

联合 GEE 与多源遥感数据的黑龙江流域沼泽湿地信息提取

宁晓刚^{1,2}, 常文涛^{1,2}, 王浩^{1,2}, 张翰超¹, 朱乾德³

1. 中国测绘科学研究院 摄影测量与遥感研究所, 北京 100036;

2. 山东科技大学 测绘科学与工程学院, 青岛 266590;

3. 南京水利科学研究院 水文水资源与水利科学国家重点实验室, 南京 210029

摘要: 湿地是地球上最重要的生态系统之一, 在维持全球生态环境安全等方面发挥着举足轻重的作用。由于湿地独特的水文特征, 传统的湿地监测需要耗费大量的人力和财力, 对于大尺度的湿地信息提取更是困难重重。随着大数据和云计算的兴起, 为大尺度和长时间序列的空间数据处理提供了契机。本文基于 Google Earth Engine (GEE) 云平台, 使用 Sentinel-1 合成孔径雷达 (SAR) 数据、Sentinel-2 光学数据以及地形数据, 探讨了红边、雷达以及地形特征对大范围区域沼泽湿地提取的重要性, 验证了利用 JM 距离寻找沼泽湿地提取最优特征组合的可行性, 结合随机森林算法对 2018 年黑龙江流域沼泽湿地进行提取。研究表明: (1) Sentinel-2 红边波段和 Sentinel-1 雷达波段以及地形数据有利于沼泽湿地信息提取, 相比植被指数和水体指数沼泽的制图精度分别提高了 7.56%, 5.04%, 4.48%; (2) 利用 JM 距离得到的分离度表明, 红边特征>其他光学特征>地形特征>雷达特征。进行特征优选后沼泽湿地的制图和用户精度分别提高了 1.45% 和 3.02%, 最终结合随机森林算法的总体精度为 91.54%, 沼泽的提取精度为 88.55%。本研究利用 GEE 云平台和多源遥感数据以及机器学习算法, 能够准确、快速、高效地提取大尺度范围沼泽湿地信息, 具有较大的应用潜力。

关键词: 遥感, Google Earth Engine, 沼泽湿地, Sentinel-1, Sentinel-2, JM 距离, 随机森林, 红边波段

引用格式: 宁晓刚, 常文涛, 王浩, 张翰超, 朱乾德. 2022. 联合 GEE 与多源遥感数据的黑龙江流域沼泽湿地信息提取. 遥感学报, 26(2): 386-396

Ning X G, Chang W T, Wang H, Zhang H C and Zhu Q D. 2022. Extraction of marsh wetland in Heilongjiang Basin based on GEE and multi-source remote sensing data. National Remote Sensing Bulletin, 26(2): 386-396 [DOI: 10.11834/jrs.20200033]

1 引言

湿地仅覆盖地球陆地表面面积的 3%—8%, 但它却是减少全球温室气体和减轻气候变化的最重要的因素之一, 并且极大地影响了生物多样性和水文联系 (Mitsch 等, 2013)。湿地的生态系统服务功能包括调蓄洪水, 改善水质, 维持海岸线稳定, 保持水生植物生产力, 维持野生动植物种群续存等 (Mitsch 和 Gosselink, 2000)。但是由于人类活动和自然过程的影响, 湿地面积正加速萎缩 (Ozesmi 和 Bauer, 2002)。因此, 对湿地进行快速

准确的监测是当前湿地研究的热点 (Shalaby 和 Tateishi, 2007)。

相对于传统的野外调查, 遥感技术具有观测范围广, 更新周期短, 信息量大等优点, 在湿地信息提取及变化监测方面发挥了重要作用。但是在在大尺度范围利用遥感数据进行准确的湿地信息提取仍具有挑战性。光学影像作为最主要的遥感数据源, 受到云雨等天气影响很大, 经常形成无效观测。云污染对重访周期较长的中高分辨率卫星数据影响更为显著, 对于大尺度区域和阴雨天气较多的南方地区更易造成数据缺失。基于微波

收稿日期: 2020-02-10; 预印本: 2020-07-04

基金项目: 国家重点研发计划 (编号: 2017YFC0404503); 国家自然科学基金 (编号: 41601529); 中国科学院流域地理学重点实验室开放基金 (编号: WSGS2017002); 南京水利科学研究院基本科研业务费专项 (编号: Y518002)

第一作者简介: 宁晓刚, 1979 年生, 男, 研究员, 研究方向为地理国情和自然资源遥感监测。E-mail: ningxg@casm.ac.cn

通信作者简介: 王浩, 1985 年生, 男, 副研究员, 研究方向为生态遥感监测与土地利用变化分析。E-mail: wanghao@casm.ac.cn

的合成孔径雷达(SAR)弥补了光学影像的不足,其对云层具有很强的穿透能力,可以不受天气影响对地表进行全天时全天候观测。雷达后向散射系数对地物的介电特性敏感,对植被含水量和土壤含水量以及地表的几何特性如粗糙度具有不同的响应,可以提供不同于光学影像的独特信息(王安琪等,2012)。此外,利用地形地貌特征对提取某些湿地类型,如灌丛湿地和森林湿地具有重要作用(严婷婷等,2014)。研究表明,光学、雷达以及地形数据的结合使用可以发挥多源遥感数据优势互补的作用,光谱信息和散射机制的结合对湿地信息提取有着巨大的潜力。刘蕾等(2013)结合Landsat TM影像、C波段的ENVISAT ASAR以及地形辅助数据对扎龙湿地进行分类,发现雷达后向散射系数和地形辅助数据对湿地分类精度的提高起到重要作用。Dubeau等(2017)使用了Landsat、L波段ALOS PALSAR以及地形数据对埃塞俄比亚的达布斯湿地进行分类,光学、雷达和地形数据的结合取得了最佳的总体分类精度。Limpitlaw和Gens(2006)研究发现,单独使用光学或SAR数据识别开阔水体、淹水植被和非淹水植被时,精度并不理想,而将两者结合时可以显著提高这3种地物的识别精度。然而在这些研究中鲜有将光学、SAR、地形特征进行横向对比,以评价它们对沼泽湿地信息提取的重要性差异。

近年来,随着欧洲空间局“哥白尼”计划(Copernicus Programme)的启动,先后发射了12 d重访周期(双星协同可达6 d)的Sentinel-1合成孔径雷达(SAR)卫星和10 d重访周期(双星协同可达5 d)的Sentinel-2多光谱卫星。这为全球湿地制图提供了前所未有的机会,高重访周期有效地减轻了云污染的影响。并且,Sentinel-2是目前唯一一个在红边范围内含有3个波段的多光谱传感器,这对监测植被健康信息非常有效(Shoko和Mutanga,2017)。Chatziantoniou等(2017)使用Sentinel-1和Sentinel-2数据与支持向量机(SVM)对希腊的科洛尼亚国家公园和沃尔维湖进行土地覆盖制图,重点关注湿地类,取得了93.85%的总体精度。Whyte等(2018)协同使用Sentinel-1和Sentinel-2数据和面向对象(OBIA)方法采用支持向量机和随机森林算法对iSimangaliso湿地公园进行土地覆盖制图,Sentinel-1和Sentinel-2的联合使用显著提高了湿地制图的精度。其他研究也表

