

宋宏利, 张晓楠. 我国区域 OpenStreetMap 土地覆被信息时空分布特征[J]. 江苏农业科学, 2017, 45(18): 274-277.
doi:10.15889/j.issn.1002-1302.2017.18.070

我国区域 OpenStreetMap 土地覆被信息时空分布特征

宋宏利^{1,2}, 张晓楠³

(1. 河北工程大学地球科学与工程学院, 河北邯郸 056038; 2. 河北省煤炭资源综合开发与利用协同创新中心, 河北邯郸 056038;
3. 河北工程大学矿业与测绘工程学院, 河北邯郸 056038)

摘要:以我国区域为研究对象, 分析了 OpenStreetMap 土地覆被信息的时空特征。结果表明, 该数据在我国东北、华北、西北、华中、东南和西南 6 个地理分区均呈聚集状态分布, 且数量表现出明显的差异性, 东南区域是土地覆被信息最为集中的区域, 该区域仅占研究区域总面积的 5%, 却集中了土地覆被信息总量的 33%; 土地覆被信息以人工地表和水体为主, 二者之和约占土地覆被信息总量的 80%, 其次为林地、草地、耕地, 裸地数量最小; 随着空间分辨率从 30 m 升尺度至 500 m, 各种类别的样本数据均随之减少, 但林地、草地及耕地的减少速度明显低于人工地表和水体; 年际及月季尺度分析表明, 除裸地外, 其余 5 种土地覆被类别均呈现较为显著的增加趋势, 这说明 OpenStreetMap 数据具有稳定的数据来源, 具有较好的现势性, 可以形成一定的时间序列。该研究为未来大尺度土地覆被遥感产品生产及验证过程中训练、参考数据的选择, 特别是人工地表和水体, 提供了一定参考, 也为土地覆被遥感产品更新完善提供了有效的数据来源。

关键词: OpenStreetMap; 土地覆被; 时空分布; 特征分析

中图分类号: S127 **文献标志码:** A **文章编号:** 1002-1302(2017)18-0274-04

2007 年《Citizens as sensors: the world of volunteered geography》一文发表, 自发地理信息的概念被首次提出^[1]。随着 WEB 2.0 和互联网 + 等技术的出现, 该数据已成为空间科学领域重要的数据来源, 新地理信息时代已经到来^[2]。土地覆被是地球表面各物质类型及其自然属性与特征的综合体^[3]。精准的土地覆被信息已成为自然资源调查、环境监测、生物多样性保护、地理国情普查、气候变化及生态建模等领域的基础数据^[4-8]。近年来, 在国内外学者的共同努力下, 自发地理信息在土地覆被相关研究领域取得了丰硕成果。Foody 等以非洲西部为研究区域, 采用 Degree Confluence Project 和 GEOWIKI 数据验证了 GLOBCOVER 产品中林地的类别精度, 研究表明自发地理信息具有和官方数据相近的精度, 可以用于土地覆被遥感产品的精度验证^[9]; Estima 等以葡萄牙为研究区域, 分别分析了 Flickr 地理照片、OpenStreetMap 数据在土地覆被产品验证中的应用, 并认为在葡萄牙区域内, Flickr 和 OpenStreetMap 数据具有数量大、时间连续等优点, 可以有效解决地面实地调查数据稀少、调查成本巨大的不足^[10-12]; 为了探索自发地理信息在土地覆被分类中的可行性, Jokar 等在未采用任何遥感和地面验证数据的情况下, 尝试单独采用 OpenStreetMap 原始数据, 并以决策树分类算法为模型进行土地覆被分类制图, 将分类结果与 GMESUA 数据进

行比较^[13]; 与 Jokar 相似, Daniel 等通过下载伦敦区域范围的 Flickr 和 Geography 数据, 在支持向量机支持下, 以相片的亮度、色度、饱和度以及边缘和 GIST 特征为分类指标, 实现对照片的土地覆被分类, 并将分类结果与 GLC2000 数据进行了比较^[14]。然而经综合分析, 上述研究多集中于非洲和欧洲区域, 而关于我国区域自发地理信息在土地覆被领域的研究则鲜有报道。本研究以我国区域为研究范围, 以 OpenStreetMap 数据为研究对象, 拟从数量、时间序列、空间分布多个方面揭示我国区域 OpenStreetMap 数据的土地覆被类别特征。该研究结果将有助于揭示 OpenStreetMap 在该区域应用于土地覆被遥感产品更新完善的可行性, 也为未来大尺度土地覆被遥感产品分类过程中训练样本及精度验证中参考样本的获取提供科学参考。

