,仅由 7 个可见光、近红外波段构成,而 MODIS 数据的空间分辨率虽仅为 250 ~ 1 000 m,但其时间分辨率为 0.25 天,拥有 36 个光谱波段,因此,MODIS COLLECTION5 土地覆被数据与 FROM - GLC - agg 数据相比,更宜于林地、灌木等物候特征显著的类别。

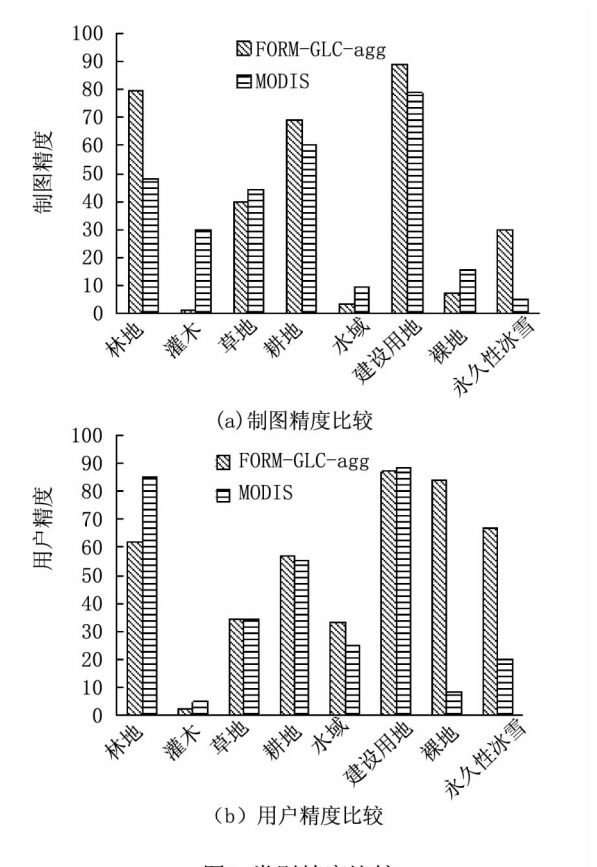
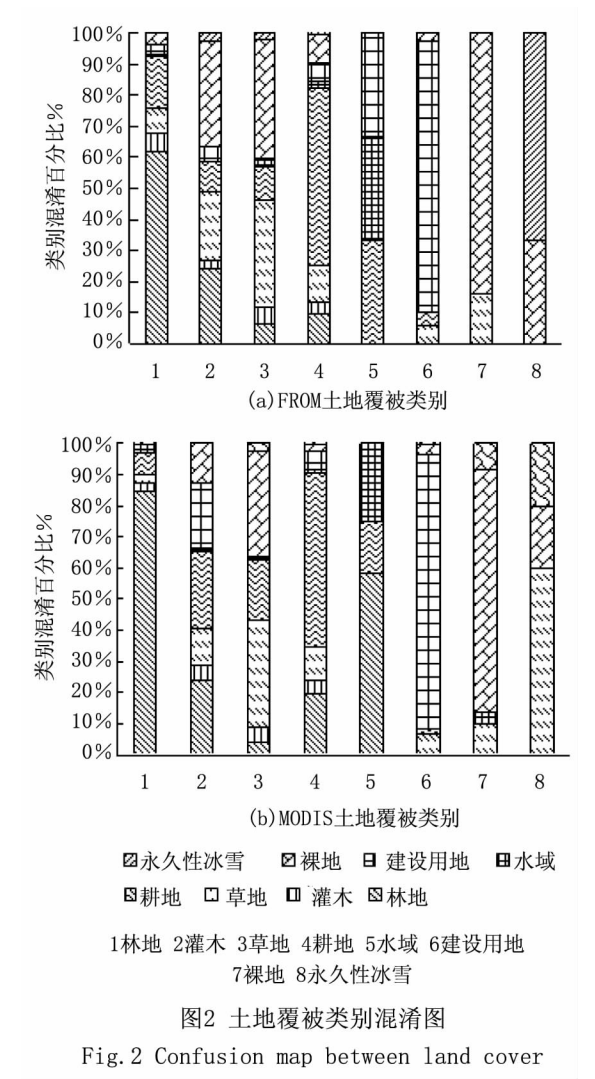


图1 类别精度比较
Fig.1 Comparison on category accuracy between FORM and MODIS

3.2 土地覆被类别混淆分析

图2揭示了 FROM - GLC - agg 数据、MODIS COLLECTION5 数据与参考数据之间的类别混淆。从图中可以看出,混淆主要发生于林地、灌木、耕地以及草地和裸地之间。例如,在参考数据中,有 115 个点的类别为耕地,但在 FROM - GLC - agg 产品中则均为林地,混淆率约占参考数据总数的 17%;草地与裸地之间的混淆也较为严重,在参考数据中有 145 个点为草地,但在 MODIS 中则为裸地,类别混淆率达到了 38%。上述类别混淆的原

