

doi: 10.6046/zrzyyg.2022413

引用格式: 冯倩, 张佳华, 邓帆, 等. 基于特征优选和时空融合算法的黄河三角洲湿地类别制图方法研究[J]. 自然资源遥感, 2024, 36(2): 39–49. (Feng Q, Zhang J H, Deng F, et al. A mapping methodology for wetland categories of the Yellow River Delta based on optimal feature selection and spatio-temporal fusion algorithm[J]. Remote Sensing for Natural Resources, 2024, 36(2): 39–49.)

基于特征优选和时空融合算法的黄河三角洲 湿地类别制图方法研究

冯倩¹, 张佳华^{2,3}, 邓帆¹, 吴贞江³, 赵恩灵¹, 郑培鑫¹, 韩杨¹

(1. 长江大学地球科学学院, 武汉 430100; 2. 中国科学院空天信息创新研究院中国科学院数字地球
重点实验室, 北京 100094; 3. 中国科学院大学地球与行星科学学院, 北京 100049)

摘要: 滨海湿地的遥感分类研究对于滨海湿地的保护和规划具有重要意义。为此, 以黄河三角洲作为研究区, 采用 2019 年 3–10 月获取的 8 景 Landsat8 OIL 作为数据源, 使用 GEE (Google Earth Engine) 云平台, 根据影像的不同特征构建了 7 种不同的分类方案; 然后, 使用随机森林分类器对不同特征集合进行分类, 并选择其中分类效果最好的用于绘制黄河三角洲地区的湿地类别图。其中 8, 9 月份数据由于受到云的污染导致质量差, 使用增强型自适应反射率时空融合模型 (enhanced spatial and temporal adaptive reflectance fusion model, ESTARFM) 算法对有云区域进行填补处理。结果表明: ① ESTARFM 时空融合模型生成的预测影像与真实影像波段表现出较好的相关性, 其 R 值均能达到 0.73 以上, 说明重构的影像可以用于本研究; ② 使用随机森林算法对研究区地物类型进行分类, 其中方案 7 通过特征优选, 分类结果总体精度达 92.28%, Kappa 系数达 0.91, 分类结果与湿地实况相吻合, 比常规方案分类精度更高。研究结果有助于了解和掌握该区域湿地不同类型的空间分布特征, 可为区域生态环境的保护和规划提供科学依据。

关键词: Landsat8; 多时相数据; 黄河三角洲湿地; 图像融合; Google Earth Engine; 随机森林

中图法分类号: TP 753 **文献标志码:** A **文章编号:** 2097-034X(2024)02-0039-11

0 引言

滨海湿地是海洋与陆地生态系统的交错过度地带, 是一个高度动态、复杂且脆弱的生态区域^[1]。滨海湿地拥有极强的降解污染的功能; 同时在气候调节、水源涵养、生物多样性和生态链完整性维护等方面都发挥着重要作用^[2]。黄河三角洲国家级自然保护区是世界上温暖带保存最完整、最年轻的湿地生态系统。近年来, 随着城镇化的加剧、区域经济发展、农田大面积开垦, 黄河三角洲湿地已成为了全球受威胁最严重的生态系统之一^[3–4]。因此, 了解和掌握该区域湿地不同类型的空间分布特征, 可为区域生态环境的保护和规划提供科学依据。

遥感技术快速发展, 被广泛用于湿地资源调查、识别与监测等研究中, 基于遥感的湿地信息提取方法主要分为人工目视解译与计算机自动分类^[5]。

目视解译的方法对解译人员的专业技术水平要求高且工作量巨大。计算机自动分类的技术对遥感分类提供了更高效、更精确的影响分类的途径。目前, 已有很多机器学习算法被广泛应用于遥感影像分类中, 例如最大似然法、神经网络、支持向量机, 以及深度学习等^[6]。其中, 随机森林是使用广泛的机器学习算法之一, 它可以进行分类与回归分析, 适合高维复杂数据集, 而且可以判别特征变量的重要性, 给出特征变量间的依赖关系, 有利于对特征变量定量分析。刘春亭等^[7]基于多源 Sentinel-1/2 影像和随机森林对不透水层提取, 精度可达到 93.37%。

但由于滨海湿地地物类型多样, 光谱异质性高, 易出现“异物同谱”、“同物异谱”的现象, 研究发现使用多时相数据可以有效解决这种现象^[8]。但光学遥感影像易受观测条件的影响, 导致影像质量差或者数据的缺失, 数据融合可以解决相关问题^[9–11], Gao 等^[12]提出增强型自适应反射率时空融合模型

收稿日期: 2022-10-24; 修订日期: 2023-03-15

基金项目: 山东省自然科学基金项目“遥感信息与模型系统模拟黄河三角洲滨海湿地动态及未来情景预测”(编号: ZR2020QF067)资助。

第一作者: 冯倩 (1996–), 女, 硕士, 主要研究方向为湿地遥感。Email: 202071360@yangtzeu.edu.cn。

通信作者: 张佳华 (1966–), 博士, 教授, 主要从事生态环境遥感等研究。Email: zhangjh@radi.ac.cn。

(ESTARFM)算法,在移动窗口的基础上,综合考虑目标像元与邻近像元的空间距离、光谱差异和时间差异,大大提高了融合精度,为填补空缺时相的数据提供了方法。Zhu 等^[13]在 STARFM 算法的基础上,提出了 ESTARFM 模型,在影像重构时考虑了地物反射率变化的时间趋势,且利用空间和光谱相似性来构建中心像元,有效提高了在异质性较强区域的融合精度。有效地弥补数据缺失的问题。

获取时序数据集可提取丰富的时序特征用于分类。但在分类前,遥感数据的多特征变量常需要进行特征选择以提升分类精度,同时可以有效地解决数据冗余问题,提高分类器的计算速度^[14-15]。张磊等^[16]基于 Sentinel-2 数据利用不同的特征组合对黄河三角洲的湿地信息进行提取,并用随机森林算法进行分类,最优特征集合的总体精度达 90.93%。杨迎港^[17]使用优化过的随机森林算法对河南省新郑市的部分区域进行地物分类,分类的总体精度达 93.44%。朱琦^[18]依托 Google Earth Engine (GEE) 平台使用多时相 Landsat8 数据对海南省尖峰岭地区热带天然林进行分类,分类精度达 91.19%。

在本地存储和处理分析数据的传统方式较难满足本研究的需要。GEE 是一个地理空间分析的云平台,它包含数据存储能力和高性能的计算能力^[19]。用户利用 GEE 不仅可以获取大量的遥感数据资源,还可以依托该平台对影像的各类地物类型快速提取分析。王渊等^[20]基于 GEE 云平台,批量

