

# 基于 GEE 平台与 Sentinel-NDVI 时序数据江汉平原种植模式提取

张紫荆<sup>1</sup>, 华 丽<sup>1\*</sup>, 郑 萱<sup>2</sup>, 李嘉麟<sup>1</sup>

(1. 华中农业大学资源与环境学院, 武汉 430000; 2. 北京大学深圳研究生院城市规划与设计学院, 深圳 518000)

**摘 要:** 快速精准地掌握作物种植模式信息对于农业产能评估、区域种植结构平衡和国家粮食安全保障具有重要的战略意义。然而, 目前尚缺乏高精度、大范围、涵盖复杂种植模式的空间数据集。因此, 本研究以中国南方重要的产粮地之一江汉平原为例, 基于 GEE 平台和 Sentinel-NDVI 数据, 构建时序物候特征集, 通过探究半自动提取大量样本的方法, 对比多粒度级联森林 (multi-grained cascade Forest, gcForest) 与深度神经网络 (Deep Neural Network, DNN) 两种深度模型分类精度, 对江汉平原 5 种作物类型的 6 种植模式进行精细识别与制图研究。结果表明: 1) 江汉平原主要包括单季稻、莲藕两种单季种植作物; 小麦-水稻/棉花、油菜-水稻/棉花等 4 种年内复种模式; 2) 基于半自动采样点 (4 000 个) 的 gcForest 模型总体精度最高, 可达到 87.25%; 两种模型基于 4 000 个采样点的分类精度相较于 400 个实地采样点分别有 8.08 和 5.5 个百分点的提升, 该半自动提取样本点的途径可有效提高分类精度。该研究证明, 基于物候的 Sentinel-NDVI 数据在复杂农业景观区域的作物种植模式制图有较大潜力。

**关键词:** 模型; 时序数据; 半自动提取; 种植模式

doi: 10.11975/j.issn.1002-6819.2022.01.022

中图分类号: S127

文献标志码: A

文章编号: 1002-6819(2022)-01-0196-07

张紫荆, 华丽, 郑萱, 等. 基于 GEE 平台与 Sentinel-NDVI 时序数据江汉平原种植模式提取[J]. 农业工程学报, 2022, 38(1): 196-202. doi: 10.11975/j.issn.1002-6819.2022.01.022 http://www.tcsae.org

Zhang Zijing, Hua Li, Zheng Xuan, et al. Extraction of cropping patterns in Jiangnan Plain based on GEE and Sentinel-NDVI time series data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(1): 196-202. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.01.022 http://www.tcsae.org

## 0 引 言

随着粮食生产需求的快速增长和生态环境的恶化, 优化作物种植模式、维持农业生态系统平衡正迅速成为重要的挑战<sup>[1-2]</sup>。研究农业生态的关键数据之一是作物类型和种植模式信息, 其对于确定作物的具体管理模式至关重要<sup>[3-5]</sup>。然而, 在区域尺度上, 关于作物类型和种植模式的时空数据较少<sup>[6]</sup>。美国农业部国家农业统计局 (National Agricultural Statistics Service, NASS) 发布的农田数据 (Cropland Data Layer, CDL) 只提供了大宗作物种植频率, 但没有包括实际的作物种植模式<sup>[7]</sup>; Leteinturier 等<sup>[8]</sup>对比利时地区进行了作物种植结构信息的收集, 但在大多数欧洲国家, 由于数据保护法律这类信息无法共享。

随着遥感技术的发展, 越来越多的学者开始了作物种植结构识别提取研究。由于中分辨率成像光谱仪 (Moderate-resolution Imaging Spectroradiometer, MODIS) 具有高时间分辨率的特点, 目前已有多项基于该数据的种植结构制图成果。Chen 等<sup>[9]</sup>利用 MODIS 数据和决策树分类器对巴西 5 种作物类型的 6 种植模式制图, 其农

