

考虑种植方式的极化 SAR 水稻植株密度反演方法*

徐静璇^{1,2}, 李坤^{1,2,3†}, 秦怡⁴, 张渤琳^{1,2}, 张风丽^{1,3}

(1 中国科学院空天信息创新研究院, 北京 100094; 2 中国科学院大学, 北京 100049;

3 中科卫星应用德清研究院, 浙江 德清 313200; 4 中国地质大学, 北京 100083)

(2023 年 1 月 4 日收稿; 2023 年 4 月 6 日收修改稿)

Xu J X, Li K, Qin Y, et al. Rice plant density inversion using polarimetric SAR considering planting methods[J]. Journal of University of Chinese Academy of Sciences, 2025, 42(2): 167-175. DOI:10.7523/j.ucas.2023.032.

摘要 植株密度是影响水稻生长发育和产量的重要因素。由于种植方式影响水稻田间分布特征, 进而引起密度的差异, 所以在水稻植株密度反演中, 需要考虑种植方式的影响。合成孔径雷达(SAR)因其全天时、全天候、穿透性等优势, 已被证明是水稻监测的重要手段之一。极化 SAR 不仅具有传统 SAR 的优势, 还对目标结构、分布特征等非常敏感, 在水稻植株密度反演中具有更大的应用潜力。基于 RADARSAT-2 全极化 SAR 数据, 充分挖掘极化信息, 并考虑插秧、撒播种植方式, 利用弹性网络模型实现水稻植株密度反演。结果表明, 极化 SAR 在水稻植株密度反演中具有较好的效果, 插秧和撒播田植株密度的 RMSE 分别为 25 和 39 株/m²; 而且与不考虑种植方式的反演结果相比, 精度提高 30% 以上。

关键词 极化 SAR; 水稻; 种植方式; 密度反演

中图分类号: S127; TP79; TP722.6 文献标志码: A DOI:10.7523/j.ucas.2023.032

Rice plant density inversion using polarimetric SAR considering planting methods

XU Jingxuan^{1,2}, LI Kun^{1,2,3}, QIN Yi⁴, ZHANG Bolin^{1,2}, ZHANG Fengli^{1,3}

(1 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

2 University of Chinese Academy of Sciences, Beijing 100049, China; 3 Deqing Academy of Satellite Applications,

Deqing 313200, Zhejiang, China; 4 China University of Geosciences, Beijing 100083, China)

Abstract Plant density is an important factor affecting rice growth and yield. Because planting methods affect the distribution characteristics of rice in the field, and then cause the difference in density, it is necessary to consider the influence of planting methods in the inversion of rice plant density. Synthetic aperture radar (SAR) has been proved to be one of the important means of rice monitoring due to its advantages of all-day, all-weather and penetrability. Polarimetric SAR not only

* 国家自然科学基金(41871272)资助

† 通信作者, E-mail: likun201032@aircas.ac.cn

has the advantages of traditional SAR, but also is very sensitive to target structure and distribution characteristics. It has greater application potential in rice plant density inversion. Based on RADARSAT-2 full polarimetric SAR data, this study fully excavates the polarimetric information, considers the planting methods of transplanting and sowing, and uses the elastic network model to realize the inversion of rice plant density. The results show that polarimetric SAR has a good effect on the inversion of rice plant density. The RMSE of plant density in transplanting and sowing fields is 25 and 39 plants/m² respectively. Compared with the inversion results without considering planting methods, the accuracy is improved by more than 30%.

Keywords polarimetric SAR; rice; planting methods; plant density inversion

水稻是我国最重要的粮食作物之一,实时高效地进行水稻生长监测,对我国农业政策的制定、生态资源的合理配置以及保障国家粮食安全具有重要意义。卫星遥感因其具有大范围同步观测、时效性高、周期性等特点已被广泛应用于农作物生长监测。水稻是一种喜高温多湿的作物,生长季内多云雨,光学遥感数据的获取存在很大困难,合成孔径雷达(synthetic aperture radar, SAR)具有全天时、全天候、可穿透云雾且大范围成像的能力^[1],已成为水稻监测的重要手段之一。极化 SAR 同时获取目标的回波强度和相位信息,能够为水稻监测提供更加丰富的信息^[2-5]。

水稻植株密度是影响水稻生长发育和产量的重要因素之一,通过实时精确估计水稻植株密度不仅能及时监测水稻的健康状况,还能为产量估算等提供数据输入和参考。目前,激光雷达、可见光成像、微波遥感等多种技术已经应用于农作物植株密度反演中。Yang 和 Lin^[6]利用机载激光雷达和航拍图像基于核密度估计原理进行作物密度估计,机载激光雷达工作灵活性高但研究区域范围受到限制,不具备大范围同步观测的能力。单捷等^[7]以高分一号卫星(GF-1)影像为数据源,选取多种遥感监测指标与冬小麦密度进行回归分析,得到密度的反演精度在拔节期最高,回归模型的决定系数为 0.543 8。邹楠^[8]利用高光谱遥感探讨不同密度下春玉米冠层机构特征,建立不同密度下春玉米冠层的高光谱模型。由于水稻喜高温多湿的特性,其生长季内多云雨,给光学遥感数据的获取带来很大困难,因此国内外学者多利用微波遥感数据进行水稻植株密度反演研究。张远^[9]基于 L 波段 ALOS/PALSAR 数据 HH 极化的雷达后向散射系数 σ^0 和改进的一阶辐射传输模型,结合遗传算法进行水稻乳熟期植株密度反演,模拟误差最大不超过 30 株/m²。Inoue 等^[10]研究