1 数据来源及预处理

1.1 数据来源

OpenStreetMap 是 2004 年成立于英国的自发地理信息在线协作网站 (<http://www.openstreetmap.org>), 该网站允许全球的自愿者通过注册, 将个人采集的具有空间参考的地理信息共享到该地图平台, 为全球用户提供免费的数据下载服务^[15]。本研究所用数据下载于 Geofabrik 网站 (<http://download.geofabrik.de/>), 该网站将 OpenStreetMap 信息按照大洲及国家进行分类, 为数据使用者提供了 pbf、bz2 和 shp 3 种数据格式。由于数据的处理和分析将在 GIS 环境中进行, 因此本研究采用的是 shp 格式的 OpenStreetMap 数据。经分析, 每个特定时间基点的 OpenStreetMap shp 格式数据均包含 buildings、landuse、natural、places、points、railways、roads 和 waterways 8 个图层, 其中 buildings、landuse 和 natural 3 个图层为 Polygon 类型, places 和 points 为 point 类型, railways 和

收稿日期: 2017-03-17

基金项目: 河北省自然科学基金 (编号: D2013402014); 河北省高等学校科学技术研究重点项目 (编号: ZD2017212)。

作者简介: 宋宏利 (1980—), 男, 河北抚宁人, 博士, 副教授, 硕士生导师, 主要研究方向为遥感产品精度验证。E-mail: songholi2003@163.com。

通信作者: 张晓楠, 硕士, 副教授, 主要研究方向为国土资源遥感。E-mail: 360217051@qq.com。

waterways 为 line 类型。鉴于本研究的目的是分析我国区域 OpenStreetMap 所包含土地覆被信息的时空分布特征,进而为大尺度土地覆被遥感产品精度验证及更新完善提供参考,而现有土地覆被遥感产品的空间分辨率主要分为 30、300、500 m 3 种类型,每个像元均代表一定的空间范围,点、线类型则与像元代表范围存在明显的尺度差异。尽管 buildings 为面图层,但经笔者分析发现,该图层所代表的地理实体空间范围较小,小于 30 m 分辨率的传感器难以识别。基于上述分析,本研究仅选择 landuse、natural 2 个面图层作为土地覆被信息提取来源。

1.2 数据预处理

Landuse 和 Natural 数据在自愿者上传数据时提供了关于类别的描述,并将其类别名称存储在属性表的 Type 字段中,如 farmland、residential、grass、reservoir 等,但该分类体系侧重于图斑的土地利用属性,与现有的土地覆被遥感产品分类体系侧重于土地覆被特征存在一定差异,如 conservation,该类别仅说明了陆表的土地利用方式为自然保护区,但并没有明确保护区内土地覆被的组成情况,类似此种类别并未纳入本研究分析范畴,而是将其从数据集中删除,以减小专题误差。在综合分析 Landuse、Natural 的类别描述及结合前人已有研究成果^[16]基础上,将 Landuse、Natural 数据的分类体系进行了整合,共分为林地、草地、耕地、水体、人工地表及裸地 6 种类别(表 1)。

表 1 土地覆被类别分类体系

编号	类别	开源街图 - 土地利用	开源街图 - 自然物
1	林地	果园, 树林	林地
2	草地	牧草地, 草地	
3	耕地	农业地, 农场, 农田	
4	水体	盐池, 水产养殖池, 流域, 水库	水体
5	人工地表	建筑, 建筑工地, 市政管理, 公共服务, 商业用地, 混凝土, 工业用地, 居民点用地	
6	裸地	裸地, 沙地, 荒地	

