

基于 MODIS 指数和随机森林的江西省早稻种植信息提取

林志坚¹, 姚俊萌¹, 苏校平², 蔡哲¹, 刘丹^{1,3*}

(1. 江西省农业气象中心, 南昌 330096; 2. 南昌市气象局, 南昌 330008; 3. 江西省气象科学研究所, 南昌 330096)

摘要: 尽早获取双季早稻的种植信息, 对政府部门掌握全省水稻生产形势及制定粮食安全保障的相关政策方针具有重要意义。传统业务服务中, 通常将水稻生长早期的多时相 MODIS 指数与阈值法相结合, 对种植信息进行提取, 但该方法主观性强, 受人为及不同地区水稻物候期差异影响大, 且存在混合像元等限制, 机器学习算法可以较好解决此问题。因此, 该研究提出一种结合水稻生长早期 MODIS 指数和随机森林的种植信息提取方法, 基于江西省早稻生长早期多时相 MODIS 增强型植被指数 (Enhanced Vegetation Index, EVI)、归一化植被指数 (Normalized Difference Vegetation Index, NDVI) 和地表水分指数 (Land Surface Water Index, LSWI) 的变化特征, 利用随机森林算法构建早稻种植区域提取模型与丰度反演模型, 提取全省早稻种植信息, 并利用 Sentinel-1A 提取的验证样区与统计资料验证。结果表明, 早稻种植区域及丰度的空间分布特征与 Sentinel-1A 提取的验证样区的空间特征基本一致, 提取模型的分类精度为 93.18%, 丰度反演模型与样本数据的平均绝对误差、均方根误差和决定系数分别为 0.07、0.10 与 0.86, 且在高丰度种植区反演效果更优。与统计资料相比, 全省早稻面积识别精度为 92.33%。该研究解决了水稻种植信息提取中阈值选取合理性、混合像元与时效性限制等问题, 为水稻生长早期种植信息的业务化提取提供一种参考方法, 具有一定应用价值。

关键词: 遥感; 算法; 早稻识别; MODIS 指数; 随机森林

doi: 10.11975/j.issn.1002-6819.2022.11.022

中图分类号: S127; TP75

文献标志码: A

文章编号: 1002-6819(2022)-10-0197-09

林志坚, 姚俊萌, 苏校平, 等. 基于 MODIS 指数和随机森林的江西省早稻种植信息提取[J]. 农业工程学报, 2022, 38(11): 197-205. doi: 10.11975/j.issn.1002-6819.2022.11.022 http://www.tcsae.org

Lin Zhijian, Yao Junmeng, Su Xiaoping, et al. Extracting planting information of early rice using MODIS index and random forest in Jiangxi Province, China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 197-205. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.11.022 http://www.tcsae.org

0 引言

早稻是江西省保障粮食总产稳定的重要作物之一。近年来, 受农业产业结构调整、土地利用变化等政策影响, 江西早稻种植面积变化已逐渐成为制约粮食稳定生产的不确定因素。据江西省统计年鉴资料, 自 2000 年以来全省早稻种植面积的标准差达到 $1.09 \times 10^5 \text{ hm}^2$, 约占总种植面积的 10%。因此, 准确监测早稻播栽面积等种植信息, 对政府部门掌握水稻早期生产形势和部署相关工作等具有重要科学意义。

在农作物遥感监测研究中, MODIS 数据借助其观测周期短、覆盖范围广、图像易获取等优势, 相较于 Landsat8、Sentinel 等高分辨率遥感影像, 更适用于大尺度的水稻种植信息提取业务^[1-3]。基于 MODIS 数据源, 前人已研究出多种水稻种植信息提取的方法。其中, 利用水稻生长早期的多时相 MODIS 归一化植被指数 (Normalized Difference Vegetation Index, NDVI)、增强型

植被指数 (Vegetation Index, EVI) 和陆表水分指数 (Land Surface Water Index, LSWI) 间关系设定阈值的物候算法具有所需影像少、操作简单等优点, 在业务化提取省级尺度的水稻种植信息方面得到广泛应用^[4-8]。但物候算法的提取精度依赖于阈值选取, 具有一定主观性与经验性, 且不同地区水稻物候期存在一定差异, 算法泛化性较弱^[1], 不利于种植信息的稳定提取。而随机森林算法具有人工干预少、不易过拟合等特点, 能够有效解决上述算法存在的问题, 近年来在水稻种植信息提取中得到成功应用^[9-11]。Wang 等^[12]基于长时间序列的遥感指数, 采用随机森林算法分别提取了松嫩平原淹水期和移栽期的水稻面积; 雷小雨等^[13]通过选取水稻秧苗期与成熟期两个图像差值特征, 突出水稻物候变化特点, 结合随机森林算法, 高精度提取研究区域水稻种植面积。张雅琦^[14]基于水稻 NDVI 序列曲线的物候统计特征参数, 运用随机森林算法, 建立水稻种植制度遥感分类模型。但运用随机森林算法开展的研究, 多需要对长时间 MODIS 指数序列或包含有水稻生长晚期的图像数据进行提取分析, 导致其不能及时为政府部门提供水稻早期生产形势的参考信息。因此本研究考虑发展一种基于水稻生长早期 MODIS 指数数据和随机森林算法的水稻种植信息提取方法。此外, 由于 MODIS 数据空间分辨率相对不高, 大范围水稻面积监测时存在混合像元问题, 从而限制了水稻的识别精度。因此, 需对提取的水稻像元进行解混, 得到水稻丰度^[15-16]。

收稿日期: 2022-03-28 修订日期: 2022-05-26

基金项目: 中国气象局创新发展专项 (CXFZ2021J062); 江西省气象科技项目 (JX2020Q02, JX2021Q03)

作者简介: 林志坚, 助理工程师, 研究方向为农业气象技术应用与推广。

Email: 1771298692@qq.com

※通信作者: 刘丹, 工程师, 研究方向为农业气象业务预报与技术应用。

Email: nuistdan@126.com

目前,用于地物丰度提取的方法主要包括线性回归法^[17]、混合像元分解法^[18-19]和随机森林回归算法^[20-21],其中随机森林回归算法能够分析特征量与地物丰度之间的非线性关系,反演精度较优^[22]。Huang 等^[23]基于时序 MODIS 数据,利用随机森林回归算法提取了中国沿海地区城市用地丰度;Hu 等^[24]利用时序 MODIS 影像与统计数据,结合随机森林回归算法成功提取黑龙江省主要农作物种植丰度。然而,该方法在农作物像元丰度提取的业务应用方面仍较少,对江西地区早稻丰度提取效果仍有待考察。