发现 C 波段 RADARSAT-2 数据的 VH 极化后向散射系数 σ^0 与植株密度具有较好的相关性。贾明权等^[11-12]利用微波散射计获取水稻后向散射系数 σ^0 ,在此基础上开展水稻微波散射机理建模研究,发现水稻植株密度与后向散射系数具有很强的相关性,并利用神经网络实现水稻植株密度反演。Zhang 等^[13]利用 RADARSAT-2 数据 HH 极化 σ^0 结合水稻微波散射模型,实现了水稻拔节、抽穗、成熟期的植株密度反演。景卓鑫等^[14]将 RADARSAT-2 数据的 HH 和 HV 后向散射系数与神经网络模型结合进行水稻密度、株高、叶面积指数的反演。这些研究都说明 SAR 在水稻植株密度反演中具有有效性,且不同极化方式对于水稻植株密度反演具有很大的应用潜力。但是,目前大多数水稻反演研究都没有考虑不同种植方式的影响。不同种植方式造成的植株田间分布特征的差异,会对其雷达响应特征造成很大的影响。张远^[9]研究指出水稻种植方式直接影响田间水稻植株分布,从而影响水稻植株密度,同时水稻植株分布造成的田间结构差异也会引起雷达响应特征的不同。因此,在水稻植株密度反演过程中,需要考虑不同种植方式的差异。

另一方面,目前大多数研究主要是利用雷达后向散射系数进行水稻植株密度反演,而后向散射系数作为雷达信号与目标相互作用的结果,易受到降雨、灌溉、下垫面等因素的干扰^[15],尤其是在水稻幼苗期植株较矮小时。全极化 SAR 不仅能够获取目标与雷达信号相互作用的结果(后向散射系数),还能够获取回波的相位信息,进而构建极化散射矩阵及二阶统计量极化相干/协方差矩阵,并利用极化分解技术来反映目标与雷达信号的相互作用机理。通过多种不同的极化分解方法,能够获得 100 多个极化特征参数,这使得数据量变得繁多复杂,若仅使用简单的阈值方法分析,

在加大任务量的同时也难以保证精度。随着计算机技术的发展,机器学习、人工智能等模型方法越来越受到研究者的青睐,其快速的数据处理能力和精准的计算能力被应用于广泛的研究领域。由于机器学习方法在处理大数据上具有较强的优势,越来越多的文献将 Ridge 回归、LASSO 回归、弹性网络、主成分回归、偏最小二乘回归、随机森林及决策树等机器学习方法引入用于解决高维数据问题^[16-18]。弹性网络 (elastic net) 模型由斯坦福大学的 Zou 和 Hastie^[19] 于 2005 年提出,作为一种同时使用 L_1 和 L_2 正则项训练的线性回归模型,结合了 LASSO 回归能筛选重要特征变量和 Ridge 回归防止多重共线性的优点^[20-21]。对于观测数量远小于预测变量,以及存在较严重的多重共线性的情况,弹性网络既能实现特征变量的选择又能进行良好的数据压缩,从而获得更加合理可靠的模型结果^[22-24]。

鉴于此,本文基于 RADARSAT-2 全极化 SAR 数据,提取雷达后向散射系数和多维极化特征参数,充分挖掘雷达极化信息;同时考虑到水稻种植方式的差异对植株密度反演的影响,在水稻密度反演过程中,首先考虑 2 类种植方式(插秧、撒播)的差异,构建随机森林模型实现 2 类种植方式水稻的分类识别;然后针对 2 类不同种植方式的水稻,分别利用弹性网络构建反演模型,实现水稻植株密度反演;最终获得研究区内插秧田和撒播田的水稻幼苗期植株密度分布图,为区域尺度的水稻植株密度反演提供一种技术手段。

1 材料与方法

1.1 研究区概况

研究区域选在江苏省淮安市金湖县附近(图 1),该地区地处江苏省中部偏西、淮河流域中下游,西邻洪泽湖,主要包含淮安市的盱眙县、洪泽县和金湖县交界处附近的各部分区域,是我国东部主要的水稻种植区之一。研究区的整体面积约为 40 km×30 km,中心点的地理坐标为 118°59'55.14"E,33°07'05"N,地势西高东低但总体较为平坦。从气候方面来看,研究区位于苏北灌溉总渠以南,属于亚热带湿润季风气候,因此受季风气候影响,四季分明,冬冷夏热,秋高气爽,春温多变,年平均气温约为 14.6℃,光热充足。雨量充沛,雨热同季,年平均降水量约为 1 085 mm,且河湖众多,水网密布,为水稻生长创造了适宜的环

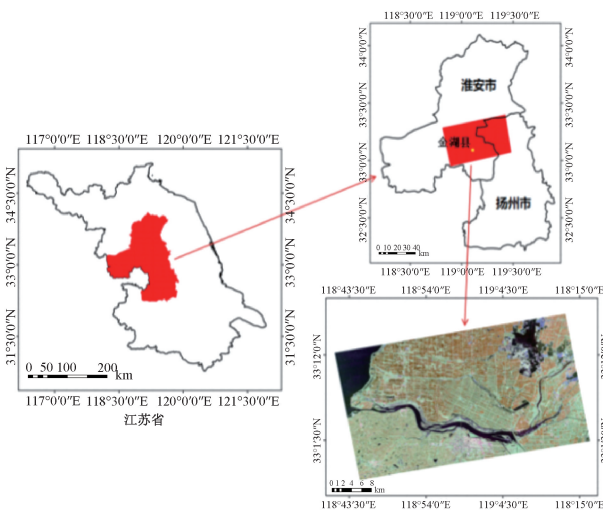


图 1 研究区位置图

Fig. 1 Location map of study area

境。由于所处的地理气候条件影响,研究区水稻种植模式主要为一年一季。经过野外实地调查发现,该地区的水稻种植方式存在插秧和撒播 2 种。

1.2 数据获取与处理

1.2.1 雷达数据预处理

本研究利用 C 波段 RADARSAT-2 全极化 SAR 数据实现水稻植株密度反演。根据研究区的水稻物候历,获取 2012 年 6 月 27 日 RADARSAT-2 全极化单视复 (single-look complex, SLC) 数据,对应于水稻幼苗期,数据空间分辨率(方位向×距离向)为 5.2 m×7.6 m,近地点和远地点入射角分别为 38°和 41°。

