

doi: 10.6046/zrzyyg.2023141
引用格式: 阳驰轶,官海翔,吴玮,等. 基于国产 GF-3 雷达影像的农田洪涝遥感监测方法[J]. 自然资源遥感,2023,35(4): 71-80. (Yang C Y, Guan H X, Wu W. Remote sensing monitoring method for flooded farmland based on domestic GF-3 radar images [J]. Remote Sensing for Natural Resources, 2023,35(4): 71-80.)

基于国产 GF-3 雷达影像的农田洪涝遥感监测方法

阳驰轶^{1,2}, 官海翔^{1,2}, 吴 玮³, 刘美玉³, 李 颖^{4,5}, 苏 伟^{1,2}

(1. 中国农业大学土地科学与技术学院,北京 100083; 2. 农业农村部农业灾害遥感重点实验室,北京 100083; 3. 应急管理部国家减灾中心,北京 100124; 4. 中国气象局河南省农业气象保障与应用技术重点开放实验室,郑州 450003; 5. 河南省气象科学研究所,郑州 450003)

摘要: 在全球气候变暖背景下,愈发频繁的洪涝灾害是造成我国粮食作物减产的主要农业灾害之一。雷达遥感技术具备全天候的对地观测能力,是快速监测区域范围内洪涝灾害信息的一种重要手段。随着人工智能领域的发展,机器学习方法广泛应用于洪涝灾害遥感监测,虽然该类算法具有较高的精度,但其训练过程往往需要大量的野外调查或遥感解译样本支持。为克服样本标记限制、提高区域尺度洪涝灾害监测的精度,本研究以 2021 年 7 月 20 日河南北部特大洪涝事件为背景,利用国产高分三号(GF-3)双极化雷达影像(HH-HV),构建了一种基于弱监督高斯混合模型(gaussian mixture model,GMM)的洪涝淹没作物监测方法,通过该方法提取了豫北部分区域农田洪涝淹没范围。通过对比 4 种典型的机器学习方法,包括随机森林、支持向量机、K 最近邻分类和平行六面体方法,发现该文构建的弱监督 GMM 方法的精度最高,其总体精度为 0.95,Kappa 系数为 0.90。该研究对于提高基于合成孔径雷达(synthetic aperture radar,SAR)遥感技术监测区域尺度作物洪涝的准确性和普适性具有重要意义。

关键词: GF-3 雷达影像; 洪涝; 农田洪涝淹没; 高斯混合模型

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 2097-034X(2023)04-0071-10

0 引言

洪涝灾害是最常见和最具毁灭性的自然灾害之一,往往造成严重的经济损失,威胁人类生命安全,近年来,洪涝更加频繁地威胁着世界上许多地区^[1]。在全球气候变化背景下,极端气候引发的自然灾害对我国农业生产影响巨大^[2]。洪涝灾害对农作物产量影响较大,会造成土壤肥力流失、农作物减产,甚至死亡等一系列问题。得益于遥感和计算机技术的快速发展,基于遥感影像的洪涝灾害监测方法能够克服传统实地调查的不足,已成为快速、大范围监测洪涝灾害的重要手段^[3]。

由于水体在近红外波段的强烈吸收,以及植被的强烈反射,光学遥感数据被广泛用于洪涝监测。王国杰等^[4]基于高分一号卫星数据,采用 ResNet, VGG, DenseNet, HRNet 这 4 种卷积神经网络方法提取地表水体信息,结果表明 DenseNet 在遥感水体识

别领域有很好的应用前景。刘宇晨等^[5]利用哨兵二号(Sentinel-2)年内长时序影像,提出了一个结合像元时间特征的大尺度、长时序的多指数组合的水体提取方法,完成了长江流域的地表水体遥感提取。Shao 等^[6]基于风云四号卫星数据多时相合成影像,结合归一化植被指数法和近红外波段信息,对南亚东北部流域的洪涝灾害进行监测。然而传统的光学影像易受云雾、降水等天气因素的强烈干扰,合成孔径雷达(synthetic aperture radar,SAR)技术具有较强的穿透能力,不受云雨的干扰。开阔水域的主要散射机制是镜面散射,水的介电常数相对恒定,由此产生的能量很少能够传回传感器,因此开阔水域的后向散射系数通常很低,利用目视解译或阈值法即可将其与非淹没区分开来^[7];然而,当水淹没在农作物冠层以下时,由于双弹散射机制的影响,反射回卫星的信号强度会增强^[8]。在平静的开阔水域中,共极化(HH/HV)是被使用最广泛的,因为它能提供最高的对比度^[9];尽管交叉极化受作物茎叶

收稿日期: 2023-05-17; 修订日期: 2023-07-29
基金项目: “十四五”国家重点研发计划项目“农情信息空天地高精度高时效智能监测系统研发与应用”之课题三“农情参数高分遥感机理模型与定量解析研究”(编号: 2022YFD2001103)和国家自然科学基金项目“多源高分光学遥感数据与作物模型同化的不确定性研究”(编号: 41805090)共同资助。
第一作者: 阳驰轶(2002-),女,本科,主要从事基于机器学习方法的自然灾害遥感监测。Email: chiyiyang@cau.edu.cn。
通信作者: 苏 伟(1979-),女,教授,博士,主要从事作物长势遥感监测、作物灾害遥感监测。Email: suwei@cau.edu.cn。

的影响存在较强的去极化效应,但由于对体积散射敏感,因此对于监测部分被淹没的农作物具有较高潜力。与单极化相比,使用双极化来监测洪涝农田可以提高分类精度^[10]。高分三号卫星(GF-3)是我国首颗有高分辨率、多成像模式、大成像幅宽的C频段、多极化SAR卫星,能够全天候、全天时监测全球海洋和陆地信息^[11]。

目前基于SAR影像进行农田洪涝淹没区提取的方法主要有阈值分割法、机器学习法、深度学习法等。郭山川等^[12]在时序哨兵一号(Sentinel-1)A数据和GEE(Google Earth Engine)平台支持下,结合阈值分割法,提出一种洪水动态监测方法以实现广域尺度的洪水事件自动检测和淹没过程动态监测。郁宗桥等^[13]利用随机森林(random forest,RF)指标重要性评估功能,在提取的138个纹理特征中筛选出得分较高的纹理特征来提取水体。Nemni等^[14]基于卷积神经网络方法,利用Sentinel-1 SAR影像数据,使用半自动技术和目视解译训练模型,实现进一步自动化和提高洪水监测速度。在这些方法中,阈值分割法利用了图像的灰度特征,计算简单、效率高,但受作物生长和洪水分布的空间异质性影响,监测结果往往会存在严重的“椒盐”现象,对于图像灰度差异较小的区域分割精度较低;机器学习法具有较高的鲁棒性,能够捕捉到洪涝区域的时空差异和变化特征,但是需要依据先验知识预先选取分类样本,对样本需求量大^[15],可能出现过度拟合、迭代次数过多等问题;深度学习的基础是神经网络,能够应对具体场景建立模型,学习能力强,适用范围广,但深度学习建立模型时也需要有大量数据支撑、建立

