

黄河三角洲典型盐沼植被时空分布数据集 (1999–2020) 研发

胡鉴芳^{1,2,3}, 宫兆宁^{1,2,3,4*}, 张成^{1,2}, 邱华昌^{1,2,3}

1. 首都师范大学资源环境与旅游学院, 北京 100048; 2. 三维信息获取与应用教育部重点实验室, 北京 100048;
3. 资源环境与 GIS 北京市重点实验室, 北京 100048; 4. 水资源安全北京市重点实验室, 北京 100048

摘要: 以盐地碱蓬、芦苇、互花米草为代表的黄河三角洲典型盐沼植被, 在维持生物多样性、提供重要栖息地、降低暴雨径流、调节气候等方面提供多种生态服务功能。研究其分布和演替趋势对整个生态系统的结构保护与管理具有重要意义。本数据集在黄河三角洲开展研究, 基于 Google Earth Engine (GEE) 大数据平台, 利用 1999–2020 年 2,068 景 Landsat TM/ETM/OLI、Sentinel-2 MSI 光学数据和 Sentinel-1 SAR 雷达数据, 结合盐沼植被的物候季相特征, 构建全生长期的多特征空间。利用递归特征消除的特征优选算法筛选出最佳特征组合, 利用随机森林算法进行 1999–2020 年黄河三角洲典型盐沼植物种群的分类, 进一步分析获得典型盐沼植被时空分布数据集。该数据集覆盖区域为黄河三角洲河口湿地, 栅格数据空间分辨率分别为 10 m。数据存储为.tif 格式, 共包含 64 个数据文件, 数据量为 172 MB (压缩为 1 个文件, 31.7 MB)。

关键词: 黄河三角洲; 盐沼植被; 1999–2020; 特征优选算法

DOI: <https://doi.org/10.3974/geodp.2022.02.07>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2022.02.07>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2022.01.06.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2022.01.06.V1>.

1 前言

河口湿地是海陆相互作用的前沿地区, 是海岸带生态系统中最重要的一部分^[1]。湿地植被的结构、布局可以反映湿地生态环境的基本特点和变化情况, 对区域生态环境变化和监测研究意义重大。河口湿地盐沼植被具有能量交换、生物栖息、改善环境等生态功能^[2], 对其分布信息的准确提取, 动态变化规律的及时掌握, 对于湿地生态环境乃至区域生态环境的保护与管理意义重大^[3]。

黄河三角洲湿地是典型的滨海湿地生态系统, 地处沿海水陆交错地带, 具有高度敏感

收稿日期: 2022-02-10; 修订日期: 2022-05-16; 出版日期: 2022-06-25

基金项目: 中华人民共和国科学技术部 (2017YFC0505900)

*通讯作者: 宫兆宁, DWF-7853-2022, 首都师范大学, gongzhn@163.com

数据引用方式: [1] 胡鉴芳 宫兆宁 张成等. 黄河三角洲典型盐沼植被时空分布数据集 (1999–2020) 研发 [J]. 全球变化数据学报, 2022, 6(2): 217–224. <https://doi.org/10.3974/geodp.2022.02.07>.
[2] 胡鉴芳 宫兆宁 张成等. 黄河三角洲典型盐沼植被时空分布数据集 (1999–2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.01.06.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.01.06.V1>.

性和脆弱性。在全球气候变暖、物种入侵、人类活动等背景下,黄河三角洲的地表形态发生显著演变,芦苇、盐地碱蓬等原生物种面临着生境破碎、面积减退的问题,而入侵物种互花米草凭借自身竞争力优势,迅速定植、拓张,河口湿地植被的分布格局频繁演替^[3]。准确有效地提取河口湿地盐沼植被信息一直是亟待解决的难题。这一难题的解决对于黄河三角洲湿地生态环境监测和保护研究意义重大^[4]。本数据集以黄河三角洲入海口处为研究区,结合盐沼植被的物候季相特征,在 GEE 数据平台,充分利用 2,068 景 Landsat TM/ETM/OLI、Sentinel-2 MSI 光学数据和 Sentinel-1 SAR 雷达数据,得到黄河三角洲典型盐沼植物种群的空间分布数据,并挖掘分析其分布特征,得到 1999–2020 年典型盐沼植物种群时空分布数据集。