在进行极化 SAR 水稻密度反演之前,首先对极化 SAR 数据进行预处理,包括辐射定标、几何校正和图像滤波等。数据预处理利用欧洲太空局 SAR 数据处理软件 SNAP 9.0 (<http://step.esa.int/main/download/snap-download/>) 和 PolSARpro 6.0 (<https://step.esa.int/main/download/polsarpro-v6-0-biomass-edition-toolbox-download/>) 完成。图像滤波采用的方法是增强 Lee 滤波,滤波窗口大小选择 5×5。经过处理后获得 HH、VV、VH、HV 极化的后向散射系数图像,像元大小为 5 m×5 m。除获取不同极化的后向散射系数以外,还生成了相干矩阵 T_3 ,并提取 58 个极化特征参数,包括 Yamaguchi3/4 分量极化分解分量、An&Yang3/4 分量极化分解分量、Freeman-Durden 极化分解分量、vanZyl 极化分解分量、Cloude-Pottier 分解分量等,详见表 1。所有极化特征参数图像的采样间隔也是 5 m×5 m。

表 1 获取的研究区雷达参数表

Table 1 Acquired radar parameters in the study area		
雷达参数	数量	具体参数
HH/HV/VH/VV 极化后向散射系数	4	HH、HV、VH、VV
Yamaguchi3/4 极化分解分量	7	Yamaguchi3 _ Dbl、Yamaguchi3 _ Odd、Yamaguchi3 _ Vol、Yamaguchi4 _ Y4 _ Dbl、Yamaguchi4 _ Y4_Hlx、Yamaguchi4 _ Y4_Odd、Yamaguchi4 _ Y4_Vol
An&Yang3/4 极化分解分量	7	An_Yang3_Dbl、An_Yang3_Odd、An_Yang3_Vol、An_Yang4_Dbl、An_Yang4_Hlx、An_Yang4_Odd、An_Yang4_Vol
Touzi 极化分解分量	16	TSVM_alpha_s、TSVM_alpha_s1、TSVM_alpha_s2、TSVM_alpha_s3、TSVM_phi_s、TSVM_phi_s1、TSVM_phi_s2、TSVM_phi_s3、TSVM_psi、TSVM_psi1、TSVM_psi2、TSVM_psi3、TSVM_tau_m、TSVM_tau_m1、TSVM_tau_m2、TSVM_tau_m3
vanZyl 极化分解分量	3	VanZyl3_Dbl、VanZyl3_Odd、VanZyl3_Vol
Freeman-Durden 极化分解分量	5	Freeman_Dbl、Freeman_Odd、Freeman_Vol、Freeman2_Ground、Freeman2_Vol
相干矩阵元素	9	T_{11} 、 T_{12} _imag、 T_{12} _real、 T_{13} _imag、 T_{13} _real、 T_{22} 、 T_{23} _imag、 T_{23} _real、 T_{33}
H/A/Alpha 极化分解特征向量	11	Alpha、anisotropy、beta、combination _ 1mH1mA、combination _ 1mHA、combination _ H1mA、combination _ HA、delta、entropy、gamma、lambda
总计	62	

1.2.2 地面实测及统计数据

研究区存在 2 种水稻种植方式,一是插秧、二是撒播。对于插秧种植,需要提前在秧田中培育水稻幼苗,待水稻种子发芽并长至幼苗期后可通过人工或机插秧的方式进行移栽。移栽后的插秧田中秧苗一般呈现行列分明的均匀分布,行距约为 30 cm,墩距约为 15 cm。插秧田的水稻品种主要为籼稻。2012 年,研究区内插秧田的种植日期大约在 6 月 20 日,成熟收割日期在 10 月 25 日前后,全生育期约为 120 d。对于撒播种植,水稻种子或小秧苗直接以撒播形式播种,因此撒播田水稻幼苗无明显行列规律,呈随机分布。撒播田的水稻品种主要为粳稻。2012 年,研究区内撒播田的种植时间大约在 6 月 16 日,成熟收割时间在 11 月 15 日前后,全生育期约为 150 d。

在获取的 RADARSAT-2 数据卫星过境的同时进行地面同步测量。为保证每块样本田能有足

够的像元进行极化特征参数的提取和分析,选择面积在 120 m×120 m 以上的样本田。在研究区内选择 43 块水稻样本田,其中包含 31 块插秧田和 12 块撒播田,调查记录每块水稻样本田的经纬度坐标、水稻品种、种植方式和植株密度(行距、墩距、株/墩、株/m²)等。实验用皮尺测量行距、墩距等参数,行距、墩距的精度为 cm 级,种植密度可以由行距、墩距计算得到,每个水稻样本田内采样 3 次,取平均值作为该点实测值。在研究区内选取均匀分布的城镇区样本点 8 个、森林区样本点 8 个、浅滩区样本点 8 个和水体样本点 9 个,记录每个样本点的经纬度和地物类型。