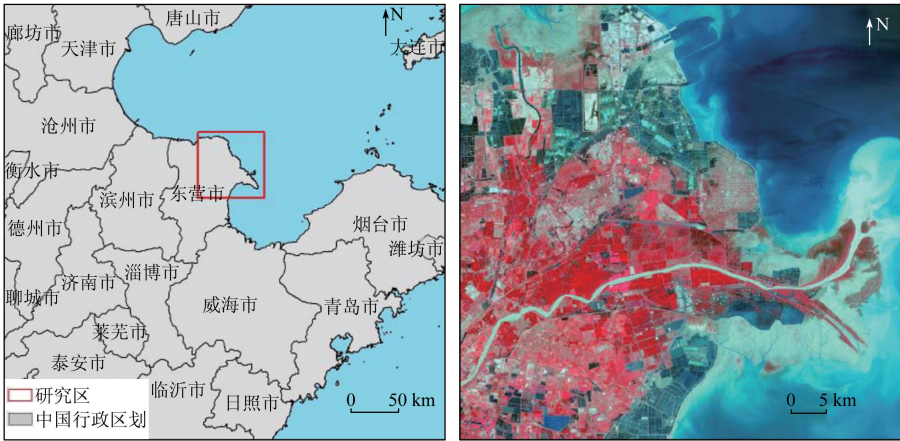
处理了大湾区 1988—2018 年近 30 a,共 3 530 景 Landsat 遥感影像,评价了对大湾区的生态质量的时空变化,极大地提高了影像处理的效率。在前人研究基础上,本文拟通过 GEE 平台,构建适合的分类方案,对黄河三角洲研究区进行快速分类,分析各个地物的分类情况,为研究区的保护和发展提供参考。

基于以上考虑,本研究的主要研究目的是:构建多种方案,找出最优的一组用于分类,以实现黄河三角洲湿地类别的高精度分类。

1 研究区概况与数据源

1.1 研究区概况

黄河三角洲位于山东省东北部,以垦利县宁海为轴点,北起套尔河口,南至淄脉河口,向东撒开的扇状地形,海拔高程低于 15 m;北部和东部分别与渤海湾、莱州湾相邻,属于暖温带半湿润大陆季风性气候区,季节分明;地势多为洼地和微斜平地,由于临海土壤盐碱化程度高,土壤和植物的种群组成较为简单,多为芦苇、怪柳及盐地碱蓬^[21]。主要以河流、水库、坑塘,滩涂等常年积水湿地为主,占总湿地面积的 63%;季节性积水湿地(芦苇沼泽、疏林沼泽、灌丛沼泽、湿草甸和水稻田等)占湿地总面积的 37%。选取 E 118°31'~119°20',N 37°31'~38°13'之间的区域作为研究区(图 1)。



(a) 黄河三角洲地理位置 (b) 研究区真彩色合成影像 (B4(R), B3(G), B2(B))

图 1 黄河三角洲地理位置及卫星影像

Fig. 1 Location and Landsat8 image of Yellow River Delta

1.2 数据源及预处理

1.2.1 Landsat 数据

本研究选取黄河三角洲 2019 年 3—10 月份获取的 8 景 Landsat8 影像为数据源(表 1),数据的下载地址为 USGS(<https://www.usgs.gov>),下载的数

据属于 Level 1T 级别产品,已完成地面控制点校正和系统辐射校正。进一步对影像进行辐射定标、大气纠正、几何校正和裁剪的预处理操作,消除大气因素对地物反射率的影响。

表 1 Landsat8 OIL 影像信息

Tab. 1 Image information of Landsat8 OIL

影像编号	获取时间	云量/%	波段数 ^①	影像质量
1	2019-03-14	<1%	6	良好
2	2019-04-15	<1%	6	良好
3	2019-05-01	<1%	6	良好
4	2019-06-02	<1%	6	良好
5	2019-07-20	<1%	6	良好
6	2019-08-21	约 90%	6	差
7	2019-09-22	约 95%	6	差
8	2019-10-24	<1%	6	良好

①: 6 个波段分别是: 蓝、绿、红、近红外、SWIR1 和 SWIR2。

表 2 Landsat 与 MODIS 数据波段信息

Tab. 2 Spectral bands and resolutions of Landsat and MODIS

Landsat 波段	波段宽度/nm	空间分辨率/m	MOD09A1 波段	波段宽度/nm	空间分辨率/m
B2 蓝	450~510	30	B3	459~479	500
B3 绿	530~590	30	B4	545~565	500
B4 红	640~670	30	B1	620~670	500
B5 近红外	850~880	30	B2	841~876	500
B6 SWIR1	1 570~1 650	30	B6	1 628~1 652	500
B7 SWIR2	2 110~2 290	30	B7	2 105~2 155	500

1.2.3 样本点数据

样本数据由 2 部分构成,分别是国家地球系统科学数据中心(<http://www.geodata.cn>)的部分样本点以及通过 Google Earth 软件目视解译一部分湿地类型,这 2 部分数据构成分类样本和验证样本分别用于分类器建立和精度验证。表 3 为地物类别的像元数量,验证集的数量是样本数量的 24%左右。

表 3 地物样本点信息

Tab. 3 Figure sample point information

地物类别	浅海水域	泥质海滩	河流	草本沼泽	灌木沼泽	水库/坑塘	水田	盐田	建筑物	养殖池	农田
分类样本数量	33 927	1 086	2 275	471	595	3 652	2 974	249	1 822	844	793
验证样本数量	8 238	285	570	128	133	929	762	67	484	222	193

1.3 湿地分类体系

结合黄河三角洲研究区湿地类型分布的具体情况,参考湿地公约、《全国湿地资源调查与监测技术规程》以及相关文献资料^[22],制订湿地分类方案如表 4 所示。该区域人类活动频繁,一定程度上影响湿地信息的提取,因此对该区域的非湿地类型也进行了提取,主要包括建筑用地与农田。

表 4 黄河三角洲湿地分类方案

Tab. 4 Categories plan of wetlands in the Yellow River Delta

一级分类	二级分类	说明
近海与海岸湿地	浅海水域	低潮时水深<6 m 的浅海水域,包括海湾海峡
	泥质海滩	由淤泥质组成的植被覆盖度<30%的淤泥质海滩
河流湿地	河流	常年有水或间歇性有水流动的河流,包括河床部分
沼泽湿地	草本沼泽	以草本植物为主的永久或季节性咸淡水沼泽,喜湿多年生草本和禾本科植物占优势,研究区典型植被如芦苇、香蒲、盐地碱蓬、互花米草等
	灌木沼泽	以灌木为主的永久性或季节性沼泽,如怪柳等
人工湿地	水库/坑塘	包括水库、坑塘、养殖池以及城市景观和娱乐水面等人工建造的静止水体
	水田	用于种植水稻田、水生作物的耕地,如水稻田、藕池
	盐田	为获取盐业资源而修建的晒盐场所或盐池