明了Sentinel-1和Sentinel-2在湿地研究方面的巨大潜力(Yesou等,201)。此外,由于数据量和机器性能的限制,融合多源数据实现省级乃至更大范围的高频次观测仍然面临很大困难。使用基于工作站的常规遥感影像处理软件在收集、存储和处理覆盖大尺度区域长时间序列的遥感数据方面效率低下(Teluguntla等,2018)。随着高级机器学习工具的开发以及强大的云计算资源的集成,有效的解决了现有方法和产品的局限性,为地理空间科学在新的时空尺度上开展更广泛的应用提供了新的机遇(Hird等,2017)。Google Earth Engine(GEE)是一个开放式的云计算平台,拥有海量历史遥感影像数据存档,借助谷歌公司强大的计算资源实现PB级遥感数据的高性能并行运算(Gorelick等,2017)。用户通过调用GEE提供的Python和JavaScript应用编程接口API(Application Programming Interface)实现对云端海量数据的访问,结合其提供的算法实现在线计算和结果可视化等操作。凭借此优势,GEE可以大大减少数据获取和处理的工作量,已经成为地理空间科学研究的重要工具(Gorelick等,2017)。孟梦等(2019)使用GEE平台生产了白洋淀湿地1987年—2017年的NDVI和NDWI数据集并对其湿地生态系统景观进行分类。Mahdianpari等(2019)利用GEE平台和Sentinel-1和Sentinel-2数据对纽芬兰岛进行湿地分类,他们发现光学和SAR的结合可以显著提高湿地分类的准确性,运用基于对象的随机森林(RF)分类可以实现88.37%的总体精度和0.85的Kappa系数。Amani等(2019)使用GEE和机器学习算法利用Landsat 8和Sentinel-1数据基于有限的实测样本绘制了加拿大3个省的湿地地图,总体准确度分别为84%,78%和82%。常文涛等(2020)研究了Sentinel-2红边特征和Sentinel-1雷达特征对小范围湿地提取的作用,但在更大的范围上特征适用性问题需要进一步研究。

综上,尽管光学、SAR和地形等多源数据结合已被用于湿地信息提取,但鲜有将光学、SAR等众多特征进行横向对比,定量评价各个特征对于沼泽湿地信息提取的作用,以寻找有利于沼泽湿地提取的最优特征组合。为此,本文借助GEE云平台,将Sentinel-1 SAR、Sentinel-2光学和地形数据以及机器学习算法相结合,对中国境内黑龙江流域沼泽湿地进行提取。设计了多种分类方案,

探讨了红边、雷达以及地形特征对大范围区域沼泽湿地提取的重要性,验证了利用JM距离寻找最优特征的可行性。鉴于全国湿地类别的特征相似性,希望未来能够将本文的方法用于更大的尺度以在全国范围内进行沼泽湿地信息提取。

2 研究区概况

本文的研究区为黑龙江流域(中国境内)($41^{\circ}42'N-53^{\circ}33'N$, $115^{\circ}25'E-135^{\circ}2'E$),文中所有黑龙江流域均特指位于中国境内的部分。黑龙江跨蒙古国、中国、俄罗斯3国,中国境内流域面积约为89万 km^2 ,占流域总面积的48%。黑龙江流域有着中国最大的沼泽湿地,河流、湖泊等水资源丰富。黑龙江流域水系以黑龙江为干流,主要支流有松花江、乌苏里江、嫩江等(崔茂欢,2006)。流域跨温带和寒温带两个气候带,年均气温在 $-8-6^{\circ}C$ 。年均降水量为250—800 mm,主要发生在6—9月,约占全年降水量的70%(鄢波等,2019)。

3 材料与方法

3.1 数据来源及预处理

研究中使用的所有遥感数据均由GEE平台在线调用和处理,具体包括Sentinel-2卫星的多光谱数据, Sentinel-1卫星的后向散射系数, SRTM数字高程数据以及训练样本数据。

根据已有研究(Lin等,2015; Song等,2011),使用训练样本做监督分类时,若训练样本与研究区域的范围和时相相对应时,大气校正并不是必须的。考虑到2018年GEE上Sentinel-2的地表反射率数据覆盖不全,因此Sentinel-2数据使用的是L1C级的大气表现反射率(TOA)产品。由于研究区范围巨大(89万 km^2)以及云污染的影响,在短时间内(例如一个月)难以获取覆盖研究区范围的数据。因此,使用2018-08—2018-10

创建了光学合成影像,选择此时间段是因为它提供了最多的无云数据并且处于植被生长期,可以保留更多的湿地植被信息。在此期间内共获得了Sentinel-2影像2942景。为了减轻云污染的影响,在合成无云影像时对云量百分比进行了限制($<20\%$)。然后使用Sentinel-2云掩膜算法对指定时间和空间范围内影像进行计算,以中值合成方法重构最小云量合成影像。受益于GEE平台的数据运算和管理机制,研究中使用的所有遥感数据均采样至10 m,并且GEE通过内嵌算法统一坐标系确保不同数据源之间的几何配准精度。

Sentinel-1数据使用的是干涉宽幅模式(IW)的地距多视产品(GRD),分辨率为10 m。通常, Sentinel-1在极地地区收集全极化即HH、HV、VH和VV数据,而在其他地区只收集VH和VV数据。因此本研究共获取2018-08—2018-10 VH极化数据433景, VV极化数据407景。然后通过调用GEE预处理模块完成轨道参数定标、热噪声去除、辐射定标和地形校正等操作。通过上述步骤将无单位的后向散射强度图像转换为后向散射系数值(以dB为单位)。虽然SAR影像不受云雨等天气条件的影响但是其受观测角度影响较大,即距离成像中心越远的像元噪声越强。因此预处理后的SAR数据逐像元计算其在时间范围内的中位数,以减小噪声的影响。

分类结果的精度与训练样本的数量、分布和代表性显著相关。根据黑龙江流域土地覆盖情况结合相关文献(方利等,2017)将研究区土地覆盖类型分为沼泽、水田、人工地表、旱地、盐碱地、林地、明水面7类,其中沼泽、水田和明水面是湿地的主要类型。采用随机采样的方法在研究区生成样本点,利用2018-08—2018-10 Google Earth等高分辨率影像对样本点进行目视解译,得到的样本总数为6649个,其中用于训练和验证的比例为7:3,分别为4673个和1976个。最终各类型样本数量如表1所示。

表1 样本采集情况

Table 1 Situation of sample collection

类型	沼泽	水田	人工地表	旱地	盐碱地	林地	明水面	总计
训练样本	820	619	622	908	488	716	500	4673
验证样本	357	260	235	363	230	305	226	1976
总计	1177	879	857	1271	718	1021	726	6649