田、种植模式和作物类型的总体精度分别可达到 90%、73%和 86%; 黄青等<sup>[10]</sup>以中国东北为研究区域, 通过建立 MODIS-NDVI 种植结构提取模型, 获得了东北区主要种植结构格局特征, 其提取结果总体精度达 87%以上。然而, 由于 MODIS 的空间分辨率有限, 难以满足耕地破碎化地区的精细化制图要求<sup>[6]</sup>, Landsat 影像 (30 m) 尽管有了更高的空间分辨率, 但其时间分辨率较低, 难以通过完整的时序数据获取作物准确的物候信息<sup>[11]</sup>。随着开源哨兵数据的出现, 上述问题得到了有效解决<sup>[12]</sup>。Liu 等<sup>[12]</sup>结合 Sentinel-2 与 Landsat 数据开发了一种基于物候的算法, 提取中国 7 个典型区域作物种植强度与种植模式, 准确率达到 89%~99%; Belgiu 等<sup>[13]</sup>基于 Sentinel-2 时间序列数据和随机森林模型, 在奥地利农业区提取了 9 种作物, 精度在 91%~95%, 证明哨兵 2 号时序数据的应用为精细作物类型数据集创造了新的可能性。然而, 目前针对农业景观异质性较高特点的中国南方地区, 仍然缺乏覆盖范围广、精度高、作物和种植模式类型全面的“一张图”, 因此有必要探究 Sentinel-2 在该地区的制图潜力。

近年来, 随着图像分析技术的不断发展, 识别提取方法也从机器学习、线性极化等浅层模型向深度模型过渡, 越来越多的深度神经网络如 CNN<sup>[14-15]</sup>、DNN<sup>[16-17]</sup>、RNN<sup>[18-19]</sup>等被用于农作物精准提取。但研究表明, 深度神经网络对训练样本有更强的依赖性, 高质量、多数量的样本对精度的影响超过了分类器本身<sup>[20-21]</sup>。2017 年, 学者周志华提出了一种深度森林模型 gcForest, 该模型由

收稿日期: 2021-08-26 修订日期: 2021-12-15

基金项目: 国家青年科学基金项目 (41601280)

作者简介: 张紫荆, 研究方向为农业遥感。

Email: Zhangzijing@webmail.hzau.edu.cn

\*通信作者: 华丽, 博士, 副教授, 研究方向为农业遥感。

Email: huali@mail.hzau.edu.cn

级联森林结构和多粒度扫描结构组成。试验通过对比深度神经网络与 gcForest 在人脸识别、运动识别、图像识别等领域的识别精度,证明 gcForest 所需要的参数较少、容易训练,在一定程度上缓和了对大量训练样本的依赖性<sup>[22]</sup>。然而,对于两类模型的对比目前仍停留在上述试验领域,针对中国作物类型复杂,样本获取困难等现状,仍较少有研究探究两种不同类型的深度模型对复杂作物种植模式识别的适用性。

江汉平原是中国南方重要的产粮区之一,其面积广阔、作物类型和轮作模式复杂,绘制高精度种植结构图仍是一个挑战。本研究基于 GEE 平台调用 Sentinel-2 L2A 级数据,建立物候特征集,对比 gcForest 和 DNN 和两类不同的深度模型在不同样本数量下的识别的精度。这项研究的目的是:1) 探究开源的高分辨率哨兵 2 号时序数据及两种深度模型,对中国复杂农田景观下作物类型与轮作模式识别与制图的潜力;2) 提供可用于相关研究高精度空间数据集,弥补当前该数据集的缺失问题,为相关政策的制定与推动提供科学依据。

## 1 研究区概况与数据源

### 1.1 研究区概况

江汉平原位于湖北省的中南部,总面积约 2.9 万 km<sup>2</sup>,除西部和东北部有小部分丘陵外,基本为平原区。全年气候温和,年平均气温为 16.3 ℃。雨量充沛,年平均降水量在 1 100~1 300 mm,湖泊众多,河湖密布。作为中国举足轻重的商品粮基地,有“湖广熟,天下足”的美誉,因其独特的地理环境与气候条件,该地区形成了复杂的种植模式,其中水稻种植多为一年两熟的耕作模式,以“虾稻养殖模式”“小麦-中稻轮种”和“油菜-中稻轮种”为主。夏季棉花种植面积较大,主要以“小麦-棉花轮种”和“油菜-棉花轮种”为主。研究区地理位置与高程概况如图 1 所示。