2 研究方法

本文提取湿地分类的流程图如图 2 所示。主要有 3 个步骤：①对 Landsat 数据和 MOD09A1 数据做

预处理,经预处理的数据作为后续的基础数据;②对于部分受云污染的影像,使用 ESTARFM 算法进行填补处理;③构建不同分类方案,使用随机森林算法提取出最优的一组用于分类。

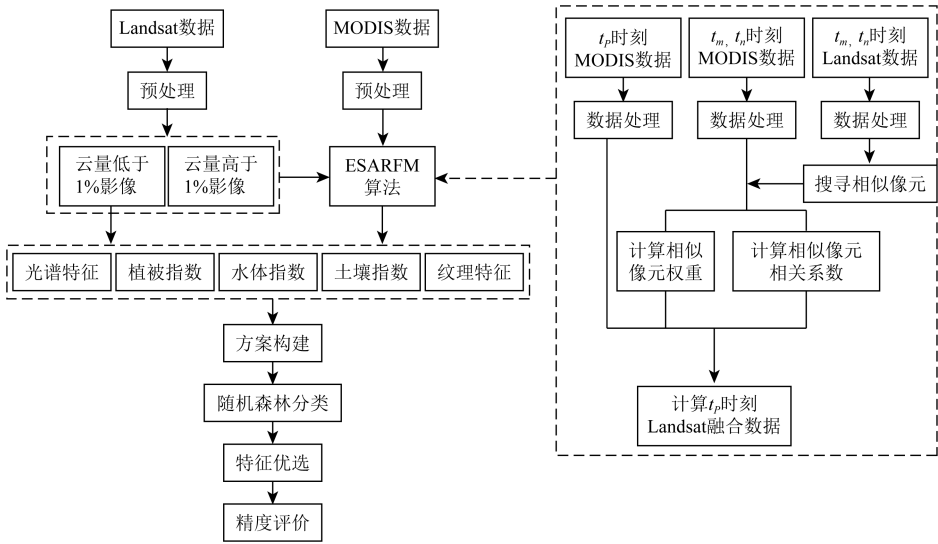


图 2 实验流程图

Fig. 2 The flowchart of the experiment

2.1 ESTARFM 时空融合模型

给定一个区域,假设在同一日期获得的遥感数据 MODIS 和 Landsat 数据在经过配准之后,它们之

间的反射率只由系统偏差引起的,则在同一区域时间 t_m 和 t_n ,MODIS 和 Landsat 数据的反射率之间的线性关系为:

$$F(x,y,t_n,B) = a \times (C(x,y,t_n,B) - C(x,y,t_m,B)) + F(x,y,t_m,B) \quad (1)$$

式中: F 为 Landsat 的反射率; C 为 MODIS 的反射率; (x,y) 表示像元的位置; t_k 为影像的采集时间; B 为影像的波段; a 和 b 为线性方程的系数,即高分辨率像元反射率和低分辨率像元反射率之间的转换系数。

只有一种土地类型的纯净像元十分稀少,大多数低分辨率影像像元都为混合像元,那么, Landsat 和 MODIS 数据之间的反射率关系可能不存在如式(1)中描述的情况。因此,引入转换系数 $v(x,y)$ 解决这一问题。它可由 Landsat 和 MODIS 影像中像元的相同端元的反射率变化通过线性回归获得,公式为:

$$F(x,y,t_n,B) = v(x,y) \times (C(x,y,t_n,B) - C(x,y,t_m,B)) + F(x,y,t_m,B) \quad (2)$$

同时,式(2)仅利用单个像元来预测像元反射率精度是不够准确的。充分利用相邻像元的信息可以获得更高的融合精度。因此,使用移动窗口方法,在移动窗口内搜寻相似像元,并将相似像元信息融入预

测像元反射率计算中,有助于提高精度,其中,相似像元是指与中心像元地物类型一致的邻域像元,此时,位于窗口中心的高分辨率像元反射率的计算公式为:

$$F(x_w,y_w,t_n,B) = \sum_{i=1}^N w_i \times v_i \times (C(x_i,y_i,t_n,B) - C(x_i,y_i,t_m,B)) + F(x_w,y_w,t_m,B) \quad (3)$$

式中: W 为移动窗口的大小,由研究区地表异质性决定; N 为地物类型的种类; (x_w,y_w) 为中心像元位置; (x_i,y_i) 为第 i 个相似像元的位置; w_i 为第 i 个相似像元的权重,它表示第 i 个相似像元对中心像元的反射率的贡献,主要由 Landsat 和 MODIS 影像在相似像元位置上的光谱相似性以及相似像元与

中心像元的距离所确定; v_i 为第 i 个相似像元的转换系数。

之后,根据在时间 $t_k(k=m,n)$ 与预测时间 t_p 之间的 MODIS 影像的反射率的变化程度来计算时间权重 T_k ,计算公式为:

$$T_k = \frac{1}{\left| \sum_{i=1}^N C(x_i, y_i, t_k, B) - \sum_{i=1}^N C(x_i, y_i, t_p, B) \right|} \times \frac{1}{\sum_{k=m,n} \left(\frac{1}{\left| \sum_{i=1}^N C(x_i, y_i, t_k, B) - \sum_{i=1}^N C(x_i, y_i, t_p, B) \right|} \right)} \quad (k = m, n)。$$

(4)

由此,预测 t_p 时刻的影像的最终计算公式如下:

$$F(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_p, B) = T_m \times F_m(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_p, B) + T_n \times F_n(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_p, B)。$$

(5)