综上,本研究从决策服务及业务应用的实效性出发,基于江西省 2020 年早稻生长早期(3 月 13 日—6 月 1 日)的多时相 MODIS 影像,将 EVI、NDVI、LSWI 等指数作为特征数据,利用随机森林算法,分别构建早稻种植区域提取模型及早稻丰度反演模型,提出一种适用于业务服务的省级尺度早稻生长早期种植信息提取方法,以期解决目前水稻生长早期种植信息提取方法中阈值选取合理性、MODIS 混合像元及提取时效性等问题。

1 研究区域与数据

1.1 研究区域

江西省地处长江中下游南岸,面积约 16.69 万 km²,其中耕地面积 2.93 万 km²。该地区属亚热带温暖湿润季风气候,四季分明,日照充足,冬暖夏热,雨量丰沛,气候条件有利水稻生长,是中国优质水稻的主产区之一。江西省水稻一般为一年两熟,双季早稻通常在 3 月中旬开始播种,4 月中下旬移栽,5 月上旬进入分蘖期、中下旬拔节,6 月孕穗抽穗,7 月上中旬成熟收割^[25-26],其中播种至拔节期的营养生长阶段为早稻生长早期,孕穗至成熟期的生殖生长阶段为早稻生长晚期。

研究选择南昌县作为早稻种植区域提取及丰度反演模型精度评价的验证区,其范围为 28.27°~28.97°N、115.82°~116.32°E。该地区种植条件优越,稻田面积较为稳定,适宜用于早稻信息提取验证^[25]。

1.2 数据

1.2.1 MODIS 数据及其预处理

本研究采用 2020 年 1 月 1 日—6 月 1 日(儒略日第 1~153 天)间共 20 景 MODIS 地表反射率融合产品(MOD09A1)作为基础数据,其可通过美国航空局网站(<https://ladsweb.modaps.eosdis.nasa.gov/search/>)下载获得。MOD09A1 产品的空间分辨率为 500 m,时间分辨率为 8 d。研究区涉及图幅编号为 h27v06、h28v05、h28v06。

利用 MODIS 影像处理工具 MRT (MODIS Reprojection Tool)对原始影像进行拼接、重采样、重投影等预处理步骤^[18],然后分别计算 EVI、NDVI 及 LSWI^[8],并利用 S-G 滤波方法^[27]对 3 种指数的时间维进行处理,最终得到共 60 景指数图像,并选取 3 月 13 日—6 月 1 日(儒略日第 73~153 天)早稻生长早期的 33 幅指数图像进行研究。

1.2.2 Sentinel-1A 数据及其处理

为验证本研究方法提取的早稻种植信息的精度,选择 Sentinel-1A 干涉宽幅模式下的地距多视影像资料制作的南昌县早稻分布与丰度值与研究结果进行对比。

Sentinel-1A 数据来源于欧洲航天局网站 (<https://scihub.copernicus.eu/>),空间分辨率为 5 m×20 m,时间分辨率为 12 d,宽幅为 250 km,包含 VV 与 VH 两种极化数据^[28],其在典型区域的地物分类应用中具有一定优势^[29]。依据梁翰月^[25]提出的方法,进行数据配准、斑点噪声滤波、地理编码、辐射定标等预处理后,利用 ENVI 软件将 2020 年 4 月 27 日、5 月 9 日、6 月 14 日的 VV 极化数据进行 RGB 假彩色合成,进而结合监督分类中的最大似然法提取早稻种植区域,精度在 98%以上。并根据种植区域提取结果,制备验证区早稻种植丰度,其计算式如下^[20]:

$$R = \frac{S}{N} \quad (1)$$

式中 R 为任意 MODIS 像元内的早稻种植丰度,表示 1 个 MODIS 像元内早稻所占比例; S 和 N 为 MODIS 像元内 Sentinel-1A 资料提取的早稻像元个数和像元总数。

1.2.3 地面观测数据

地面观测数据主要用于构建早稻种植区域提取模型与丰度反演模型,并作为精度评价标准。本研究中土地利用类型分为早稻、其他作物、林地、水体及人造建筑等 5 类,其中早稻及其他作物样点数据来源于江西省 17 个农业气象观测站点、江西省农业气象中心实地调查及“江西微农”平台农情反馈功能获取的作物种植区域地理信息,获取时间为 2020 年 3 月中旬至 7 月中旬早稻生长发育期间。林地、水体及人造建筑样本则通过 Google Earth 高空间分辨率影像进行目视解译选取,且为满足 MODIS 数据的精度要求,所选地物类型在 MODIS 像元中丰度须超过 80%。本研究共用到 2 120 个样点数据(图 1)、分别为早稻 453 个、林地 426 个、其他作物 398 个、水体 421 个、人造建筑 422 个。其中,用于早稻种植区域提取的训练样本与验证按 3:1 的比例随机抽取。

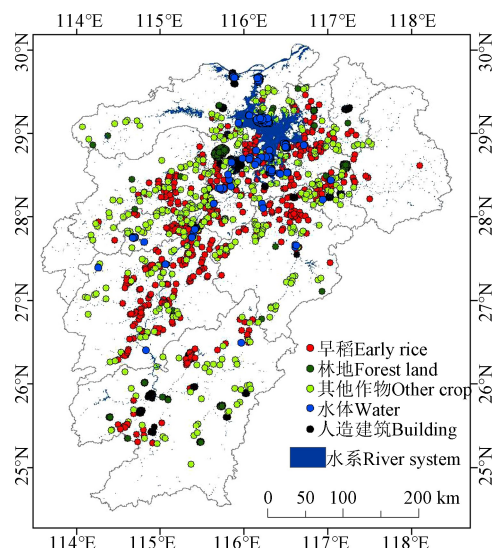


图 1 研究区地面观测站点分布

Fig.1 Spatial distribution of field samples in the study area

早稻样本点的实际丰度值是在 ENVI 软件中通过目视解译法标定早稻种植区域的基础上,利用 ROI 统计功能计算对应 MODIS 像元内早稻所占面积比例得到。考虑到丰度的物理意义,早稻丰度值的范围控制在 0~1^[18]。所有早

稻样本点的实际丰度都将用于构建丰度反演模型。此外，由于随机森林建模过程中袋外样本的存在，因此，所有早稻样本点的实际丰度值也可用于评估模型精度^[30]。

2 研究方法

2.1 技术路线

本研究通过 4 个主要步骤提取早稻种植信息(图 2)。首先基于 MODIS 地表反射率数据，得到 2020 年江西省早稻播种至拔节期的 EVI、NDVI 及 LSWI 等指数数据，并研究各地物指数随时间变化的差异，分析利用 3 种指

数提取水稻种植信息的可行性，进而构建特征数据集。其次，基于特征数据集及地物分类数据随机抽取训练样本，结合随机森林分类算法建立早稻种植区域提取模型，实现对江西省早稻种植区域的识别与提取。然后，基于特征数据集与早稻样本点的丰度值，利用随机森林回归算法建立早稻丰度反演模型，并结合提取的早稻种植区域，得到早稻丰度图。最后，利用验证样本、Sentinel-1A 提取的早稻验证样区及统计数据分别对早稻种植区域提取模型与丰度反演模型的样本精度、空间精度及面积提取精度进行验证。