过程复杂,对硬件的要求较高,在数据不平衡、算法偏差的情况下可能出现“歧视”现象。

GMM基于高斯混合分布的假设,能够对不同类别地物间的复杂混合进行概率估计,适用于具有高空间异质性的农田环境;其次GMM的计算复杂度相对较低,可在区域尺度上实现作物动态快速监测。然而,目前还没有学者探讨GMM在作物洪涝监测中的适用性。针对阈值分割法可能出现的精度不高问题,传统的机器学习算法和深度学习法所需样本标记量大的问题,本文基于双极化GF-3雷达影像,构建了一种弱监督GMM的洪涝监测方法来提取研究区农田洪涝淹没范围。

本研究基于GMM的弱监督分类和RF、支持向量机(support vector machine,SVM)、K最近邻(k-nearest neighbor,KNN)分类、平行六面体分类(parallelepiped classification,PC)的监督分类共5种机器学习方法,在“7·20”河南暴雨事件的背景下,应用Sentinel-2影像目视解译选取样本集作为输入,利用国产GF-3 SAR的HH,HV这2种极化方式的影像对豫北地区进行农田洪涝提取,对比农田洪涝中2种极化方式的后向散射特征,分析评估不同方法的水体提取精度,为洪涝灾害的农田水体提取提供参考,结果表明,构建的GMM弱监督分类模型可以在少样本的情况下提高农田水体监测精度。

1 研究区概况及数据源

1.1 研究区概况

研究区位于河南省北部,见图1,地理坐标E113°

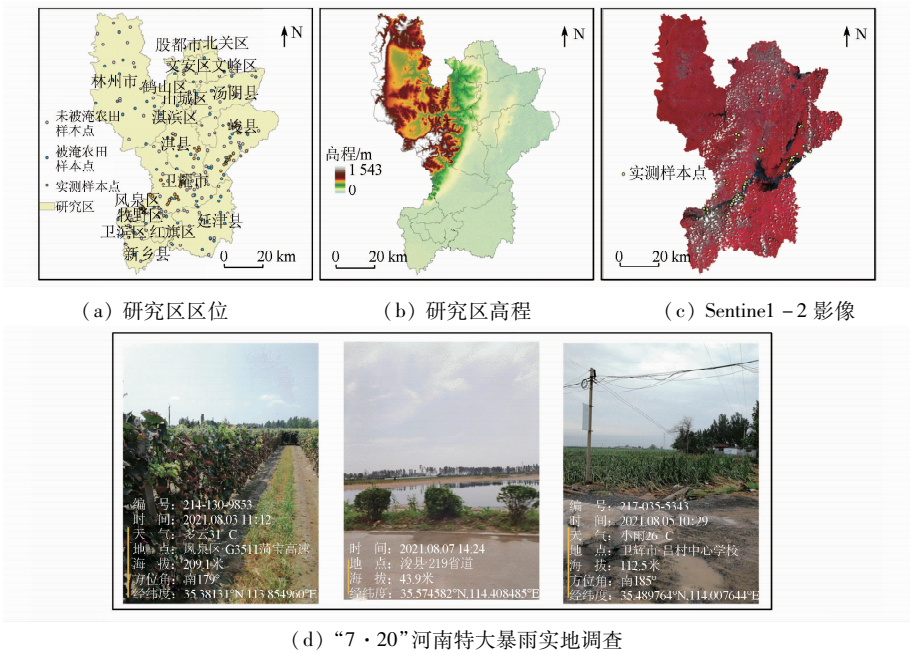


图1 研究区区位、高程、Sentinel-2 影像与河南“7·20”特大暴雨实地调查

Fig.1 Study area location ,study area elevation ,Sentinel-2 image and field investigation of “July 20” rainstorm in Henan Province

38'~114°43',N35°36'~36°22',覆盖安阳市、鹤壁市、新乡市 3 个地级市,共包括 18 个市区县,邻近太行山,总体地势西高东低,总面积为 7 610 km²,温带大陆性季风气候,雨热同期,全年 10℃以上活动积温在 4 000~5 000℃左右,年均降雨量 580 mm,降水量在季节分配上极不均匀,与冬、夏季风的进退大致相同。研究区种植的主要农作物是玉米、小麦,东部种植了很多花生,玉米和花生在五月底左右播种,九月底左右收获,小麦在十月中旬播种,次年五月底六月初收获。

台风“烟花”于 2021 年 7 月中旬开始影响中国的东部和中部地区,7 月 21 日开始,强降雨北移,河南省北部洪涝灾害严重,其中受洪涝影响最严重的 4 个城市是郑州市、洛阳市、新乡市和鹤壁市^[16]。

本研究使用了基于 Sentinel-2 假彩色合成影像进行人工目视解译后得到的样本点来训练和验证农田洪涝淹没范围和未淹没范围。根据 2021 年 8 月对研究区进行野外调查获得的实测样本点,在

Sentinel-2 假彩色合成影像中,可以看出水体呈现黑色,植被呈现红色。水体呈现黑色是因为水体在电磁波谱的近红外波段有强烈吸收作用,植被呈现红色是因为其在近红外波段反射强烈^[17]。根据此原理在 Sentinel-2 假彩色合成影像中选择研究所需样本点,其中被淹农田样本来自完全和不完全淹没的农田区域,最终获得被淹农田样本点和未被淹农田样本点各 110 个,采用随机抽样方法选择被淹农田样本点和未被淹农田样本点各 40 个作为验证样本,剩余为训练样本。

根据 2021 年 7 月 15 日—31 日 20 时安阳市、新乡市、鹤壁市 3 个地级市和浚县、淇县 2 个县的气象站点监测的降水量,见图 2。从图 2 可以看出,总体上 5 地 17 和 18 日出现小规模降雨,大暴雨、特大暴雨从 19 日持续到 23 日,后逐渐消退,21 日降雨量达到顶峰,5 个观测站中淇县 21 日降雨量最大,超过 350 mm。

