引文格式：罗福正,杨柳,卢彬,等. 基于多时相Sentinel-1影像的江南地区水稻信息提取[J]. 地理空间信息, 2024, 22(11): 73-77. doi:10. 3969/j. issn. 1672-4623.2024.11.014

基于多时相 Sentinel-1 影像的江南地区水稻信息提取

罗福正[1,2]，杨　柳[1,2]，卢　彬[1,2]，黄永生[1,2]
（1. 青海省地理空间和自然资源大数据中心，青海 西宁 810000；2. 青海省地理空间信息技术与应用重点实验室，青海 西宁 810000）

摘　要：为准确提取多云雨江南地区水稻种植信息，以江苏省丹阳市访仙镇为研究区，以多时相Sentinel-1数据为主、Sentinel-2数据为辅开展研究。通过计算水稻关键生育期内不同地物与水稻间的J-M距离，引入水稻移栽期内的归一化水体指数（NDWI）和收割完成后的归一化裸土指数（NDBI），获取水稻种植范围的最优特征组合；再以K近邻、随机森林和支持向量机（SVM）为基础建立分类模型，并对比验证了最优特征组合下不同模型的分类精度。结果表明，引入NDWI和NDBI的分蘖拔节期、孕穗抽穗期的VV极化后向散射系数以及孕穗抽穗期、乳熟成熟期VH极化后向散射系数为水稻提取的最优组合，采用SVM算法建立的分类模型的水稻识别精度最高，总体精度为0.916，Kappa系数达到0.828。

关键词：水稻提取；水稻生育期；SVM；Sentinel-1；Sentinel-2
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）11-0073-05

Rice Information Extraction in the Southern Yangtze River Region Based on Multi-temporal Sentinel-1 Images

LUO Fuzheng[1,2], YANG Liu[1,2], LU Bin[1,2], HUANG Yongsheng[1,2]
(1. Qinghai Geospatial and Natural Resources Big Data Center, Xining 810000, China;
2. Qinghai Provincial Key Laboratory of Geospatial Information Technology and Application, Xining 810000, China)

Abstract: To accurately extract rice cultivation information in the cloudy and rainy southern Yangtze River region, taking Fangxian Town, Danyang City, Jiangsu Province as the research area, we primarily utilized multi-temporal Sentinel-1 data to carry out research while supplementing with Sentinel-2 data. We computed the J-M distance between different terrain objects and rice during critical growth stages, and introduced the normalized difference water index (NDWI) during the rice transplanting period and the normalized difference bare soil index (NDBI) after harvesting to obtain the optimal feature combination for identifying rice cultivation areas at first. And then, we built classification models based on the K nearest neighbors, random forest and support vector machine (SVM) algorithms. Finally, we compared and validated the classification accuracy of different models under the optimal feature combination. The results reveal that the feature combination incorporating NDWI and NDBI during the tillering and heading stages in VV-polarized backscatter coefficients, as well as during the heading and maturing stages in VH-polarized backscatter coefficients, provide the optimal configuration for rice extraction. SVM algorithm achieved the highest accuracy in rice identification, with an overall accuracy of 0.916 and Kappa coefficient of 0.828.

Key words: rice extraction, rice growth stage, SVM, Sentinel-1, Sentinel-2

水稻是全球最重要的粮食作物之一。准确获取水稻种植面积对于农业管理、产量预测和粮食供应规划至关重要[1-2]。相较于传统调查方法，卫星遥感技术可实现快速准确的大范围水稻信息提取，为农业管理和科学决策提供重要支持[3-4]。水稻与其他地物间的光谱差异使得多光谱遥感成为水稻信息提取最常用的手段之一。苗翠翠[5]等采用S-G滤波法和决策树分类器对多时相MODIS NDVI数据进行时间序列重构，并建立模型对江苏省水稻进行识别和监测；钟仕全[6]等通过分析HJ-1B卫星水稻作物的光谱反射特性，构建了水稻作物识别模型，总体精度在90%以上；Wei P[7]等基于水稻全生育周期Sentinel-2卫星影像，利用基于像元的时间序列分析法准确提取了水稻种植区域。然而，云雨天气将大大降低多光谱遥感提取水稻的效率和准确度。近年来合成孔径雷达（SAR）因其独有的全天候观测能力被越来越广泛地应用于农作物尤其是水稻的遥感监测识别领域。Bazzi H[8]等基于Sentinel-1 SAR数据，采用决策树和随机森林算法对法国南部卡