式中: T_m 和 T_n 分别为时刻所计算的融合影像的时间权重; F_m 和 F_n 分别为时刻的通过式(3)计算得到的融合影像的中心像元值。

表 5 Landsat8 特征集描述
Tab. 5 Description of the feature set from Landsat8

特征类别	特征名称	特征描述/公式
光谱特征	波段 (band)	蓝,绿,红,近红,中红 1,中红 2
植被/水体指数	归一化植被指数 (normalized difference vegetation index, NDVI)	$\frac{R_{nir} - R_{red}}{R_{nir} + R_{red}}$
	比值植被指数 (ratio vegetation index, RVI)	$\frac{R_{nir}}{R_{red}}$
	差值植被指数 (differential vegetation index, DVI)	$R_{nir} - R_{red}$
	归一化水体指数 (normalized difference water index, NDWI)	$\frac{R_{green} - R_{nir}}{R_{green} + R_{nir}}$
	盐分指数 2 (salinity index 2, SI2)	$\sqrt{R_{green}^2 + R_{red}^2 + R_{nir}^2}$
盐分/土壤指数	盐分指数 3 (salinity index 3, SI3)	$\sqrt{R_{green}^2 + R_{red}^2}$
	盐分指数 (salinity index, SI-T)	$(\frac{R_{red}}{R_{nir}}) \times 100$
	优化型土壤调节植被指数 (soil adjusted vegetation index, SAVI)	$\frac{R_{nir} - R_{red}}{R_{nir} + R_{red} + 0.6}$
	土壤亮度指数 (soil brightness index, SBI)	$\sqrt{R_{red}^2 + R_{nir}^2}$
K-T 变换	亮度 (brightness)	$0.302\ 9R_{blue} + 0.278\ 6R_{green} + 0.473\ 3R_{red} + 0.559\ 9R_{nir} + 0.508\ 0R_{swirl} + 0.187\ 2R_{swir2}$
	绿度 (greenness)	$-0.294\ 1R_{blue} - 0.243\ 0R_{green} + 0.542\ 4R_{red} + 0.727\ 6R_{nir} - 0.717\ 0R_{swirl} - 0.168\ 0R_{swir2}$
	湿度 (wetness)	$0.151\ 1R_{blue} + 0.197\ 3R_{green} + 0.328\ 3R_{red} + 0.340\ 7R_{nir} - 0.711\ 7R_{swirl} - 0.455\ 9R_{swir2}$
纹理特征	方差 (GLGM_Variance)	$\sum_i \sum_j p(i, j) \times (i - Mean)^2$
	对比度 (GLGM_Contrast)	$\sum_i \sum_j p(i, j) \times (i - j)^2$
	熵 (GLGM_Entropy)	$\sum_i \sum_j p(i, j) \times \ln p(i, j)$
	相关性 (GLGM_Correlation)	$\sum_i \sum_j \frac{(i - Mean) \times (j - Mean) p(i, j)^2}{Variance}$
	二阶矩 (GLGM_Second Moment)	$\sum_i \sum_j p(i, j)^2$

实验选择 8 景影像的 6 个波段的反射率作为光谱特征;对影像进行预处理后提取植被指数、水体指数、盐分指数与土壤指数;同时,影像纹理特征计

2.2 特征提取与分类

2.2.1 特征选择

本研究根据研究区地物特征选取遥感影像的多时相光谱特征、植被指数、水体指数、土壤指数、盐分指数、K-T 变化,纹理特征构建特征集。表 5 详细描述了各种特征及其表达式。

算采用最常用的灰度共生矩阵法 (gray level co-occurrence matrices, GLCM)^[23]。对影像进行主成分分析发现,影像第一主成分包含了 90% 的影像,为减

少数据冗余,以影像第一主成分作为纹理分析的输入影像,基于 GLCM 计算相关统计量,本研究选取角二阶矩、相关性、对比度、熵和方差 5 个冗余度较小的二阶统计量。

2.2.2 分类方案

本研究共设计以下 7 种方案进行对比研究(表 6)。设置不同方案主要有以下 2 方面的目的:①研究不同特征变量对湿地信息提取的影响,确定不同特征变量的重要性程度;②通过方案之间的对比,探索提高湿地分类精度的最佳方法。

表 6 实验方案信息

Tab. 6 The information of experimental Programs

方案	特征组合
方案 1	多时相光谱特征
方案 2	多时相光谱特征+多时相植被指数/水体指数
方案 3	多时相光谱特征+多时相土壤指数/盐分指数
方案 4	多时相光谱特征+多时相 K-T 变化特征
方案 5	多时相光谱特征+多时相纹理特征
方案 6	多时相光谱特征+多时相植被/水体指数+多时相盐分/土壤指数+多时相 K-T 变化特征+多时相纹理特征
方案 7	特征优选组合

2.2.3 分类方法与精度验证

随机森林算法是 Breiman^[24]等提出的一种机器学习算法,它不仅可以实现分类与分析,在特征选择方面也发挥着重要的作用。以本研究为例,大约 76%的样本作为训练样本集,约 24%的样本数据为

验证样本,这部分数据称为袋外数据(out-of-bag, OOB)。通过 OOB 数据产生的袋外数据误差(out-of-bag-error)可以计算特征变量的重要性(variable important, VI),从而进行特征优选。本研究将影像的总特征个数的算术平方根作为不同方案的特征个数。在总特征数确定的基础上,通过大量的实验发现,决策树数量大于 150 时,所有方案的 OOB 误差逐渐收敛并趋于稳定。因此,本文生成决策树的数量为 150。

本研究精度验证采用混淆矩阵(confusion matrix)来比较分类结果进行精度评估。本文主要选取总体精度、Kappa 系数、生产者精度和用户精度作为评价指标对各个方案进行评价。

3 结论

3.1 影像融合结果

图 3 为 ESTARFM 时空融合模型的预测影像与真实影像对比图。本研究采用定性与定量结合的方法对影像融合效果评价。通过目视解译发现基本上保留了原有的光谱信息,各类地物类型都能很好表征出来,与原始影像差异较小。其中 8 月份的目视效果比九月份的较好,九月份预测影像的红波段受到的影响较大,水田在假彩色波段组合显示中为浅红色,色调低于原始影像,亮度基本保持一致。