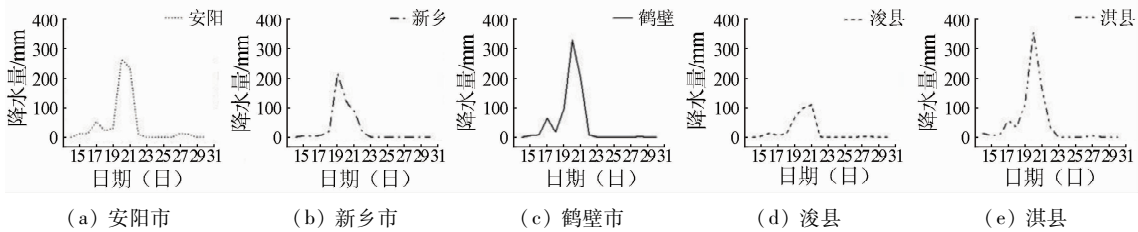


图 2 豫北部分气象观测站 7 月 15—31 日降水量

Fig.2 Precipitation of some meteorological observation stations in Northern Henan from July 15 to 31

1.2 数据源与预处理

1.2.1 数据源

本研究采用的研究数据包括国产 GF-3 SAR 数据、土地覆被类型数据、Sentinel-2 数据、降水数据。GF-3 SAR 数据下载自中国资源卫星应用中心(<https://data.cresda.cn/#/2dMap>),成像时间为洪涝发生后(2021 年 7 月 27 日),10 m 分辨率,条带成像模式为精细条带(FSII),L1A 级产品,极化方式为双极化的 HH 和 HV。基于 Sentinel-1 和 Sentinel-2 数据的 10 m 分辨率的 2021 年土地覆被类型数据来源于欧空局(https://developers.google.cn/earth-engine/datasets/catalog/ESA_WorldCover_v100)。Sentinel-2-Level-1C 级产品数据获取于欧空局,成像时间为 2021 年 7 月 27 日。降水数据来自气象站点的监测数据,包括安阳市、鹤壁市、新乡市、浚县、淇县共 5 个观测站,采集时间为 2021 年 7 月 15 日—31 日的 20 时。

1.2.2 数据预处理

在 PIE-SAR 软件中对 GF-3 影像进行预处

理,依次进行复数据转换、多视处理、滤波操作、地理编码、DB 转换的操作。PIE-SAR 软件内已做辐射定标处理。幅度特征是 SAR 影像反映不同后向散射强弱的特征,幅度计算公式为:

$$A = \sqrt{r^2 + i^2}, \tag{1}$$

式中: A 为 SAR 影像中像点的幅度; r 为 SAR 影像中像点复数中的实数; i 为虚数。

在本研究中,GF-3 SAR 影像在方位向和距离向视数为 3×3 时被重采样为 10 m 分辨率的规则网格。首先,基于 3 m×3 m 的滤波窗口,采用增强型 Lee 滤波(EnLee 滤波)降低影像噪声;然后,利用外部的数字高程模型(digital elevation model,DEM)数据对 GF-3 SAR 影像的 L1 级产品数据进行地形辐射校正,生成 L2 级地理编码产品;再进行 DB 转换输出雷达后向散射系数,获得对应地物绝对的后向散射值;最后,利用土地覆被类型数据作为掩模来剔除研究区内非农田的地物类型,以获得研究区范围内预处理完的 SAR 影像数据。

2 研究方法

本研究首先将 HH, HV 极化的 GF-3 影像数据进行预处理, 获得 HH 和 HV 通道的后向散射系数;

然后, 将训练样本连同 HH 和 HV 极化的后向散射作为输入, 通过 GMM 的弱监督分类和 RF, SVM, KNN, PC 分类的监督分类共 5 种分类方法来提取研究区的农田洪涝淹没范围; 最后基于验证样本评价上述 5 种分类方法的精度。技术流程见图 3。

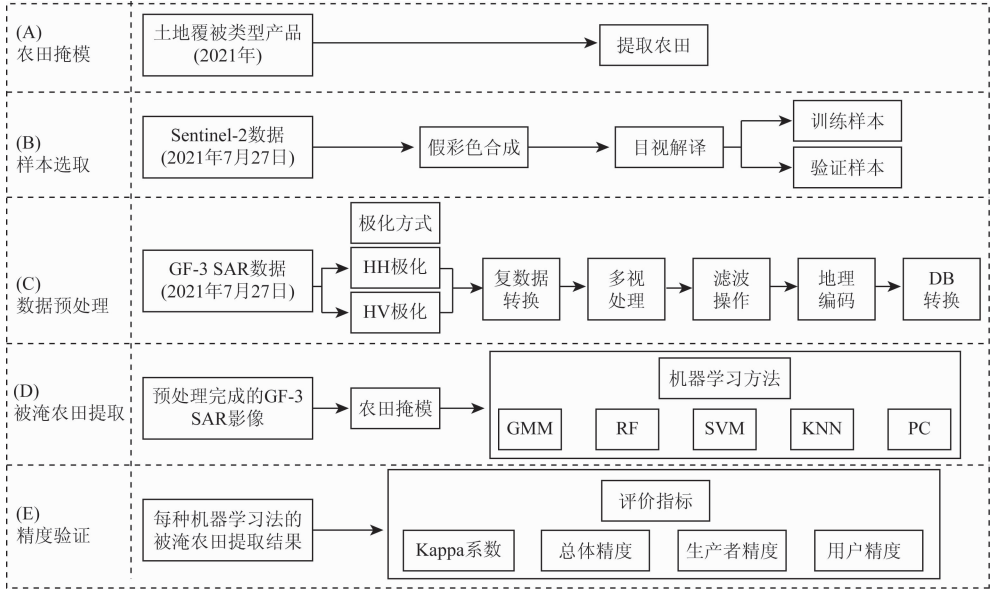


图 3 技术流程

Fig. 3 Technique process

2.1 高斯混合模型弱监督分类算法

GMM 采用若干个高斯分布的线性叠加来表示, 将一个事物分解为若干个基于高斯概率密度函数形成的模型^[18]。GMM 在遥感领域主要应用于遥感影像分割、遥感时序拟合等。样本数据 X 为一维数据的高斯模型称为单高斯模型, 遵从以下概率密度函数:

$$p(x | \theta) = (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\}, \quad (2)$$