收稿日期：2023-08-24。
项目来源：国家自然科学基金面上项目（2020-ZJ-927）。
第一作者简介：罗福正（1987—），本科，工程师，研究方向为遥感测绘，E-mail：paper202308@163.com。

玛格地区水稻进行了识别，准确率均在96%以上；郑少兰[9]等采用单时相Sentinel-1A SAR数据和Landsat8卫星资料对广东省五华县水稻面积进行提取，总体精度达到93.83%；刘龙威[10]等基于时间序列Sentinel-1A极化数据，采用光谱相似性方法通过阈值提取了水稻种植面积，总体精度达到97.34%；查东平[11]等通过分析水稻不同物候期SAR数据组合与其他地物间的J-M距离确定最优组合，再采用随机森林法建立的提取模型，总体精度可达94.3%。为分析水稻各生育阶段极化特征差异，对比不同监督学习模型在水稻提取上的性能表现，得到适用于江南地区的最优提取方案，本文以水稻关键生育期多时相Sentinel-1 SAR极化数据为基础，开展水稻与其他地物可分离度分析，筛选最优极化特征组合；同时避开6—9月江南云雨高峰期，引入6月初水稻移栽期和10月底水稻收割完成后的Sentinel-2光谱数据，分别计算归一化水体指数（NDWI）和归一化裸土指数（NDBI）；再采用多种机器学习方法建立水稻提取模型，并验证了各模型的提取精度，为利用Sentinel-1 SAR数据进行水稻提取提供了新思路。

1 研究区概况与数据来源

本文选取江苏省丹阳市访仙镇作为研究区。研究区地貌以丘陵和平原为主；属亚热带季风气候，夏季高温多雨、湿润且多光照。6—10月农作物种植以一季水稻为主，5月底—6月中旬为移栽期，6月中旬—7月底为分蘖拔节期，8月为孕穗抽穗期，9月进入乳熟成熟期，10月上中旬水稻开始收割。本文采用遥感目视解译和野外调查的方式获取的水体、水稻、草地、林地、建筑地物样本数量分别为100个、150个、90个、100个和110个。

本文采用的Sentinel-1卫星数据从欧空局官网下载，时间范围为2022年6—10月，共获得5幅覆盖研究区内不同生育期的影像，产品类型包含VH与VV两种极化方式；Sentinel-2卫星数据选用6月初和10月底云像元比低于10%的S2影像，分别获得3幅和5幅。

2 研究方法

2.1 数据预处理

为最大限度地消除SAR数据中的噪声信号，本文对Sentinel-1卫星数据进行预处理，主要包括热噪声去除[12]、轨道校正、辐射定标、Refined Lee滤波[13]等，均通过欧空局SNAP软件Sentinel-1 Toolbox实现。为尽可能地消除云和云阴影对Sentinel-2数据质量的影响，利用云概率数据集[14]“QA60”波段数据将数值为10和11以及云概率大于50%的云像元剔除。由于在水稻移栽期种植地块以水体特征为主，而收割完成后地块裸土特征显著，因此本文计算了NDWI[15]和NDBI[16]。