1/个

3.2 研究方法

3.2.1 技术路线

本文的技术路线如图1所示。首先对2018-08—2018-10 Sentinel-2 影像进行最小云量合成，结合 SRTM DEM 数据提取的坡度和高程特征以及 Sentinel-1 雷达数据提取的 VV 和 VH 后向散射系数特征进行波段合成、研究区裁剪。为了说明红边、雷达和地形特征的作用以及有利于沼泽湿地提取的最佳特征设计6种实验方案：光谱特征+植被指数+水体指数（方案1），光谱特征+红边特征（方案2），光谱特征+地形特征（方案3），光谱特征+雷达特征（方案4），光谱特征+植被指数+水体指数+红边特征+地形特征+雷达特征（方案5），特征优选（方案6）。利用样本点进行随机森林分类和精度评价，对精度最高的实验方案进行分类器参数调优，最后得到分类结果。

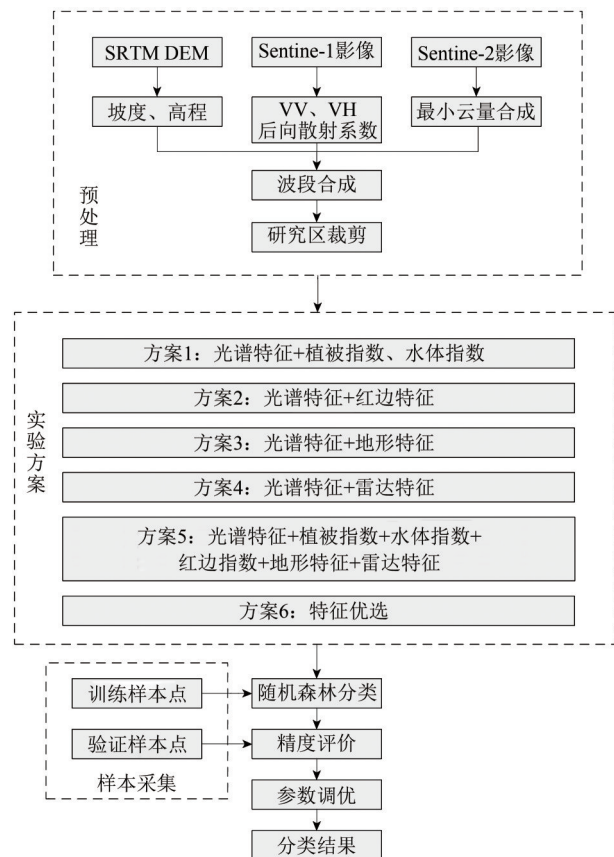


图1 技术路线

Fig. 1 Technical scheme

3.2.2 特征集构建

选取光谱特征、植被指数、水体指数、红边

指数、地形特征以及雷达后向散射系数构建特征集，各个特征的详细说明见表2。需要说明的是，Sentinel-2包括3个红边波段，为了说明红边波段的作用将红边波段与红边指数一起作为红边特征。

表2 特征说明

Table 2 Description of the features

特征变量	简称	特征说明
光谱特征	B	B1—B4、B8、B8A、B9—B12
植被指数	NDVI	$(B8A-B4)/(B8A+B4)$
	EVI	$2.5 \cdot (B8A-B4)/(B8A+6 \cdot B4-7.5 \cdot B2+1)$
	GNDVI	$(B8A-B3)/(B8A+B3)$
水体指数	NDWI	$(B3-B8A)/(B3+B8A)$
	MNDWI	$(B3-B11)/(B3+B11)$
红边指数	NDVIre1	$(B8A-B5)/(B8A+B5)$
	NDVIre2	$(B8A-B6)/(B8A+B6)$
	NDVIre3	$(B8A-B7)/(B8A+B7)$
	NDre1	$(B6-B5)/(B6+B5)$
	NDre2	$(B7-B5)/(B7+B5)$
	Clre	B7/B5-1
地形特征	Bre	B5、B6、B7
	Elevation	高程
	Slope	坡度
雷达特征	VV	VV极化后向散射系数
	VH	VH极化后向散射系数

注：B1—B4、B8、B8A、B9—B12分别为 sentinel-2 影像的波段1到波段4、波段8、波段8A、波段9到波段12。

3.2.3 特征优选

采用JM距离（Jeffries-Matusita distance）对沼泽与其他地类的可分性进行定量分析，并据此确定最佳特征组合。JM距离是衡量类别间分离度的有效工具（张猛和曾永年，2015），表达式为

$$JM = 2(1 - e^{-B}) \quad (1)$$

式中， B 表示某一特征的巴氏距离。在样本满足正态分布的前提下，不同类别样本的巴氏距离 B （Bhattacharyya distance）为

$$B = \frac{1}{8} (e_1 - e_2)^2 \frac{2}{\delta_1^2 + \delta_2^2} + \frac{1}{2} \ln \left(\frac{\delta_1^2 + \delta_2^2}{2\delta_1^2 \cdot \delta_2^2} \right) \quad (2)$$

式中， e_k 表示某类特征的均值， δ_k^2 表示某类特征的方差（张猛等，2017）。JM距离表示样本间的可分离程度，其值在0—2，值越大表示分离性越好。

3.2.4 随机森林分类

本研究采用随机森林（RF）算法用于黑龙江流域湿地分类。RF是一种非参数分类器，由一组

决策树组成，能够处理高维遥感数据 (Belgiu 和 Drăguț, 2016)。与其他的机器学习算法如决策树 (DT) 和支持向量机 (SVM) 相比，它更加稳健并且易于使用 (Rodriguez-Galiano 等, 2012)。随机森林的构建过程为：首先使用 bootstrap 非参数抽样方式从原始训练样本集中有放回的随机抽取样本生成训练样本集合进行训练得到决策树模型。假设每个样本有 M 个特征，之后在决策树的每个结点处从 M 个特征中随机抽取 m ($m \ll M$) 个进行结点分裂；由于随机森林是一种集成学习方法，不容易出现过拟合现象，因此在决策树构建过程中不需要进行剪枝 (何云 等, 2019)。重复上述步骤 k 次得到由 k 个决策树模型组成的随机森林，每个分类样本的分类结果由 k 个决策树通过多数表决的方式决定。

随机森林的效果通过两个参数来调整，即决策树的个数 k 和结点使用的特征数 m 。根据之前的研究 (Belgiu 和 Drăguț, 2016)，当特征数使用以下值时 (树的个数固定为 500)：(1) 特征总数的三分之一；(2) 特征总数的平方根；(3) 特征总数的二分之一；(4) 特征总数的三分之二；(5) 所有特征，分类精度的变化很小。因此根据文献 (Breiman, 2001) 的建议，将结点特征数设置为特征总数的平方根，研究中调用 GEE 平台提供的随机森林算法并对决策树的个数进行调整。