式中: μ 为数据的期望; σ 为数据标准差。

GMM 可以看作 S 个单高斯混合模型组合而成的模型, 它的概率密度函数也可看成是 S 个服从均

值为 μ_i 、方差为 Σ_i 的子成分的高斯分布的加权平均和, 此时的样本数据 $X = [X_1, X_2, \dots, X_m]^T$ 是 m 维数据, $x = [x_1, x_2, \dots, x_m]^T$ 是 X 的一个实例, 表达式如下:

$$p(x | \theta) = \sum_{i=1}^S \alpha_i \varphi(x | \theta_i), \quad (3)$$

式中: α_i 为各成分属于第 i 个子模型的概率, 满足:

$$\alpha_i \geq 0, \sum_{i=1}^S \alpha_i = 1, i = 1, 2, \dots, S, \quad (4)$$

参数 θ 为每个子模型的期望、方差(或协方差)、在混合模型中发生的概率, 公式为:

$$\varphi(x | \theta_i) = (2\pi)^{-\frac{m}{2}} |\sum_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (x - \mu_i)^T \sum_i^{-1} (x - \mu_i)\right\}. \quad (5)$$

在本实验中需要分离被淹农田与未被淹农田, S 设置为 2, x 是选取的训练样本, 子模型被分为属于被淹农田和未被淹农田 2 类, 所以 GMM 模型可以简化如下:

$$p(x | \theta) = \alpha_1 \varphi(x | \theta_1) + \alpha_2 \varphi(x | \theta_2), \quad (6)$$

式中涉及到的参数 $\alpha_1, \alpha_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2$ 分别为训练样本属于子模型的概率、训练样本的均值和方差。本研究采用 EM 算法迭代 GMM 参数, 每次迭代分为 2 步: E 步, 求期望 $E(Q_{ji} | X, \theta)$, 计算每一个数据 j

来自子模型 i 的概率, 公式为:

$$Q_{ji} = \frac{\alpha_i \varphi(x_j | \theta_i)}{\sum_{i=1}^K \alpha_i \varphi(x_j | \theta_i)}, \quad (7)$$

式中: $j = 1, 2, \dots, 140$; $i = 1, 2$ 。M 步, 求极大, 计算新一轮迭代的模型参数 α_i, μ_i 和 Σ_i , 公式为:

$$\alpha_i = \frac{\sum_{j=1}^{140} Q_{ji}}{N}, \quad (8)$$

$$\mu_i = \frac{\sum_{j=1}^{140} (Q_{ji} x_j)}{\sum_{j=1}^{140} Q_{ji}}, \quad (9)$$

$$\Sigma_i = \frac{\sum_{j=1}^{140} Q_{ji} (x_j - \mu_i) (x_j - \mu_i)^T}{\sum_{j=1}^{140} Q_{ji}}, \quad (10)$$

式中: $i = 1, 2, \dots, 140$, 且在计算 Σ_i 时, μ_i 应用此轮迭代更新后的值。重复计算 E 步和 M 步直至收敛, 即满足 $\|\theta_{i+1} - \theta_i\| < \varepsilon$, ε 为一很小的正数, 可以任意小, 此时认为取到比较合理的模型参数。

2.2 监督分类方法

1) RF。RF 的基本原理是利用自助法 (bootstrap) 重采样技术, 本研究中从原始训练样本集 $R = \{R(1), R(2), \dots, R(140)\}$ 中有放回的重重复随机抽取 q 个样本 $R_1, R_2, \dots, R_q$, 再对这 q 个样本分别建立决策树, 这些决策树构成的集合组成 RF, 通过每棵决策树对样本的预测结果投票输出最终农田分类结果。

2) SVM。SVM 最早是由 Vapnik^[19] 提出的一种二分类模型, 是一种基于统计学习理论的机器学习方法, 是在高维特征空间上间隔最大的线性分类器, 基本原理是求解能够正确划分训练数据集且集合间隔最大的分离超平面^[20], 得到最优的分类结果, 适合少量样本时进行分类。

3) KNN 分类。KNN 最初是由 Cover^[21] 提出的一种经典的简易惰性学习方法, 是最近邻分类的扩展, 在特征空间中的所有训练样本 N 中寻找待分类样本附近的 K 个最近的样本, 再按照投票法等分类决策规则把农田分为被淹与未被淹的 2 类。

4) PC。PC 是一种传统的监督分类方法^[22], 根据训练样本的亮度值不同划分成 n 维的平行六面体数据空间, 规则由标准差阈值决定, 待分类波段像元的光谱值属于任一训练样本的范围, 则被划分到对应类别。

2.3 精度评价指标

为对比不同方法的精度, 基于提取的农田洪涝淹没结果及验证样本建立混淆矩阵, 计算 4 种精度评价指标, 包括总体精度 (overall accuracy, OA)、Kappa 系数、生产者精度 (producer accuracy, PA)、用户精度 (user accuracy, UA), 计算公式为:

$$OA = \frac{\sum_{i=1}^n x_{ii}}{N}, \quad (11)$$

$$Kappa = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^n (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^n (x_{i+} x_{+i})}, \quad (12)$$

$$PA = \frac{x_{ii}}{x_{+i}} \times 100\%, \quad (13)$$

$$UA = \frac{x_{ii}}{x_{+i}} \times 100\%, \quad (14)$$

式中: n 为类别, 此研究中 $n = 2$; N 为类别个数的总和即总验证样本个数; x_{ii} 为混淆矩阵对角线元素; x_{i+} 为类别的列总和; x_{+i} 为类别的行总和; x_{11} 为验证样本中被标记为被淹的农田且结果中被提取为被淹的农田的样本; x_{12} 为验证样本中被标记为被淹的农田且结果中被提取为未被淹的农田的样本; x_{21} 为验证样本中被标记为未被淹的农田且结果中被提取为被淹的农田的样本; x_{22} 为验证样本中被标记为未被淹的农田且结果中被提取为未被淹的农田的样本。

3 结果与分析

3.1 后向散射特征分析

分别统计被淹农田未被淹农田样本在 HH 和 HV 极化的后向散射特征, 见图 4。从图中可以看出, HV 极化中被淹农田与未被淹农田的可分性高于 HH 极化, 因为 HH 极化对二次散射敏感, 在洪涝场景下, 仅有水位较高, 且植被保持垂直结构时才会构成此类散射机制。研究区农田农作物主要为玉米, 由于此次洪涝较为严重, 玉米大多被完全淹没, 因此上述类别的被淹玉米在研究区内并不占据主导; 而 HV 对于体积散射较为敏感, 虽然被淹没玉米区域以镜面反射为主, 但未被淹玉米的体积散射贡献仍可作为区分其与被淹玉米的有效机制, 这导