1.3 研究方法

本研究的总体技术路线如图 2 所示。首先对获取的 RADARSAT-2 全极化 SAR 数据进行预处理,提取雷达后向散射系数和 multidimensional polarization feature parameters;然后在水稻密度反演过程中考虑不同水稻种植方式的差异,构建基于随机森林的不同种植方式水稻分类识别方法;在此基础上,针对 2 类种植方式(插秧、撒播)的水稻田,分别进行斯皮尔曼相关性分析获得植株密度雷达特征参数,并利用弹性网络构建水稻密度反演模型,实现插秧田和撒播田的水稻植株密度反演。

1.3.1 基于随机森林的不同种植方式水稻田分类识别方法

随机森林模型具有处理非线性和高维度数据、去噪和检测特征重要性等优点。随机森林模型作为包含多棵决策树的分类器,其输出类别是由各个树输出类别的众数决定的^[25-26]。随机森林模型主要包括从样本中随机选取小集合样本组成训练集和从样本特征中随机选取子特征组成集合两层“随机”处理。本文使用的随机森林模型基于分类回归树(classification and regression tree,

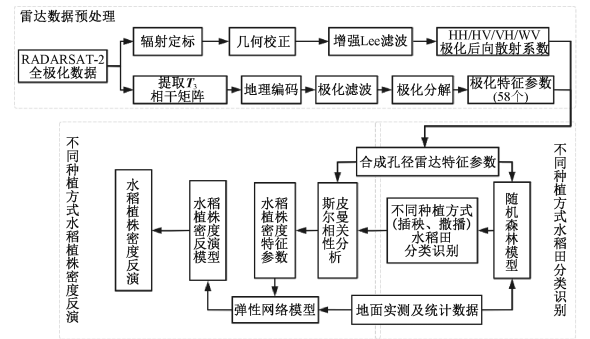


图 2 总体技术路线图

Fig. 2 Flow chart of technical method

CART)构建,采用基尼系数(Gini)进行特征选择,将多棵分类树组成随机森林强分类器^[27]。假设对于样本集 D ,每个样本含有 m 个特征,在样本中有放回的选取子样本集 $d_{11}, d_{22}, \dots, d_{ij}$,且 $d_{ij}, d_{2j}, \dots, d_{ij} \subseteq D, d_{i1}, d_{i2}, \dots, d_{ij} \subseteq M$, 则得到随机森林分类器结果公式如下

$$L = \text{Mode}[f(d_{ij})] \tag{1}$$

式中: $f(d_{ij})$ 为样本集 d_{ij} 在 CART 树中的弱分类器结果, Mode 为多棵树的众数结果, L 则为随机森林强分类器结果。

利用雷达数据预处理得到的后向散射系数和极化分解参数构建随机森林模型进行水稻种植方式分类识别。研究区地面实验获取插秧水稻田、撒播水稻田、城镇、森林、水体、浅滩 6 种地物,共 53 个样本点,每个样本点有雷达特征参数 62 个。将样本数据按照 6 : 4 分配用于随机森林模型的反复训练和测试。样本点中的雷达特征参数作为自变量,对应的地物类型作为因变量,首先构建一个包含 100 棵决策树的随机森林模型,将建模集进行可放回的样本和雷达参数抽取,将每次抽取的特征子集以 Gini 系数缩小最快的特征参数排序方式进行决策树拟合,最终构成含有 100 棵分类树的随机森林模型。将测试自变量输入构建的随机森林模型,得到 100 棵决策树分类结果,取众数作为最终识别类型,对比验证模拟结果和实际观测结果,循环训练,直至所有验证集能进行一次识别。为了判断在样本点有限的情况下采用的模型是否过拟合,对比模型在训练集和测试集上的拟合精度,发现模型在测试集上的精度比训练集上的低 0.2~0.3 左右,在可接受的范围内。利用构建的随机森林模型得到分类结果后,采用众值滤波方法对分类结果进行后处理,去除结果中小碎斑,并将非水稻类别全部掩膜掉,得到最终的插秧、撒播水稻田分布图。

1.3.2 斯皮尔曼相关分析

为提高水稻密度反演的精度,在构建反演模型之前先进行斯皮尔曼相关分析,筛选出与水稻植株密度相关性较高的雷达特征参数。斯皮尔曼相关系数使用单调函数描述 2 个变量间的相关性,是衡量变量间依赖性的非参数指标,常用希腊字母 ρ 表示。对于 2 个随机变量 X 和 Y ,样本容量为 n , n 个原始数据被转换为等级数据,斯皮尔曼相关系数 $\rho^{[28]}$ 可根据下式计算

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{2}$$

式中: x_i 为由原始数据 X 转换而成的等级数据, \bar{x} 为 X 等级数据的平均值, y_i 为由原始数据 Y 转换而成的等级数据, \bar{y} 为 Y 等级数据的平均值。斯皮尔曼相关系数能在不考虑变量的总体分布形态和样本容量大小的情况下很好地表达无量纲或者量纲不同的 2 个变量的相关性。

通过计算水稻植株密度与雷达后向散射系数以及多维极化特征参数的斯皮尔曼相关系数,分析雷达特征参数与水稻植株密度的相关性,并将斯皮尔曼相关系数 $|\rho| > 0.5$ 的雷达特征参数称为植株密度雷达特征参数,用于后续水稻植株密度反演模型的构建。

1.3.3 基于弹性网络的不同种植方式水稻植株密度反演

雷达图像处理得到的后向散射系数和极化分解参数都各自具有特定的物理意义,同时每个参数表达的都是地物结构和水体极化散射信息,具有一定的相关性。因此本研究选择弹性网络模型进行水稻密度反演模型的构建,该模型适合多特征参数下的拟合反演,能提高重要性强的特征参数的权重,同时降低重要性弱的特征参数的权重,甚至降为 0。