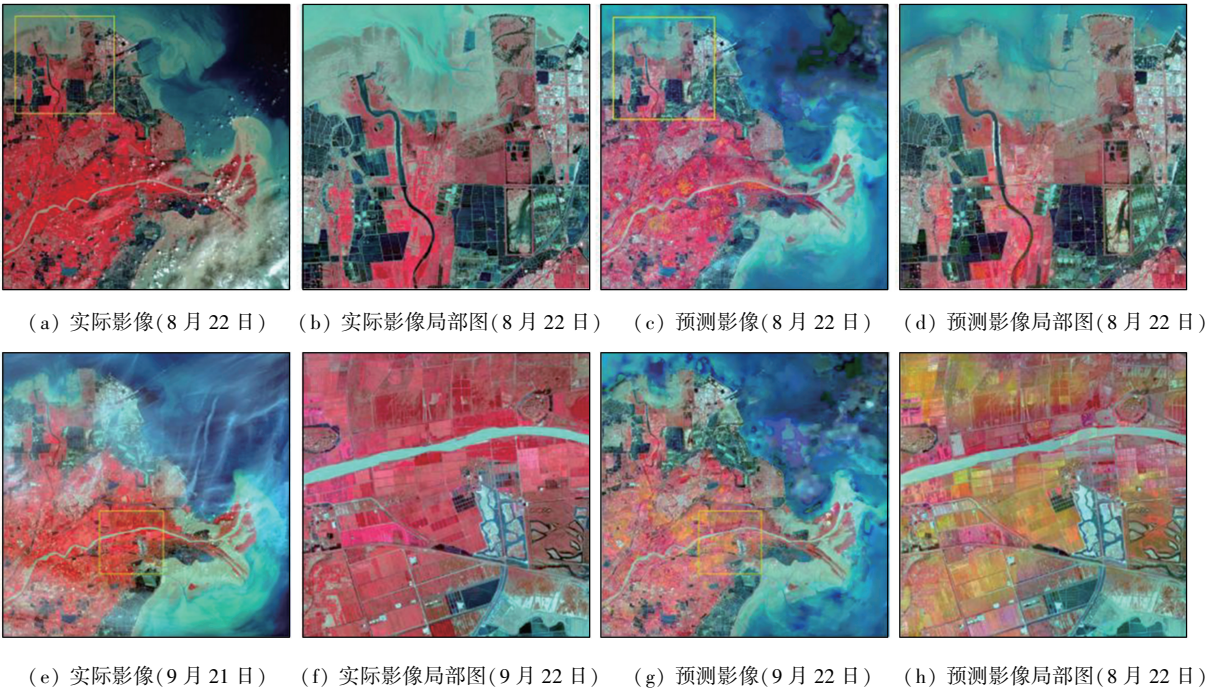


图 3 Landsat8 原始影像与预测影像及局部图

Fig. 3 Original and predicted images and local maps of Landsat8

进一步验证预测影像与原始影像的相关性,选取研究区当天原始影像不受云污染的部分区域与该

区域的预测影像区域进行相关性分析对比,对比其各波段以及部分相关指数的相关性, R 表示融合后的预测影像与真实影像反射率数据之间的相关性。由表 7 可知,2 个区域的 R 值均在 0.73 以上,这表明 ESTARFM 模型获得的预测影像与真实影像相关

性较好;8 月份的各个波段和指数的 R 值均高于 9 月份,其中红光波段的 R 值为 0.73,山东省秋季作物收割时间为 8—11 月初,这或与秋季作物收割有一定的关系。目视解译与相关性分析表明,预测影像具有实用性。

表 7 原始 Landsat8 OLI 影像与 ESTARFM 融合后结果相关性分析

Tab.7 Comparison of original Landsat8 OLI image with ESTARFM

指数	日期	蓝	绿	红	近红	SWIR1	SWIR2	NDVI	NDWI	SAVI	SI2	SI3
相关系数 R	8 月 21 日	0.87	0.85	0.79	0.88	0.87	0.81	0.85	0.90	0.78	0.85	8.85
	9 月 22 日	0.77	0.79	0.73	0.81	0.79	0.73	0.78	0.80	0.84	0.79	0.79

3.2 方案 7 特征确定

将方案 1—6 作为对比实验,探究不同特征变量对湿地信息提取精度的影响。以方案 6 作为基础,采用随机森林算法对 184 个特征变量进行重要性评价,选择分类精度最高的特征子集构成方案 7。为了更加直观清晰地呈现高重要性的特征变量,选取前 54 个重要性得分较高的特征生成重要性分布图(图 4),结合特征变量个数,根据特征数量重要性得分顺序,每 5 个特征变量为一组的加入,分析特征变量数与分类精度和 Kappa 系数的关系(图 5)。图

中,WL-COR-3 为 3 月份的纹理特征中的相关性指数(GLGM_Correlation);GK-T-3 为 3 月份的 GK-T 变化的绿度指数;NDVI-3 为 3 月份的归一化水体指数;SWIR1-3 为 3 月份中红 1 波段,以此类推。

根据图 4,5 可知:不同的特征变量的重要性得分差异较大,特征变量的个数影响分类精度。其中 SI3-7 特征的重要性得分最高,高达 6.92,BK-T-7 特征重要性得分最低(0.15),几乎不产生影响。图 5 表明,开始时,随着特征变量的增加,分类精度呈现急速上升的趋势,当加入 30 个特征,分类精度也迅速达到 84.26%,Kappa 系数达到了 0.82,这是因为前期特征变量的重要性评分高,特征之间相关性小和冗余特征少,从而可以迅速提高分类结果的精度;之后,特征变量数从 30 个增加至 54 个时,分类精度仍呈现上升的趋势,但增长速度不如之前,分类精度从 84.26%逐步达到 92.28%,Kappa 系数也达到了 0.91;后来,特征变量数从 54 个增加至 148 个时,分类精度逐渐呈现波动的下降的趋势。因为后期冗余特征和不相关特征增加,降低了分类器的性能,导致分类精度降低。同时,图 5 也表明,当特征变量的个数达到 54 个时,分类精度和 Kappa 分别为 92.28%和 0.91,均达到最大值。因此将重要性评分排在前 54 个特征作为方案 7 的最终结果参与精度评价。

3.3 特征优选分析

进一步分析每个月份特征变量可知:不同月份特征变量提供不同程度的贡献(表 8)。特征变量最多的是 3 月份,特征变量数有 10 个,该月份包含所有类别的特征变量;八月份的特征变量数有 7 个,主要是植被指数与土壤、盐分指数和纹理特征,九月份的特征值只有一个(SI2),这与融合后的预测影像的质量有一定的关系,9 月份的预测影像的质量不如 8 月份。该月份对于湿地的提取优选特征变量数也对应的少。

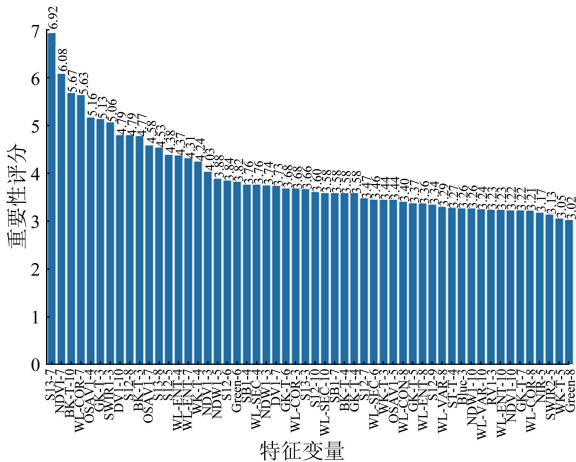


图 4 特征重要性分布

Fig.4 The distribution of characteristic importance

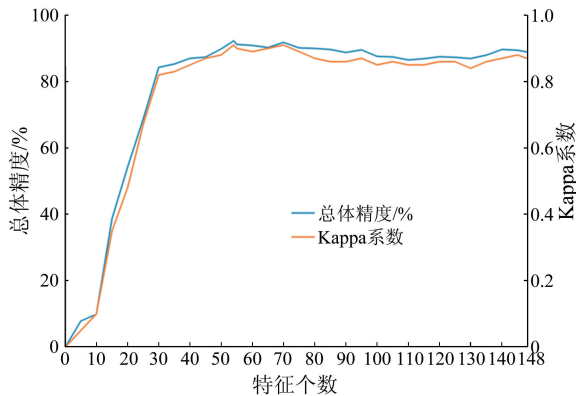


图 5 特征变量个数与分类精度和 Kappa 系数

Fig.5 The relation between the number of characteristic variables and the classification accuracy and the Kappa coefficient