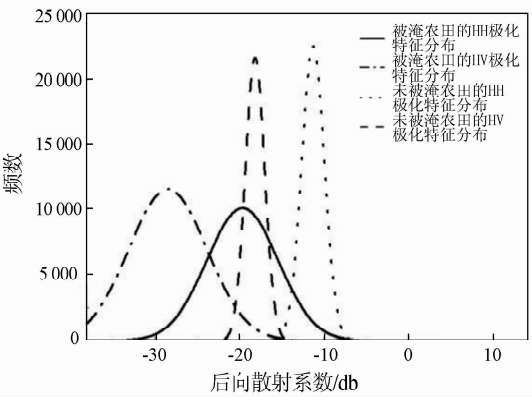


图 4 被淹农田与未被淹农田在 HH、HV 极化的后向散射系数
Fig. 4 Backscatter coefficients of HH and HV images of flood-affected farmland and non-disaster-stricken farmland

致 HV 比 HH 具有更高的敏感性。

3.2 提取农田洪涝淹没范围对比与精度验证

为对比机器学习方法的性能,选取 4 种场景,分别为建筑物稀疏区域的被淹农田、建筑物密集区域的被淹农田、以未被淹农田为主的农田区域、有河流存在的山区,见图 5。对于完全被淹的农田,其表面被镜面反射主导,导致低后向散射系数,而建筑物密

集区内建筑物与开阔水域之间的相互作用会产生二次散射,导致该地区 SAR 影像的后向散射系数上升。因此,影像中农田与建筑物边缘存在的亮斑,可能会干扰农田洪涝识别;此外,山区地形起伏较大,会使 SAR 影像产生几何位移、顶底倒置和局部形变等失真现象,因此通常需要额外的辅助数据或处理步骤消除。

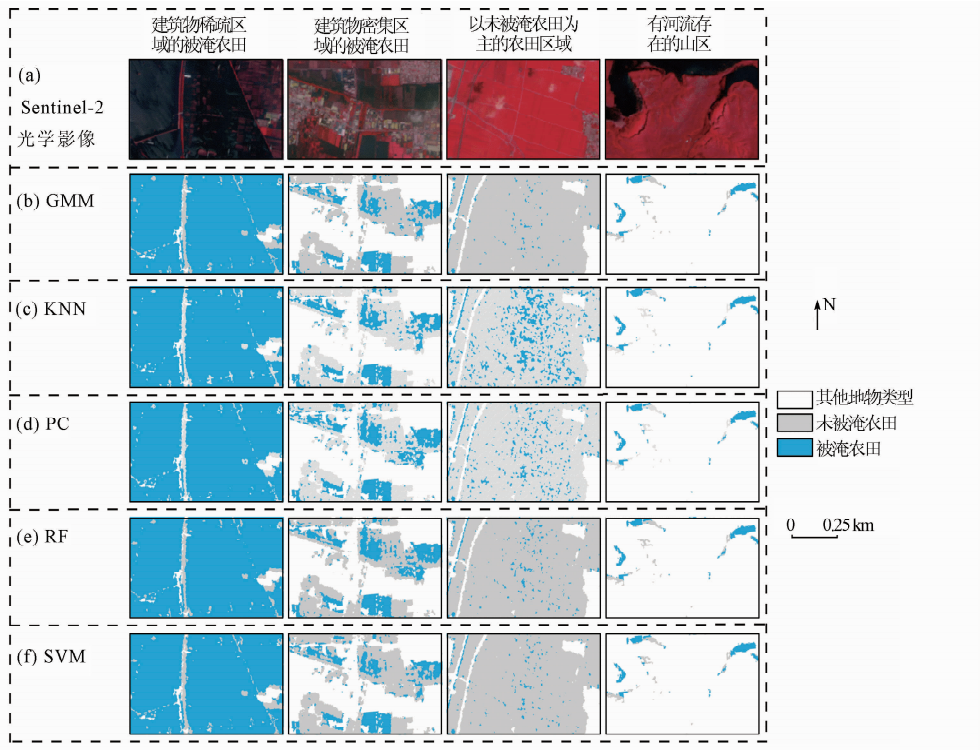


图 5 Sentinel-2 光学影像和分类结果的局部图

Fig. 5 Partial plot of Sentinel-2 optical imagery and classification results

图 6 为 5 种机器学习方法结果精度验证,根据图 6,GMM 的 OA 为 0.95,Kappa 系数为 0.90,OA 和 Kappa 系数均最高。GMM(图 5(b))方法所得结果的斑点噪声少于其他 4 种方法,被淹农田空间分布连贯。GF-3 影像中存在着很多小面积的被淹农

田像元与未被淹农田像元亮度值接近的情况,这部分地物在分类时容易被混淆,所以 PC(图 5(d))分类结果存在大量零星的斑点噪声。RF(图 5(e))、SVM(图 5(f))把部分未被淹农田、山体阴影误分为被淹农田,也有一些遗漏的被淹农田,这类农田为轻度淹没,淹没水位较低,绝大部分已下渗为地下水。在 SVM 分类中,不同训练样本被映射到不同的特征空间并用于构建分类超平面,然而,在某些情况下,SVM 可能无法发现相应的决策边界,例如,部分未被淹农田、山体阴影的特征与被淹农田的反射特征区分度较小,这可能导致将其误分为被淹农田;KNN(图 5(c))、PC 错分、漏分现象多且明显,各类精度指标都较低(图 6),这些错分、漏分现象多与 KNN、PC 算法本身原理简单有关。其中 KNN 存在大量、小面积的错分现象和许多漏分现象,在非洪涝区把许多农田误分为被淹农田,因为 KNN 算法对样本的依赖性高,往往需要大量训练样本以保证运行分类算法时结果收敛,注重考虑像元领域间的空间

