

# 基于时序 Sentinel-2A 影像的玉米秸秆覆盖区智能识别研究

陶万成<sup>1,2</sup>, 张颖<sup>1,2</sup>, 谢茈萱<sup>1,2</sup>, 王新盛<sup>1,2</sup>, 董镜<sup>1,2</sup>,  
张明政<sup>1,2</sup>, 苏伟<sup>1,2\*</sup>, 李佳雨<sup>1,2</sup>, 轩阜<sup>1,2</sup>

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

**摘要** 秸秆还田是减少土壤侵蚀、增加土壤有机碳的重要措施,对黑土地保护具有重要意义。区域范围内玉米秸秆覆盖区的准确、快速识别,对监测保护性耕作实施、农业补贴政策的制定具有重要作用。以实施保护性耕作的典型区吉林省四平市为研究区,基于GEE(Google Earth Engine)云平台,结合2020年5月—11月的Sentinel-2时序遥感影像,依据玉米生长季和收获后的秸秆状态构建光谱特征和指数特征,指数特征包括归一化差值植被指数(NDVI)和归一化差值秸秆指数(NDRI)。为避免数据冗余,对时序特征值按大小排序,同时利用分位法以0%,25%,50%,75%,100%分位选取分位(QT)特征,进而构建数据集。应用参数优化后的随机森林方法对按照7:3划分的样本集进行训练和验证,然后对数据集分类,结合连通域标定法去除分类过程中产生的细小连通域,进一步优化全局结果。通过Kappa和整体精度(OA)定量和定性评价,实验结果表明:(1)基于不同特征集组成数据集的分类模型(M1/M2/M3/M4/M5)定量评价结果均优于90%,其中所设计数据集的分类模型M5效果最好,Kappa和OA分别为97.41%和97.91%,相比于未加入QT特征集的分类模型M2的Kappa和OA分别提升4.52%和3.64%,同时M5识别结果可以有效保留边缘细节信息;(2)针对不同时间尺度的QT特征集,利用5月—11月时序遥感影像的QT特征集分类模型M5\_6/M5可以极大地抑制其他作物秸秆的影响,相比仅利用11月时序影像QT特征的M5\_1模型分类结果的Kappa和OA分别提升了3.9%和3.12%;(3)基于M5模型,结合连通域标定法的分类模型M6的Kappa和OA分别为96.76%和97.36%,仅次于M5模型识别结果,模型M6在保证较高精度的同时避免了细碎图斑,优化了分类可视化效果。该研究提出的M6模型适用于识别研究区玉米秸秆覆盖区,该方法能够在GEE云平台环境下快速执行,适合推广应用于东北地区秸秆覆盖区。

**关键词** 秸秆覆盖区; GEE云平台; 时序 Sentinel-2A 影像; 随机森林; 连通域

**中图分类号:** TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)06-1948-08

## 引言

黑土是宝贵的自然资源,是肥力最高、最适宜农耕和最具生产潜力的土壤。黑土地保护性耕作模式是保护黑土地,促进农业可持续发展的重要措施。保护性耕作是指作物收获后地面上秸秆覆盖度大于30%的耕作和种植系统<sup>[1]</sup>,针对黑土地的保护我国东北地区正积极实施、推广保护性耕作模式<sup>[2]</sup>。秸秆还田不仅有利于资源合理利用,还能避免因秸秆露天焚烧造成的环境污染<sup>[3]</sup>,与传统耕作方式相比,长期实施秸秆还田有助于降低土壤容重,提高土壤孔隙度、土壤有

机碳和土壤大团聚体百分比等<sup>[4]</sup>,对土壤保护与资源可持续利用起着重要作用。在亟需对黑土地实施保护措施的形势下,究竟哪些地块、有多少黑土地实施了秸秆覆盖,对黑土地保护相关科学研究和当地政府政策制定与实施具有重要意义。

传统作物秸秆覆盖区地面调查和识别的方法费时、费力、成本高,难以在区域范围内实施。遥感是快速获取区域范围内地表信息的技术手段,特别是欧空局(European Space Agency, ESA)发布的Sentinel-2卫星影像具有对秸秆纤维素和木质素敏感的短波红外波段(SWIR),是快速、准确监测区域范围内黑土地上作物秸秆覆盖的有效数据源。已有许多研究者根据秸秆以2100 nm处的强吸收谷与木质素、纤维

收稿日期: 2021-06-01, 修订日期: 2021-08-01

基金项目: 国家自然科学基金项目(42171331), 中国农业大学2115人才工程项目资助

作者简介: 陶万成, 1993年生, 中国农业大学土地科学与技术学院博士研究生 e-mail: B20203210937@cau.edu.cn

\* 通讯作者 e-mail: suwei@cau.edu.cn

素的高度相关性,利用多光谱影像、地面实测秸秆高光谱数据构建光谱指数展开秸秆覆盖度研究<sup>[5-7]</sup>。秸秆覆盖度研究一般仅利用作物收获后单一时相影像估算,无法表征不同秸秆覆盖类型的实际秸秆覆盖量,例如,高留茬秸秆覆盖从覆盖度的角度分析其覆盖度并不大,但实际上是 4~50 cm 的留茬直立在地表,其实际覆盖量并不小。因此,本研究进行秸秆覆盖区识别,是进一步开展地块尺度秸秆覆盖的基础,也是对秸秆覆盖类型精细分类(如覆盖还田、粉碎还田、高留茬等)的前提,通过遥感影像直接获取高精度的玉米秸秆覆盖区更具有实际意义。目前,利用遥感技术的秸秆覆盖区识别研究还较少,且没有综合考虑作物生长期和收获后的状态信息。本工作基于作物不同状态信息对玉米秸秆覆盖区开展精细、智能识别研究。