2 研究方法

首先参照已有研究,将我国区域分为东北、华北、东南、华中、西南和西北 6 个地理分区^[17],以 2016 年 12 月的 OpenStreetMap 数据为研究对象,分析该时间基点土地覆被信息在不同地理分区样本总数及类别总数的数量特征;其次,为分析在不同空间分辨率尺度下,研究区域土地覆被信息数量的变化情况,在 GIS 软件支持下生成覆盖研究区域的边长为 30、300、500 m 的空间格网,并将空间格网与 OpenStreetMap 数据进行空间叠置分析,然后分析不同分辨率情形下每个格网所包含的 OpenStreetMap 土地覆被类别的组成情况,如果 1 个格网内存在面积 > 65% 的某种土地覆被类别,则该网格被保留,并将相应网格的属性值定义为该土地覆被类别;如果 1 个网格内不存在面积 > 65% 的类别,即该格网无绝对优势类别,从遥感意义上定义为混合像元,则该网格被删除,不参与计算分析。土地覆被遥感产品的更新完善只要求对变化区域进行更新,因此,数量庞大且现势性强的土地覆被信息成为该

过程的重要基础,为此,本研究分别分析了在年际尺度(2013 年、2014 年、2015 年、2016 年)和月际尺度(2016 年 1—12 月)下,OpenStreetMap 土地覆被信息的变化特征。

3 结果分析

3.1 OpenStreetMap 土地覆被类别分布特征分析

图 1 和表 2 揭示了我国区域 OpenStreetMap 土地覆被信息的空间分布特征。最邻近指数(nearest neighbor index, NNI)表达了地理实体空间分布的聚集或离散状态,如果 $NNI > 1$,则地理实体呈离散分布,如果 $NNI < 1$,则地理实体呈聚集分布。从表 2 中可以看出,OpenStreetMap 土地覆被信息在我国区域的西南、华中、华北、东北、西北及东南地区的最邻近指数均小于 1,说明在上述 6 个区域土地覆被信息均为聚集分布。西北地区的 NNI 值最小,仅为 0.28,说明该区域的土地覆被信息聚集特征最为明显,从图 1 中可以看出,土地覆被信息主要集中于陕西省中部、甘肃省中南部、新疆北部及西藏南部区域,而在昆仑山以南至雅鲁藏布江以北的广大区域则鲜有 OpenStreetMap 土地覆被数据;东南区域的 NNI 为 0.31,仅次于西北区域,从图 1 可以看出土地覆被信息主要集中于海南岛沿海区域及长江三角洲和珠江三角洲区域,其中杭州、苏州、无锡、扬州、广州、厦门、上海、东莞、英德、三亚等城市的样本总量约占该区域样本量总和的 85%;在华中区域,样本主要聚集于江西中北部、安徽中东部、湖南东部及湖北东部区域,而在区域西部、东南部则样本稀少。尽管东北、华北和西南 3 个区域的 NNI 相对较小,区域内土地覆被信息分布相对均匀,但仍存在局部区域的异质性,图 1 表明在黑龙江省北部至内蒙古北部,从东向西的广大区域土地覆被样本较为稀少,该区域样本主要集中于吉林省、辽宁省及内蒙古中南部区域;在华北区域,样本在河北省北部的张家口、承德及河南省南部区域分布较为稀少。

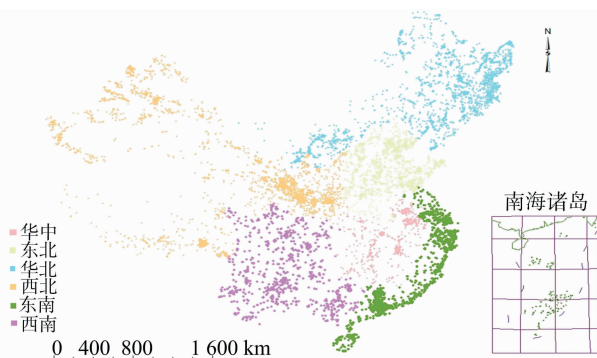


图 1 我国区域 OpenStreetMap 土地覆被信息空间分布

表 3 是根据表 1 的分类体系从 OpenStreetMap 的 Landuse 和 Natural 2 个图层中提取的各地理分区样本总数及类别样本比例。从表 3 中可以看出,样本在我国不同地理区域的数量存在一定差异,其中东南区域和东北区域的样本比例差异最大,东南区域仅占研究区总面积的 5.36%,为 6 个地理分区中空间面积最小的分区,但其区域内的样本数量则为 63 708 个,占样本总数的 33.10%;东北区域面积比例为 19.30%,仅次于西北区域的 47.59%,但其区域内的样本总数仅为 9.27%,与东南区域相差 23.83 百分点;华北、华中 2