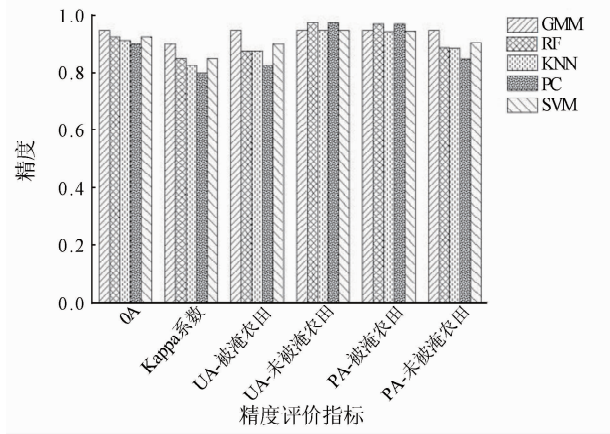


图 6 5 种机器学习方法结果精度验证

Fig. 6 Accuracy verification of five machine learning methods

关系,对于小面积、破碎的洪涝区域敏感度低。

3.3 洪涝范围的空间分布分析

图 7 为 Sentinel-2 光学影像和 5 种机器学习方法的农田洪涝淹没全局图,总体来看,研究区西北部山地较多,被淹没相对较少,被淹农田主要集中在研究区的中部和北部,大部分在蓄滞洪区,此次河南省暴雨的国家蓄滞洪区共有 9 处,涉及安阳市、鹤壁市、新乡市的汤阴县、内黄县、安阳县、浚县、淇县、卫辉市等 6 个市区县。河流距离对洪涝发生有显著影

响,距离河流越近,越容易发生洪涝,尤其是大江大河沿岸地区^[23],研究区属于卫河流域,是此次暴雨事件重灾区,卫河流经研究区的新乡县、淇县、浚县、汤阴县,河流很多地方流水湍急,洪涝发生时淹没影响较大,中下游洼地多,部分被扩建为正式行洪、滞洪区,安阳地区和汤阴县有大型水库可一定程度上减缓洪涝影响。安阳市南部有郑庄山,北部有灵寿山,山地较多,相比鹤壁市和新乡市,高程值较大,因此洪涝淹没情况也较轻。

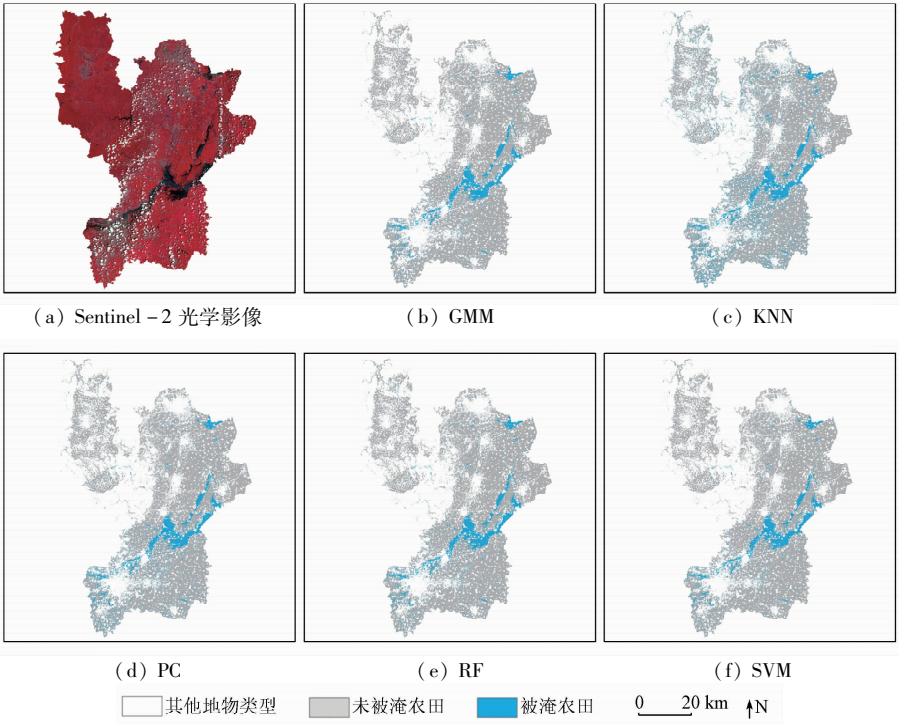


图 7 Sentinel 2 光学影像和分类结果的全局图
Fig. 7 Global view of Sentinel 2 optical imagery and classification results

GMM 估计的各市区县的被淹农田面积见表 1,被淹农田总面积为398 km²,占研究区农田总面积

表 1 各市区县农田洪涝淹没面积
Tab. 1 The area of farmland affected by the flood disaster in each urban, district and county

市区县	被淹农田 面积/km ²	总农田面 积/km ²	被淹农田占总农田 面积比例/%
新乡县	22. 61	268. 46	8
凤泉区	13. 44	53. 70	25
卫滨区	3. 72	23. 77	16
汤阴县	24. 92	473. 67	5
北关区	0. 80	18. 81	4
浚县	115. 57	732. 32	16
淇县	40. 54	255. 82	16
红旗区	3. 65	56. 32	6
淇滨区	4. 32	148. 73	3
殷都区	0. 78	25. 76	3
牧野区	9. 01	26. 50	34
龙安区	4. 68	146. 00	3
文峰区	10. 11	86. 56	12
延津县	22. 41	713. 63	3
卫辉市	100. 13	459. 42	22
山城区	1. 73	75. 60	2
鹤山区	0. 96	41. 70	2
林州市	18. 05	609. 47	3

的 9. 44%,农田淹没最严重的 3 个地区是牧野区、凤泉区和卫辉市,3 个地区都地处新乡市内,其中卫辉市农田洪涝淹没面积最大,有 100. 13 km²,而安阳市总体淹没水平相比其他 2 市较低,GMM 分类结果与实际农田受灾情况基本相符。

4 讨论

4. 1 不同场景下的被淹农田提取

对于不同场景,SAR 影像提取洪涝作物的能力可能会受到不同因素的影响。在以农作物为主的耕作区,建筑物与开阔水域之间的二次散射和影像中的亮斑都会影响洪涝的提取;此外,建筑物产生的阴影可能会被误认为是水体,从而导致洪涝范围的过度监测。

由于 SAR 传感器的侧式几何结构,地物可能会因为城市建筑和较高大的植被等自然特征的相对位置而被错分。而雷达阴影的产生是由于相邻建筑物

的位置和 SAR 传感器的侧式几何结构,雷达波束无法照射到表面,导致这些区域在影像上看起来很暗,在干燥的情况下却被误分类为洪涝区,造成洪涝的过度监测。

与平原地区相比,山地和丘陵地区的地形更加复杂,地势凹凸不平。这使得洪涝覆盖的地表形态更加多样和复杂,雷达回波对不同斜率和高程的地形反射效果不同,导致 SAR 影像中洪涝区的后向散射信号强度的高度空间异质性^[24]。