玉米是我国东北黑土地上主要种植的一种作物;吉林省四平市内有针对黑土地保护创建的“梨树模式”区域,广泛存在少耕、免耕等保护性耕作方式,又大量存在翻耕等传统耕作方式,非常适合开展秸秆覆盖区识别研究。由于玉米秸秆覆盖地块的反射率较高,且具有大面积连通的自然特性,在 Sentinel-2A 影像上表现为亮度偏高特征,呈现区域规则连续分布。根据秸秆覆盖大于 30% 时为保护性耕作的规定,遥感影像可视作秸秆覆盖区和背景两类。随机森林通过给定样本的学习训练形成分类规则,具有分析复杂地理信息系统分类

特征的能力,适用于分类和高维数据的回归分析,同时具备更高的精度和稳健性。在作物倒伏识别<sup>[8]</sup>和土地覆盖分类<sup>[9]</sup>等方面,随机森林都具有很好的分类效果和处理速度,因此受到广泛关注。基于像素的随机森林算法分类结果中普遍会出现细小图斑。为克服不足,应用形态学滤波去除图斑。

鉴于此,本研究基于 GEE 云平台<sup>[10-11]</sup>,综合考虑玉米收获前、后的状态信息,结合研究区内 2020 年 5 月—11 月份的时序 Sentinel-2A 影像创建的光谱指数和分位(quantile, QT)特征构建数据集,应用优化后的随机森林方法准确识别玉米秸秆覆盖区域;根据秸秆覆盖大范围连通的特性,利用连通域标定法优化分类结果,实现玉米秸秆覆盖区制图。

## 1 实验部分

### 1.1 研究区概况

研究区为吉林省四平市,地理范围为 42°31'—44°09'N, 123°17'—125°49'E,总面积约 1.4 万 km<sup>2</sup>,地理位置如图 1(a)所示。四平市处在黑土地和吉林省“黄金玉米带”核心地区,从 2007 年开始探索并实施以“秸秆覆盖、条带休耕”为主要内容的保护性耕作。开展区域范围内的玉米秸秆覆盖区准确识别,可应用于评估保护性耕作模式的实施程度,为监测保护性耕作推广,相关政策拟定提供可靠的数据支持。

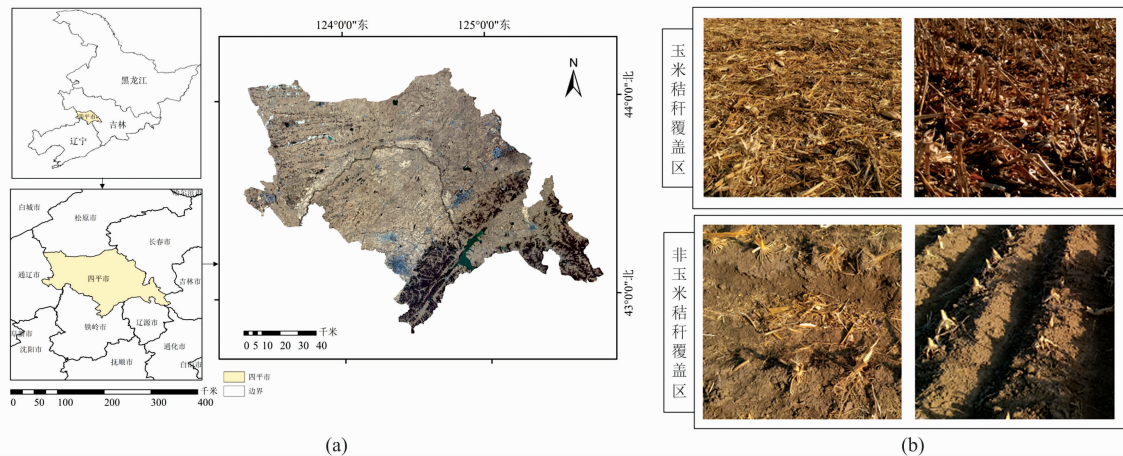


图 1 研究区位置与 Sentinel-2A 影像(日期: 2020 年 11 月 4 日)(a)与秸秆覆盖区照片(b)

Fig. 1 Study area and Sentinel-2A acquired on November 4, 2020 (a), photos of cornresidue cover areas (b)

### 1.2 数据源

#### 1.2.1 地面调查数据

2020 年 11 月 3 日—11 日,在研究区进行了田间玉米秸秆覆盖实地调研。11 月为研究区内玉米收获后秸秆覆盖的稳定时期,存在大量的样本地块。根据 Sentinel-2A 影像的最高空间分辨率(10 m),在样本地块中随机选取 10 个分布均匀、大小为 1 m×1 m 的样方,采用拉绳法测量。秸秆调查时使用两根以 0.1 m 间隔作为标记的 1 m 刻度绳,垂直摆放在样方中,计算标记点与地面秸秆相交点数量与总标记数比值,统计 10 个样方的均值;依据保护性耕作的定义<sup>[1]</sup>对秸秆覆盖区进行识别,图 1(b)中当统计的样方均值大于 30% 时为秸秆覆盖区,主要由机械收获后产生,包括低和高留茬;

低于 30% 时为非覆盖区,主要是人工收获后产生的根茬。为了识别玉米秸秆覆盖区,采用实地调查和目视解译相结合的手段获取样本数据集,在剔除异常值后保留了 646 个样本:包括玉米秸秆覆盖、水稻秸秆覆盖、建筑用地、树林和水体样本,样本数量分别为 200, 103, 118, 137 和 88。

#### 1.2.2 Sentinel-2A 遥感影像

从 Google Earth Engine(GEE)云平台获取时序 Sentinel-2A 影像,时间分辨率为 5 d,光谱包括可见光、近红外、红边、水汽、卷云和短波红外波段,各波段分辨率分别为 10、20 和 60 m。为避免低分辨率波段对分类任务的影响,舍弃空间分辨率为 60 m 的水汽、卷云和可见光中的气溶胶波段,选择 B2(Blue), B3(Green), B4(Red), B5(Red-edge1), B6

(Red-edge2), B7 (Red-edge3), B8 (Nir), B8A (Red-edge4), B11 (SWIR1) 和 B12 (SWIR2) 较高空间分辨率波段。