Zou 和 Hastie^[19]提出的弹性网络模型结合了 Ridge 回归和 LASSO 回归。Ridge 回归^[29]是在最小二乘法的平方和后添加 L_2 正则项,使每个参数都存在且权重值根据参数重要性赋予不同的值,提高模型稳定性,但只适合处理较少参数的数据。LASSO 回归^[30]添加的是 L_1 正则项,能压缩特征参数的权重甚至为 0,但也会造成过度压缩,且 LASSO 回归要求样本量大于特征参数量。弹性网络代价函数同时使用 L_1 和 L_2 正则项,能够实现多维信息特征和防止模型过拟合。弹性网络的表达式如下:

$$\hat{\beta}(\text{ElasticNet}) = \underset{\beta}{\operatorname{argmin}} \{ \|Y - X\beta\|^2 + \lambda [\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2] \} \tag{3}$$

式中: $\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2$ 为弹性网络的正则项。当 $\alpha = 0$ 时,表达式保留 L_2 正则项为 Ridge 回归,当 $\alpha = 1$ 时,表达式保留 L_1 正则项为 LASSO 回归。

研究区水稻密度样本经过斯皮尔曼相关性筛选得到植株密度雷达特征参数集,其中雷达特征参数作为自变量,对应的植株密度作为因变量。首先构建一个空弹性网络,并对每个自变量赋予初始权重,将训练自变量代入弹性网络得到模拟植株密度值,对比模拟值和对应训练样本对的因变量,将误差分配到每个权重值,使模型模拟值逐步趋近实际观测值,如此循环训练所有的训练集,直到模型误差减小到可接受的范围或者达到设定的训练次数,则得到训练好的密度反演模型。将样本数据按照 6 : 4 分配代入弹性网络模型中进行反复训练和测试,最后可得到较稳定的植株密度反演模型。

最终获得插秧水稻田植株密度反演模型和撒播水稻田植株密度反演模型如下

$$\gamma = -0.091T_{11} - 0.821\text{VanZyl3_Odd} + 57.1\text{HH} - 21.6\text{VV}.$$

(4)

式中: γ 为插秧田植株密度, T_{11} 、 VanZyl3_Odd 、 HH 、 VV 均为雷达特征参数。由式(4)可知,除 T_{11} 、 VanZyl3_Odd 、 HH 、 VV 这 4 个特征参数外,其余的特征参数权重均被赋为 0,并且 HH 、 VV 权重值较大,远大于 T_{11} 和 VanZyl3_Odd 的权重值,表明插秧水稻田植株密度反演中雷达后向散射系数占主导地位, T_{11} 和 VanZyl3_Odd 特征参数起到修饰辅助的作用。

$$\gamma = -0.114\gamma + 2.38\text{TSVM_alpha_s1} + 1.12\text{TSVM_alpha_s2}.$$

(5)

式中: γ 为撒播田植株密度, γ 、 TSVM_alpha_s1 、 TSVM_alpha_s2 均为雷达特征参数,由式(5)可知,除 γ 、 TSVM_alpha_s1 、 TSVM_alpha_s2 这 3 个特征参数外,其余的特征参数权重均被赋予为 0,并且 TSVM_alpha_s1 、 TSVM_alpha_s2 权重值较大,大于 γ 权重值,表明撒播水稻田植株密度反演中 Touzi 极化分解分量占主导地位。

2 结果与分析

2.1 不同种植方式水稻田的分类识别

利用基于随机森林的不同种植方式水稻分类识别方法,得到研究区插秧、撒播 2 类水稻田的分布图,如图 3 所示。通过与地面实验数据对比分析,插秧水稻田识别精度约为 74%,撒播水稻田识别精度约为 71%,总体识别精度约为 75%。从分布图可以看出,在本研究区中,2 类种植方式水稻田的分布具有一定的规律性,其中插秧水稻田

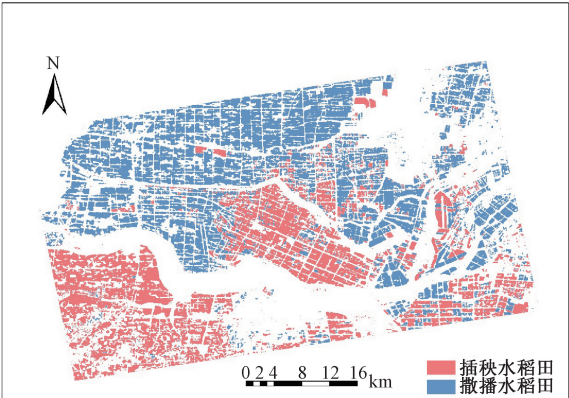


图 3 研究区插秧和撒播水稻田分布图

Fig. 3 Distribution map of transplanting and sowing rice fields in the study area

主要分布在南部区域,撒播水稻田主要分布于北部区域。

2.2 水稻植株密度雷达特征参数

通过斯皮尔曼相关分析,选取与水稻植株密度相关性大于 0.5 的特征参数作为水稻植株密度雷达特征参数,最终得到插秧水稻田植株密度雷达特征参数共计 12 个,如表 2 所示;撒播植株密度雷达特征参数共计 10 个,如表 3 所示。从表 2 可以看出,插秧水稻田植株密度雷达特征参数多为与面散射相关的极化特征参量,这是因为插秧水稻田植株呈现墩状行列分布特征,下垫面裸露比较多,尤其是在水稻生长发育前期。从表 3 可以看出,撒播水稻田植株密度雷达特征参数多为与体散射相关的极化特征参量,这是因为撒播水稻田植株呈现单株随机分布特征,下垫面裸露面积较小,体散射较强。