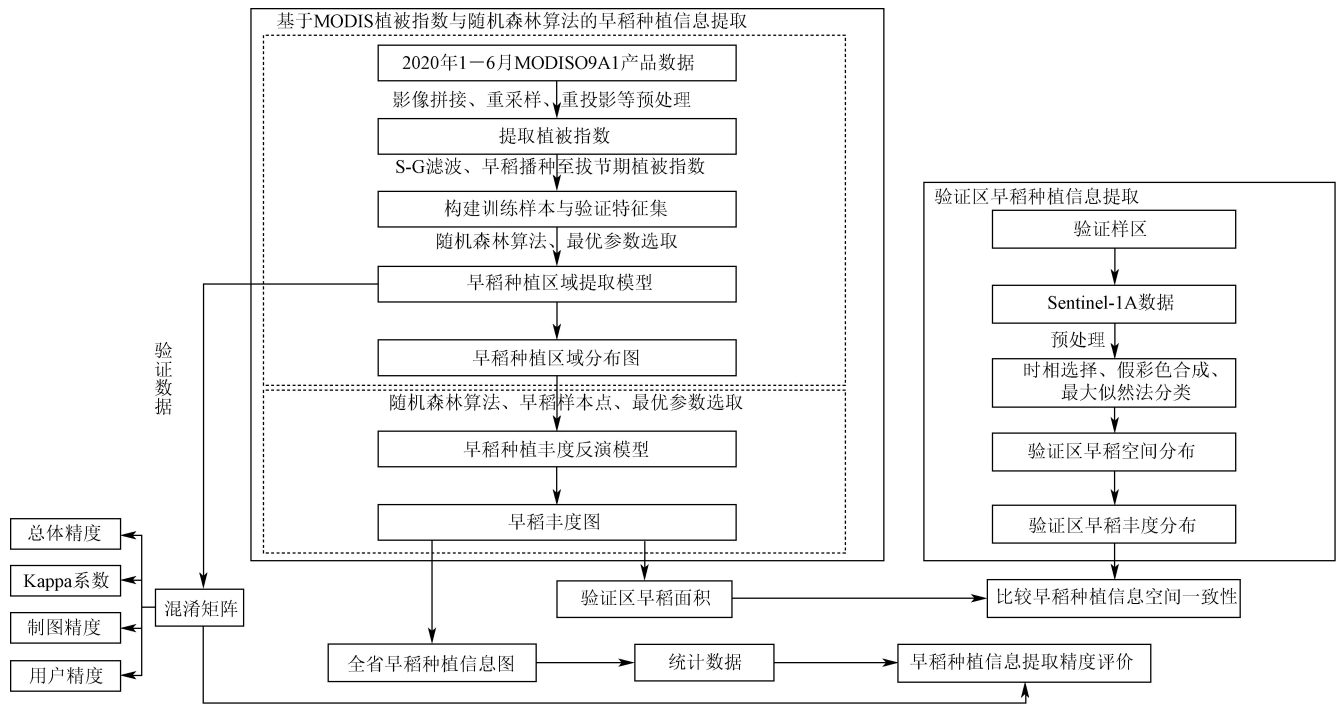


图 2 技术流程图

Fig.2 Flow chart of algorithm

2.2 随机森林模型构建

2.2.1 随机森林算法原理

随机森林算法是由 Breiman^[30]提出的基于多棵决策树集成的算法，既可用于构建分类模型也可用于构造回归模型^[31-32]。近年来，随机森林算法因其计算速度快、稳定性强、精度较高等优点^[27,33]，在农作物种植信息提取领域得到广泛应用。由于本研究需要分别提取早稻种植区域与种植丰度，因此可通过随机森林分类与回归算法分别构建模型，进而提取江西省早稻种植信息。此外，在随机森林模型构建过程中，未被用于训练的样本可组成 m 个袋外样本，产生袋外误差 (Out of Bag error, OOBError)，用于确定模型参数^[34]。对于早稻丰度反演模型，在每棵决策树建成后，算法能自动对袋外样本值进行预测^[22]，在决策树量足够的情况下，可得到所有样本因变量的平均预测结果。该预测过程近似于进行了大量计算的 k 折交叉验证，其结果可用于评估模型精度^[30]。

2.2.2 模型参数确定

随机森林算法中，决策树的树量 (ntrees) 与各决策树节点随机选取的特征量个数是影响模型精度的两个重要参数，其中特征量个数通常取特征量的平方根^[10]，而

ntrees 达到适合的数值时，OOBError 将不再明显提升^[34]。因此本研究根据 ntrees 与 OOBError 的关系图选择两个模型的最优 ntrees。

如图 3a 所示，对于早稻种植区域提取模型，随着决策树的增加，OOBError 逐渐减小，且在 ntrees 大于 500 后，无明显增益，因此早稻种植区域提取模型的决策树量设为 500。与种植区域提取模型训练结果类似，早稻丰度反演模型的 OOBError 随着决策树的增加逐渐减小 (图 3b)，且当 ntrees 为 800，OOBError 最小，因此早稻丰度反演模型的决策树量设为 800。

2.3 精度评估方法

本研究采用混淆矩阵对早稻种植区域提取模型精度进行验证，其可衍生得到制图精度、用户精度、总体分类精度和 Kappa 系数等评价因子^[27]。对于早稻丰度反演模型，利用对所有早稻样本点丰度的预测结果，结合平均绝对误差 (Mean Absolute Error, MAE)、均方根误差 (Root Mean Square Error, RMSE)、决定系数 (R^2) 评估模型精度^[33]。此外，早稻种植区域与丰度的空间分布均采用验证区的 Sentinel-1A 资料提取结果对比验证；早稻种植面积则采用统计年鉴资料进行精度评估。