利用 GEE 获取研究区内 2020 年的云占比小于 10% 的 Sentinel-2A 影像; 同时选取 20 个实地调查的玉米秸秆覆盖样本点, 统计年内归一化差值植被指数 (normalized difference vegetation index, NDVI) 和归一化差值秸秆指数 (normalized difference residue index, NDRI) 特征曲线和置信区间, 如图 2 所示。S1 为玉米出苗期, H 为玉米收获期, S2 为玉米秸秆覆盖期, 横坐标为年积日 (day of year, DOY), 图 2 (a) 纵坐标为 NDVI 均值 (NDVI<sub>M</sub>) 和 NDVI<sub>M</sub> ± 标准差 (Std), 图 2 (b) 纵坐标为 NDRI<sub>M</sub> ± Std。图 2 (a) 中 DOY 为 204 时达到 NDVI<sub>M</sub> 最大值 0.864。图 2 (b) 中 NDRI 和相应时间区间的 NDVI 趋势相似, 但值范围相差较大。为降低计算

量, 选取从 5 月 13 日到 11 月 29 日的 182 景云占比小于 10% 的 Sentinel-2A 作为影像源, 充分合理地利用时间序列特征。根据研究需求, 对获取影像进行波段筛选、拼接和裁剪等预处理。

1.3 方法

基于时序 Sentinel-2A 的玉米秸秆覆盖区识别包括 3 个步骤。首先, 基于 Sentinel-2A 时序影像计算光谱和指数特征, 同时利用分位法获取分位 QT 特征, 进而构建数据集; 然后, 对样本数据按 7 : 3 随机划分为训练集和测试集, 结合 RF 模型分类出玉米秸秆覆盖区域, 而基于像素的 RF 分类结果中极易产生细小图斑, 研究采用连通域标定法优化全局; 最后, 通过评价指标定量分析结果。具体技术流程如图 3 所示。

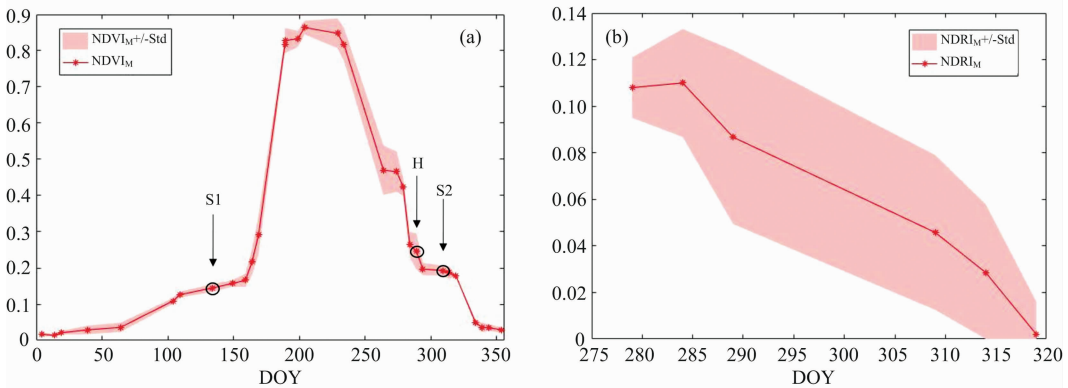


图 2 研究区 2020 年内玉米种植区时序 NDVI 和 NDRI 变化曲线

(a): 时序 NDVI 曲线; (b): H-S2 期间 NDRI 曲线

Fig. 2 Time-Sequence NDVI and NDRI change curves of corn planting areas in the study area in 2020

(a): Temporal NDVI variation curve; (b): NDRI variation curve during H-S2

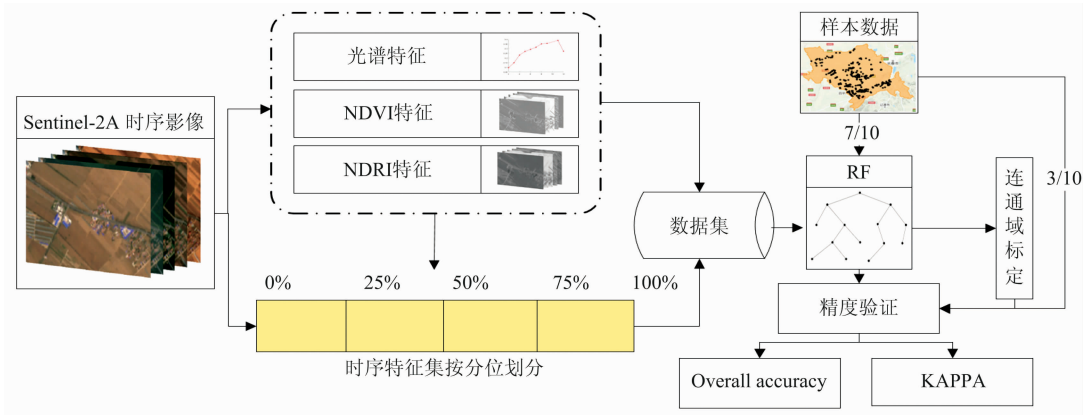


图 3 玉米秸秆覆盖区识别技术流程图

Fig. 3 Flowchart of identification technology of corn residue cover area

1.4 数据集构建

1.4.1 光谱特征

对研究区内 2020 年 11 月 4 日获取的 Sentinel-2A 多光谱反射率影像进行统计分析得知, 建筑物光谱特征较明显、容易识别, 因此重点分析玉米秸秆覆盖区、水稻秸秆覆盖区、树林和水体样本光谱特征。统计样本内各地类波段反射率值并绘制箱线图, 结果如图 4 (a-d)。

从图 4 (a, b) 可以看出, 玉米秸秆和水稻秸秆的光谱反射率区间范围为 0.10~0.45, 图 4 (c, d) 水体和树林光谱反射率区间范围分别为 0.00~1.05 和 0.00~0.30。通过对比发现, 水体和树林相对秸秆反射率明显较低, 易于分类。水稻秸秆和玉米秸秆的光谱反射率变化趋势和区间范围基本相同, B2—B8 波段反射率呈上升趋势, B8A—B12 波段呈下降趋势, 仅利用单期遥感影像很难进行地类判别。鉴于此, 考