3.2.5 精度评价

利用混淆矩阵和从 Google Earth 等高分辨影像采集的样本点 (表 1) 对黑龙江流域湿地分类进行

精度验证。计算总体精度、Kappa 系数、制图精度和用户精度，总体精度反映了算法的总体效果，通过正确分类的样本数占验证样本总数的比例来衡量。Kappa 系数表示地面真实数据和预测值之间的一致性程度。制图精度表示该类别的地面真实参考数据 (验证样本) 被正确分类的概率。用户精度表示在该次分类中，在分类图上，落在该类别上的检验点，被正确分类为该类别的比率 (Congalton, 1991)。

4 结果与分析

4.1 特征优选分析

基于 GEE 平台取得每个类别的样本均值计算 JM 距离对特征集内所有特征进行可分性分析，本研究关注的是不同特征对沼泽湿地信息提取的重要性，因此只计算沼泽同其他土地覆盖类型的 JM 距离。利用 JM 距离作为衡量不同特征沼泽同其他地类的分离程度的标准，按照分离度由高到低依次增加特征，观察沼泽的制图精度变化 (图 2)。当特征数为 21 时沼泽的制图精度达到最大，之后随着特征继续增加沼泽精度开始趋于稳定并有所下降。因此使用前 21 个特征作为方案 6 使用的特征，具体为 B6、B11、B4、B2、NDVIre2、MDWI、B5、NDVIre1、MNDWI、B9、CIre、B7、Elevation、B3、NDVI、Slope、VH、B8A、B8、VV、NDre2。开始随着所用特征的增加，沼泽制图精度呈快速增长趋势，当使用特征数为 10 时即能达到最优精度的 90.74%。

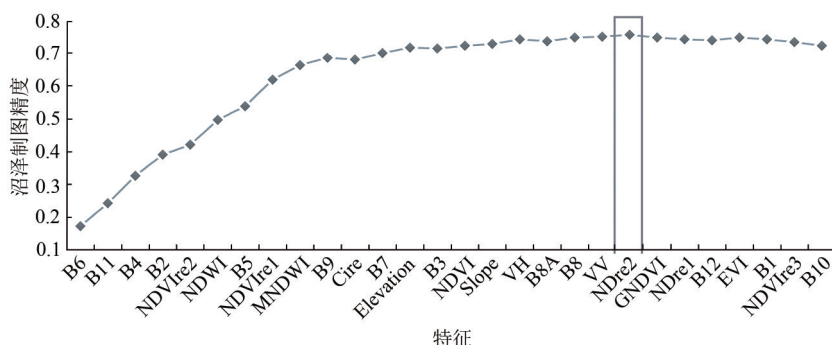


图2 沼泽制图精度随特征变化情况

Fig. 2 Variation of marsh producer accuracy with features

对前 21 个特征的 JM 距离进行分类统计分析，将光谱特征、植被指数和水体指数作为其他光学特征类别，得到表 3。4 类特征中分离度最高的是

红边特征，其次是其他光学特征、地形特征、雷达特征。红边特征中分离度最高的是红边波段 B6，分离度最低的是红边指数 NDre2。红边特征与植被

的生化信息(如叶面积指数、叶绿素含量等)信息相关(Dubeau等, 2017), 相比其他特征更能凸显植被种间差异, 有利于植被茂密繁杂的沼泽区域的识别。雷达特征中, VH交叉极化后向散射系数能获取更多的植被结构信息, 分离度上要高于VV极化后向散射系数。雷达特征在四类特征中的分离度最低, 但仍对沼泽湿地信息提取起到积极作用。

4.2 6种方案分类结果比较

图3展示了6种分类方案的分类结果, 经过特

征优化的方案6取得了最好的分类效果, 分类结果碎片化的现象明显少于其他几种分类方案。

表3 不同类型特征的Jeffries-Matusita(JM)距离
Table 3 Jeffries-Matusita(JM) distance between different types of features

特征类型	分离度最高的特征/排名	分离度最低的特征/排名	平均JM距离
红边特征	B6(1)	NDre2(21)	1.715
其他光学特征	B11(2)	B8(19)	1.704
地形特征	Elevation(13)	Slope(16)	1.624
雷达特征	VH(17)	VV(20)	1.586

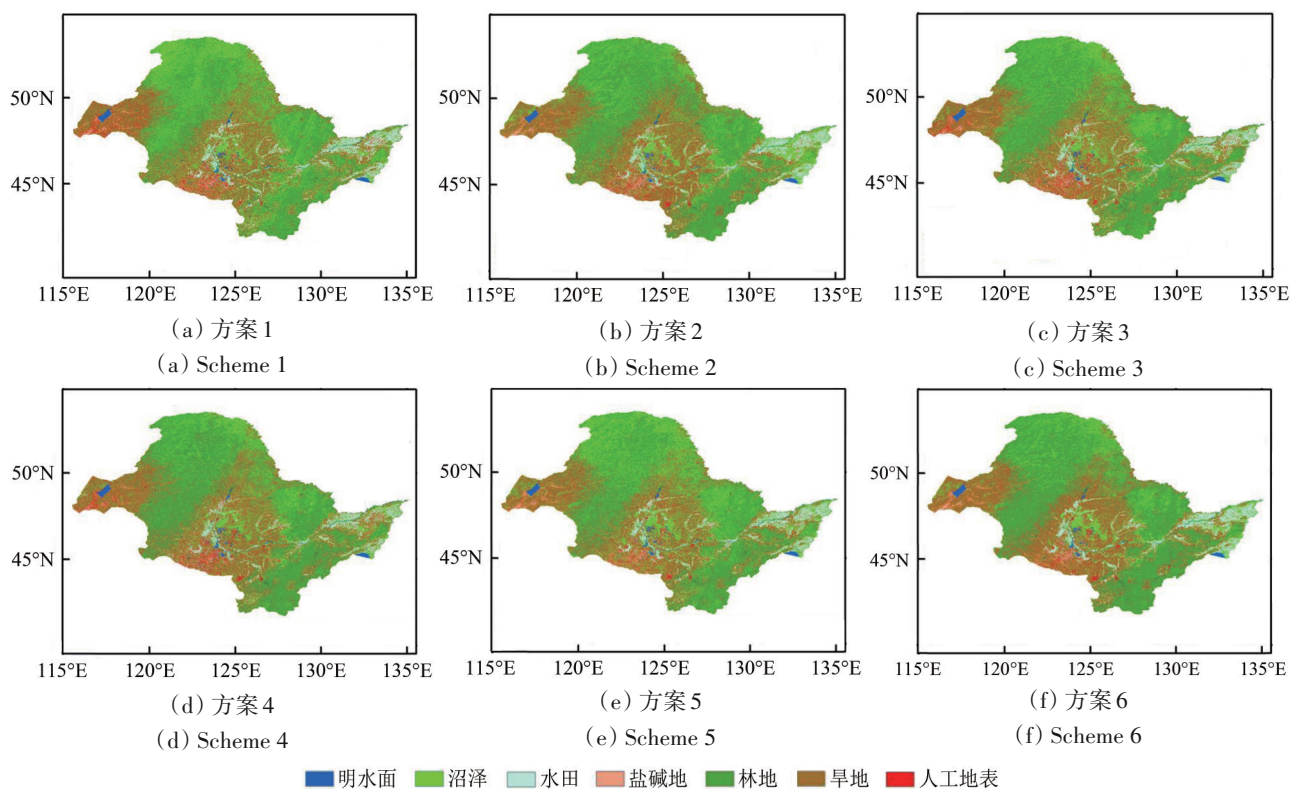


图3 6种实验方案的分类结果

Fig. 3 Classification results of six schemes

6种分类方案的分类精度如表4所示, 综合来看在所有类别中明水面的分类精度最高, 沼泽的分类精度最低。与植被指数和水体指数相比, 红边特征、雷达特征和地形特征对沼泽提取的作用更大。红边特征、雷达特征和地形特征分别使沼泽的制图精度和用户精度提高了7.56%, 5.01%; 5.04%, 5.39%; 4.48%, 4.22%。虽然所有特征都参与分类的方案5取得了较好的效果, 但是没有考虑到信息冗余问题。方案6通过JM距离对所有特征进行优选, 不仅取得了比所有28个特征参与分类的方案5更好的精度而且有效的降低了数据量,