因主要是由于各种类别间的生命形态相似性,导致了林地、耕地、草地、灌木等类别的光谱特征十分相似,这增加了从遥感图像本身识别的难度。



3.3 类别精度空间分布特征

图3和图4揭示了 FORM - GLC - agg 500 m 数据的类别精度在研究区域的分布特征及各精度区间的面积百分比。类别精度取值区间为 0 ~ 1, 其中 0 代表该位置的类别精度最低,即该位置 FORM - GLC - agg 的类别与参考数据不一致;相反,1 表示类别精度最高,即该位置的类别与参考数据一致。从图中可以看出,类别精度具有显著的空间集聚特征,其中 0.8 ~ 1 的高值区域主要分布于中国的西北、东北和华北中南部,该区域地表土地覆被类别较为单一,以裸地、耕地和林地为主,该区域面积约占研究区总面积的 36.77%,说明 FORM - GLC - agg 数据的类别精度总体较高;0.4 ~ 0.8 的中值区域主要分布于精度高值区域周

围,符合土地覆被类别的区域化变量特征,该区域土地覆被类别主要由草地、林地组成,该区域约占研究区总面积的 50%; $0 \sim 0.4$ 的低值区域主要分布于中国的青藏高原和中国的西南部,该区域地表景观呈现破碎分布特征,异质性较强,增加了遥感数据识别土地覆被类别的难度,但该区域所占比重较小,仅占研究区总面积的 20%。

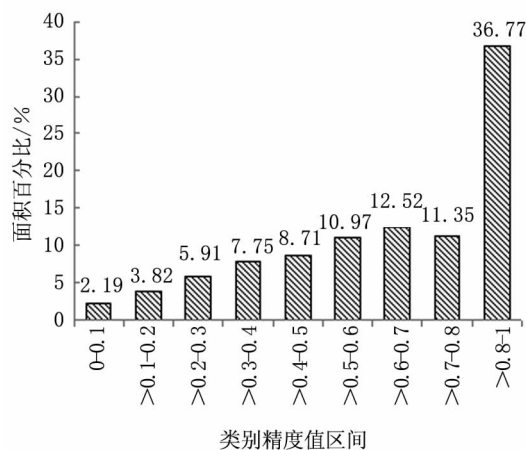


图3 类别精度所占面积百分比

Fig.3 Area percent of category accuracy

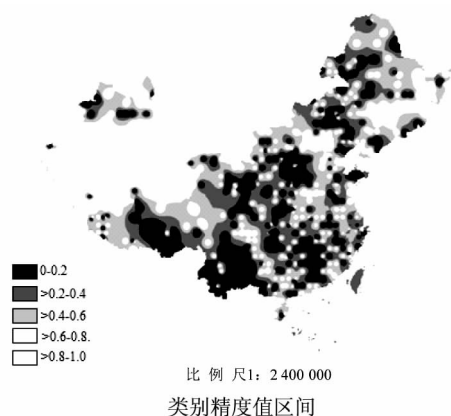


图4 类别精度所占面积百分比

Fig.4 Area percent of category accuracy

4 结论

1) 研究表明,无论是 FORM - GLC - agg 500 m 数据还是 MODIS 500 m 数据,在林地与灌木、灌木与耕地以及草地与裸地之间均存在显著的混淆现象,这说明对于通过遥感技术精确获取陆表土地覆被类别仍存在较大的困难,应该深入研究异物同谱问题,进一步加强基于光谱特征空间区分相似土地覆被类别算法的研究。

2) FORM - GLC - agg 500 m 数据的类别精度指示值预测结果表明,精度高值区域主要分布于

地表景观均质性较强区域,主要分布于林地、耕地和裸地较为集中的中国东北、华北南部和西北地区;而精度值较低区域主要分布于青藏高寒区、中国西南部等地,其地表景观破碎较为严重,因此如何精确提取景观异质性区域土地覆被类别是今后亟待解决的问题。

参考文献:

- [1] 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(2):12-14.
- [2] 李晓兵. 国际土地利用-土地覆盖变化的环境影响研究[J]. 地球科学进展,1999,14(4):395-400.
- [3] 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.
- [4] SUTHERLAND W J, ADAMS W M, ARONSON R B, et al. One hundred questions of importance to the conservation of global biological diversity[J]. Conservation Biology,2009,23(3):557-567.
- [5] LAMBIN E F, GEIST H J, LEPELERS E. Dynamics of land-use and land-cover change in tropical regions[J]. Annual Review of Environment and Resources,2003,28(1):205-241.
- [6] FEDDEMA J J, OLESON K W, BONAN G B, et al. The importance of land-cover change in simulating future climates[J]. Science,2005,310(5754):1674-1678.
- [7] BONAN G B, OLESON K W, VERTENSTEIN M, et al. The land surface climatology of the community land model coupled to the NCAR community climate model[J]. Journal of Climate,2002,15(22):3123-3149.
- [8] RUNNING S W, NEMANI R, ANN H, et al. A continuous satellite-derived measure of global terrestrial primary production[J]. Bioscience,2004,54(6):547-560.
- [9] ZHANG K, KIMBALL J S, MU Q Z, et al. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005[J]. Journal of Hydrology,2009,379(1):92-110.
- [10] FOLEY J A, DEFRIES R, ASNER G P, et al. Global consequences of land use[J]. Science,2005,309(5734):570-574.
- [11] 俞乐,王杰,李雪草,等. 基于多源数据集成的多分辨率全球地表覆盖制图[J]. 中国科学:地球科学,2014(8):5.
- [12] FRIEDL M A, MCIVER D K, HODGES J C F, et al. Global land cover mapping from MODIS: algorithms and early results[J]. Remote Sensing of Environment,2002,83(1):287-302.

(责任编辑 王利君)