考虑到不同作物的物候和结构特性，本研究使用 7 月 22 日 Sentinel-2A 影像提高水稻和玉米地类的分离度，即在作物生长季中玉米种植区的 NDVI 值大于水稻。为了判断该期影像是否有助于区分不同作物，利用样本统计绘制出玉米和水稻的光谱反射率箱线图，结果如图 5(a, b)。图 5(a) 中玉米 B2—B5 波段反射率明显低于图 5(b) 水稻，B7—B8A 反射率高于水稻，其他波段反射率相似，其原因可能与两种农作物

的冠层结构相关，这些特征有利于分类。因此，光谱特征集是结合 7 月 22 日和 11 月 4 日的 Sentinel-2A 影像建立。

1.4.2 指数特征

为区分玉米秸秆覆盖区、水稻秸秆覆盖区、裸地和其他相似地类，分别构建了差值植被指数 (NDVI) 和秸秆指数 (NDRI)。其中，NDVI 可以较好地反映出不同农作物的生长状况，因此用于区别玉米和水稻种植区域。NDRI 光谱指数与

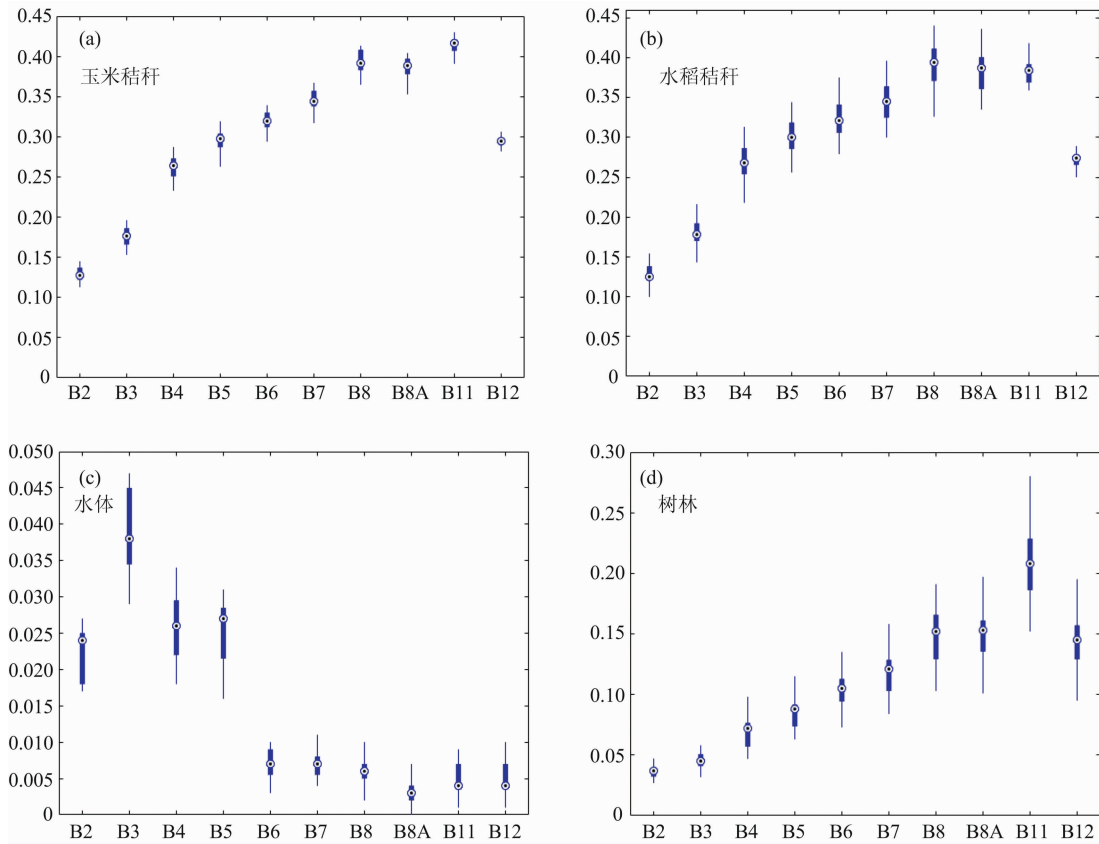


图 4 不同地物类型光谱反射率特征 (2020 年 11 月 4 日 Sentinel-2A 影像)

(a): 玉米秸秆; (b): 水稻秸秆; (c): 水体; (d): 树林

Fig. 4 Spectral reflectance characteristics of different ground types (Sentinel-2A images on 4 November 2020)

(a): Corn stalk residue; (b): Rice stalk residue; (c): Water; (d): Woods

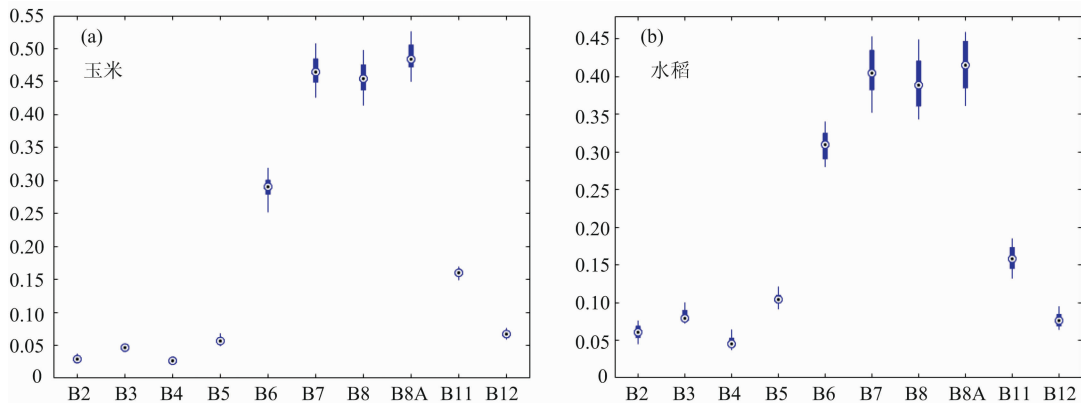


图 5 不同作物光谱反射率特征 (2020 年 7 月 22 日 Sentinel-2A 影像)