表 2 土地覆被信息空间分布最邻近指数

地区	西南地区	华中地区	华北地区	东北地区	西北地区	东南地区
最邻近指数	0.48	0.37	0.4	0.35	0.28	0.31

表 3 土地覆被信息在不同地理区域的比例分布

区域	面积比例 (%)	样本 (%)	土地类别					
			林地 (%)	草地 (%)	耕地 (%)	水体	人工地表 (%)	裸地 (%)
东北区域	19.30	9.27	15.68	9.68	12.67	7.32	10.16	0.90
华北区域	6.81	12.61	14.76	21.58	6.41	7.14	17.09	37.84
东南区域	5.36	33.10	32.71	35.81	32.06	43.01	22.54	41.44
华中区域	6.97	11.88	9.98	10.95	3.83	16.38	8.36	5.41
西南区域	13.97	12.21	13.33	4.63	8.50	11.46	14.17	10.81
西北区域	47.59	20.92	15.54	17.35	36.52	14.69	27.68	3.60

个区域的面积比例基本相同,其区域内样本比例分别为 12.61% 和 11.88%,相差较小;西北区域面积比例最大,占研究区总面积的 47.59%,其区域内样本比例为 20.92%,仅次于东南区域。

相对于总体样本比例,类别样本比例在东南区域和西北区域则表现出了不同的优势类别。在东南区域,林地和草地、水体和裸地相对于其他区域具有明显的优势,其类别比例分别为 32.71%、35.81%、43.01% 和 41.44%;而耕地和建设用地则在西北区域具有较为明显的比例优势,其中耕地的样本比例为 36.52%,建设用地则为 27.68%;东北、华北、华中区域尽管没有明显优势类别,但其区域内的林地、草地和水体的样本比例较高,分别为 15.68%、21.58% 和 16.38%,三者均仅低于东南地区。

3.2 不同分辨率条件下 OpenStreetMap 土地覆被信息特征分析

图 2 揭示了在 30、300、500 m 分辨率尺度下 6 种土地覆被类别的数量变化情况,从图 2 中可以看出水体和人工地表的数量在我国区域具有明显优势,裸地的数量最少,林地、耕地、草地数量介于水体、人工地表和裸地之间。例如在 30 m 分辨率尺度下,水体和人工地表二者之和约占土地覆被信息总数的 85%,其中水体的样本数为 83 123 个,占样本总数的 42%,人工地表的样本数为 83 652 个,占样本总数的 43%;林地、耕地、草地的样本数分别为 16 560、6 706、7 650 个;裸地仅为 161 个。随着空间分辨率的降低,样本数量均呈现下降趋势,但类别的减少比率表现出不同的特征。例如当分辨率从 30 m 降为 300 m 时,水体和人工地表的数量减少趋势最为明显,二者分别减少为 25 515 个和 27 985 个,分别减少了 69% 和 67%,明显大于林地、草地;当分辨率由 300 m 降为 500 m 时,6 种土地覆被类别样本的数量继续减少,但其减少的速率明显变缓,除人工地表外,其他 5 种类别的样本变化趋于平稳状态。上述结果表明自愿者上传的关于人工地表和水体类别的自发地理信息多集中于较小地理实体,覆盖的空间范围较小,样本数量受空间分辨率变化的影响较为明显,而林地、草地、耕地等类别的实体面积相对较大,受空间分辨率变化的影响较小。由于裸地样本数量较少,其变化也明显低于其他几种类别。

3.3 OpenStreetMap 土地覆被信息时序特征分析

图 3 揭示了我国区域 OpenStreetMap 土地覆被信息在年

际尺度的变化趋势。结果表明,从 2013—2016 年 4 年间,除裸地外,5 种土地覆被信息样本数量均有增加,人工地表和水体的样本数量增加较为明显,其中人工地表 4 年间的增长率分别为 67%、52% 和 72%,水体的增长率分别为 58%、31% 和 41%,2013—2015 年间人工地表样本数量一直低于水体数量,但 2016 年人工地表数量接近水体,二者样本数达到 85 000 个。林地、草地和耕地数量 4 年间增长较为平缓,林地 2013 年样本数量为 3 200 个,2016 年为 16 560 个,4 年间增加了 13 360 个,增长了 4 倍;草地 2013 年为 1 848 个,2016 年为 7 650 个,增加了 5 802 个,增长了 3 倍;耕地 2013 年为 1 047 个,2016 年为 6 706 个,增加了 5 659 个,增长了 5.4 倍。裸地在 2013—2015 年间均没有任何信息,自 2016 年开始,自愿者开始上传关于裸地数据,其数量由 2016 年 1 月的 3 个,增加至 2016 年 12 月的 161 个。