在实际应用中,需要结合多种数据和方法,如遥感影像、地理信息系统、气象数据等进行不同场景的数据处理和分析,消除干扰因素,提高农田洪涝识别的精度和可靠性。

4.2 5 种机器学习法在农田洪涝提取上的优缺点

4.2.1 监督分类方法在农田洪涝提取上的优缺点

本研究利用 GF-3 SAR 影像 HH 和 HV 2 种极化方式的数据,建立弱监督 GMM 分类,与其他 4 种机器学习方法(RF,SVM,KNN,PC)对比,结果表明 GMM 分类结果的精度最高,对比其余 4 种机器学习方法在农田洪涝提取的优缺点如下:

1)RF 能够有效缓解决策树模型的过拟合问题,能输出特征重要程度,减少冗余信息对分类结果的影响。但是它的训练速度比较慢,对于非线性结果计算量会增大,且容易受到 SAR 斑点噪声的影响。

2)SVM 在处理小样本数据时具有较好的适应性和泛化性,处理高维度数据时表现较好,因而适用于处理多维遥感数据。由于 SVM 需要进行数据预处理和特征选择,并且随着范围的增大,计算量也增大,因而不适用于实时提取被淹农田。

3)KNN 可以处理非线性特征,对于决策边界分布不规则的情况有较好的适应性,且不需要预先训练模型,可直接应用于实际的农田洪涝场景。然而 KNN 对于噪声和异常值较为敏感,需要对 K 值合理设定,增加了人工干预过程。

4)PC 使用阈值分类,是一种基于阈值分割的方法,阈值选择影响分类结果的准确性,分类精度相对复杂模型较低,对于特征较少的数据分类不佳,在使用多光谱遥感数据进行被淹农田提取时,PC 可能会受到植被遮挡和阴影的影响。

4.2.2 GMM 在农田洪涝提取上的优缺点

GMM 方法可以有效解决其他机器学习方法遇到的样本量小、样本分布不均、斑点噪声等问题,它在农田洪涝提取的优点如下:

1)GMM 属于软分类模型的范畴,可以给出每个像素被分类为被淹农田的概率,因此可以估计混合

像素中淹没作物的比例。

2)GMM 处理偏态数据能力强,在实际中,被淹农田的分布并不均匀,所以可能出现数据偏态分布情况,GMM 可以拟合各种分布可能,从而更好地适应这些数据。GMM 使用不同的协方差矩阵来表示不同方向上的方差,可更好地适应各向异性的数据的形状差异。

3)GMM 支持多模态特征识别,可以对 SAR 影像中的几何、纹理和强度等多模态特征进行检测和分类,提高分类精度。

但是在本研究中仍然发现了一些 GMM 的不足:

1)GMM 对初始值很敏感,如果训练样本选取不当,可能使 EM 算法求出的解收敛到局部最优,而不是全局最优。

2)本研究采用的 GMM 模型需要预先确定混合高斯子分量个数,并且随着高斯成分个数的增加,算法拟合数据能力增强,模型变复杂,可能出现大量冗余参数。未来可以采用适当的评价准则来辅助判断选取的混合高斯子分量的数目。

5 结 论

本文运用国产 GF-3 SAR 卫星数据,对比分析 GMM,RF,SVM,KNN,PC 共 5 种机器学习方法,对研究区农田洪涝灾害进行监测,得到结论如下:

1)在农田洪涝灾害场景中,基于镜面反射的主导机制,HV 极化比 HH 极化敏感,更适用于提取被淹农田。

2)有高分辨率和敏感性能的国产 GF-3 SAR 卫星在暴雨天农田洪涝实时提取中体现较大优势,准确性和鲁棒性得到了保障。

3)在相同的少量训练样本前提下,对比其他机器学习方法(RF,SVM,KNN,PC),本研究使用的 GMM 弱监督分类来提取农田洪涝淹没范围时表现最优,OA 为 0.95,Kappa 系数为 0.90。

在未来的研究中,可以考虑以下方面的改进:

1)结合高分辨率的 SAR 影像数据和地理信息系统数据,提供更准确全面的农田洪涝灾害监测信息,进一步提高分类精度和稳定性。建立统一的数据采集和处理流程标准,提高数据质量和结果的可靠性。

2)改善模型参数的选择和调整方式,自动选取最优的模型参数,保证在简化人为干扰模型的前提下进一步提高被淹农田提取精度。

3)深入研究基于低洼田、山地田和平原田等不

同场景的农田洪涝淹没提取,增强模型在具体实际中的适应性。

参考文献 (References) :

[1] Tong X H, Luo X, Liu S G, et al. An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO – SkyMed radar imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 136: 144 – 153.

[2] 徐宗学, 刘琳, 杨晓静. 极端气候事件与旱涝灾害研究回顾与展望 [J]. 中国防汛抗旱, 2017, 27 (1) : 66 – 74.

Xu Z X, Liu L, Yang X J. Study on extreme climate events and drought/flood disasters: Review and prospect [J]. China Flood & Drought Management, 2017, 27 (1) : 66 – 74.

[3] 曹罗丹, 李加林. 基于遥感与 GIS 的浙江省洪涝灾害综合风险评估研究 [J]. 自然灾害学报, 2015, 24 (4) : 111 – 119.

Cao L D, Li J L. Integrated risk assessment of flood disaster in Zhejiang Province based on RS and GIS [J]. Journal of Natural Disasters, 2015, 24 (4) : 111 – 119.

[4] 王国杰, 胡一凡, 张 森, 等. 深度卷积神经网络的遥感影像水体识别 [J]. 遥感学报, 2022, 26 (11) : 2304 – 2316.

Wang G J, Hu Y F, Zhang S, et al. Water identification from the GF – 1 satellite image based on the deep convolutional neural networks [J]. National Remote Sensing Bulletin, 2022, 26 (11) : 2304 – 2316.

[5] 刘宇晨, 高永年. Sentinel 时序影像的长江流域地表水体提取 [J]. 遥感学报, 2022, 26 (2) : 358 – 372.

Liu Y C, Gao Y N. Surface water extraction in Yangtze River basin based on sentinel time series image [J]. National Remote Sensing Bulletin, 2022, 26 (2) : 358 – 372.

[6] Shao J L, Gao H, Wang X, et al. Application of Fengyun – 4 satellite to flood disaster monitoring through a rapid multi – temporal synthesis approach [J]. Journal of Meteorological Research, 2020, 34 (4) : 720 – 731.