(a): 玉米; (b): 水稻

Fig. 5 Spectral reflectance characteristics of different crops (Sentinel-2A image on 22 July 2020)

(a): Corn; (b): Rice

NDVI 相似, 同样是利用双波段进行计算, 该光谱指数与玉米秸秆覆盖度显著相关<sup>[12]</sup>, 其计算公式如式(1)

$$NDRI = \frac{\rho_{B12} - \rho_{B4}}{\rho_{B12} + \rho_{B4}} \quad (1)$$

式(1)中,  $\rho_{B12}$  是短波红外 2 反射率,  $\rho_{B4}$  为红波段反射率。在玉米秸秆中含有大量的纤维素和木质素, 反射率在 2 100 和 2 300 nm 附近有强吸收, 即 Sentinel-2A 影像的 B12 波段反射率。使用 Sentinel-2A 的 B12 和 B4 波段反射率构建的 NDRI 可以最大限度地减少绿色植被的影响。综上分析, 在光谱特征集的基础上, 利用 NDVI 和 NDRI 建立指数特征集。

### 1.4.3 分位特征

分位 QT 特征的建立主要基于四分位数的思想。在统计学中, 将所有数值按大小排列并分成四等份, 处于三个分割点位置的数值为四分位数。同理, 四分位数法可应用于时序

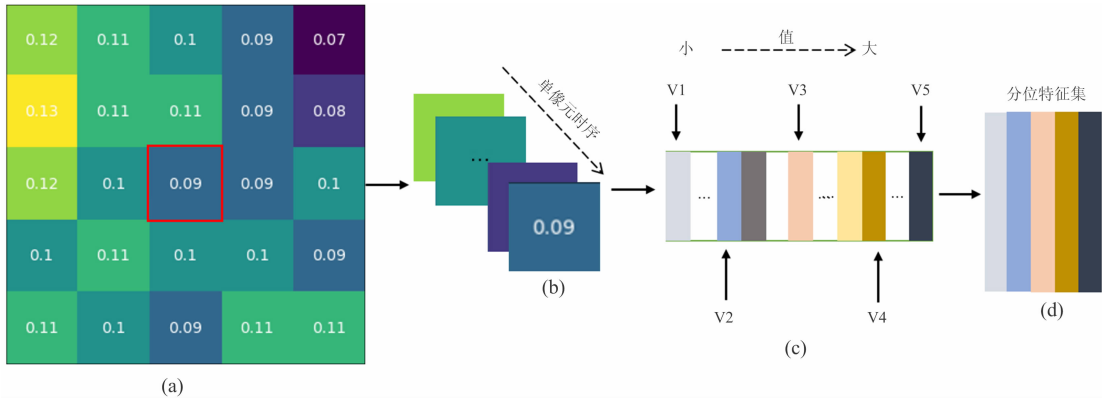


图 6 分位 QT 特征构建示意图

(a): 蓝波段热图; (b): 单像素时序数据; (c): 时序数据抽选; (d): 分位数据集

Fig. 6 QT feature construction schematic diagram

(a): Blue wave heat map; (b): Single pixel time series data; (c): Time series data selection; (d): Quantile dataset

### 1.5 随机森林秸秆覆盖区识别

随机森林监督分类算法是一种基于多棵 CART 决策树组合构成的集成学习方法。为了完全划分变量空间, 利用自助抽取输入样本和节点随机分裂技术构建多棵决策树。通过单个决策树的多数投票策略获得预测结果, 有效避免单个模型或参数引起的误差, 具有更高的精度和鲁棒性<sup>[13]</sup>。由于其分类的优越性, 该分类器得到遥感界的广泛认可。

针对所设计的数据集, 具有高维度特性, 随机森林可以将每个维度的重要性输出到分类中, 帮助特征选择以提高效率。因此, 本研究使用嵌入在 GEE 中的随机森林算法对玉米秸秆覆盖区进行分类。根据分类结果的准确性调整决策树数量以期生成最佳的分类模型。

### 1.6 连通域标定和评价指标

研究区中秸秆覆盖地块一般具有大面积连通的天然特征, 而本研究基于像素级别对玉米秸秆覆盖区进行分类, 结果中会出现细碎的图斑, 因此有必要借助连通域标定法对结果全局优化。连通域标定法示意图见图 7, 假定图 7(a)为分类结果中某一区域, 标签 1 表示为玉米秸秆覆盖区, 标签 0 表示背景类, 绿色区域为设定的 4 连通域滑动窗口。通过滑动窗口寻找标签为 1 的连通域, 保留图 7(b)左上角最大连通

遥感影像以构建 QT 特征, 流程如图 6 所示。首先, 通过 GEE 云平台获取时序影像集, 图 6(a)中为 11 月 4 日影像中蓝波段热图, 颜色越深反射率值越小, 相反则反射率值越大; 以中间红框内的像素为例, 抽取该位置时间维度上的数据, 如图 6(b); 对获取的数据按照从小到大进行排序, V1—V5 分别对应 0%, 25%, 50%, 75% 和 100%, 依次选择对映分位上的反射率值, 如图 6(c); 最后, 按照上述方法对影像中像素遍历处理获取光谱分位集, 如图 6(d)。

QT 特征还可以应用于时序指数特征集, 与计算光谱分位集相似, 首先获取时序指数特征集, 然后应用 QT 特征提取指数分位集。QT 特征集构建包括光谱分位集和指数分位集, 这种提取特征的方式利用时序特性的同时避免了数据冗余, 极大地降低了计算量。