$$\mathrm{NDWI}=\frac{B_2-B_4}{B_2+B_4} \tag{1}$$

$$\mathrm{NDBI}=\frac{B_{11}-B_8}{B_{11}+B_8} \tag{2}$$

式中，B_2、B_4、B_{11}、B_8分别为Sentinel-2影像的蓝色波段、近红外波段、短波红外波段和近红外波段。

2.2 J-M距离

J-M距离[17]是一种衡量遥感图像不同类别间差异性的常用方法。在图像分类任务中，J-M距离可作为一种有效的分类准则，帮助确定不同类别之间的相似程度，从而辅助分类决策。其计算公式为：

$$J=2(1-e^{-B}) \tag{3}$$

$$B=\frac{1}{8}(m_1-m_2)^2\frac{2}{\delta_1^2+\delta_2^2}+\frac{1}{2}\ln\left(\frac{\delta_1^2+\delta_2^2}{2\delta_1\delta_2}\right) \tag{4}$$

式中，B为不同地类样本的巴氏距离[18]；m_i为地类特征平均值；δ_i为某地类特征的统计方差。

2.3 识别模型算法

K近邻算法[19]（KNN）、支持向量机[20]（SVM）和随机森林[21]（RF）是遥感领域3种常用的机器学习算法。KNN算法是一种简单而直观的分类和回归算法，利用与新数据点最近的K个邻居来预测新数据点类型。SVM算法能搜寻最大程度分隔不同类别的超平面，通过在特征空间中寻找最优超平面来实现分类。RF算法通过构建多个决策树并集成其结果来提高分类的准确性和稳定度。为节约计算成本同时防止出现过拟合[22]，本文将决策树数目设置为100个，特征数设置为自动选择，最大树深度设置为5层；KNN算法中K值设置为50；SVM算法中核函数选择径向基核函数，惩罚系数设置为100，其他参数默认。由于样本量较少，本文在模型训练阶段采用5折交叉验证方法，即将数据均分为5份，逐次将其中4份作为训练集，剩余1份作为验证集，以此循环5次，尽可能避免因数据集较少而引起的过拟合现象。

2.4 精度评价方法

本文采用精确率（P）、召回率（R）和F_1值[23]3个指标对水稻提取精度进行评价。P是指被识别为水稻的像元中真实为水稻的比例；R是指被正确识别的水稻占水稻样本的比例；F_1值是综合评价识别模

型性能的一个重要评价指标，是 P 和 R 的调和平均值，具体计算公式为：

$$P=\frac{TP}{TP+FP} \tag{5}$$

$$R=\frac{TP}{TP+FN} \tag{6}$$

$$F_1=2\times\frac{P\times R}{P+R} \tag{7}$$

$$OA=\frac{TP+TN}{TP+FP+TN+FN} \tag{8}$$

式中，TP 为被正确识别的水稻样本数量；FP 为非水稻样本被错误识别为水稻的数量；FN 为水稻样本被错误识别为非水稻的数量。

3　研究结果与分析

3.1　散射特征分析

按照水稻生育周期对不同地物样本进行时序划分，并计算时序后向散射系数均值。各地物不同极化数据统计结果和变化情况见图1、2，均值统计见表1。从各地物VH极化后向散射均值来看，水稻各生育期中水体的VH散射系数均最低，数值稳定在-27 dB左右；建筑和林地较接近，数值在-15 dB左右。水稻从移栽到收割后，经历了泥水混合相态到特征再到裸土特征的转变，VH极化散射系数随时间呈上升趋势，由移栽期-18.85 dB逐步升至收割后的-17.72 dB。由箱线图统计结果来看，水稻在各生育期与其他地物间的分离程度均较低，尤其是与草地的数值分布情况基本一致，直接利用阈值法进行区分难度较大。由不同生育期各地类VV散射系数箱线图和均值折线图可知，由于VV极化穿透性强于VH极化[24]，其接收的地物后向散射能量中能包含更多的植被下垫面特征，因此各地物的VV极化后向散射系数均不同程度的低于VH极化。水体的VV极化后向散射系数在水稻不同生育期总体变化不大，位于-20~-17 dB之间。林地与建筑之间的分离度比VH极化有明显提高。水稻在移栽、分蘖拔节和孕穗抽穗期与其他地物的分离程度明显优于生长发育末期。随着水稻的生长发育，其叶片和冠层散射逐渐增强，导致VV极化后向散射系数逐渐升高，乳熟成熟期的水稻样本VV极化散射系数均值基本能与林地持平。