特征数减少到21个, 减少了信息冗余提高运算效率。通过特征优化后, 沼泽的制图精度和用户精度相比方案5分别提高了1.40%和4.60%, 总体精度和Kappa系数提高了0.86%和0.02。因此采用方案6作为最终的湿地分类方案。

4.3 最佳分类方案下的沼泽湿地提取精度评价

对方案6进行随机森林分类, 对每个结点决策树的个数进行了调整以寻找最佳值, 首先将树的个数设置为1记录其沼泽的制图精度, 然后将树的个数设置为5, 之后以5为步长, 观察树的个数每

增加5其精度的变化(图4)。由图4看出,增加树的个数,沼泽的生产者精度快速增长,当树的个数为20时,沼泽的生产者精度达到最大,之后随着树的个数增加,沼泽的生产者精度开始缓慢下

降。因此最优参数为决策树的个数为20棵。经过随机森林调参后沼泽的制图精度提高了13.16%,总体精度提高了7.99%。

表4 分类精度统计

Table 4 Classification accuracy statistics

类别	沼泽	水田	人工地表	旱地	盐碱地	林地	明水面	总体精度/%	Kappa
方案1	PA/%	62.75	88.08	69.79	70.25	84.78	66.23	75.20	0.71
	UA/%	64.55	93.09	73.54	64.07	77.08	70.88		
方案2	PA/%	70.31	88.85	85.53	71.90	86.96	56.07	77.53	0.73
	UA/%	69.56	92.03	84.45	69.60	87.72	63.81		
方案3	PA/%	67.23	85.77	74.89	74.10	80.43	78.03	78.34	0.74
	UA/%	68.77	93.31	76.86	69.51	78.72	78.29		
方案4	PA/%	67.79	86.15	82.55	72.73	87.39	78.69	80.31	0.76
	UA/%	69.94	92.95	82.55	69.66	84.45	78.69		
方案5	PA/%	74.23	90.00	88.51	77.13	88.70	72.79	82.69	0.79
	UA/%	71.24	92.13	88.14	73.88	91.89	78.72		
方案6	PA/%	75.63	90.76	89.36	76.58	83.47	79.34	83.55	0.81
	UA/%	75.84	89.05	85.36	73.74	90.14	83.73		

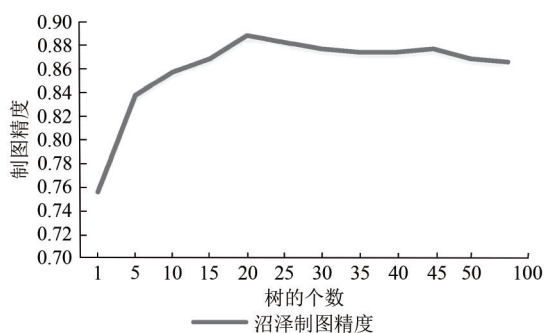


图4 随机森林树的个数与沼泽制图精度的关系

Fig. 4 Relationship between the number of random forest trees and the producer accuracy of marsh

表5显示了方案6最终的混淆矩阵,其中对角线表示制图精度。利用JM距离进行特征优选和随机森林调参后分类精度显著提高,水田、人工地表、盐碱地和明水面的生产者精度均在90%以上。沼泽、旱地和林地的精度相对较低,这主要是由于旱地上农作物种植时间存在差异,不同地区甚至是同一地区的旱地存在有植被覆盖和无植被覆盖两种地表覆盖差异,而本文在采集样本点时没有对此加以区分,导致旱地的训练样本包含有植被覆盖的旱地和无植被覆盖的旱地2类,由于这3种地类的光谱相似性导致相互混淆。利用最佳分类方案进行湿地分类总精度为91.54%,Kappa系

数为0.90,沼泽的制图精度和用户精度分别为88.79%和88.30%,水田的制图精度和用户精度分别为94.23%和96.45%,明水面的制图精度和用户精度分别为98.67%和98.23%。