域(黄色区域), 删除右下角最小连通域(黄色区域)。如果图像中存在大量目标区域, 则需标定所有的连通域, 通过设置阈值对全局处理。

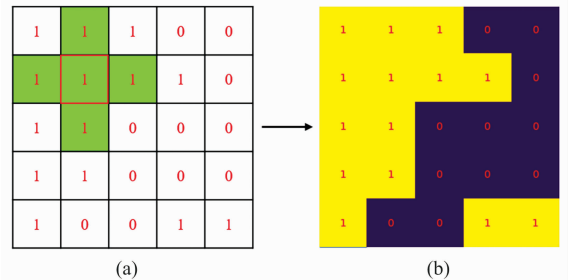


图 7 连通域标定示意图

(a): 分类区域数值图; (b): 像素连通域示意图

Fig. 7 Connected domain calibration schematic diagram

(a): Numerical diagram of classification area;

(b): Schematic diagram of connected domain of pixels

Kappa 系数是衡量分类结果空间一致性的指标, 明确揭示了分类结果的空间变化。总体精度(overall accuracy, OA)以预测正确与总体数量之间的比值反映分类结果的质量<sup>[14]</sup>。

本研究选择上述指标来定量定性评价分类结果。

## 2 结果与讨论

基于时序 Sentinel-2A 影像设计的数据集，采用随机森林算法识别玉米秸秆覆盖区。为确定算法中最佳决策树数量，研究对比 5, 10, 20, 30 和 40 决策树对分类模型的影响，具体见表 1。从 5 到 30 整体呈上升趋势，30 达到最大值并趋于稳定，当决策树数量设置为 40 时，评价精度略有下降。因此，实验统一标准，将决策树数量均设置为 30。

表 1 不同决策树的分类模型定量评价  
Table 1 Quantitative evaluation of classification models for different decision trees

决策树数量	Kappa/%	OA 总体分类精度/%
5	93.97	95.07
10	94.54	95.61
20	96.65	95.76
30	97.41	97.91
40	96.98	97.54

### 2.1 基于不同特征集组合分类

在相同训练样区条件下，通过不同的特征集组合识别玉米秸秆覆盖区，定量评价结果见表 2。以光谱特征集为基础，模型 M5 精度最高，Kappa/OA 为 97.41%/97.91%，比 M1, M2 和 M4 模型的 Kappa/OA 分别提高了 5.84%/4.69%，

4.52%/3.64% 和 1.29%/1.04%，表明相比于指数特征集，加入 QT 特征集可以明显改善模型精度；以 QT 特征集为基础，M4 比 M3 模型提高 1.95%/1.56%，表明与指数特征集相比，加入光谱特征集更有利于分类；以指数特征集为基础，M3 比 M2 模型提高 1.28%/1.04%，表明加入 QT 特征集优于光谱特征集。

表 2 基于不同特征集的分类定量评价结果  
Table 2 Classification quantitative evaluation results based on different feature dataset

模型	光谱特征集	指数特征集	QT 特征集	Kappa/%	OA/%
M1	+	-	-	91.57	93.22
M2	+	+	-	92.89	94.27
M3	-	+	+	94.17	95.31
M4	+	-	+	96.12	96.87
M5	+	+	+	97.41	97.91

本研究使用五种不同特征集组合模型，测试结果见图 8，图 8(a)和图 9(a)分别为 2020 年 11 月 4 日和 7 月 22 日真彩色合成图，分类结果中蓝色和白色区域分别表示玉米秸秆覆盖区和背景。根据图 8(b)–(f)对比发现，M5 模型一定程度上降低细小图斑的产生，同时保留道路、农田等边缘细节信息，效果最好；M1 和 M2 模型分类结果较差，易产生噪声，将道路等地类分为秸秆覆盖区；M3 和 M4 模型分类结果相对较好，保留了部分边缘细节信息。实验结果与定量评价结果基本一致，QT 特征的加入极大改善了分类可视化结果。

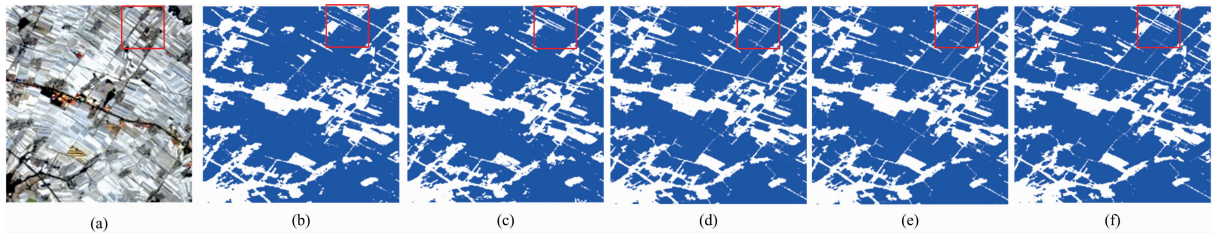


图 8 基于不同特征集的分类可视化结果

(a): 原图 1\_w; (b): M1 结果; (c): M2 结果; (d): M3 结果; (e): M4 结果; (f): M5 结果

Fig. 8 Visualization results of classification based on different feature dataset

(a): The original sub-image 1\_w; (b): The result of M1; (c): The result of M2; (d): The result of M3; (e): The result of M4; (f): The result of M5

### 2.2 基于不同时间尺度的特征分类

通过上述实验证明了 QT 特征的重要性，为了评估不同时间尺度生成的 QT 特征集对识别秸秆覆盖区的影响，本研究基于 M5 模型，对利用不同 QT 特征集的 M5\_1—M5\_6 模型定量分析，具体信息见表 3。

根据表 3，利用 5 月—11 月影像创建的 QT 特征 M5\_6/M5 模型分类效果最好。其中，M5\_2 比 M5\_1 模型的 kappa/OA 低 0.97%/0.77%，可能是 10 月为玉米收获期，秸秆地块中包括收获与未收获，收获的地块又包括秸秆成捆或散放在地里等，信息多样，识别困难；M5\_3 相比于 M5\_2 模型精度低，分析认为 9 月份研究区受到台风影响，缺少质量好的影像。而 M5\_4, M5\_5 和 M5\_6 精度相继提升表明长

时序影像创建的 QT 特征集有利于识别秸秆覆盖。

表 3 基于不同时间尺度特征分类定量评价结果  
Table 3 Quantitative evaluation results based on different time scale feature classification