表 2 插秧水稻田植株密度雷达特征参数

Table 2 Radar characteristic parameters of plant density in transplanting rice field

参数类型	参数名称	相关性系数
An&Yang3/4 极化分解分量	AnY3_Odd	0.59
	AnY4_Odd	0.59
Freeman-Durden	Freeman_Odd	0.53
极化分解分量	Freeman2_Ground	0.67
H/A/Alpha 极化分解分量	λ	0.75
相干矩阵元素	T_{11}	0.63
	T_{12_real}	0.62
vanZyl 极化分解分量	VanZyl3_Odd	0.59
Yamaguchi3/4 极化分解分量	Yamaguchi3_Odd	0.58
	Yamaguchi4_Y4_Odd	0.57
雷达后向散射系数	HH	0.86
	VV	0.52

表 3 撒播水稻田植株密度雷达特征参数

Table 3 Radar characteristic parameters of plant density in sowing rice field

参数类型	参数名称	相关性系数
Freeman-Durden 极化分解分量	Freeman2_Vol	0.657
	TSVM_alpha_s	0.714
Touzi 极化分解分量	TSVM_alpha_s1	0.543
	TSVM_alpha_s2	0.714
	T _{12_imag}	0.771
相干矩阵元素	T _{13_imag}	0.543
	T _{23_imag}	0.714
H/A/Alpha 极化分解特征向量	γ	0.543
Yamaguchi3/4 极化分解分量	Yamaguchi4_Y4_Vol	0.600
雷达后向散射系数	HV	0.600

2.3 水稻植株密度的反演

利用 1.3.3 节所述的方法,进行 2 类不同种植方式水稻田植株密度反演,得到研究区插秧田植株密度分布图(图 4(a)),撒播田植株密度分布图(图 4(b))。

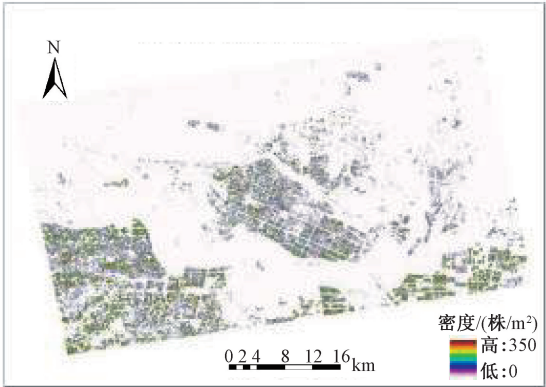
基于整个研究区 2 类水稻田植株密度反演结果,分别计算出插秧和撒播水稻田的平均植株密度。插秧田和撒播田整体植株密度平均值分别约为 173 和 182 株/m²。水稻幼苗期植株密度插秧田整体小于撒播田,这与地面实地调查的结果相符。插秧田水稻植株以墩为单位行列分布,植株之间的间距较大,整体田间植株密度相对较小;撒播田直接将水稻种子撒播在田里,水稻植株随机分布,单株水稻独立存在,植株间相距很近,没有墩的概念,田间植株密度相对较大。另外,根据野外实地调查可知,研究区内撒播田水稻种植时间大约在 6 月 16 日,插秧田的种植日期大约在 6 月

20 日,虽然时间仅相差 4 天,但是插秧的水稻秧苗已在苗床中生长一段时间,而撒播田是将水稻种子直接撒播进田里,所以 2 类水稻田的物候期存在一定的差异,也会造成二者植株密度的差异。

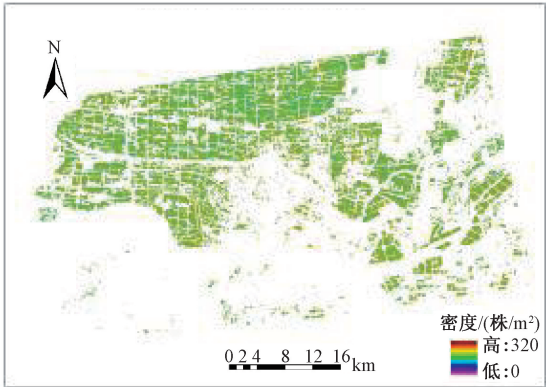
2.4 水稻植株密度反演结果的精度验证

利用野外地面实验采集的水稻植株密度值,对水稻植株密度反演结果进行精度验证。选取研究区内均匀分布的插秧水稻田样本点 16 个,撒播水稻田样本点 6 个,将地面实测记录的水稻植株密度值与反演图像上对应样点的反演植株密度值进行对比分析,对水稻植株密度反演模型精度进行验证,分别如图 5(a)和 5(b)所示。表明构建的植株密度反演模型结果较好,插秧田的植株密度反演模型的决定系数(R^2)为 0.878 3,均方根误差(root mean square error, RMSE)为 25 株/m²,模型模拟密度与实测密度最大偏差 60 株/m²,最小偏差 1 株/m²;撒播田的植株密度反演模型的决定系数(R^2)为 0.802 1, RMSE 为 39 株/m²,模型模拟密度与实测密度最大偏差 67 株/m²,最小偏差 3 株/m²。

为进一步定量评价水稻植株密度反演中考虑不同种植方式对反演结果的影响,这里将反演结果与不考虑种植方式的影响、直接进行水稻植株密度反演的结果(图 5(c))进行对比分析。结果表明,不考虑种植方式的影响、直接进行水稻植株密度反演的 RMSE 为 96 株/m²,考虑种植方式差异的水稻植株密度反演方法能够使密度反演的精度提高约 41%(插秧)和 32%(撒播)。因此在水稻植株密度反演过程中考虑不同的种植方式是十分必要的。