2 数据集元数据简介

《黄河三角洲典型盐沼植被时空分布数据集(1999–2020)数据集》^[5]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

3 数据研发方法

本文选取了 1999–2020 年所有可用的质量较好的(云量<30%) Landsat 系列数据、Sentinel-2 光学数据以及 Sentinel-1 SAR 雷达数据。具体包括 197 景 Landsat-7ETM+影像、62 景 Landsat-8 OLI 影像、1,294 景 Sentinel-2 MSI 影像和 515 景 Sentinel-1 SAR 影像。GEE 云平台(<https://earthengine.google.com/>)整合了大量的遥感影像数据和其他观测数据资源,可以在线调用全球尺度 1972 年至今的 Landsat 系列卫星遥感数据,2000 年至今的 MODIS 数据及其系列产品,2014 年至今的 Sentinel 系列卫星遥感数据以及高程、气象、土地利用等数据资料^[7]。本研究所用影像资料均可在 GEE 平台实现调用。其他辅助分析数据包括近年来课题组多次对黄河三角洲自然保护区进行野外踏勘的实测样点数据,盐沼植被不同生长状态的光谱数据使用 ASD FieldSpec 3 背挂式光谱仪采集,采样光谱范围为 350–2,500 nm。同时,为避免环境变化问题导致样本点选取的偏差,结合 Google Earth 高分辨率遥感影像目视解译结果,补充高精度样本点数据。

本文以 GEE 云平台作为支撑,通过代码编写实现长时间序列 Landsat7/8 数据、Sentinel-2 数据的辐射定标、大气校正, Sentinel-1 数据的辐射定标、地形校正等预处理。结合湿地植物的物候季相特征和生活史差异,以年均尺度计算多种时间序列植被指数,构成时-谱维度特征集。并从 Sentinel-1 数据中提取后向散射系数构建时-空维度特征集。考虑到黄河三角洲近年来的植被分布动态变化及卫星数据的质量和可用性,结合 Sentinel-1 系列数据进行 2014–2020 年时-空-谱多维度特征协同分类,基于 Landsat 历史时序数据进行 1999–2013 年时-谱维度特征提取。

多维度特征提取充分利用植被的时序特征、光学特征、空间纹理特征,可以有效提高地物分类精度,但同时较高维数的特征集带来的数据冗余问题仍然带来了一定程度的困扰。

表 1 《黄河三角洲典型盐沼植被时空分布数据集（1999–2020）》元数据简表

条目	描述
数据集名称	黄河三角洲典型盐沼植被时空分布数据集（1999–2020）
数据集短名	SaltMarshVegYRD1999–2020
作者信息	胡鉴芳 AHA-1158-2022，首都师范大学，2190902115@cnu.edu.cn 宫兆宁 DWF-7853-2022，首都师范大学，gongzhn@163.com 张成 EIE-6796-2022，首都师范大学，chzhang_hn@163.com 邱华昌 DNC-0167-2022，首都师范大学，qiu_huachang@163.com
地理区域	黄河三角洲入海口处河口湿地
数据年代	1999–2020 年
时间分辨率	年
空间分辨率	30 m、10 m
数据格式	.shp、.tif、.xlsx
数据量	172 M（压缩为一个文件，31.7 MB）
数据集组成	64 个数据文件组成
基金项目	中华人民共和国科学技术部（2017YFC0505900）
数据计算环境	ENVI、ArcGIS
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《全球变化数据学报（中英文）》发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[6]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

本文采用递归特征消除法（Recursive feature elimination, RFE）对特征集进行降维处理，剔除冗余波段，获取最佳特征组合。在本文中 RFE 算法的实现流程可概括如下：