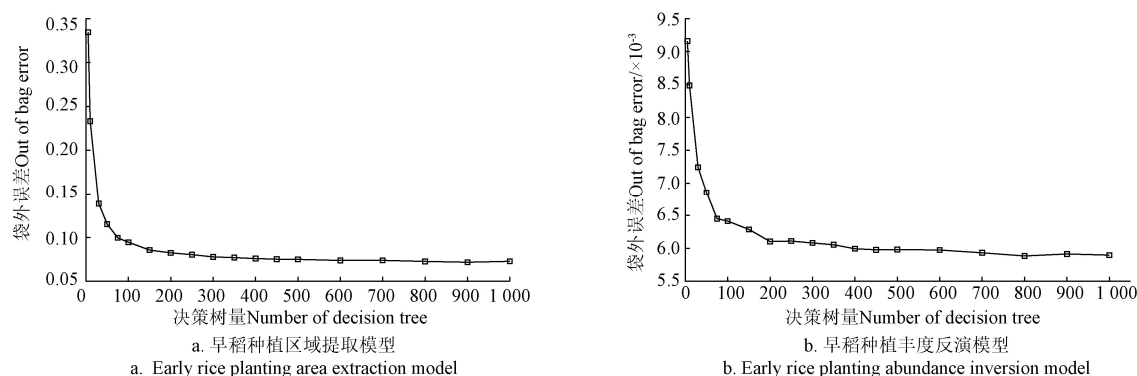


图3 早稻种植区域提取模型与早稻丰度反演模型的袋外误差随决策树树量的变化

Fig.3 Change of out of bag error value of the early rice planting area extraction model and abundance inversion model with the number of decision tree

3 结果与分析

3.1 早稻生长早期 MODIS 指数作为特征量的可行性分析

为合理分析水稻生长早期不同地物 MODIS 指数间的变化差异,评估利用早稻播种至拔节期 EVI、NDVI 与 LSWI 等 3 种指数各 11 个时相资料进行种植信息提取的可行性,以构建随机森林算法的特征样本集。图 4 展示了各类地物样本点的 3 种指数平均值随时间的变化曲线。对于 EVI 指数 (图 4a), 相比其他地物, 早稻在播种移栽期 (儒略日第 73~113 天) 通常需要大量的水, 此时 EVI 值较低, 维持在 0.2 左右, 大于水体与建筑, 但小于林地与其他作物; 随着水稻开始分蘖 (儒略日第 121~153 天), EVI 值迅速增加, 至拔节末达 0.5 左右, 该特征可用于区分水稻及其他地物^[35], 但在分蘖初期 (儒略日第 121~129 天), 水稻与其他作物的 EVI 值相近 (0.3 左右); 拔节期 (儒略日第 145~153 天) 则与林地相近,

易造成像元混淆误差。对于 NDVI 指数 (图 4b), 不同类型地物在拔节期 (儒略日第 145~153 天) 之间存在一定差异, 其值维持在 0.6 左右, 大于其他作物、人造建筑及水体, 但小于林地, 该特征能够改善 EVI 指数在早稻拔节期的地物混淆误差。但对比图 3a, NDVI 仍无法改善分蘖初期 (儒略日第 121~129 天) 早稻与其他作物的混淆误差 (数值均在 0.45 左右)。对于 LSWI 指数 (图 4c), 水稻在分蘖初期 (第 121~129 天) 数值快速增加, 在 0.4~0.5 之间, 与其他地物数值呈现一定差异, 尽管在一些时期早稻的 LSWI 指数值与其他地物相近, 但同时引入 EVI 与 NDVI 可以实现 MODIS 指数间特征的互补。因此, 本研究将 3 种指数各 11 个时相资料, 共 33 个特征量作为早稻种植区域提取模型的特征数据集, 在理论上能够较好地区分不同地物类型。同时, 为保证输入特征集的一致性, 早稻丰度反演模型也采用上述特征集进行构建。

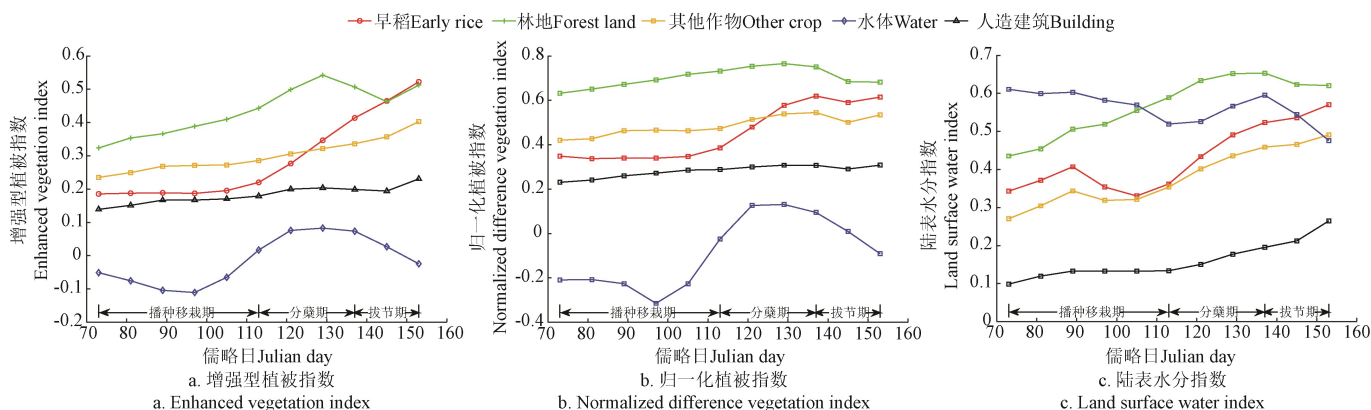


图4 早稻播种至拔节期 5 类地物类型样本点的各指数的时间变化曲线

Fig.4 Time series profile of indexes for 5 specify land-cover types during sowing to jointing of early rice

3.2 早稻种植区域提取模型精度评价

确定特征数据集及模型参数后,可构建早稻种植区域提取模型识别早稻像元,并利用验证样本进行精度评价。如表 1 所示,模型总体分类精度为 93.18%, Kappa 系数为 0.915。从不同地物类型来看,模型提取的早稻制图精度和用户精度分别达到 92.04%和 91.23%,误分主要

出现在早稻与其他作物之间。模型对林地、水体与人造建筑的分类精度整体较高,用户精度与制图精度大部分达到 95%以上。对于其他作物类型,其制图精度与用户精度均低于其他类别地物,分别为 82.83%与 85.42%,主要误分也出现在其与早稻之间。

利用梁翰月^[25]的方法,基于 Sentinel-1A 数据提取南

昌县早稻种植区域，并与随机森林模型结果进行对比（图 5），可以发现，两种数据都可反映南昌县西北、东部以及西南存在大面积早稻种植区域的空间分布特征。但相

比更高分辨率的 Sentinel-1A 数据，MODIS 数据在验证区的早稻种植区域空间提取结果相对粗糙，无法呈现亚像元下早稻种植信息特征。

表 1 研究区不同地物类型验证数据分类精度评价结果
Table 1 Evaluation of classification precision for different land type of the validate the data in the study area.