(a) 插秧水稻田



(b) 撒播水稻田

图 4 研究区水稻田植株密度分布

Fig. 4 Plant density distribution of rice fields in the study area

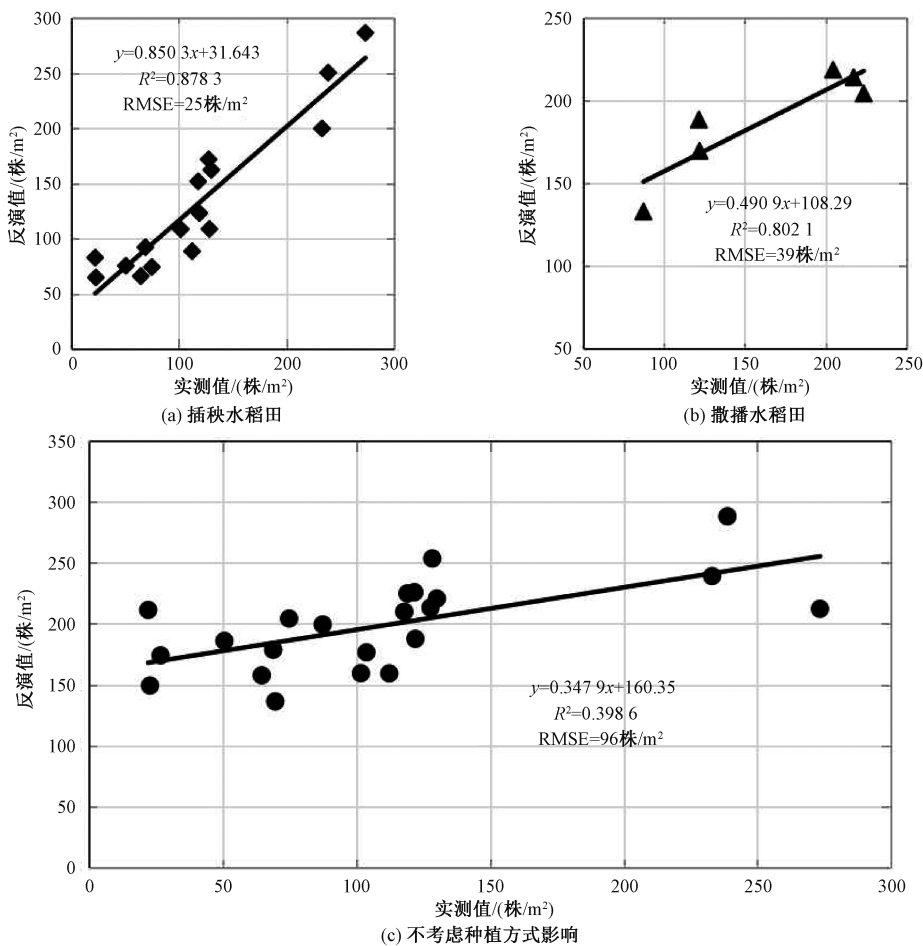


图 5 水稻植株密度反演精度

Fig. 5 Inversion precision of rice plant density

3 讨论与结论

不同种植方式(如插秧、撒播等)会导致水稻植株田间分布特征存在较大差异,进而对植株密度反演产生较大的影响。鉴于此,提出一种考虑插秧和撒播不同种植方式的极化 SAR 水稻植株密度反演方法,提高反演精度,为水稻生产潜力评估、长势监测与估产等应用提供准确的输入。首先基于随机森林进行不同种植方式水稻田分类识别;然后针对 2 类不同种植方式的水稻,分别利用弹性网络构建反演模型,实现水稻植株密度反演;最终获得研究区内插秧田和撒播田的水稻幼苗期植株密度分布图。通过与实测密度值对比分析,插秧田和撒播田植株密度反演 RMSE 分别为 25 和 39 株/ m^2 ,而不考虑种植方式的影响、直接进行水稻植株密度反演 RMSE 为 96 株/ m^2 ,考虑种植方式差异的水稻植株密度反演方法能够使密度反演的精度提高约 41% 和 32%。结果表明,基于

极化 SAR 特征参数对于水稻植株密度反演具有很大的应用潜力,而且在水稻植株密度反演中考虑种植方式的差异是非常必要的,能够大大提高反演精度,为水稻生产潜力评估、长势监测与估产等应用提供准确的输入。

通过对比分析水稻植株密度模型反演结果与地面实测结果,认为造成水稻植株密度反演误差的原因可能为:1)地面实测记录的水稻样本田地理坐标经纬度存在偏差,影响后续的建模与验证;2)实测样本数量的限制对弹性网络模型的训练和测试造成较大影响。此外本研究构建的水稻植株密度反演模型在本研究区具有较好的效果,模型的精度需要在其他区域进一步进行验证和优化。在后续研究中将尝试通过加入水稻微波散射模型或者改进的物理传输模型,同时引入更多的实测水稻参数用于反演模型的训练,以期提高模型反演精度。

参考文献

[1] Karthikeyan L, Pan M, Wanders N, et al. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms [J]. *Advances in Water Resources*, 2017, 109: 106-120. DOI: 10.1016/j.advwatres.2017.09.006.

[2] 化国强, 李晨, 杨沈斌, 等. 利用 Radarsat-2 数据基于比值检测的水稻制图[J]. *江苏农业学报*, 2012, 28(6): 1451-1458.