图 4 揭示了我国区域 OpenStreetMap 土地覆被信息在月季尺度的变化趋势。结果表明,耕地的增长速度最快,由 2016 年 1 月的 3 129 个增加为 2016 年 12 月的 6 706 个,增加了 1 倍;林地的增长数据仅次于耕地,其数量由最初的 9 605 个,增长为 12 月的 16 560 个,增长了 72%;草地由最初的 4 831 个,增长为 7 650 个,增加了 58%;相对于上述 4 种类别,水体的增长速度最小,由最初的 59 091 个,增长为 12 月的 83 123 个,增加了 41%;裸地信息相对较少,在 2016 年 1—6 月均为 3 个,6 月间并无增长,该类别样本数量在 7 月为 137 个,于 12 月增加到 161 个。

4 结论与讨论

本研究以我国区域为研究对象,分析了 OpenStreetMap 土地覆被信息在研究区域的空间分布、数量、空间分辨率及时间尺度特征,结果表明:(1)从空间分布角度,OpenStreetMap 在我国的东北、华北、华中、东南、西南及西北 6 个区域均呈现聚集分布特征,其中西北区域的聚集现象最为明显,其最邻近距离指数仅为 0.28。该聚集特征与 Estima 在葡萄牙大陆的研究结果较为相似,OpenStreetMap 数据多集中于人口聚集或人类行为频繁区域,如经济发达的区域中心城市或区域旅游热点区域;OpenStreetMap 在我国 6 个地理分区的数量分布并不均匀,呈现明显的差异性,其中东南区域是土地覆被信息数据最为集中区域,该区域仅占研究区域总面积的 5%,却集中了 OpenStreetMap 土地覆被信息总量的 33%,东北区域面积比例