3.2　生育期时相优选

为尽可能地避免特征冗余造成的计算资源浪费，本文利用J-M距离量化评价水稻与其他地物间的分离度，从而达到时相优选的效果。本文计算了水稻VV、VH极化数据在5个生育期与其他地物之间的

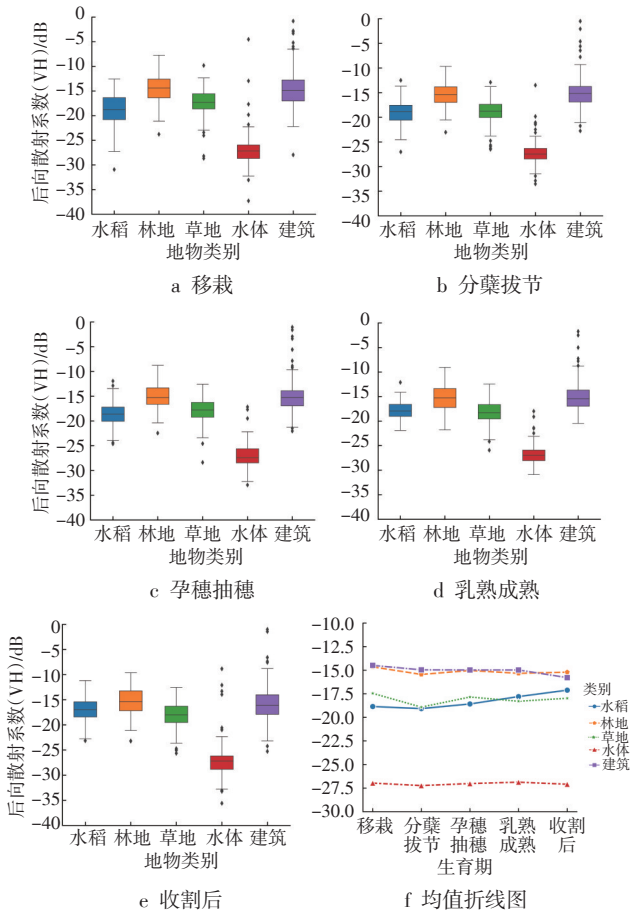


图1　不同生育期各地类VH散射系数箱线图和均值折线图

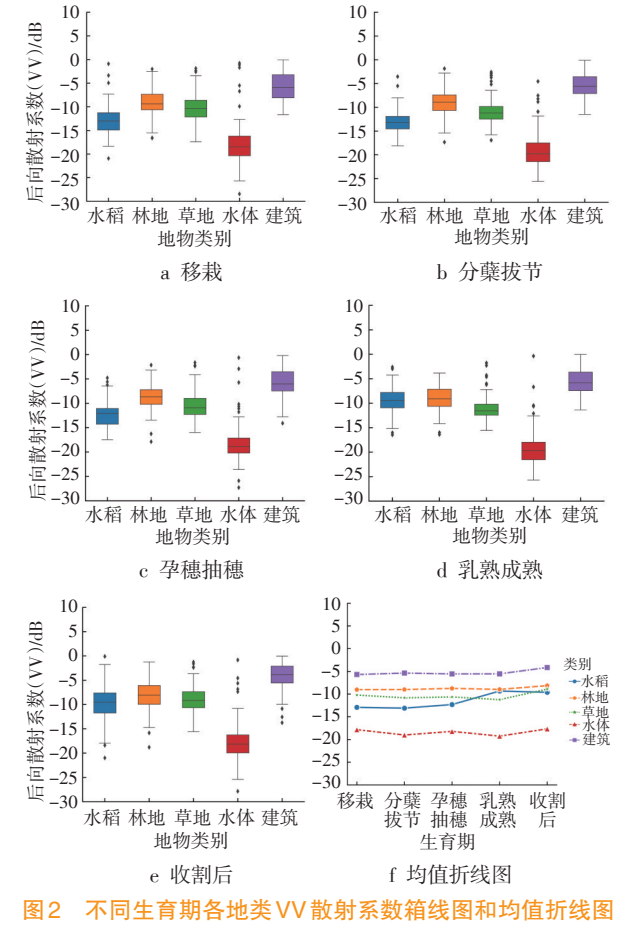


图2　不同生育期各地类VV散射系数箱线图和均值折线图

表1 各地物不同水稻生育期VV、VH极化后向散射系数/dB

极化类型	地物类别	移栽期	分蘖拔节期	孕穗抽穗期	乳熟成熟期	收割后
VH极化	水稻	−18.85	−19.08	−18.58	−17.80	−17.12
	林地	−14.65	−15.45	−15.04	−15.33	−15.21
	草地	−17.46	−18.92	−17.85	−18.29	−17.98
	水体	−26.97	−27.23	−27.02	−26.87	−27.09
	建筑	−14.49	−14.97	−14.98	−14.98	−15.79
VV极化	水稻	−12.93	−13.10	−12.31	−9.34	−9.62
	林地	−9.03	−9.01	−8.76	−8.98	−8.15
	草地	−10.22	−10.83	−10.64	−11.23	−8.92
	水体	−17.84	−19.01	−18.22	−19.26	−17.66
	建筑	−5.70	−5.39	−5.57	−5.55	−4.16