[3] Bouvet A, Le Toan T. Use of ENVISAT/ASAR wide-swath data for timely rice fields mapping in the Mekong River Delta [J]. *Remote Sensing of Environment*, 2011, 115(4): 1090-1101. DOI: 10.1016/j.rse.2010.12.014.

[4] 赵路生. 基于 ASAR 数据的水稻制图及生物量参数反演[D]. 成都: 电子科技大学, 2013.

[5] 董彦芳, 庞勇, 孙国清, 等. ENVISAT ASAR 数据用于水稻监测和参数反演[J]. *武汉大学学报(信息科学版)*, 2006, 31(2): 124-127. DOI: 10.3321/j.issn:1671-8860.2006.02.008.

[6] Yang Y, Lin Y. Crop detection and density estimation combing LiDAR points cloud with remote sensing image[C]// *Proc SPIE 7494, MIPPR 2009: Multispectral Image Acquisition and Processing*. October 30, 2009, Yichang, China. SPIE. 2009: 494-500. DOI: 10.1117/12.834012.

[7] 单捷, 孙玲, 王志明, 等. GF-1 影像遥感监测指标与冬小麦长势参数的关系[J]. *江苏农业学报*, 2019, 35(6): 1323-1333. DOI: 10.3969/j.issn.1000-4440.2019.06.008.

[8] 邹楠. 基于玉米冠层结构特点的遥感监测模型研究[D]. 新疆维吾尔自治区石河子: 石河子大学, 2017.

[9] 张远. 微波遥感水稻种植面积提取、生物量反演与稻田甲烷排放模拟[D]. 杭州: 浙江大学, 2008.

[10] Inoue Y, Sakaiya E, Wang C Z. Capability of C-band backscattering coefficients from high-resolution satellite SAR sensors to assess biophysical variables in paddy rice [J]. *Remote Sensing of Environment*, 2014, 140: 257-266. DOI: 10.1016/j.rse.2013.09.001.

[11] Jia M Q, Tong L, Zhang Y Z, et al. Rice biomass estimation using radar backscattering data at S-band[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(2): 469-479. DOI: 10.1109/JSTARS.2013.2282641.

[12] 贾明权. 水稻微波散射特性研究及参数反演[D]. 成都: 电子科技大学, 2013.

[13] Zhang Y, Liu X H, Su S L, et al. Retrieving canopy height and density of paddy rice from Radarsat-2 images with a canopy scattering model[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 28: 170-180. DOI: 10.1016/j.jag.2013.12.005.

[14] 景卓鑫, 张远, 王珂靖, 等. 基于 Radarsat-2 雷达数据的水稻参数反演[J]. *江苏农业科学*, 2013, 41(11): 61-65. DOI: 10.15889/j.issn.1002-1302.2013.11.066.

[15] 邵芸, 郭华东, 范湘涛, 等. 水稻时域散射特征分析及其应用研究[J]. *遥感学报*, 2001, 5(5): 340-345. DOI: 10.3321/j.issn:1007-4619.2001.05.004.

[16] Gu S H, Kelly B, Xiu D C. Empirical asset pricing via machine learning [J]. *The Review of Financial Studies*, 2020, 33(5): 2223-2273. DOI: 10.1093/rfs/hhaa009.

[17] Jiang F W, Tang G H, Zhou G F. Firm Characteristics and Chinese Stocks [J]. *Journal of Management Science and Engineering*, 2018, 3(4): 259-283. DOI: 10.3724/SP.J.1383.304014.

[18] 李斌, 邵新月, 李玥阳. 机器学习驱动的基本面量化投资研究[J]. *中国工业经济*, 2019(8): 61-79. DOI: 10.19581/j.cnki.ciejournal.2019.08.004.

[19] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, 67(2): 301-320. DOI: 10.1111/j.1467-9868.2005.00503.x.

[20] 周亚. 基于 XGBoost 与弹性网络回归的集成模型对上证指数日极差的预测分析[J]. *中国物价*, 2021(5): 68-71.

[21] 陈洪芳, 孙梦阳, 高毅, 等. 利用弹性网络算法求解大型三坐标测量机几何误差的方法[J]. *中国机械工程*, 2022, 33(9): 1077-1083. DOI: 10.3969/j.issn.1004-132X.2022.09.009.

[22] 卢颖. 广义线性模型基于 Elastic Net 的变量选择方法研究[D]. 北京: 北京交通大学, 2011.

[23] 李光早, 王士同. 基于稀疏表示和弹性网络的人脸识别[J]. *计算机应用*, 2017, 37(3): 901-905. DOI: 10.11772/j.issn.1001-9081.2017.03.901.

[24] 黄登香. Elastic Net 方法在几类模型变量选择中的应用[D]. 南宁: 广西大学, 2014.

[25] 姚明煌. 随机森林及其在遥感图像分类中的应用[D]. 福建泉州: 华侨大学, 2014.

[26] 孙永平, 王立峰, 张震伟, 等. 基于随机森林回归的火电机组供电煤耗遗传优化模型[J]. *信息通信技术与政策*, 2021, 47(3): 76-82. DOI: 10.12267/j.issn.2096-5931.2021.03.013.

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

[28] 王晓燕, 李美洲. 浅谈等级相关系数与斯皮尔曼等级相关系数[J]. *广东轻工职业技术学院学报*, 2006, 5(4): 26-27. DOI: 10.3969/j.issn.1672-1950.2006.04.008.

[29] Hoerl A E, Kennard R W. Ridge regression: biased estimation for nonorthogonal problems [J]. *Technometrics*, 1970, 12(1): 55-67. DOI: 10.1080/00401706.1970.10488634.

[30] Tibshirani R. Regression shrinkage and selection via the lasso [J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1996, 58(1): 267-288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.