表 8 优选特征分布表

Tab. 8 Distribution list of optimal features

月份	优选特征	特征数量
3 月	Blue-3,SWIR1-3,NDVI-3,NDWI-3,RVI-3,GK-T-3,BK-T-3,WK-T-3,SI3-3,WL-COR-3	10
4 月	WK-T-4,BK-T-4,GK-T-4,OSAVI-4,SBI-4,ST-T-4,WL-ENT-4,WL-SEC-4	8
5 月	NIR-5,SWIR2-5,NDWI-5,SI2-5,OSAVI-5,GK-T-5,WK-T-5	7
6 月	Green-6,GK-T-6,WL-SEC-6	3
7 月	NDVI-7,DVI-7,SI3-7,OSAVI-7,SBI-7,SI2-7,GK-T-7,WL-COR-7,WL-ENT-7	9
8 月	Green-8,SI2-8,SI3-8,WL-CON-8,WL-ENT-8,WL-VAR-8,WL-COR-8	7
9 月	SI2-9	1
10 月	DVI-10,NDWI-10,NDVI-10,SI2-10,BK-T-10,WL-SEC-10,WL-VAR-10,WL-ENT-10	8

3.4 提取分类结果及精度评价

基于多时相 Landsat 遥感数据,提取多时相光谱特征、植被指数和水体指数、土壤指数、K-T 变化和纹理特征确定 7 种不同的实验方案,基于 GEE 云平台,使用随机森林算法完成黄河三角洲湿地类别信息提取的研究。7 种方案的提取结果如图 6 所示。通过目视解译对比原始影像可以看出:方案 1—3 以及方案 6 的湿地信息提取不理想,不同类别之间

存在较多错分的情况。例如,方案 2 中,草本沼泽与建筑用地被错分为灌木沼泽;方案 3 中部分草本沼泽被错分为建筑用地,部分河流被错分为灌木沼泽;所有方案在海陆交界地带都存在河流和浅海水域 2 种类别不同程度上的混分现象。根据目视解译对比 7 个方案的效果来看,方案 4、方案 5、方案 7 的提取结果较好。分类精度统计见表 9,表中 PA,UA 分别代表各类别的生产者精度和用户精度。

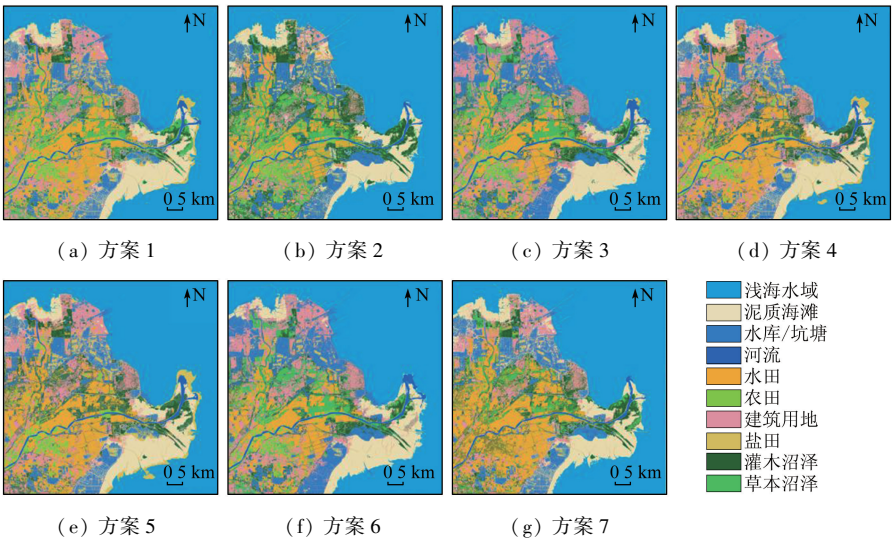


图 6 不同方案分类结果

Fig. 6 Classification results of different plans

表 9 分类精度统计

Tab. 9 The statistics of classification accuracy

类别	方案 1		方案 2		方案 3		方案 4		方案 5		方案 6		方案 7	
	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%
浅海水域	95.94	97.35	99.15	96.41	98.31	94.13	99.54	98.35	99.36	97.13	97.50	96.86	99.12	94.88
泥质海滩	81.07	84.69	88.02	86.08	85.79	86.31	95.48	88.61	93.44	87.20	94.69	85.55	98.60	96.98
水库坑塘	80.85	70.03	75.16	81.01	79.07	80.73	83.16	88.91	75.16	85.68	77.06	90.21	83.67	90.33
河流	86.37	89.07	86.92	96.18	87.46	93.93	93.80	96.70	91.26	89.52	88.52	96.41	89.36	99.86
水田	91.89	89.52	93.30	89.51	94.28	85.79	94.84	90.83	90.03	94.42	90.87	92.17	96.86	90.02
农田	93.79	75.57	89.40	97.91	82.72	99.72	90.14	95.31	91.76	95.78	93.91	81.30	84.90	96.12
建筑用地	62.53	79.68	70.69	74.55	70.45	69.63	81.10	81.30	85.70	66.65	76.24	69.00	86.37	86.30
盐田	76.51	84.65	89.06	82.07	87.43	88.62	90.95	90.76	82.22	85.51	93.53	85.92	96.00	97.77
草本沼泽	54.71	73.11	84.28	57.51	74.05	59.39	74.83	77.93	66.87	70.95	83.54	72.93	87.74	86.33
灌木沼泽	64.30	81.72	70.51	82.77	63.36	86.17	70.76	89.77	74.43	84.23	53.84	83.50	71.49	70.85
总精度/%	85.11		89.05		87.39		91.90		90.01		88.89		92.28	
Kappa 系数	0.83		0.86		0.85		0.91		0.88		0.87		0.91	

由表 9 可知,以多时相光谱特征为基础的方案 1 总体精度为 85.11%,Kappa 系数为 0.83。在多时相光谱特征的基础之上加入不同的特征变量对湿地信息提取会产生不同的影响。分别加入植被指数和水体指数、土壤指数和盐分指数、K-T 变化和纹理特征之后,方案 2—5 分类精度略有上升,总体精度分别达到 89.05%,87.39%,91.90%和 90.01%,Kappa 系数分别为 0.86,0.85,0.91 和 0.88。所有特征均加入分类精度在方案 1 的基础上提高的并不高,其总体精度为 88.89%,Kappa 系数为 0.87,总体精度不如方案 1 和方案 3。这说明:基于多光谱特征分别加入本研究的其他特征均会提高湿地信息提取的精度,所有特征参与分类时,会由于信息冗余反而在一定程度上会导致分类精度降低。基于随机森林算法进行特征筛选之后的方案 7 总体精度有一定的提高,相较于前 6 个方案中提取效果最好的方案 4 而言,总体精度提高了 0.89%,Kappa 系数并未增长。但方案 7 的优选的特征进行分类时,保留了对分类有用的特征数据,剔除掉了冗余信息,可以避免冗余的特征变量对湿地提取的干扰,使所有类型的生产者精度和用户精度都有一定程度的改善。