- 步骤 1：训练一个随机森林；
- 步骤 2：计算置换的重要性测度；
- 步骤 3：消除不相关变量；
- 步骤 4：重复步骤 1 至 3，直到获得最后的特征集合，给定的变量置换重要性预测强烈依赖于它与其他变量的相关性。

随机森林（Random Forest, 简称 RF）于 2001 年首次提出^[8]，已被广泛应用于湿地监测等遥感研究中^[8,9]。本文将优选的特征数据集输入随机森林算法，实现典型盐沼植被的高精度提取。不同物种对环境的不同响应导致生态位分化现象，通过对 1999–2020 年共 13 期的植被分布数据空间分析得到 1999–2020 年盐地碱蓬和芦苇的分布频率图，并结合首次发现互花米草以来，2008–2020 年共 10 期的植被分布数据得到 2008–2020 年互花米草的分布

频率图,其中同种植被年际出现频度越高则该植被在此区域分布频率越高。本研究认为年际间分布频率大于等于70%的区域为适合典型盐沼植被分布的生态位。

4 数据结果与验证

4.1 数据集组成

本数据集由以下数据组成:

- (1) 1999–2020 年黄河三角洲盐沼植被空间分布数据;
- (2) 1999–2020 黄河三角洲互花米草分布频率数据;
- (3) 1999–2020 黄河三角洲盐地碱蓬分布频率数据;
- (4) 1999–2020 黄河三角洲芦苇分布频率数据。

其中,在黄河三角洲盐沼植被空间分布数据中,字段“value”代表不同盐沼植被分类,value 值为 1 代表互花米草;value 值为 2 代表盐地碱蓬,value 值为 3 代表芦苇。

4.2 数据结果

基于以上建立的多维度特征提取方法及随机森林分类算法,结合 GEE 大数据处理平台,利用 2,068 景 Landsat TM/ETM+/OLI 数据、Sentinel-2 MSI 光学影像以及 Sentinel-1 SAR 雷达数据,得到 1999–2020 年黄河三角洲典型盐沼植被时空分布数据集。该数据集的空间分辨率分别为 30 m、10 m,大小为 143 MB。

黄河三角洲湿地同时受到海洋和黄河径流泥沙的共同作用,呈现出频繁的蚀淤交替变化,其生态系统敏感而脆弱,植被演替频繁^[10]。图 1 展示了黄河三角洲典型盐沼植被长时序分布图,虽然黄河三角洲湿地仍然处于高度动态变化的过程,但其植被分布已呈现较为明显的梯度分布特征,互花米草入侵前,研究区主要盐沼植被芦苇和盐地碱蓬由内陆向海呈带状分布。芦苇群落的发育多沿河伸展,垂直河流向海方向,植被向着更适合盐生环境的方向演替,盐地碱蓬成为滩海过渡地带的先锋物种。随着入侵物种对河口湿地生态系统的干扰,黄河三角洲典型盐沼植被的发育演化具有了新的特点。

2008 年开始在研究区北岸滨海边缘区域观测到零星互花米草斑块,其凭借较强环境适应性侵占海草床成功定植,并通过较强种间竞争能力占据潮间带空生态位实现快速扩张,结合图 2 的统计结果发现,截至 2013 年其面积已达到 1,513.16 hm²。互花米草极强的扩散能力使其占据了黄河三角洲滨海湿地大部分有利生境,加上近年来频繁人类活动影响,导致盐地碱蓬、芦苇等原生植被群落的分布结构发生改变,原生湿地出现不同程度的退化。芦苇主要沿黄河现行河道以及黄河故道两侧分布,沿着现行河道两侧向东北方向延伸,沿黄河故道向内陆有所消退,虽然其在黄河入海口新淤积光滩处与互花米草竞争中处于劣势,但其主体分布较为稳定。盐地碱蓬虽然是淤泥质潮滩和重盐碱地段的先锋植物,但黄河三角洲河口湿地淤进蚀退转换快,盐地碱蓬群落随潮汐往复长势较差,稳定性不足,发生大面积萎缩现象。互花米草种群对外部环境变化的适应能力较强,到 2015 年在黄河入海流路东南侧已经形成大面积的单建群种植被群落,其分布总面积达到 2,459.09 hm²。于 2017 年在最南部光滩开始定植,并沿东西方向在光滩上快速扩张,至 2020 年在南部滨海光滩处形成了东西向分布的四个互花米草扩张斑块。