	5月—6月	7月	8月	9月	10月	11月	Kappa/%	OA/%
M5_1	-	-	-	-	-	+	93.51	94.79
M5_2	-	-	-	-	+	+	92.54	94.02
M5_3	-	-	-	+	+	+	92.15	93.71
M5_4	-	-	+	+	+	+	93.89	95.17
M5_5	-	+	+	+	+	+	95.46	96.35
M5_6	+	+	+	+	+	+	97.41	97.91

### 2.3 结合连通域标定法分类

在 M5 模型基础上, 结合连通域标定法构建模型 M6 对全局优化, 设置连通域阈值为 30, M6 分类结果的 Kappa/OA 为 96.76%/97.36%。相比于 M5 模型精度有所降低, 但仍高于其他模型, 同样适用于玉米秸秆覆盖区识别。选取两组测试结果对比如图 9 和图 10。

根据图 9(c), M5[图 9(d)]和 M6[图 9(b)]相比于 M5\_1 模型分类结果可以很好的保留细节信息, 进一步证明 QT 特征设计的有效性; M6 相比于 M5 模型在保留绝大部分细节信息的同时, 抑制了细碎图斑的产生。图 10(a)是夏季的子图块, 其中浅绿色为水稻地块, 图 10(b)是 11 月初秸秆覆盖的子图块通过对比图 10(c) M5\_1, 图 10(d) M5 和图 10(e) M6 分类结果发现, 在仅利用 11 月影像的 M5\_1 模型容易将水稻秸秆覆盖区分为玉米秸秆覆盖区, 而利用 5 月—11 月份

影像的 M5 模型可很好地将两者区分开。此外, M5 模型结果中有相对较少的细碎图斑, 结合连通域标定的 M6 模型在优化细碎图斑后得到更理想的分类结果, 见图 10(e)。

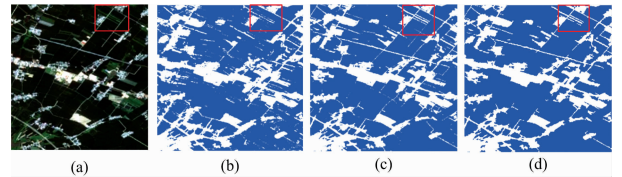


图 9 结合连通域标定分类结果 1

(a): 原图 1\_S; (b): M5\_1 结果; (c): M5 结果; (d): M6 结果

Fig. 9 Combined with the connected domain calibration classification result 1

(a): The original sub-image 1\_S; (b): The result of M5\_1; (c): The result of M5; (d): The result of M6

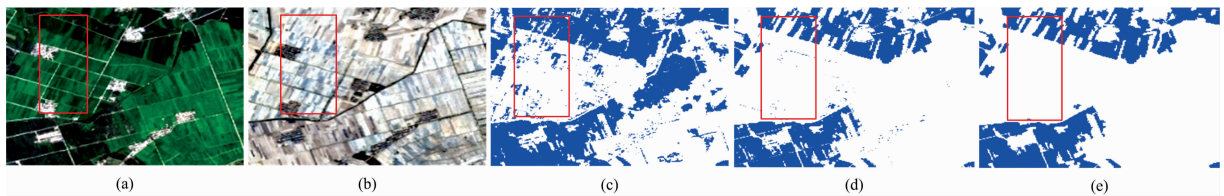


图 10 结合连通域标定分类结果 2

(a): 原图 2\_S; (b): 原图 2\_W; (c): M5\_1 结果; (d): M5 结果; (e): M6 结果

Fig. 10 Combined with the classification result of connected domain calibration 2

(a): The original sub-image 2\_S; (b): The original sub-image 2\_W; (c): The result of M5\_1; (d): The result of M5; (e): The result of M6

通过实验对比 M6 模型可视化效果最好, 定量评价结果仅次于 M5 模型, 因此研究使用 M6 模型对整个研究区进行预测, 得到的分类结果见图 11。其中, 左侧为整个研究区识别结果, 右侧为研究区框选出的子区域识别结果, 图 11(a)主要包括秸秆和建筑, 图 11(b)主要包括秸秆, 建筑, 水域和树林。从不同区域识别结果可以看出秸秆覆盖区边缘信息保持完好, 几乎没有细碎图斑, 证明了所提出方法的有效性, 适用于玉米秸秆覆盖区识别。

### 3 结论

以吉林省四平市为例, 基于 GEE 云平台, 首先对时序 Sentinel-2A 影像预处理, 利用光谱, 指数和 QT 特征构建数据集。然后应用最佳决策树个数的随机森林分类器分析不同特征组合, 不同时序尺度的 QT 特征和是否结合连通域标定对识别玉米秸秆覆盖区的影响。最终采用可视化效果最好, 精度较高的 M6 模型对研究区影像预测, 主要研究结论如下:

(1) 利用不同特征集组合的分类模型所设计的分类模型 M5 定量评价结果最优, 通过模型对比表明加入 QT 特征能够有效提升精度, 保留边缘细节信息, 在一定程度上降低了噪声出现的概率, 优于未加入 QT 特征的分类结果。

(2) 基于不同时序尺度的 QT 和光谱特征集的分类模型, 实验利用 5 月—11 月时序影像内 QT 特征比较短时序内影像的 QT 特征分类效果好, 表明长时序特征有利于玉米秸秆覆盖区识别, 同时可以避免其他作物秸秆覆盖区域的干扰。

(3) 在模型 M5 分类结果的基础上, 结合连通域标定法的 M6 模型对全局连通域进行识别, 通过设置阈值优化结果, 实验结果表明该方法在保证较高玉米秸秆覆盖区域识别精度的同时去除了细碎图斑。

通过本研究提出 M6 模型可以准确的从 Sentinel-2A 影像中识别出玉米秸秆覆盖区, 对当地监测保护性耕作实施和相关服务政策的拟定提供便利。后续研究主要有两点: ① 将着重挖掘影像空间域纹理特征在玉米秸秆覆盖识别中的应用潜力。② 尝试多源遥感影像融合进一步优化结果。