J-M距离（表2），可以看出，水稻在各生育期与草地、林地的VV极化、VH极化后向散射系数的分离度均小于1.4，可分离性较差。在VV极化方式上，建筑与水稻在移栽、分蘖拔节和孕穗抽穗期内，J-M距离均在1.8以上，具有良好的可分离性。水体与水稻在乳熟成熟期VH极化后向散射系数间的J-M距离达到1.891，可分离度较高。综合来看，移栽期、分蘖拔节期和孕穗抽穗期的VV散射系数以及分蘖拔节期、孕穗抽穗期和乳熟成熟期的VH散射系数能较好地分离水稻与其他地物。

表2 不同生育期水稻与其他地物J-M距离（VV/VH）

地物类别	移栽期	分蘖拔节期	孕穗抽穗期	乳熟成熟期	收割后
建筑	1.855/1.122	1.804/1.316	1.860/1.241	1.257/1.244	1.459/0.934
林地	1.216/1.197	1.400/1.296	1.247/1.302	0.805/1.174	0.864/0.946
水体	0.476/0.960	0.816/1.541	0.745/1.573	1.476/1.891	0.884/1.475
草地	1.012/0.875	1.021/0.808	0.901/0.839	0.952/0.877	0.854/0.834

本文按照数量对上述6个预选特征依次进行随机组合，分别计算水稻的平均最优J-M距离，并找出最优组合所对应特征；再添加NDWI和NDBI信息，并对比分析J-M距离的变化结果。由表3可知（1~5分别代表从移栽期到收割完成5个水稻生育期），无指数特征加入时，平均J-M距离在1.35~1.59之间；引入NDWI和NDBI后，J-M距离出现不同程度的上升，由于NDWI可有效解决草地与水稻不易分离的难题，加入后J-M距离提升至1.55以上；同时引入NDWI和NDBI后，分蘖拔节期、孕穗抽穗期的VV极化后向散射系数与孕穗抽穗期、乳熟成熟期VH极化后向散射系数的特征组合能显著提高水稻与其他地物之间的分离水平，J-M距离达到1.821。