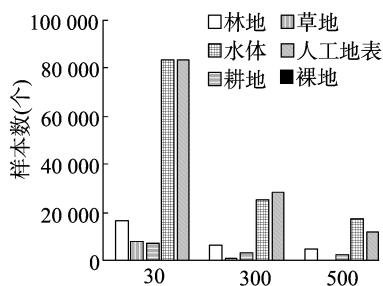


图2 不同分辨率条件下土地覆被信息数量

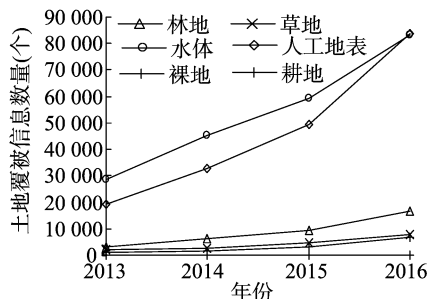


图3 OpenStreetMap 土地覆被信息年际变化

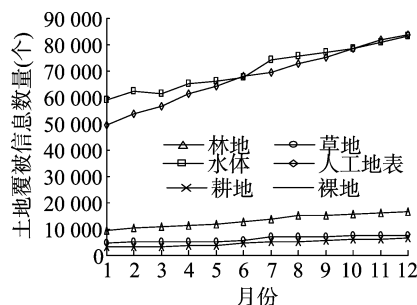


图4 OpenStreetMap 土地覆被信息月际变化

为 19.30%,但其区域内的土地覆被信息数量仅为 9.27%。(2)我国区域 OpenStreetMap 土地覆被信息以人工地表和水体为主,其次为林地、草地、耕地,裸地数量最小。截至 2016 年 12 月,研究区域人工地表的数量为 83 652 个,水体为 83 123 个,二者之和约占土地覆被信息总量的 80%,林地、草地和耕地的数量分别为 16 560、7 650、6 706 个,裸地仅为 161 个。随着空间分辨率从 30 m 升尺度为 500 m,各种类别的样本数据均随之减少,但林地、草地及耕地的减少速度明显低于人工地表和水体。年际及月季尺度表明,除裸地外,其余 5 种土地覆被类别均呈现增加趋势,这说明,OpenStreetMap 数据具有较好的现势性,可以形成一定的时间序列;(3)由于自发地理信息在数量及现势性方面的优势,该数据已被认为是未来土地覆被遥感产品精度验证及更新完善过程中的重要数据来源^[18]。但由于众包数据是由自愿者自发上传,目前尚无统一的框架约束,其数据质量一直备受争议。针对此问题,笔者下一步将以国家基础地理信息中心研发的 30 m 分辨率土地覆被遥感产品 GlobeLand30 为参考数据,对 OpenStreetMap 数据进行精度验证,揭示该数据各种土地覆被类别的精度特征,以深入分析 OpenStreetMap 在土地覆被遥感产品验证中的可行性。

致谢:感谢河北省煤炭资源综合开发与利用协同创新中心及河北省一流学科“地质资源与地质工程”为本研究的完成提供的资金支持。

参考文献:

- [1] Goodchild M F. Citizens as sensors: the world of volunteered geography[J]. *GeoJournal*,2007,69(4):211-221.
- [2] 李德仁,邵振峰. 论新地理信息时代[J]. *中国科学 F 辑:信息科学*,2009,39(6):579-587
- [3] 廖安平,陈利军,陈 军,等. 全球陆表水体高分辨率遥感制图[J]. *中国科学:地球科学*,2014,44(8):1634-1645.
- [4] 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.
- [5] Lafontaine J H, Hay L E, Viger R J, et al. Effects of climate and land cover on hydrology in the Southeastern US: Potential impacts on watershed planning[J]. *Jawra Journal of the American Water Resources Association*,2015,51(5):1235-1261
- [6] Gagné S A, Eigenbrod F, Bert D G, et al. A simple landscape design framework for biodiversity conservation[J]. *Landscape and Urban Planning*,2015,136:13-27.
- [7] Mantyka-Pringle C S, Visconti P, Marco M D, et al. Climate change modifies risk of global biodiversity loss due to land-cover change[J]. *Biological Conservation*,2015,187:103-111.
- [8] Gessesse B, Bewket W, Bräuning A. Model-based characterization and monitoring of runoff and soil erosion in response to land use/land cover changes in the Modjo watershed, Ethiopia[J]. *Land Degradation & Development*,2015,26(7):711-724.
- [9] Foody G M, Boyd D S. Using volunteered data in land cover map validation: Mapping West African forests[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*,2013,6(3):1305-1312.
- [10] Estima J, Fonte C C, Painho M. Comparative study of Land Use/Cover classification using Flickr photos, satellite imagery and Corine Land Cover database[C]. 17th AGILE international conference on geographic information science,2014:3-6.
- [11] Estima J, Painho M. Exploratory analysis of OpenStreetMap for land use classification[C]. Proceedings of the second ACM SIGSPATIAL international workshop on crowd sourced and volunteered geographic information. ACM,2013:39-46.
- [12] Estima J, Painho M. Investigating the potential of OpenStreetMap for land use/land cover production: a case study for Continental Portugal[M]. Springer International Publishing,2015:273-293.
- [13] Arsanjani J J, Helbich M, Bakillah M, et al. Toward mapping land-use patterns from volunteered geographic information[J]. *International Journal of Geographical Information Science*,2013,27(12):2264-2278.
- [14] Leung D, Newsam S. Land cover classification using geo-referenced photos[J]. *Multimedia Tools and Applications*,2015,74(24):11741-11761.
- [15] 马 超,孙 群,徐 青,等. 自发地理信息可信度及其评价[J]. *地球信息科学学报*,2016,18(10):1305-1311.
- [16] Johnson B A, Iizuka K. Integrating OpenStreetMap crowd sourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: case study of the Laguna de Bay area of the Philippines[J]. *Applied Geography*,2016,67:140-149.
- [17] 陆 苗,吴文斌,张 莉,等. 不同耕地数据集在我国区域的对比研究[J]. *中国科学:地球科学*,2016,46(11):1459-1471.
- [18] 陈 军,张 俊,张宏伟,等. 地表覆盖遥感产品更新完善的研究动向[J]. *遥感学报*,2016,20(5):991-1001.