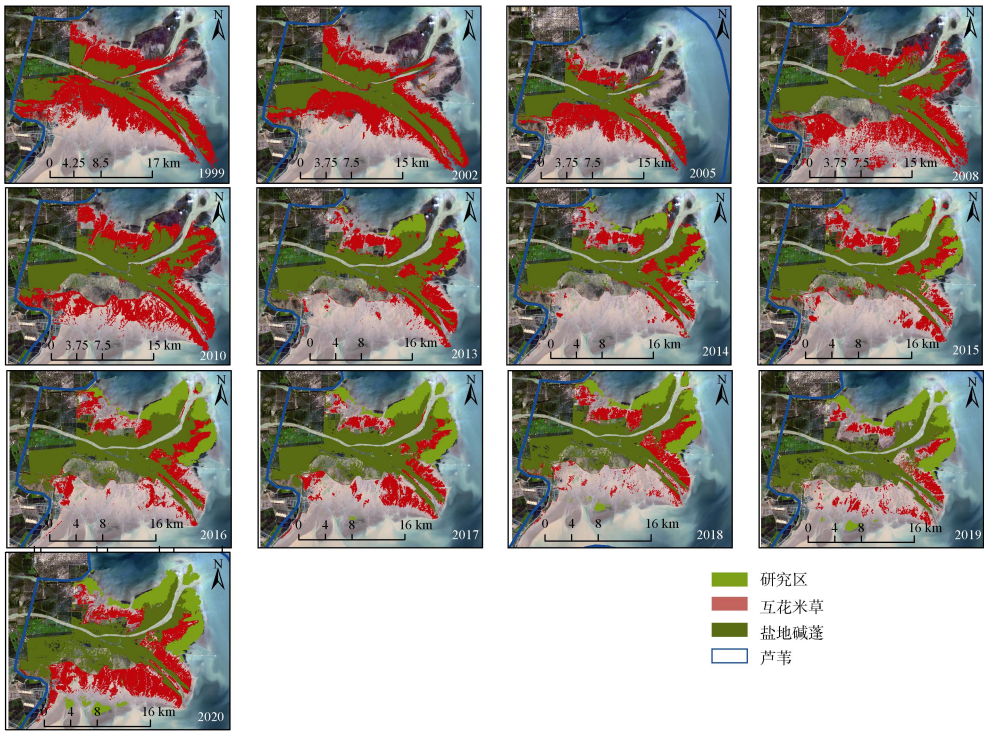


图 1 1999–2020 年 13 期典型盐沼植被的空间分布图

图 3 展示了三种典型盐沼植被分布频率图，其中红色区域年际间植被出现频度较高，代表最适合生长的生态位，黄绿色生长频率较低，多为生长不稳定的退化区域或新扩张区域。三种植被中互花米草的分布范围最小，但在其分布范围内，互花米草年际出现频度均较高，分布范围多为稳定生长区域以及新扩张区域。这主要得益于互花米草强大的繁殖能力以及对水盐环境的适应能力。互花米草通过有性繁殖产生种子进入新生境，形成互花米草斑块以实现种群空间拓植，一旦定植成功，互花米草通过无性繁殖迅速生长分蘖，使互花米草斑块不断扩大融合，呈峰面状连续扩散抢占适宜生态位。沿海滩涂广阔的光滩上遍布盐地碱蓬，但分布频率普遍较低，大部分区域年际间波动较大或发生明显退化，稳定分布区域主要位于潮滩的中部。虽然盐地碱蓬是淤泥质潮滩和重盐碱地段的先锋植物，但其植株低矮、根茎较弱，难以承受环境胁迫且缺乏竞争能力，造成从个体指标到群落指标的双重抑制影响。互花米草的入侵更是导致适宜盐地碱蓬生长的生境明显萎缩。芦苇具有较广的生长生态幅度，在淡水、咸水、咸淡水交互区域均有分布。其沿河岸成狭长条带状分布，一直延伸到入海口附近，分布区域也大部分为稳定生长的区域。根据盐沼植被近年来分布频率图，本研究将分布频率大于等于 70% 的区域，划分为适合典型盐沼植被分布的生态位，该区域盐沼植被年际间波动较小，在研究时间内同种植被出现频度达到 70%。如图 4 所示，芦苇稳定生长区域的面积占比最大，达到 7,720.41 hm^2 ，其次为互花米草，达到 1,859.66 hm^2 ，盐地碱蓬生长面积最小，为 859.26 hm^2 。