本研究得到结论如下:以多时相光谱数据为基础,单独引入不同类别特征对湿地信息提取的精度影响不同,但都对湿地信息提取的精度产生积极的影响。相较而言,加入纹理特征对分类精度贡献最高;对比方案 1—7 可知,采用随机森林算法进行特征优选之后的分类精度有了明显的改善,而且参与分类的特征数达到 54 时,分类精度达到最高。以上分析说明,本研究基于随机森林算法的特征选择方法,能够保留地物重要的特征信息,在降低数据维度的同时,分类精度仍能保持较高水平,从而缩短模型运算时间,提高分类效率。

4 讨论

本研究主要优势在于研究不同月份的特征变量对湿地信息提取的贡献,达到时相与特征统筹兼顾的效果。ESTARFM 算法在填补缺失数据方面,起到了至关重要的作用,对 8,9 月份缺失数据进行补充,进而 8,9 月份的特征变量对分类起到了重要的作用。

本研究仍有不足之处,后续主要从以下 3 个方面进行研究补充:①该研究是基于像素尺度的,在草本沼泽、灌木沼泽分布区一定程度上存在“椒盐现象”,后续可结合面向对象的方法,对两者进行比较研究,以便进一步提高分类精度;②重点探究了

特征变量对湿地分类所带来的影响,但并未从不同的特征类别深入研究,后续的工作重点将不同类别的特征对湿地提取的贡献作为重点研究;③使用 ESTARFM 算法可以填补数据的缺失,但缺失数据的生成受前后 2 个时刻影像的质量、时间间隔以及算法参数设置等的影响,后续工作需要重点研究预测数据质量问题的因素。

参考文献 (References):

[1] 张晓龙,李培英,李 萍,等. 中国滨海湿地研究现状与展望[J]. 海洋科学进展,2005,23(1): 87-95.
Zhang X L, Li P Y, Li P, et al. Present conditions and prospects of Study on coastal wetlands in China[J]. Advances in Marine Science, 2005, 23(1): 87-95.

[2] 刘润红,梁士楚,赵红艳,等. 中国滨海湿地遥感研究进展[J]. 遥感技术与应用,2017,32(6): 998-1011.
Liu R H, Liang S C, Zhao H Y, et al. Progress of Chinese coastal wetland based on remote sensing[J]. Remote Sensing Technology and Application, 2017, 32(6): 998-1011.

[3] 朱纹君,韩 美,魏丹妮,等. 黄河三角洲人地关系协调度时空演变及其驱动机制[J]. 水土保持研究,2023,30(1): 388-396.
Zhu W J, Han M, Wei D N, et al. Spatiotemporal evolution and driving mechanism of human - environment coordination in the Yellow River Delta[J]. Research of Soil and Water Conservation, 2023, 30(1): 388-396.

[4] 刘海红,刘胤序,张春华,等. 1991—2016 年黄河三角洲湿地变化的遥感监测[J]. 地球与环境,2018,46(6): 590-598.
Liu H H, Liu Y X, Zhang C H, et al. Remote sensing monitoring of wetland changes in the Yellow River Delta during 1991-2016[J]. Earth and Environment, 2018, 46(6): 590-598.

[5] 钟燕飞,吴 浩,刘寅贺. 湿地遥感制图研究现状与展望[J]. 中国科学基金,2022,36(3): 420-431.
Zhong Y F, Wu H, Liu Y H. Research status and prospects of remote sensing for wetland mapping[J]. Bulletin of National Natural Science Foundation of China, 2022, 36(3): 420-431.

[6] 祁 欣. 基于机器学习的遥感数据处理方法综述[J]. 品牌与标准化,2022(s1): 148-150.
Qi Y. Survey of Remote sensing data processing methods based on machine learning[J]. Brand & Standardization, 2022(s1): 148-150.

[7] 刘春亭,冯权洸,金鼎坚,等. 随机森林协同 Sentinel-1/2 的东营市不透水层信息提取[J]. 自然资源遥感,2021,33(3): 253-261. doi:10.6046/zrzyyg.2020310.
Liu C T, Feng Q L, Jin D J, et al. Application of random forest and Sentinel-1/2 in the information extraction of impervious layers in Dongying City[J]. Remote Sensing for Natural Resources, 2021, 33(3): 253-261. doi:10.6046/zrzyyg.2020310.

[8] 黄 佩,普军伟,赵巧巧,等. 植被遥感信息提取方法研究进展及发展趋势[J]. 自然资源遥感,2022,34(2): 10-19. doi:10.6046/zrzyyg.2021137.
Huang P, Pu J W, Zhao Q Q, et al. Research progress and development trend of remote sensing information extraction methods of vegetation[J]. Remote Sensing for Natural Resources, 2022, 34(2):

10–19. doi:10. 6046/zrzyyg. 2021137.

[9] Wu Z J,Zhang J H,Deng F, et al. Fusion of GF and MODIS data for regional – scale grassland community classification with EVI2 time–series and phenological features[J]. Remote Sensing, 2021, 13(5): 835.

[10] 史飞飞,高小红,肖建设,等. 基于集成学习和多时相遥感影像的枸杞种植区分类[J]. 自然资源遥感, 2022, 34(1): 115–126. doi:10. 6046/zrzyyg. 2021064.

Shi F F,Gao X H,Xiao J S, et al. Classification of wolfberry planting areas based on ensemble learning and multi–temporal remote sensing images[J]. Remote Sensing for Natural Resources, 2022, 34(1): 115–126. doi:10. 6046/zrzyyg. 2021064.

[11] 程筱茜,洪友堂,陈劲松,等. 基于 ESTARFM 的内陆湖泊遥感动态变化监测[J]. 国土资源遥感, 2020, 32(3): 183–190. doi:10. 6064/zrzyyg. 2020. 03. 24.

Cheng X Q,Hong Y T,Chen J S, et al. A study of the long–term remote sensing dynamic monitoring of inland based on ESTARFM [J]. Remote Sensing for Land and Resources, 2020, 32(3): 183–190. doi:10. 6064/zrzyyg. 2020. 03. 24.

[12] Gao F,Masek J,Schwaller M, et al. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance[J]. IEEE Transactions on Geoscience and Remote sensing, 2006, 44(8): 2207–2218.

[13] Zhu X L,Chen J,Gao F, et al. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions [J]. Remote Sensing of Environment, 2010, 114(11): 2610–2623.

[14] 高 瑞,王志勇,周晓东,等. 利用多时相遥感监测与分析黄河三角洲湿地变化动态[J]. 测绘通报, 2021(4): 22–27.