### 1.2 数据介绍及预处理

#### 1.2.1 Sentinel-2 数据介绍及预处理

Sentinel-2 卫星的轨道类型为太阳同步轨道,高度为 786 km,采用推扫式成像模式,成像幅宽为 290 km,覆

盖可见光、近红外、短波红外等 13 个波段。空间分辨率达到 10~60 m,双星时间分辨率可达 2~5 d。

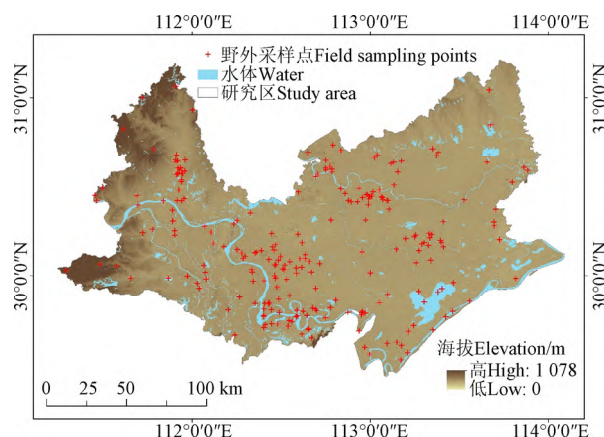


图 1 研究区数字高程和野外采样点分布

Fig.1 Digital elevation and field sampling points distribution of study area

本文基于 GEE 平台直接调用 Sentinel-2 源数据,Level L2A 属于经过正射校正、几何精校正和大气校正的大气表观反射率产品,数据的投影为 WGS84 UTM Zone49,按照研究区的范围,对数据进行裁剪和计算形成研究区内 Sentinel-NDVI 数据集。选取的影像时间节点涵盖 2019 年(4 月、5 月、7 月、8 月、11 月、12 月)和 2020 年(1 月、3 月、4 月、5 月)共 10 个时期的影像,且均满足研究区云量低于 10%的条件。由于研究区覆盖面积较大,各特征时期影像由该时期范围内影像拼接而成。

#### 1.2.2 野外采样数据介绍及预处理

结合 Google Earth 高清地图,借助手持全球定位系统(Global Positioning System, GPS)获取研究区内具体作物类型和土地利用类型的准确经纬度坐标,包括单季中稻、莲藕、油菜、小麦、棉花、林灌草(含灌木草地)、水体、不透水表层采样点共 400 个,按照 7:3 的比例将样本随机分为训练样本和精度验证点。

在布置野外调查点的同时,详细记录不同作物的物候信息。本研究涉及到的江汉平原主要作物物候日历见图 2。

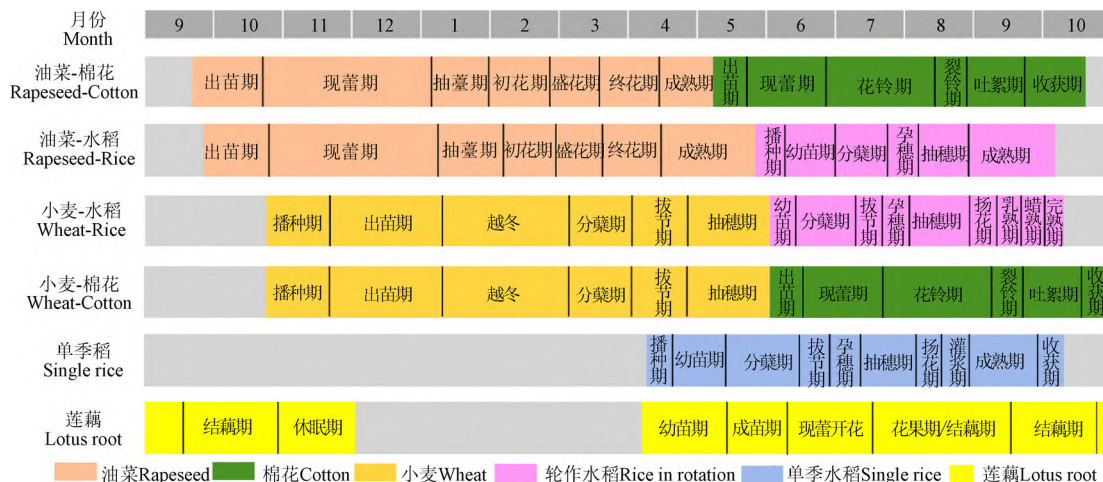


图 2 研究区主要种植模式物候日历

Fig.2 Phenology calendar of the main cropping patterns in study area

## 2 研究方法

本文的研究基于以下步骤: 1) 基于 GEE 平台, 利用 Sentinel-2 NDVI 数据和样本点数据, 构建物候时序曲线; 2) 根据曲线在关键节点的差异构建特征时期 NDVI 时序数据集, 同时基于野外采样点数据和半自动提取样本数据, 分别利用 gcForest 和 DNN 进行训练, 提取作物类型和种植模式; 3) 进行精度分析, 采用用户精度、Kappa 系数、总体精度等评价指标。