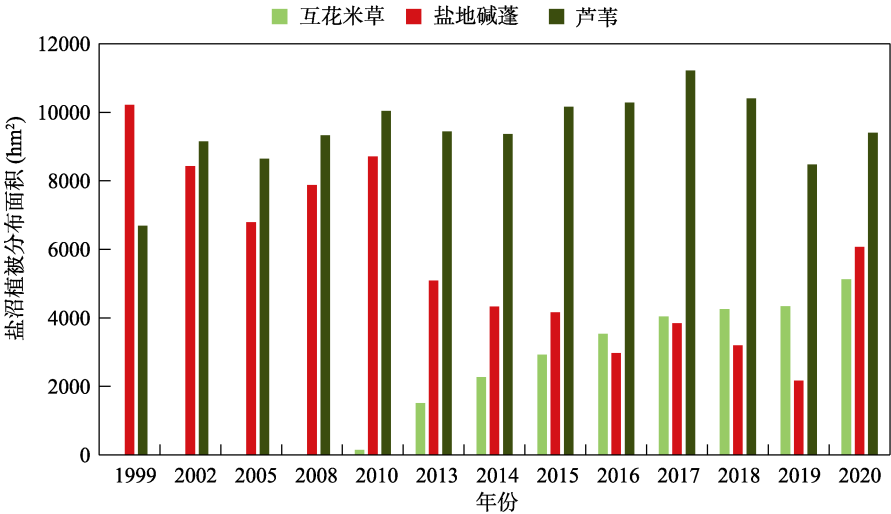


图 2 1999–2020 年典型盐沼植被面积统计图

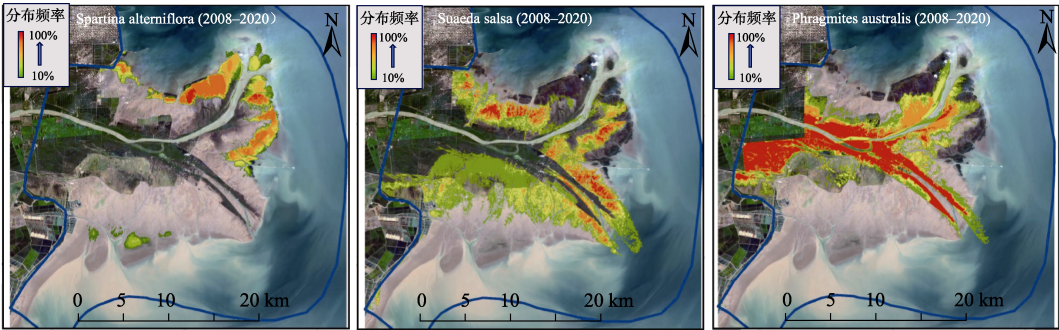


图 3 黄河三角洲典型盐沼植被分布频率图

4.3 数据结果验证

黄河三角洲典型盐沼植被分类数据的精度验证主要采用较为广泛的精度评价方法，混淆矩阵（Confusion Matrix）的方法。在混淆矩阵中，验证样本的值与提取的数据值进行对比，构建基于实测数据与提取数据的误差矩阵，通过计算分类正确的像元数与总像元数之间的数量关系，得到各类提取精度。本研究采用被广泛应用的总体精度（Overall Accuracy, OA）和 kappa 系数对典型盐沼植被的提取精度进行评价分析。其中，总体精度为分类正确的像元数量总和占总像元数量的比值，代表整体像元的分类精度。kappa 系数可以用来评价分类质量，而且能够克服只使用上述精度分析指标所带来的对采样样本及方法的依赖，用于衡量分类精度。kappa 系数计算公式如下：

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{1}$$

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_c \times b_c}{n \times n} \tag{2}$$

式中， p_o 是每一类正确分类的样本数量之和除以总样本数，也就是总体分类精度， a_1, a_2, \dots ,

a_c 为每一类的真实样本个数， b_1, b_2, \dots, b_c 为预测出来的每一类的样本个数， n 为总样本数。

本文基于 GEE 平台，对每一年的时序影像进行分类。分类时随机选取 70% 的样本，建立随机森林模型，利用剩下 30% 样本进行结果验证。图 5 展示了 1999–2020 年典型盐沼植被分类的总体精度与 kappa 系数，总体精度范围为 86%–92.4%，kappa 系数范围为 0.68–0.88。可见本研究所用分类方法取得了较为满意的精度，且在长时间序列数据中运用时效果较为稳健。