项目 Item	混淆矩阵 Confusion matrix					制图精度 Producer accuracy/%	用户精度 User accuracy/%
	早稻 Early rice	林地 Forest land	其他作物 Other crops	水体 Water	人造建筑 Building		
早稻 Early rice	104	1	8	0	0	92.04	91.23
林地 Forest land	0	104	2	0	0	98.11	97.20
其他作物 Other crops	8	2	82	5	2	82.83	85.42
水体 Water	2	0	1	102	0	97.14	93.58
人造建筑 Building	0	0	3	2	100	95.24	98.04
总体分类精度 Overall accuracy/%					Kappa 系数 Kappa coefficient	0.915	

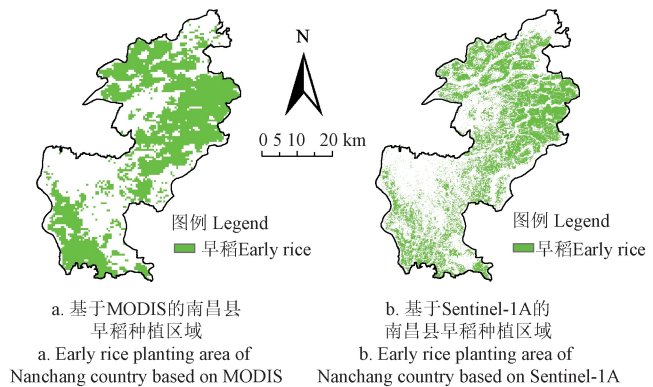


图 5 MODIS 与 Sentinel-1A 提取的 2020 年南昌县早稻分布
Fig.5 Distribution of early rice derived from the MODIS and the Sentinel-1A product in Nanchang County in 2020

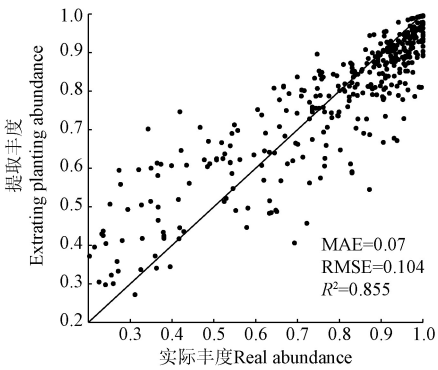
上述结果说明，随机森林分类模型能够较好区分早稻与不同地物，也能再现验证区早稻种植分布特征，但对早稻与其他作物的识别存在一定的错分现象，且在部分早稻零散种植区出现漏分。其原因可能是，各地区早稻发育进度存在一定差异，其与不同地物的指数易出现“同谱异物”现象^[36]。此外，MODIS 分辨率相对 Sentinel-1A 较低，当像元以早稻种植为主时，指数特征更加接近于早稻，易将整个像元预测为早稻种植区域，导致提取的早稻面积大于实际面积，因而无法准确获取亚像元的早稻种植信息。因此，在利用 MODIS 指数获取更加准确的全省早稻种植面积时，需进一步对早稻像元做丰度反演。

3.3 早稻种植丰度反演模型精度评价

图 6 是基于早稻种植丰度反演模型的精度验证。由图可知，模型对样本点的早稻丰度预测较好，与实际丰度的 MAE、RMSE、 R^2 分别为 0.07、0.104 与 0.855，且高丰度 (>0.50) 预测结果更优，MAE 与 RMSE 分别降为 0.06 与 0.08。

丰度反演模型的空间分布精度，如图 7 所示，模型提取的南昌县早稻种植丰度与 Sentinel-1A 提取结果具有良好的空间一致性，能够体现早稻在像元中的混合比例，特别是对高丰度区反演结果较好。但在低丰度地区，模型反演结果存在一定高估现象，说明丰度反演模型对低

丰度早稻像元敏感度较低。上述结果也与图 6 结果吻合，其原因可能是高丰度早稻样本点占比较高，因此在高值区反演结果较好。结合 3.2 节早稻种植区域提取结果，得到随机森林算法估算的南昌县早稻种植面积为 $5.65 \times 10^4 \text{ hm}^2$ ，对比统计年鉴结果 ($5.21 \times 10^4 \text{ hm}^2$)，可知基于 MODIS 数据与随机森林算法提取的验证区早稻面积精度为 91.6%。



注：MAE 为平均绝对误差；RMSE 为均方根误差； R^2 为决定系数。
Note: MAE is mean absolute error; RMSE is root mean square error; R^2 is determination coefficient.

图 6 基于早稻种植丰度反演模型的精度验证
Fig.6 Accuracy validation based on the early rice planting abundance inversion model

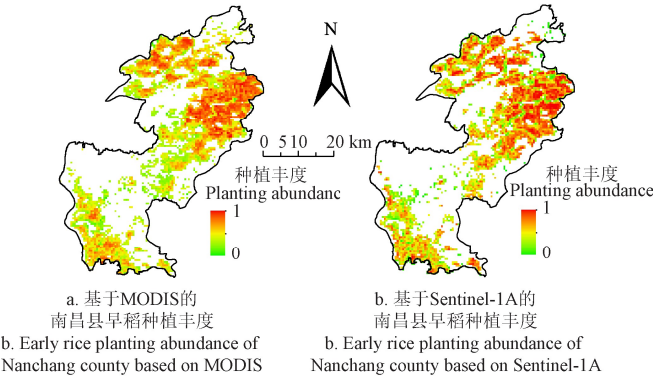


图 7 MODIS 与 Sentinel-1A 提取的 2020 年南昌县早稻丰度
Fig.7 Early rice planting abundance in 2020 derived from the MODIS and the Sentinel-1A product in Nanchang County

3.4 江西省早稻面积提取结果

利用早稻种植区域提取模型及早稻丰度反演模型，得到 2020 年江西省早稻种植分布及丰度图。由图 8 可知，江西省早稻主要种植区域为环鄱阳湖地区及赣江沿岸地区，赣南地区则较为零散且高丰度种植区较少，该特征也与江西省水稻种植的分布特征相似^[10,35]。根据上述方法，提取得到的 2020 年全省早稻面积为 $1.124\times10^6\text{ hm}^2$ ，与江西省统计局统计结果相比 ($1.218\times10^6\text{ hm}^2$)，精度为 92.33% (表 2)。

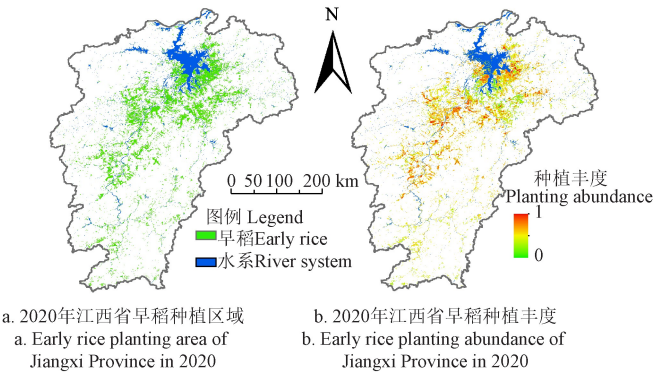


图 8 2020 年江西省早稻种植空间分布及早稻丰度图
Fig.8 Spatial distribution of early rice planting area and planting abundance in Jiangxi Province in 2020