表5 方案6混淆矩阵

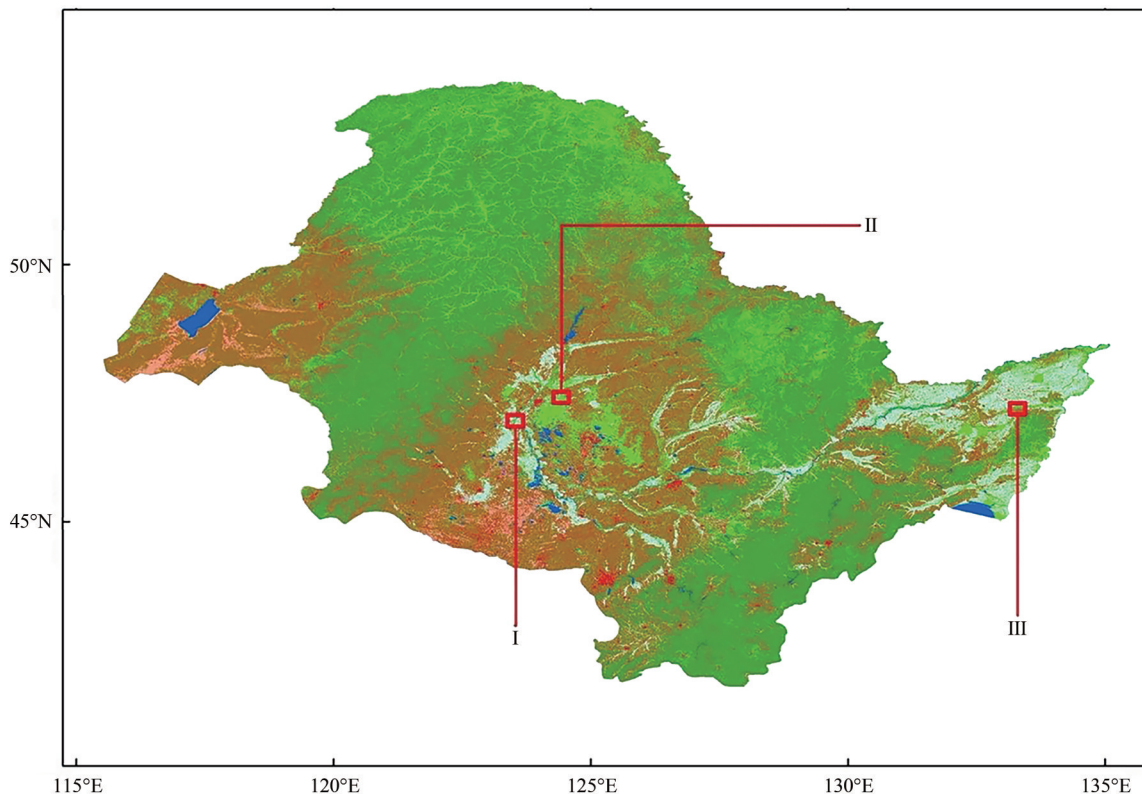
Table 5 Confusion matrix of scheme 6

	沼泽	水田	人工地表	旱地	盐碱地	林地	明水面
沼泽	317	3	0	22	0	14	3
水田	0	245	0	8	0	1	0
人工地表	0	3	220	8	9	0	0
旱地	25	9	9	319	8	18	0
盐碱地	0	0	6	1	213	0	0
林地	11	0	0	5	0	272	0
明水面	4	0	0	0	0	0	223
制图精度/%	88.80	94.23	93.62	87.88	92.61	89.18	98.67
用户精度/%	88.30	96.46	91.67	82.22	96.82	94.44	98.24
总体精度/=91.54				Kappa系数=0.90			

图5展示了本文的分类结果图,以及几处细节的分类效果,图5(b)和图5(c)为扎龙湿地,图5(d)为三江平原湿地。最终的分类结果边界清晰,基本没有破碎化的现象,对比光学影像绝大多数土地覆盖类型都得到了正确区分。图5(b)中,靠近河流的区域被正确识别为了沼泽湿地。

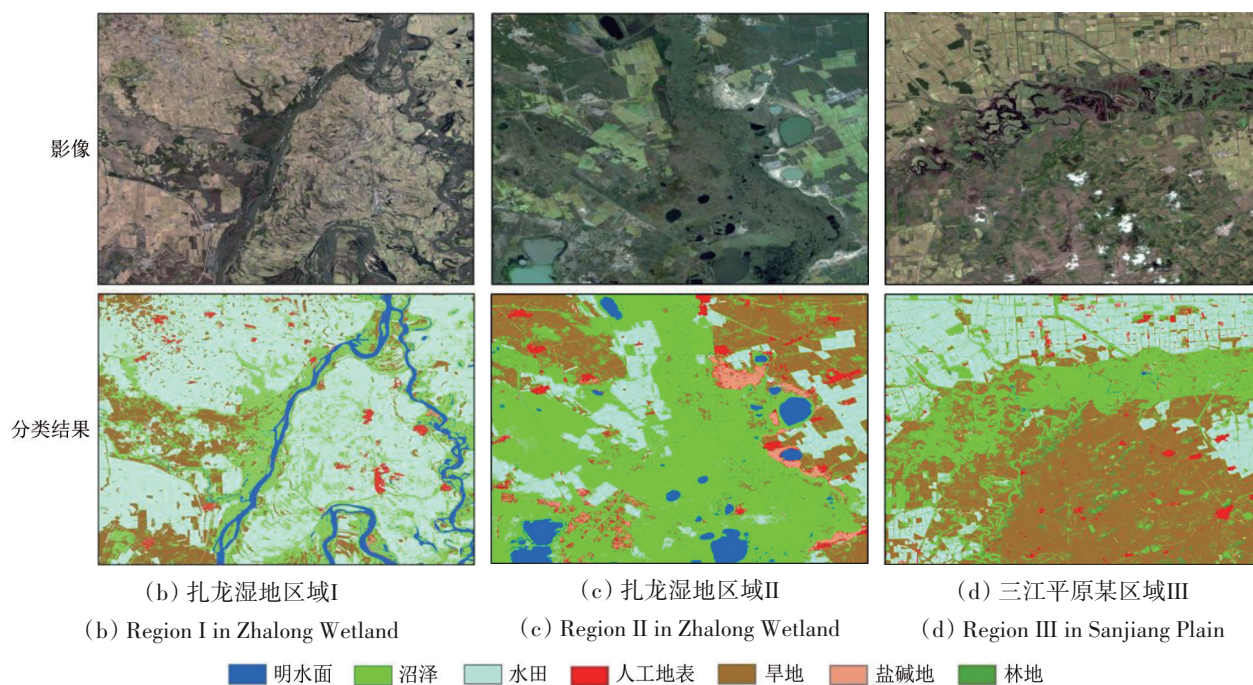
图5 (c) 位于扎龙国家级自然保护区, 主要土地覆盖类型为草本沼泽, 本文中大部分区域识别正确, 沼泽中分布的明水面、水田和盐碱地等也都

清晰可见。在图5 (d) 中也是如此, 沼泽湿地边界清晰, 比较符合实际情况, 提取效果较好。



(a) 本文分类结果图

(a) Classification result map of this paper



(b) 扎龙湿地区域I

(b) Region I in Zhalong Wetland

(c) 扎龙湿地区域II

(c) Region II in Zhalong Wetland

(d) 三江平原某区域III

(d) Region III in Sanjiang Plain

■ 明水面 ■ 沼泽 ■ 水田 ■ 人工地表 ■ 旱地 ■ 盐碱地 ■ 林地

图5 本文局部 Sentinel-2 影像及其对应的分类结果

Fig. 5 The local Sentinel-2 images and their corresponding classification results of this paper

5 结 论

本文依托于 Google Earth Engine 强大的计算能力以及哨兵卫星收集的高空间和高时间分辨率遥感数据, 考虑了多种分类方案, 基于 GEE 平台实现了大尺度区域沼泽湿地信息提取, 探讨了红边、雷达以及地形特征对沼泽湿地提取的作用, 验证了利用 JM 距离寻找最优特征的可行性。结果表明: (1) 红边特征对沼泽湿地类别的辨别能力最好, 红边特征、雷达特征和地形特征分别使沼泽的制图精度和用户精度提高了 7.56%, 5.01%; 5.04%, 5.39%; 4.48%, 4.22%。(2) 利用 JM 距离横向对比了不同特征对沼泽湿地信息提取的重要性, 就沼泽同其他类型的分离度而言, 红边特征>其他光学特征>地形特征>雷达特征。(3) 利用 JM 距离探索了有利于沼泽湿地信息提取的最优特征组合, 并将特征数由 28 减少到 21。具体为 B6、B11、B4、B2、NDVIre2、MDWI、B5、NDVIre1、MNDWI、B9、CIre、B7、Elevation、B3、NDVI、Slope、VH、B8a、B8、VV、NDre2, 相比特征优化前沼泽湿地的制图和用户精度分别提高了 1.40% 和 4.60%。在最优特征基础上进行随机森林调参最终得到的总体精度为 91.54%, Kappa 系数为 0.90, 沼泽湿地的提取精度达到 88.55%。JM 距离为特征优选提供理论支持, 相比其他特征选择方法更为直观合理, 得到最优特征组合后方便后续直接使用。需要注意的是训练样本应满足与影像时相相对应且具有代表性, 否则将影响基于样本特征计算的 JM 距离的结果。