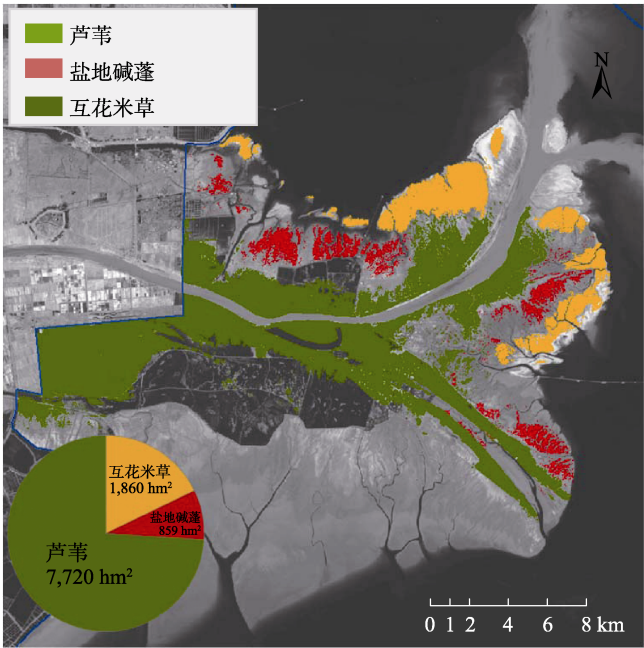


图 4 1999–2020 年黄河三角洲典型盐沼植被稳定生态位分布图

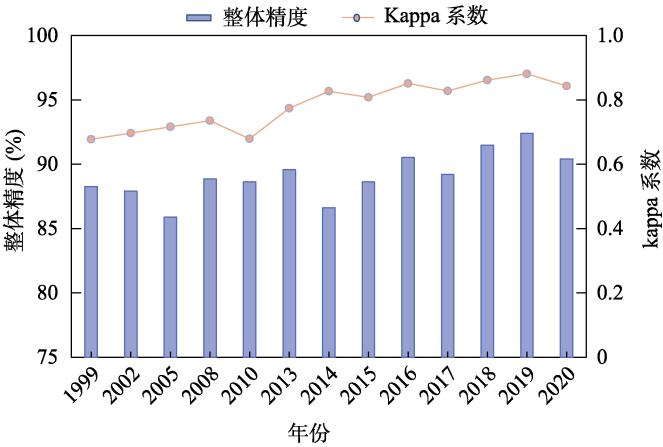


图 5 1999–2020 年黄河三角洲典型盐沼植被分类精度图

5 讨论和总结

本研究以 GEE 平台为基础,充分考虑不同盐沼植被形态特征、生活史差异的基础上,利用 2,068 景 Landsat TM/ETM/OLI、Sentinel-2 MSI 光学数据和 Sentinel-1 SAR 雷达数据,构建全生长期的多特征空间并利用递归特征消除的特征优选算法消除冗余特征,采用随机森林的分类算法得到 1999–2020 年黄河三角洲典型盐沼植物种群的分类数据,进一步分析获得典型盐沼植被时空分布数据集。尽管黄河三角洲是中国暖温带地区最年轻的新生湿地生态系统,植被群落尚未完全发育和演替,但就结果来看,黄河三角洲滨海湿地沿河向海方向,呈现出较为明显的层次结构。互花米草的快速入侵抢占盐地碱蓬生态位,引起局部地貌变化,直接或间接对原有生态系统造成影响。通过分析近 21 年来黄河三角洲典型盐沼植被时空分布特征,为大范围河口脆弱区的植被信息提取提供理论和方法支撑,并为黄河三角洲保护区的生态修复和管理提供科学依据。

作者分工: 宫兆宁对数据集的开发做了总体设计;张成、胡鉴芳采集和处理了数据、实践模型和算法、进行数据验证;胡鉴芳撰写了数据论文等。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 何彦龙. 中低潮滩盐沼植被分异的形成机制研究[D]. 上海: 华东师范大学, 2014.
- [2] 王雪宏, 栗云召, 孟焕等. 黄河三角洲新生湿地植物群落分布格局[J]. 地理科学, 2015, 35(8): 1021–1026.
- [3] Zhang, C., Gong, Z. N., Qiu, H. C., et al. Mapping typical Salt-marsh species in the Yellow River Delta wetland supported by temporal-spatial-spectral multidimensional features [J]. *Science of the total environment*, 2021, 783: 147061. DOI: 10.1016/j.scitotenv.2021.147061.