[7] Cohen J, Riihimäki H, Pulliainen J, et al. Implications of boreal forest stand characteristics for X – band SAR flood mapping accuracy [J]. Remote Sensing of Environment, 2016, 186: 47 – 63.

[8] Jedlovac G. Advances in geoscience and remote sensing [M]. London: Intech, 2009: 679 – 708.

[9] Brisco B, Touzi R, Van der Sanden J J, et al. Water resource applications with RADARSAT – 2: A preview [J]. International Journal of Digital Earth, 2008, 1 (1) : 130 – 147.

[10] Hess L L, Melack J M. Remote sensing of vegetation and flooding on Magela Creek Floodplain (Northern Territory, Australia) with the SIR – C synthetic aperture radar [J]. Hydrobiologia, 2003, 500: 65 – 82.

[11] 张庆君. 高分三号卫星总体设计与关键技术 [J]. 测绘学报, 2017, 46 (3) : 269 – 277.

Zhang Q J. System design and key technologies of the GF – 3 satellite [J]. Acta Geodaetica et Cartographica Sinica, 2017, 46 (3) :

269 – 277.

[12] 郭山川, 杜培军, 蒙亚平, 等. 时序 Sentinel – 1A 数据支持的长江中下游汛情动态监测 [J]. 遥感学报, 2021, 25 (10) : 2127 – 2141.

Guo S C, Du P J, Meng Y P, et al. Dynamic monitoring on flooding situation in the middle and lower reaches of Yangtze River region using Sentinel – 1A time series [J]. National Remote Sensing Bulletin, 2021, 25 (10) : 2127 – 2141.

[13] 郁宗桥, 王育红, 刘文宋, 等. 顾及 GF – 3 影像纹理特征的洪涝范围提取方法 [J]. 测绘通报, 2022, 542 (5) : 26 – 31.

Yu Z Q, Wang Y H, Liu W S, et al. Flood extent extraction method based on the texture features of GF – 3 images [J]. Bulletin of Surveying and Mapping, 2022, 542 (5) : 26 – 31.

[14] Nemni E, Aylett B J, Belabbes S, et al. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar Imagery [J]. Remote Sensing, 2020 (12) : 2532.

[15] Lyu W T, Yu Q Z, Yu W X. Water extraction in SAR images using GLCM and Support Vector Machine [C] // IEEE 10th International Conference on Signal Processing Proceedings, IEEE, 2010: 740 – 743.

[16] Zhang S C, Ma Z M, Li Z H, et al. Using CYGNSS data to map flood inundation during the 2021 extreme precipitation in Henan Province, China [J]. Remote Sensing, 2021, 13 (24) : 5181.

[17] Li Z P, Liu Y L, Cheng Y F, et al. Sentinel – 2A MSI and Landsat 8 OLI provide complementary information for estimating chlorophyll – a concentration in a turbid and eutrophic reservoir [J]. Remote Sensing of Environment, 2016, 173: 236 – 247.

[18] 李艳玲, 王加俊. 基于高斯混合模型的纹理图像的分割 [J]. 微电子学与计算机, 2004 (4) : 63 – 65, 68.

Li Y L, Wang J J. Texture image segmentation based on Gaussian Mixture Models [J]. Microelectronics & Computer, 2004 (4) : 63 – 65, 68.

[19] Vapnik V. Pattern recognition using generalized portrait method [J]. Automation and Remote Control, 1963 (24) : 774 – 780.

[20] Cortes C, Vapnik V N. Support – vector networks [J]. Machine Learning, 1995 (20) : 273 – 297.

[21] Cover T, Hart P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13 (1) : 21 – 27.

[22] Schowengerdt R A. Remote sensing: Models and methods for image processing [M]. San Diego: Academic Press, 1997: 411 – 412.

[23] 黄敏敏. 基于 Seminel – 1/2 数据的寿光市洪涝灾害监测与预测研究 [D]. 南京: 南京信息工程大学, 2022.

Huang M M. Research on flood disaster monitoring and prediction in Shouguang City based on Seminel – 1/2 data [D]. Nanjing: Nanjing University of information Science & Technology, 2022.

[24] Wang L H, Ma H, Li J L, et al. An automated extraction of small – and middle – sized rice fields under complex terrain based on SAR time series: A case study of Chongqing [J]. Computers and Electronics in Agriculture, 2022 (200) : 107232.

Remote sensing monitoring method for flooded farmland
based on domestic GF-3 radar images

YANG Chiyi^{1,2}, GUAN Haixiang^{1,2}, WU Wei³, LIU Meiyu³, LI Ying^{4,5}, SU Wei^{1,2}

(1. College of Land Science and Technology, China Agricultural University, Beijing 100083, China; 2. Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture, Beijing 100083, China; 3. National Disaster Reduction Center of China, Ministry of Emergency Management of the People's Republic of China, Beijing 100124, China; 4. Henan Agrometeorological Support and Applied Technique Key Laboratory, China Meteorological Administration, Zhengzhou 450003, China; 5. Henan Institute of Meteorological Sciences, Zhengzhou 450003, China)

Abstract: Against the backdrop of global warming, increasingly frequent floods become a primary agricultural disaster that causes reduced crop production in China. Radar remote sensing technology, possessing all-weather earth observation capabilities, serves as a critical means for rapid monitoring of regional flood information. With the advancement in artificial intelligence, machine learning methods have been extensively applied in the remote sensing-based monitoring of floods. Despite the high accuracy of their algorithms, their training processes often entail extensive field investigations or numerous samples for remote sensing image interpretation. This study aims to overcome sample labeling limitations and improve regional flood monitoring accuracy. Based on the catastrophic flood that occurred in northern Henan on July 20, 2021, this study constructed a flooded crop monitoring method based on the weakly supervised Gaussian mixture model (GMM) using domestic GF-3 HH-HV radar images. Then this method was applied to extract the flooding range of farmland in some areas of northern Henan. Compared to four typical machine learning methods, i.e., random forest (RF), support vector machine (SVM), K-nearest neighbor classification, and parallelepiped classification, the weakly supervised GMM in this study enjoyed the highest accuracy, with overall precision of 0.95 and a Kappa coefficient of 0.90. This study holds great significance for enhancing the accuracy and universality of regional crop flooding monitoring based on remote sensing technology and synthetic aperture radars (SARs).

Keywords: GF-3 radar image; flood; farmland flooding; Gaussian mixture model

(责任编辑: 李 瑜)