Gao R,Wang Z Y,Zhou X D, et al. Monitoring and analysis of wetland change dynamics in the Yellow River Delta using multi–temporal remote sensing[J]. Bulletin of Surveying and Mapping, 2021 (4): 22–27.

[15] 程丽娜,钟才荣,李晓燕,等. Sentinel–2 密集时间序列数据和 Google Earth Engine 的潮间带湿地快速自动分类[J]. 遥感学报, 2022, 26(2): 348–357.

Cheng L N,Zhong C R,Li Xiao Yan, et al. Rapid and automatic classification of intertidal wetlands based on intensive time series Sentinel–2 images and Google Earth Engine[J]. National Remote Sensing Bulletin, 2022, 26(2): 348–357.

[16] 张 磊,宫兆宁,王启为,等. Sentinel–2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313–326.

Zhang L,Gong Z N,Wang Q W, et al. Wetland mapping of Yellow River Delta wetlands based on multi–feature optimization of Sentinel–2 images[J]. National Remote Sensing Bulletin, 2019, 23 (2): 313–326.

[17] 杨迎港,刘 培,张合兵,等. 基于特征优选随机森林算法的 GF –2 影像分类[J]. 航天返回与遥感, 2022, 43(2): 115–126.

Yang Y G,Liu P,Zhang H B, et al. Research on GF–2 image classification based on feature optimization random forest algorithm [J]. Spacecraft Recovery & Remote Sensing, 2022, 43(2): 115–126.

[18] 朱 琦,郭华东,张 露. 基于多时相 Landsat8 影像的海南岛热带天然林类型遥感分类[J]. 自然资源遥感, 2022, 34(2): 215 – 223. doi:10. 6046/zrzyyg. 2021156.

Zhu Q,Guo H D,Zhang L, et al. Classification of tropical natural forests in Hainan Island based on multi–temporal Landsat8 remote sensing images[J]. Remote Sensing for Natural Resources, 2022, 34(2): 215–223. doi:10. 6046/zrzyyg. 2021156.

[19] Gorelick N,Hancher M,Dixon M, et al. Google Earth Engine: Planetary – scale geospatial analysis for everyone[J]. Remote Sensing of Environment, 2017, 202: 18–27.

[20] 王 渊,赵宇豪,吴健生. 基于 Google Earth Engine 云计算的城市群生态质量长时序动态监测——以粤港澳大湾区为例[J]. 生态学报, 2020, 40(23): 8461–8473.

Wang Y,Zhao Y H,Wu J S. Dynamic monitoring of long time series of ecological quality in urban agglomerations using Google Earth Engine cloud computing: A case study of the Guang dong–Hong Kong–Macao Greater Bay Area, China[J]. Acta Ecologica Sinica, 2020, 40(23): 8461–8473.

[21] 徐振田,张 莎,白 雲,等. 基于 Landsat 数据的黄河三角洲湿地提取及近 30 年动态研究[J]. 海洋湖沼通报, 2020(3): 70–79.

Xu Z T,Ail S,Zhang S, et al. Mapping the wetland in Yellow River Delta and its dynamics in recent 30 years based on Landsat data [J]. Transactions of Oceanology and Limnology, 2020(3): 70–79.

[22] 牟晓杰,刘兴土,阎百兴,等. 中国滨海湿地分类系统[J]. 湿地科学, 2015, 13(1): 19–26.

Mou X J,Liu X T,Yan B X, et al. Classification system of coastal wetlands in China[J]. Wetland Science, 2015, 13(1): 19–26.

[23] Bradley A,Jackway P,Lovell B. Classification in scale–space: Applications to texture analysis[C]//Int. Conf. on Information Processing in Medical Imaging, 1995: 375–376.

[24] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.

A mapping methodology for wetland categories of the Yellow River Delta based on optimal feature selection and spatio–temporal fusion algorithm

FENG Qian¹, ZHANG Jiahua^{2,3}, DENG Fan¹, WU Zhenjiang³, ZHAO Enling¹, ZHENG Peixin¹, HAN Yang¹
(1. School of Geosciences, Yangtze University, Wuhan 430100, China; 2. CAS Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; 3. College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Exploring the remote sensing–based classification of coastal wetlands is significant for their conservation

and planning. Hence, this study investigated the Yellow River Delta with the 8-view Landsat8 OIL images from March to October 2019 as the data source. It constructed seven classification schemes based on different features of the images on the Google Earth Engine (GEE) cloud platform. Then, it employed the random forest classifier to classify different feature sets, with the scheme exhibiting the best classification effects selected for mapping the wetland categories of the Yellow River Delta. Considering poor data quality in August and September due to cloud contamination, this study filled in the cloudy zones using the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) algorithm. The results show that: ① The predicted images generated from the ESTARFM manifested a high correlation with the real image bands, with R values above 0.73, suggesting that the reconstructed images could be used in this study; ② The random forest algorithm was used to classify the surface feature types in the study area. Through optimal feature selection, the classification results of Scheme 7 demonstrated an overall accuracy of 92.28%, higher than those of conventional schemes, with a Kappa coefficient of 0.91, aligning with the actual wetland conditions. The results of this study can assist in deeply understanding the spatial distributions of different wetlands in the area, and provide a scientific basis for the conservation and planning of the regional ecological environment.

Keywords: Landsat8; multitemporal data; Yellow River Delta wetland; image fusion; Google Earth Engine; random forest

(责任编辑: 李 瑜)

=====

下 期 要 目

孙禧勇	积雪遥感监测产品研究与应用进展
王 昊	西藏自治区建材类非金属矿山与社会经济的相关性分析
易孜芳	耕地土壤质量评价元素含量高光谱反演方法研究
尹 滔	高分辨率遥感技术在川西毛垭坝地区活动断裂研究中的应用
宋嘉鑫	基于后验概率空间变化向量分析的 NSCT 高分辨率遥感影像变化检测
罗 维	基于多源卫星遥感影像的广西中南部地区甘蔗识别及产量预测
夏 浪	深度语义分割网络无人机遥感松材线虫病变色木识别研究
陈 刚	基于机载 LiDAR 技术的植被茂密区小型滑坡识别与评价
温 泉	基于深度学习的遥感图像水体提取研究综述
潘金胤	基于遥感绿色指数的城市生态环境质量评价研究——以南京市为例
李 薇	改进 Transformer 的高光谱图像地物分类方法——以黄河三角洲为例
武德宏	金沙江滑坡群 InSAR 探测与形变因素分析
麻卫峰	地基激光雷达单木树冠体积提取球坐标积分法
李志新	用于建筑物高度反演的复杂阴影测量方法
冯思维	基于光学遥感的内陆地表水体提取综述
郭彭浩	高频域多深度空洞网络的遥感图像全色锐化算法
王 涛	基于 InVEST 模型的长三角生态绿色一体化示范区生境质量时空演变特征分析