表 2 研究区早稻主产县(市)面积及提取精度
Table 2 Early rice planting area and extracted accuracy of the main production county in study area

行政区 District	面积 Area/hm ²		精度 Accuracy/%	所在区域 Region
	统计 Statistical	提取 Extracted		
江西省 Jiangxi Province	1 217 500	1 124 100	92.33	—
鄱阳县 Poyan county	76 200	72 000	94.49	赣北
南昌县 Nanchang county	52 100	56 500	91.55	赣北
丰城市 Fengcheng city	72 000	68 700	95.42	赣北
余干县 Yugan county	62 300	66 300	93.58	赣北
临川区 Linchuan district	40 700	43 100	94.10	赣中
泰和县 Taihe county	38 500	39 800	96.62	赣中
吉安县 Jian county	31 200	30 100	96.47	赣中
兴国县 Xingguo county	22 600	17 600	77.88	赣南
宁都县 Ningdu county	22 200	12 200	54.95	赣南
于都县 Yudu county	19900	13900	69.85	赣南
信丰县 Xinfeng county	16500	15000	90.90	赣南

此外，进一步对比江西省不同地市水稻主产县的早稻统计面积与提取结果，可以发现，赣北和赣中地区的南昌县、丰城市、余干县、临川区、泰和县等早稻主产县的面积提取精度均高于 90%；而赣南地区主产县早稻面积提取面积偏少，且精度低于赣北和赣中地区。

上述结果表明，基于早稻播种至拔节期的 MODIS 指数数据，运用随机森林算法能够提取江西省早稻种植区域、种植丰度及面积等信息，解决了阈值选取合理性、MODIS 数据存在混合像元及前人研究中水稻信息提取时效性等问题。通过对全省及不同区域县市早稻面积的估算和验证，表明利用该方法提取江西省早稻种植信息，特别是赣北和赣中早稻主产区域是可行的。但在赣南等地区，其提取面积与统计数据仍存在一定误差。从其他地区

种植区域提取与丰度反演结果来看，算法本身不是造成误差过大的主要原因。对比赣南与其他地区地形后，可以发现，宁都县、于都县等赣南主产县多以丘陵、山地等地形为主，且水稻种植多在梯田及山间耕地，分布较为破碎，使得像元中早稻比例偏低，无法准确识别^[36-37]。不过，由于赣南地区早稻种植面积在全省占比为 14.84%，低于其国土面积在全省的占比 (23.59%)，且早稻种植面积大于 $1.5\times10^4\text{ hm}^2$ 的县仅有 5 个，因此利用本研究方法估算江西省早稻面积，特别是主产区早稻面积的影响有限。

4 结 论

本研究利用水稻生长早期 MODIS 的 3 种指数数据，结合随机森林算法构建早稻种植区域提取与丰度反演模型，提取了江西省早稻种植信息，并利用 Sentinel-1A 资料进行精度验证。结果显示，利用随机森林算法能够有效提取江西省的早稻像元。利用验证样本评估分类精度可知，早稻分类的制图精度和用户精度分别达到 92.04% 和 91.23%，模型分类精度为 93.18%，Kappa 系数为 0.915，分类精度高。同时，对比 Sentinel-1A 资料，模型提取结果可基本再现验证区早稻种植的空间分布，特别是对成片的早稻种植区域提取更具优势。利用随机森林方法及 MODIS 指数构建的早稻丰度反演模型也能够对验证区早稻种植丰度进行有效估算，实现对早稻种植面积的提取。结合早稻种植区域提取模型与丰度反演模型得到的 2020 年江西省早稻种植面积为 $1.124\times10^6\text{ hm}^2$ ，与统计结果相比，精度达到 92.33%。该方法在早稻种植区域提取的基础上考虑了利用随机森林方法获取混合像元的早稻种植丰度，降低了混合像元对面积估算的影响，并取得较高精度，且操作简单，计算高效。因此，本研究结果能够用于开展江西省早稻种植信息提取的业务应用与决策服务。

值得提出的是，水稻生长早期，不同类型遥感指数的时序变化特征是随机森林模型构建的理论基础，但江西省纬度跨越相对较大，不同地域早稻生育期也存在一定差异，增加了早稻种植信息遥感监测的难度。其次，江西省耕地情况复杂，特别是赣南地区多为梯田及山间耕地，其对稻田的遥感指数值影响较大，造成该地区提取精度低于赣北和赣中地区。在后续研究工作中，可进一步增加波段反射率及其他相关指数特征，并选取最优特征，以提升模型运算效率与精确度。

[参 考 文 献]

[1] 陈雨思, 李丹, 黎臻, 等. 多时相 MODIS 影像的黑龙江省水稻种植面积提取[J]. 农业工程学报, 2020, 36(23): 201-208.

Chen Yusi, Li Dan, Li Zhen, et al. Rice planting area extraction based on multi-temporal MODIS images in Heilongjiang Province of China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(23): 201-208. (in Chinese with English abstract)