尽管本文沼泽湿地提取取得了较好的精度, 但仍存在一些不足, 后续研究将从以下 3 个方面进行改进: (1) 利用沼泽与旱地、林地和多时相上展现出的物候特征差异减少单时相中三者的混淆; (2) 更深入的探讨有利于沼泽湿地提取的特征及不同特征间的相互影响, 以期更好的优化所用特征; (3) 探索基于 GEE 平台实现面向对象的方法以改善基于像元方法产生的“椒盐现象”。

参考文献(References)

Amani M, Brisco B, Afshar M, Mirmazloumi S M, Mahdavi S, Mirzadeh S M J, Huang W M and Granger J. 2019. A generalized supervised classification scheme to produce provincial wetland invento-

ry maps: an application of Google Earth Engine for big geo data processing. *Big Earth Data*, 3(4): 378-394 [DOI: 10.1080/20964471.2019.1690404]

Belgiu M and Drăguț L. 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 24-31 [DOI: 10.1016/j.isprsjprs.2016.01.011]

Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5-32 [DOI: 10.1023/A:1010933404324]

Chang W T, Wang H, Ning X G and Zhang H C. 2020. Extraction of Zhalong wetlands information based on images of sentinel-2 red-edge bands and sentinel-1 radar bands. *Wetland Science*, 18(1): 10-19 (常文涛, 王浩, 宁晓刚, 张翰超. 2020. 融合 Sentinel-2 红边波段和 Sentinel-1 雷达波段影像的扎龙湿地信息提取. *湿地科学*, 18(1): 10-19) [DOI: 10.13248/j.cnki.wetlandsci.2020.01.002]

Chatziantoniou A, Psomiadis E and Petropoulos G P. 2017. Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning. *Remote Sensing*, 9(12): 1259 [DOI: 10.3390/rs9121259]

Congalton R G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1): 35-46 [DOI: 10.1016/0034-4257(91)90048-B]

Cui M H. 2006. Status quo of wetlands of Heilongjiang River valley and the protection. *Forest Inventory and Planning*, 31(1): 75-78 (崔茂欢. 2006. 黑龙江流域湿地现状及保护. *林业调查规划*, 31(1): 75-78) [DOI: 10.3969/j.issn.1671-3168.2006.01.020]

Dubeau P, King D J, Unbushe D G and Rebelo L M. 2017. Mapping the Dabus wetlands, Ethiopia, using random forest classification of Landsat, PALSAR and topographic data. *Remote Sensing*, 9(10): 1056 [DOI: 10.3390/rs9101056]

Fang L, Wang W J, Jiang W G, Chen M, Wang Y, Jia K and Li Y S. 2017. Spatio-temporal variations of vegetation cover and its responses to climate change in the Heilongjiang Basin of China from 2000 to 2014. *Scientia Geographica Sinica*, 37(11): 1745-1754 (方利, 王文杰, 蒋卫国, 陈民, 王永, 贾凯, 李延森. 2017. 2000~2014 年黑龙江流域(中国)植被覆盖时空变化及其对气候变化的响应. *地理科学*, 37(11): 1745-1754) [DOI: 10.13249/j.cnki.sgs.2017.11.017]

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R. 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202: 18-27 [DOI: 10.1016/j.rse.2017.06.031]

He Y, Huang C, Li H, Liu Q S, Liu G H, Zhou Z C and Zhang C C. 2019. Land-cover classification of random forest based on Sentinel-2A image feature optimization. *Resources Science*, 41(5): 992-1001 (何云, 黄翀, 李贺, 刘庆生, 刘高焕, 周振超, 张晨晨. 2019. 基于 Sentinel-2A 影像特征优选的随机森林土地覆盖分类. *资源科学*, 41(5): 992-1001) [DOI: 10.18402/resci.2019.05.15]

Hird J N, DeLancey E R, McDermid G J and Kariyeva J. 2017. Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sensing*, 9(12): 1315 [DOI: 10.3390/rs9121315]