- [4] 张贵花, 王瑞燕, 赵庚星等. 基于物候参数和面向对象的滨海生态脆弱区植被遥感提取[J]. 农业工程学报, 2018, 34(4): 209–216.
- [5] 胡鉴芳, 宫兆宁, 张成. 黄河三角洲典型盐沼植被时空分布数据集 (1999–2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.01.06.V1> <https://cstr.science.org.cn/CSTR:20146.11.2022.01.06.V1>.
- [6] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [7] 吴跃, 周忠发, 赵馨等. 基于遥感计算云平台高原山区植被覆盖时空演变研究——以贵州省为例[J]. 中国岩溶, 2020, 39(2): 196–205.
- [8] Breiman, L. Random Forests [J]. *Machine Learning*, 2001, 45(1): 5–32.
- [9] 柯元楚, 史忠奎, 李培军等. 基于 Hyperion 高光谱数据和随机森林方法的岩性分类与分析[J]. 岩石学报, 2018, 34(7): 2181–2188.
- [10] 张晓龙. 现代黄河三角洲滨海湿地环境演变及退化研究[D]. 青岛: 中国海洋大学, 2005.
- [11] 宗敏, 韩广轩, 栗云召等. 基于 MaxEnt 模型的黄河三角洲滨海湿地优势植物群落潜在分布模拟[J]. 应用生态学报, 2017, 28(6): 1833–1842. DOI: 10.13287/j.1001-9332.201706.017.
- [12] 陈柯欣, 丛丕福, 曲丽梅等. 黄河三角洲互花米草、碱蓬种群变化及扩散模拟[J]. 北京师范大学学报(自然科学版), 2021, 57(1): 128–134.
- [13] 王树祥, 韩留生, 杨骥等. 一种改进的融合多指标荒漠化等级分类方法[J]. 测绘通报, 2021(4): 8–12. DOI: 10.13474/j.cnki.11-2246.2021.0102.