[2] 邓刚, 唐志光, 李朝奎, 等. 基于 MODIS 时序数据的湖

- 南省水稻种植面积提取及时空变化分析[J]. 国土资源遥感, 2020, 32(2): 177-185.
- Deng Gang, Tang Zhiguang, Li Chaokui, et al. Extraction and analysis of spatiotemporal variation of rice planting area in Hunan Province based on MODIS time-series data[J]. Remote Sensing For Land & Resources, 2020, 32(2): 177-185. (in Chinese with English abstract)
- [3] Gumma M K, Nelson A, Thenkabail P S, et al. Mapping rice areas of South Asia using MODIS multitemporal data[J]. Journal of Applied Remote Sensing, 2011, 5(1): 53547.
- [4] 平跃鹏, 臧淑英. 基于 MODIS 时间序列及物候特征的农作物分类[J]. 自然资源学报, 2016, 31(3): 503-513.
- Ping Yuepeng, Zang Shuying. Crop identification based on MODIS NDVI time-series data and phenological characteristics[J]. Journal of Natural Resources, 2016, 31(3): 503-513. (in Chinese with English abstract)
- [5] Qiu B W, Lu D F, Tang Z H, et al. Automatic and adaptive paddy rice mapping using Landsat images: Case study in Songnen Plain in Northeast China[J]. Science of the Total Environment, 2017, 598: 581-592.
- [6] Zhang M, Zhang H Q, Li X Y, et al. Classification of paddy rice using a stacked generalization approach and the spectral mixture method based on MODIS time series[J]. IEEE journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2264-2275.
- [7] 李根, 景元书, 王琳, 等. 基于 MODIS 时序植被指数和线性光谱混合模型的水稻面积提取[J]. 大气科学学报, 2014, 37(1): 119-126.
- Li Gen, Jing Yuanshu, Wang Lin, et al. Extraction of paddy planting areas based on MODIS vegetation index time series and linear spectral mixture model[J]. Transactions of Atmospheric Sciences, 2014, 37(1): 119-126. (in Chinese with English abstract)
- [8] Xiao X M, Stephen B, Liu J Y, et al. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images[J]. Remote Sensing of Environment, 2005, 95(4): 480-492.
- [9] Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31.
- [10] 卿娟. 江西省水稻种植制度遥感评价[D]. 南昌: 江西师范大学, 2020.
- Qing Juan. Remote Sensing Evaluation of Rice Planting System in Jiangxi Province[D]. Nanchang: Jiangxi Normal University, 2020. (in Chinese with English abstract)
- [11] 刘杰, 刘吉凯, 安晶晶, 等. 基于时序 Landsat 8 OLI 多特征与随机森林算法的作物精细分类研究[J]. 干旱地区农业研究, 2020, 38(3): 281-288.
- Liu Jie, Liu Jikai, An Jingjing, et al. Precise crop classification based on multi-features from time-series Landsat 8 OLI images and random forest algorithm[J]. Agricultural Research in the Arid Areas, 2020, 38(3): 281-288. (in Chinese with English abstract)
- [12] Wang Y, Zang S Y, Tian Y. Mapping paddy rice with the random forest algorithm using MODIS and SMAP time series[J]. Chaos, Solitons and Fractals, 2020, 140: 110116.
- [13] 雷小雨, 卓莉, 叶涛, 等. 基于时差特征与随机森林的水稻种植面积提取[J]. 遥感技术与应用, 2016, 31(6): 1140-1149.
- Lei Xiaoyu, Zhuo Li, Ye Tao, et al. A paddy rice planting area extraction method using random forest based on multi-temporal differences[J]. Remote Sensing Technology and Application, 2016, 31(6): 1140-1149. (in Chinese with English abstract)
- [14] 张雅琦. 基于物候特征与随机森林的水稻种植制度提取分析[D]. 北京: 中国地质大学, 2021.
- Zhang Yaqi. Extraction and Analysis of Rice Planting System Based on Phenological Characteristics and Random Forest[D]. Beijing: China University of Geosciences, 2021. (in Chinese with English abstract)
- [15] 王利民, 刘佳, 姚保民, 等. 综合 NDVI 时序特征的冬小麦混合像元分解及面积估算[J]. 中国农学通报, 2019, 35(27): 23-33.
- Wang Limin, Liu Jia, Yao Baomin, et al. Winter wheat mixed pixel decomposition based on NDVI time-order characters and area estimation[J]. Chinese Agricultural Science Bulletin, 2019, 35(27): 23-33. (in Chinese with English abstract)
- [16] 王连喜, 徐胜男, 李琪, 等. 基于决策树和混合像元分解的江苏省冬小麦种植面积提取[J]. 农业工程学报, 2016, 32(5): 182-187.
- Wang Lianxi, Xu Shengnan, Li Qi, et al. Extraction of winter wheat planted area in Jiangsu province using decision tree and mixed-pixel methods[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(5): 182-187. (in Chinese with English abstract)
- [17] 熊勤学, 胡佩敏. 基于 HJ 卫星混合像元分解法的湖北省四湖地区夏收作物种植信息提取[J]. 长江流域资源与环境, 2014, 23(6): 869-874.
- Xiong Qinxue, Hu Peimin. Extracting planting information of summer harvesting crops in Shihu region from HJ CCD data using unmixing algorithm data[J]. Resources and Environment in the Yangtze Basin, 2014, 23(6): 869-874. (in Chinese with English abstract)
- [18] 刘轶青, 陶建斌, 陈曦, 等. 一种基于时序遥感影像混合像元分解的耕地种植强度估算方法: 以湖北省为例[J]. 华中师范大学学报(自然科学版), 2020, 54(1): 96-103.
- Liu Yiqing, Tao Jianbin, Chen Xi, et al. A method for estimating cropping intensity based on mixed pixel unmixing using time series remote sensing images: A case study of Hubei Province[J]. Journal of Huazhong Normal University (Natural Sciences), 2020, 54(1): 96-103. (in Chinese with English abstract)