表3 不同数量最优特征组合和添加NDWI、NDBI后各地物与水稻间J-M距离

最优特征组合	J-M距离			
	无指数特征	NDWI	NDBI	NDWI+NDBI
VV2	1.362	1.537	1.375	1.732
VV2+VH4	1.395	1.623	1.505	1.741
VV2+VH4+VV3	1.522	1.652	1.535	1.753
VV2+VH4+VV3+VH3	1.583	1.721	1.632	1.821
VV2+VH4+VV3+VH3+VV1	1.532	1.554	1.536	1.614
VV2+VH4+VV3+VH3+VV1+VH2	1.502	1.551	1.448	1.615

3.3 水稻提取精度评价

将最优特征组合样本设置为训练集，进行归一化处理；再利用Python Sklearn库包根据KNN、RF和SVM算法构建水稻识别模型。模型构建与测试验证的具体步骤为：①输入SAR极化最优特征组合；②输入NDWI和NDBI特征参数；③测试集采用随机抽样的方式，由实地调查结果和目视解译确定，抽样次数为10次，每次获取水稻样本和非水稻样本比例为1∶3，总数为400个；④取5次抽样的平均值作为模型识别精度，得到模型测试分类结果。由图3可知，在无指数特征输入的情况下，由于水稻与草地的极化特征较接近，3种算法模型均不同程度地将大面积水稻误判为草地，分类准确性较低；引入NDWI特征后，由于稻田前期的水体特征较明显，NDWI增加了水体特征权重，3种模型均有效减少了草地误判点，但却将水稻误判为水体，KNN与RF算法的误判率分别达到11.3%和9.5%；同时将最优极化特征组合与NDBI指数作为模型输入时，上述现象明显减少，但NDBI的引入却使得弱植被特征水稻地块被误判为建筑，水稻破碎现象严重；同时引入NDWI和NDBI特征并叠加最优极化组合建立模型后，3种模型的水稻识别准确率均明显

增加。由表4可知，各模型的总体精度均在0.85以上。由于在处理少量样本时SVM算法具有核技巧和间隔最大化特性，因此表现出更优越的模型泛化性能，在识别水体、林地、草地等面积小且分布不均的土地类型时能较准确的表达；识别水稻时能有效分离田块与田埂，模型精确率、召回率、F_1值均优于其他两种算法，总体精度为0.916，Kappa系数达到0.828。

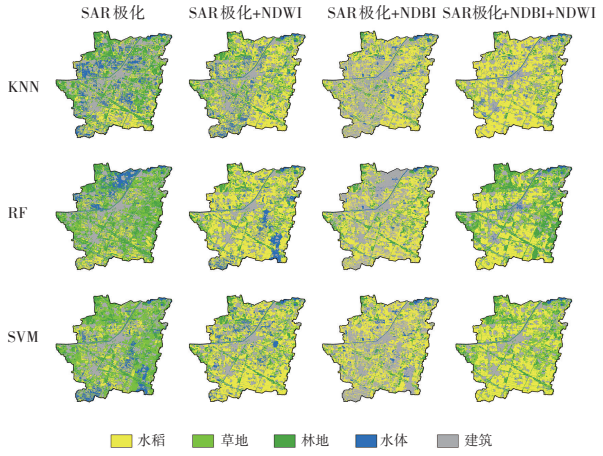


图3　不同模型和特征参量下的分类结果

审图号：GS(2024)0650

表4　不同模型引入极化与指数特征后的水稻识别精度比较

分类算法	P	R	F_1值	总体精度
KNN	0.915	0.832	0.871	0.892
RF	0.925	0.845	0.883	0.902
SVM	0.950	0.856	0.900	0.916