- Limpitlaw D and Gens R. 2006. Dambo mapping for environmental monitoring using Landsat TM and SAR imagery: case study in the Zambian Copperbelt. *International Journal of Remote Sensing*, 27(21): 4839-4845 [DOI: 10.1080/01431160600835846]
- Lin C S, Wu C C, Tsogt K, Ouyang Y C and Chang C I. 2015. Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery. *Information Processing in Agriculture*, 2(1): 25-36 [DOI: 10.1016/j.inpa.2015.01.003]
- Liu L, Zang S Y, Na X D and Pei X Y. 2013. Wetland mapping in the Zhalong natural reserve using optical and radar remotely sensed data and ancillary topographical data. *Geography and Geo-Information Science*, 29(1): 36-40 (刘蕾, 臧淑英, 那晓东, 裴雪原. 2013. 兼容光学雷达影像及地形辅助数据的扎龙湿地遥感分类. *地理与地理信息科学*, 29(1): 36-40) [DOI: 10.7702/dly-dlxxkx20130108]
- Mahdianpari M, Salehi B, Mohammadimaneh F, Homayouni S and Gill E. 2019. The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the Google earth engine cloud computing platform. *Remote Sensing*, 11(1): 43 [DOI: 10.3390/rs11010043]
- Meng M, Tian H F, Wu M Q, Wang L and Niu Y. 2019. Evolution characteristics analysis of wetland landscape pattern based on Google Earth Engine platform: a case study on Baiyangdian. *Journal of Yunnan University (Natural Science Edition)*, 41(2): 416-424 (孟梦, 田海峰, 邱明权, 王力, 牛铮. 2019. 基于Google Earth Engine平台的湿地景观空间格局演变分析: 以白洋淀为例. *云南大学学报(自然科学版)*, 41(2): 416-424) [DOI: 10.7540/j.ynu.20170715]
- Mitsch W J, Bernal B, Nahlik A M, Mander Ü, Zhang L, Anderson C J, Jørgensen S E and Brix H. 2013. Wetlands, carbon, and climate change. *Landscape Ecology*, 28(4): 583-597 [DOI: 10.1007/s10980-012-9758-8]
- Mitsch W J and Gosselink J G. 2000. The value of wetlands: importance of scale and landscape setting. *Ecological Economics*, 35(1): 25-33 [DOI: 10.1016/S0921-8009(00)00165-8]
- Ozesmi S L and Bauer M E. 2002. Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, 10(5): 381-402 [DOI: 10.1023/A:1020908432489]
- Rodriguez-Galiano V F, Ghimire B, Rogan J, Chica-Olmo M and Rigol-Sanchez J P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67: 93-104 [DOI: 10.1016/j.isprsjprs.2011.11.002]
- Shalaby A and Tateishi R. 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography*, 27(1): 28-41 [DOI: 10.1016/j.apgeog.2006.09.004]
- Shoko C and Mutanga O. 2017. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129: 32-40 [DOI: 10.1016/j.isprsjprs.2017.04.016]
- Song C H, Woodcock C E, Seto K C, Lenney M P and Macomber S A. 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects?. *Remote Sensing of Environment*, 75(2): 230-244 [DOI: 10.1016/S0034-4257(00)00169-3]
- Teluguntla P, Thenkabail P S, Oliphant A, Xiong J, Gumma M K, Congalton R G, Yadav K and Huete A. 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144: 325-340 [DOI: 10.1016/j.isprsjprs.2018.07.017]
- Wang A Q, Zhou D M and Gong H L. 2012. A research on method of wetland vegetation identification and classification based on radar backscatter characteristics. *Remote Sensing Information*, (2): 15-19 (王安琪, 周德民, 宫辉力. 2012. 基于雷达后向散射特性进行湿地植被识别与分类的方法研究. *遥感信息*, (2): 15-19) [DOI: 10.3969/j.issn.1000-3177.2012.02.003]
- Whyte A, Ferentinos K P and Petropoulos G P. 2018. A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms. *Environmental Modelling and Software*, 104: 40-54 [DOI: 10.1016/j.envsoft.2018.01.023]
- Yan B, Zhang X, Xia Z Q, Guo M C, Liu C X and Luo Y Y. 2019. Characteristics of precipitation variation in Heilongjiang River Basin. *Journal of Yangtze River Scientific Research Institute*, 36(7): 14-17, 22 (鄢波, 张潇, 夏自强, 郭勉辰, 刘宸希, 罗云英. 2019. 黑龙江流域降水变化特征分析. *长江科学院院报*, 36(7): 14-17, 22) [DOI: 10.11988/ckyyb.20171298]
- Yan T T, Bian H F, Liao G X, Sheng L X, Zhang J S and Gao M H. 2014. Research status of methods for mapping forested wetlands based on remote sensing. *Remote Sensing for Land and Resources*, 26(2): 11-18 (严婷婷, 边红枫, 廖桂项, 盛连喜, 张吉顺, 高明辉. 2014. 森林湿地遥感信息提取方法研究现状. *国土资源遥感*, 26(2): 11-18) [DOI: 10.6046/gtzyyg.2014.02.03]
- Yesou H, Pottier E, Mercier G, Grizonnet M, Haouet S, Giros A, Faivre R, Huber C and Michel J. 2016. Synergy of Sentinel-1 and Sentinel-2 imagery for wetland monitoring information extraction from continuous flow of sentinel images applied to water bodies and vegetation mapping and monitoring//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Beijing: IEEE: 162-165 [DOI: 10.1109/IGARSS.2016.7729033]
- Zhang M and Zeng Y N. 2015. Mapping paddy fields of Dongting Lake area by fusing Landsat and MODIS data. *Transactions of the Chinese Society of Agricultural Engineering*, 31(13): 178-185 (张猛, 曾永年. 2015. 基于多时相Landsat数据融合的洞庭湖区水稻面积提取. *农业工程学报*, 31(13): 178-185) [DOI: 10.11975/j.issn.1002-6819.2015.13.025]
- Zhang M, Zeng Y N and Zhu Y S. 2017. Wetland mapping of Dongting Lake Basin based on time-series MODIS data and object-oriented method. *Journal of Remote Sensing*, 21(3): 479-492 (张猛, 曾永年, 朱永森. 2017. 面向对象方法的时间序列MODIS数据湿地信息提取——以洞庭湖流域为例. *遥感学报*, 21(3): 479-492) [DOI: 10.11834/jrs.20176129]

Extraction of marsh wetland in Heilongjiang Basin based on GEE and multi-source remote sensing data

NING Xiaogang^{1,2}, CHANG Wentao^{1,2}, WANG Hao^{1,2}, ZHANG Hanchao¹, ZHU Qiande³

1. Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping, Beijing 100036, China;

2. College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China;

3. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

Abstract: Wetland is an important ecosystem on the planet and plays a pivotal role in maintaining global ecological environment security. Traditional wetland monitoring requires a lot of manpower and financial resources due to the unique hydrological characteristics of wetlands, and extracting large-scale wetland information is also difficult. Compared with traditional field surveys, remote sensing technology, which has the advantages of wide observation range, short update cycle, has played an important role in large-scale wetland information extraction. However, remote sensing monitoring of marsh wetland mainly uses optical images traditionally, which are severely affected by weather such as clouds and rain, making large-scale marsh wetland extraction challenging. The use of radar and terrain data can combine the spectral information and scattering mechanism, which has great potential for marsh wetland information extraction. Nevertheless, there are few studies that evaluate the differences of optical, SAR, and topographic features in importance for the extraction of marsh wetland information. The rise of big data and cloud computing has enabled large-scale and long time series spatial data processing. On the basis of the Google Earth Engine (GEE) cloud platform, this study uses Sentinel-1 synthetic aperture radar data, Sentinel-2 optical data, and terrain data to explore their importance to the extraction of marsh wetland at large scale, and verify the feasibility of JM distance to find the optimal feature combination to the extraction of marsh wetland. Random forest algorithm is also used to extract marsh wetlands in Heilongjiang Basin in 2018. In order to explore the importance of red edge, radar and topographic features and the best features conducive to marsh wetland extraction, six experimental schemes are designed. Scheme one uses the combination of spectral feature, vegetation index and water index. Scheme two uses the combination of spectral feature and red edge feature. Scheme three uses the combination of spectral feature and terrain feature. Scheme four uses the combination of spectral feature and radar feature. Scheme five uses the combination of spectral feature, vegetation index, water index, red edge feature, terrain feature and radar feature. Scheme six uses all features, which are optimized by JM distance. The research shows that (1) Sentinel-2 red edge bands and Sentinel-1 radar bands and terrain data are conducive to marsh wetland information extraction. Compared with vegetation indexes and water indexes, the producer accuracy of marsh increased by 7.56%, 5.04%, and 4.48%. (2) The separation obtained using JM distance shows that the order is red-edge features > other optical features > terrain features > radar features. The marsh wetland in scheme six has the highest mapping accuracy and user accuracy. After using the JM distance to select features, the producer and user accuracies of the marsh wetland increased by 1.45% and 3.02%, respectively. The overall accuracy of the combined random forest algorithm was 91.54%, and the accuracy of marsh extraction was 88%. This study uses the GEE cloud platform, multisource remote sensing data, and machine learning algorithms to accurately, quickly, and efficiently extract large-scale marsh wetland information. This method has great application potential.

Key words: remote sensing, Google Earth Engine, marsh, Sentinel-1, Sentinel-2, JM distance, random forest, red-edge bands

Supported by National Key Research and Development Project of China (No. 2017YFC0404503); National Natural Science Foundation of China (No. 41601529); Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences (No. WSGS2017002); Central Public-Interest Scientific Institution Basal Research Fund (No. Y518002)