- [19] 蓝金辉, 邹金霖, 郝彦爽, 等. 高光谱遥感影像混合像元分解研究进展[J]. 遥感学报, 2018, 22(1): 13-27.
Lan Jinhui, Zou Jinlin, Hao Yanshuang, et al. Research progress on unmixing of hyperspectral remote sensing imagery[J]. Journal of Remote Sensing, 2018, 22(1): 13-27. (in Chinese with English abstract)
- [20] 孙兴亮, 郝晓华, 王建, 等. 基于光谱-环境随机森林回归模型的 MODIS 积雪面积比例反演研究[J]. 冰川冻土, 2022, 44(1): 1-12.
Sun Xingliang, Hao Xiaohua, Wang Jian, et al. Research on retrieval of MODIS fraction snow cover based on spectral environmental random forest regression model[J]. Journal of Glaciology and Geocryology, 2022, 44(1): 1-12. (in Chinese with English abstract)
- [21] Lopatin J, Dolos K, Hernández H J, et al. Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile[J]. Remote Sensing of Environment, 2016, 173: 200-210.
- [22] 胡琼. 基于时序MODIS影像的农作物遥感识别方法研究[D]. 北京: 中国农业科学院, 2018.
Hu Qiong. Research on Methods of Crop Identification Using MODIS Time Series Data[D]. Beijing: Chinese Academy of Agricultural Sciences, 2018. (in Chinese with English abstract)
- [23] Huang X, Schneider A, Friedl M A. Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights[J]. Remote Sensing of Environment, 2016, 175: 92-108.
- [24] Hu Q, Yin H, Friedl M A, et al. Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates[J]. Remote Sensing of Environment, 2021, 258: 112365.
- [25] 梁瀚月. 基于 Sentinel-1 数据的江西省南昌县早稻信息提取研究[D]. 成都: 成都理工大学, 2017.
Liang Hanyue. Extraction of Early Paddy Rice Information in Nanchang County of Jiangxi Province Based on Sentinel-1 Data[D]. Chengdu: Chengdu University of Technology, 2017. (in Chinese with English abstract)
- [26] 段里成, 张坤, 郭瑞鸽, 等. 江西省双季早稻直播适宜播种期的区域划分[J]. 中国农业气象, 2021, 42(2): 112-122.
Duan Licheng, Zhang Kun, Guo Ruige, et al. Regional division of suitable direct sowing date of double season early rice in jiangxi province[J]. Chinese Journal of Agrometeorology, 2021, 42(2): 112-122. (in Chinese with English abstract)
- [27] 张馨予, 蔡志文, 杨靖雅, 等. 时序滤波对农作物遥感识别的影响[J]. 农业工程学报, 2022, 38(4): 215-224.
Zhang Xinyu, Cai Zhiwen, Yang Jingya, et al. Impacts of temporal smoothing methods on crop type identification[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(4): 215-224. (in Chinese with English abstract)
- [28] 刘警鉴, 李洪忠, 华瑾, 等. 基于 Sentinel-1A 数据的临高县早稻面积提取[J]. 国土资源遥感, 2020, 32(1): 191-199.
Liu Jingjian, Li Hongzhong, Hua Cui, et al. Extraction of early paddy rice area in Lingao County based on Sentinel-1A data[J]. Remote Sensing for Land and Resources 2020, 32(1): 191-199. (in Chinese with English abstract)
- [29] 姚杰鹏, 杨磊库, 陈探, 等. 基于 Sentinel-1, 2 和 Landsat 8 时序影像的鄱阳湖湿地连续变化监测研究[J]. 遥感技术与应用, 2021, 36(4): 760-776.
Yao Jiepeng, Yang Leiku, Chen Tan, et al. Consecutive Monitoring of the Poyang Lake Wetland by Integrating Sentinel-2 with Sentinel-1 and Landsat 8 Data[J]. Remote Sensing Technology and Application, 2021, 36(4): 760-776. (in Chinese with English abstract)
- [30] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [31] 王耀民, 陈皓锐, 陈俊英, 等. 光谱指数筛选方法与统计回归算法结合的水稻估产模型对比[J]. 农业工程学报, 2021, 37(21): 208-216.
Wang Yaomin, Chen Haorui, Chen Junying, et al. Comparison of rice yield estimation model combining spectral index screening method and statistical regression algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(21): 208-216. (in Chinese with English abstract)
- [32] 黄翀, 许照鑫, 张晨晨, 等. 基于 Sentinel-1 数据时序特征的热带地区水稻种植结构提取方法[J]. 农业工程学报, 2020, 36(9): 177-184.
Huang Chong, Xu Zhaoxin, Zhang Chenchen, et al. Extraction of rice planting structure in tropical region based on Sentinel-1 temporal features integration[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(9): 177-184. (in Chinese with English abstract)
- [33] 刘峻明, 周舟, 和晓彤, 等. 频率直方图与植被指数结合的冬小麦遥感产量估测[J]. 农业工程学报, 2021, 37(23): 145-152.
Liu Junming, Zhou Zhou, He Xiaotong, et al. Estimating winter wheat yield under frequency histogram and vegetation index using remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(23): 145-152. (in Chinese with English abstract)
- [34] 杨北萍, 陈圣波, 于海洋, 等. 基于随机森林回归方法的水稻产量遥感估算[J]. 中国农业大学学报, 2020, 25(6): 26-34.
Yang Beiping, Chen Shengbo, Yu Haiyang, et al. Remote sensing estimation of rice yield based on random forest regression method[J]. Journal of China Agricultural University, 2020, 25(6): 26-34. (in Chinese with English abstract)
- [35] 陈建军, 黄淑娥, 景元书. 基于 EOS/MODIS 资料的江西

- 省水稻长势遥感监测[J]. 江苏农业科学, 2012, 40(6): 302-305.
- Chen Jianjun, Huang Shu'e, Jing Yuanshu. Study on monitoring of paddy rice growth based on EOS/MODIS data in Jiangxi Province[J]. Jiangsu Agricultural Sciences. 2012, 40(6): 302-305. (in Chinese with English abstract)
- [36] 耿利宁, 景元书, 杨沈斌, 等. 基于 FastICA 算法和 MODIS 数据的水稻面积提取[J]. 大气科学学报, 2015, 38(6): 819-826.
- Geng Lining, Jing Yuanshu, Yang Shenbin, et al. Extracting the paddy rice area from MODIS imagery by FastICA algorithm[J]. Transactions of Atmospheric Sciences, 2015, 38(6): 819-826. (in Chinese with English abstract)
- [37] Mosleh M K, Hassan Q K, Chowdhury E H. Application of remote sensors in mapping rice area and forecasting its production: A review[J]. Sensors (Basel), 2015, 15(1): 769-791.

Extracting planting information of early rice using MODIS index and random forest in Jiangxi Province, China

Lin Zhijian¹, Yao Junmeng¹, Su Xiaoping², Cai Zhe¹, Liu Dan^{1,3*}

(1. Jiangxi Agricultural Meteorological Center, Nanchang 330096, China; 2. Nanchang Meteorological Bureau, Nanchang 330008, China; 3. Jiangxi Provincial Institute of Meteorological Science, Nanchang 330008, China)

Abstract: Extracting planting information of rice as early as possible can be great significance in provincial agricultural production for the food security. MODIS data has been demonstrated to be superior in extracting planting information of rice at large scale due to short observation period and wide swath and easy image acquisition. The method of MODIS index during early growth period of rice combined with threshold value was usually used in the conventional provincial decision-making service to ensure the timeliness of service and convenient operation. However, the method is largely influenced by human and phenology difference of rice in various regions leading to highly subjective and poor stability. In addition, the mixed pixels were likely causing misestimation of the MODIS product. Random forest algorithm can make up the deficiency of threshold method due to the characteristics of less manual intervention and difficult over-fitting. In the study, an extraction method of rice planting information was proposed using MODIS index and random forest during early growth period of rice. Firstly, Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI) and Land Surface Water Index (LSWI) were selected to extract planting information based on the variation characteristics of MODIS index from seeding to jointing stage in Jiangxi Province, and then feature data sets were constructed to model. The models were built to extract planting area of early rice and to inverse planting abundance using random forest algorithm. Finally, the accuracy of the planting and the abundance maps of early rice were validated by the verification samples from the measured points, validation sample region obtained by the Sentinel-1A image and the statistical data from Jiangxi Provincial Bureau of Statistics. The results showed that using MODIS index during early growth period of rice and random forest was an effective way to extract plant information. The classification accuracy of early rice planting area extraction model was 93.18% with the Kappa coefficient of 0.915, and the mapping accuracy and user accuracy were 92.04% and 91.23%, respectively when matching the verification samples. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), decision coefficient of abundance inversion model were 0.07, 0.104 and 0.855, respectively, and the better performance of the abundance inversion model was achieved in high abundance planting areas. The spatial distribution characteristics of early rice planting area and abundance were consistent with validation sample region. Compared with the statistical data, the model accuracy of early rice area was 92.33%. This method can ensure the timeliness of the service and solve the problem that the extraction of rice planting area in the conventional provincial decision-making service is greatly affected by the problems of rationality of threshold selection and mixed pixel, also have no complex operation. The finding can provide a reference to extract planting area of early rice during early growth period.

Keywords: remote sensing; algorithm; early rice identification; MODIS index; random forest