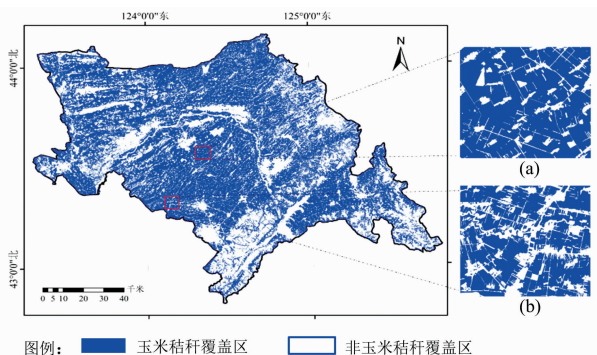


图 11 玉米秸秆覆盖区识别结果

(a): 分类结果子图 1; (b): 分类结果子图 2

Fig. 11 Identification results of corn residue cover area

(a): The sub-image 1 of classification result;

(b): The sub-image 2 of classification result

## References

- [ 1 ] Daughtry C S T, Doraiswamy P C, Hunt Jr E R, et al. *Soil and Tillage Research*, 2006, 91(1-2): 101.
- [ 2 ] Shen Y, McLaughlin N, Zhang X, et al. *Scientific Reports*, 2018, 8(1): 1.
- [ 3 ] Du Z, Ren T, Hu C. *Soil Science Society of America Journal*, 2010, 74(1): 196.
- [ 4 ] Watts D B, Runion G B, Balkcom K S. *Field Crops Research*, 2017, 201: 184.
- [ 5 ] Sindelar M, Blanco-Canqui H, Jin V L, et al. *Soil Science Society of America Journal*, 2019, 83(1): 221.
- [ 6 ] Jin X, Ma J, Wen Z, et al. *Remote Sensing*, 2015, 7(11): 14559.
- [ 7 ] Mupangwa W, Thierfelder C, Cheesman S, et al. *Renewable Agriculture and Food Systems*, 2020, 35(3): 322.
- [ 8 ] Guan H X, Liu H J, Meng X T, et al. *Remote Sensing*, 2020, 12(19): 3149.
- [ 9 ] Rodriguez-Galiano V F, Ghimire B, Rogan J, et al. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 67: 93.
- [10] Hu Y, Hu Y. *Remote Sensing*, 2019, 11(5): 554.
- [11] Venkatappa M, Sasaki N, Shrestha R P, et al. *Remote Sensing*, 2019, 11(13): 1514.
- [12] Gelder B K, Kaleita A L, Cruse R M. *Agronomy Journal*, 2009, 101(3): 635.
- [13] Bao Y, Meng X, Ustin S, et al. *Catena*, 2020, 195: 104703.
- [14] Du Z, Yang J, Ou C, et al. *Remote Sensing*, 2019, 11(7): 888.

# Intelligent Recognition of Corn Residue Cover Area by Time-Series Sentinel-2A Images

TAO Wan-cheng<sup>1,2</sup>, ZHANG Ying<sup>1,2</sup>, XIE Zi-xuan<sup>1,2</sup>, WANG Xin-sheng<sup>1,2</sup>, DONG Yi<sup>1,2</sup>, ZHANG Ming-zheng<sup>1,2</sup>, SU Wei<sup>1,2\*</sup>, LI Jia-yu<sup>1,2</sup>, XUAN Fu<sup>1,2</sup>

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 and Rural Affairs, Beijing 100083, China

**Abstract** Crop residue covering is an important way to reduce soil erosion and increase soil organic carbon, which is very important for black soil protection. Therefore, the accurate and rapid identification of corn residue cover area plays an important role in local government monitoring and promoting conservation tillage. The study area is located in Siping City, Jilin Province. Moreover, the time-series Sentinel-2A images collected from GEE (Google Earth Engine) cloud platform are used to capture spectral index based on the characteristics of the corn growing season and after harvest. Index features include Normalized Difference Vegetation Index (NDVI) and Normalized Difference Residue Index (NDRI). The time series feature values are sorted by size, and the quantile method is used to select quartile (QT) features at 0%, 25%, 50%, 75%, and 100% to construct datasets. On this basis, the random forest method after parameter optimization is applied to train and verify the sample datasets divided according to 7 : 3, and then the datasets are classified, combined with the connected domain calibration method to remove the small connected domains generated in the classification process, and further optimize the global result. Through the quantitative and qualitative evaluation of Kappa and Overall Accuracy (OA), the experimental results show that: (1) The quantitative evaluation results of the classification model (M1/M2/M3/M4/M5) based on the dataset composed of the different feature are superior 90%. Among them, the classification model M5 of the dataset designed in this paper has the best performance, of which Kappa and OA are 97.41% and 97.91%, respectively. Compared with the classification model M2 without the QT feature, the Kappa and OA are increased by 4.52% and 3.64%, respectively. At the same time, the M5 recognition result can effectively retain edge detail information; (2) For QT feature of different time scales, using the QT feature classification model M5\_6/M5 of time series remote sensing images from May to November can greatly restrain another crop residue. Compared with the Kappa and OA of the M5\_1 model classification result using only the QT features of the time series images in November, the Kappa and OA increased by 3.9% and 3.12%, respectively; (3) Based on the M5 model, the Kappa and OA of the classification model M6 combined with the connected domain calibration method are 96.76% and 97.36%, respectively, second only to the recognition results of the M5 model. The model M6 avoids fine-grained patches while ensuring high accuracy and optimizing the classification visualization effect. Therefore, the M6 model proposed in this paper is suitable for identifying areas covered by corn residue in the study area. This method can be quickly implemented in the GEE cloud platform environment and is suitable for popularization and application in a corn residue covered areas in Northeast China.

**Keywords** Crop residue cover area; GEE cloud platform; Time series Sentinel-2A image; Random forest; Connected domain

\* Corresponding author

(Received Jun. 1, 2021; accepted Aug. 1, 2021)