4　结　语

本文基于水稻关键生育期多时相Sentinel-1极化数据、移栽期和收割后的Sentinel-2光谱数据，采用KNN、RF和SVM算法分别建立模型提取江苏省丹阳市访仙镇水稻种植范围，得到结论如下：

1）各地物极化数据与水稻作物间的J-M距离因时间和极化方式的不同而体现出明显差异。移栽期、分蘖拔节期和孕穗抽穗期的VV极化数据以及分蘖拔节期、孕穗抽穗期和乳熟成熟期的VH极化数据能较好地分离水稻与其他地物，J-M距离均在1.4以上，将其作为预选特征。

2）对预选特征随机进行生育期数据组合，根据平均最优J-M距离筛选最优组合。同时引入NDWI和NDBI后，分蘖拔节期、孕穗抽穗期的VV极化后向散射系数与孕穗抽穗期、乳熟成熟期VH极化后向散射系数的特征组合能显著提高水稻与其他地物间的分离水平，J-M距离达到1.821。

3）KNN、RF和SVM算法的精确率、召回率、F_1值和总体精度均存在一定差异，其中SVM算法的精度最优，精确率、召回率、F_1值在3种算法中均处于首位，总体精度达到0.916，Kappa系数达到0.828。

本文对比分析了KNN、RF和SVM算法在水稻提取方面的性能差异，SVM算法在处理少量样本时表现出更优越的模型泛化性能，但在实际应用中，由于农田种植制度和结构差异，在处理水稻与其他农作物并存的复杂场景时需要收集大量样本进行训练，以减少错漏分现象。

参考文献

[1] Kuenzer C, Knauer K. Remote Sensing of Rice Crop Areas[J]. International Journal of Remote Sensing, 2013, 34(6):2 101-2 139

[2] 吴炳方,张淼,曾红伟,等. 全球农情遥感速报系统20年[J]. 遥感学报,2019,23(6):1 053-1 063

[3] 邬明权,王长耀,牛铮. 利用多源时序遥感数据提取大范围水稻种植面积[J]. 农业工程学报,2010,26(7):240-244

[4] Xiao X, Boles S, Liu J, et al. Mapping Paddy Rice Agriculture in Southern China Using Multi-temporal MODIS Images[J]. Remote Sensing of Environment, 2005, 95(4):480-492

[5] 苗翠翠,江南,彭世揆,等. 基于NDVI时序数据的水稻种植面积遥感监测分析:以江苏省为例[J]. 地球信息科学学报,2011,13(2):273-280

[6] 钟仕全,莫建飞,陈燕丽,等. 基于HJ-1B卫星遥感数据的水稻识别技术研究[J]. 遥感技术与应用,2010,25(4):464-468

[7] Wei P, Ye H, Qiao S, et al. Early Crop Mapping Based on Sentinel-2 Time-series Data and the Random Forest Algorithm[J]. Remote Sensing, 2023, 15(13):3 212

[8] Bazzi H, Baghdadi N, Hajj M E, et al. Mapping Paddy Rice Using Sentinel-1 SAR Time Series in Camargue, France[J]. Remote Sensing, 2019, 11(7):887

[9] 郑少兰,刘龙威. 单时相Sentinel-1A卫星SAR数据水稻信息提取[J]. 地理空间信息,2020,18(4):61-64

[10] 刘龙威,魏瑄. 基于时间序列Sentinel-1A数据的水稻识别[J]. 地理空间信息,2019,17(11):83-86

[11] 查东平,蔡海生,张学玲,等. 基于多时相Sentinel-1水稻种植范围提取[J]. 自然资源遥感,2022,34(3):184-195

[12] Mascolo L, Lopez-Sanchez J M, Cloude S R. Thermal Noise Removal from Polarimetric Sentinel-1 Data[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19:1-5

[13] 曾玲方,李霖,万丽华. 基于Sentinel-1卫星SAR数据的洪水淹没范围快速提取[J]. 地理信息世界,2015,22(5):100-103

（下转第86页）