### 2.1 时序 NDVI 曲线与数据集构建

基于 GEE 平台提取多时序 NDVI 影像, 计算各种种植模式样本点平均 NDVI 值得到物候时序曲线 (图 3)。可以看到各种种植模式在不同时期展现出不同的变化特征。1) 4 种一年两熟的轮作种植模式中, 冬季作物为油菜与小麦, 区分油菜与小麦的关键节点为 3 月中旬。3 月中旬处于盛花期油菜的 NDVI 值有明显的波谷, 而小麦在该时期仍有较高的 NDVI, 能很明显的区分油菜和小麦; 2) 夏季轮作作物主要有棉花和水稻, 区分棉花与水稻的关键节点在 6—7 月。此时, 棉花处于出苗和现蕾期, 现蕾期前棉花的生长速度较慢, 但后期增长较快, 而水稻则相对匀速增长, 棉花与水稻的 NDVI 上升趋势和数值差异有所不同, 由此可以区分棉花和水稻; 3) 单季种植模式的作物有单季稻和莲藕, 11 月莲藕的值明显高于其他作物, 基本在 0.5 以上; 单季稻在 7 月有且只有一个明显的峰值。

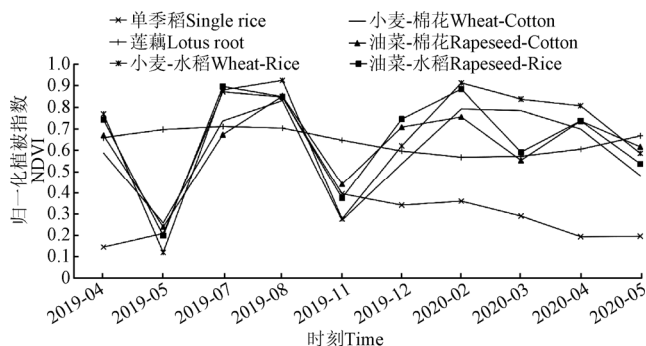


图 3 主要种植模式 NDVI 光谱曲线

Fig.3 NDVI(Normalized Difference Vegetation Index) spectral curves of main cropping patterns

## 2.2 分类模型

### 2.2.1 深度森林——多粒度级联森林

多粒度级联森林 (multi-Grained Cascade Forest, gcForest) 是 Zhou 等受到深度神经网络的启示提出的新的决策树集成方法。模型将训练分成级联森林和多粒度扫描两部分, 在此基础上进行深度集成, 已被证实在小样本集上仍有较好的分类效果<sup>[22]</sup>。本研究基于大区域范围农业样本难以获取的现状, 利用 Anaconda3 软件下的 Spyder 编译器, 通过编程使用默认模型进行 gcForest 分类, 添加 4 种基本分类器: XGBClassifier、RandomForestClassifier、ExtraTreesClassifier 和 LogisticRegression。

### 2.2.2 深度神经网络

深度神经网络 (Deep Neural Network, DNN) 由 Krizhevsky 等<sup>[23]</sup>于 2017 年提出, 是指具有多层结构的人工神经网络。DNN 内部的神经网络层可分为 3 类: 输入层、隐藏层和输出层, 每层各神经元遵循“输入-输出”的结构。相比传统浅层网络, DNN 可以更好的学习样本特征, 捕获影像的几何信息, 在分类应用中取得较优效果<sup>[16,24]</sup>。基于 Anaconda3 软件下的 Spyder 编译器进行 DNN 分类, 其包含 1 个隐藏层和 1 个 Dropout 层, 隐藏层提取 32 个维度的特征, Dropout 设置 0.2 的随机失活率。

### 2.3 半自动提取大量样本点及模型精度评价

大区域尺度下的农业精准监测一直都受到样本缺乏, 实地采集耗时耗力等问题的制约<sup>[25]</sup>。本研究探究了在实地样本分类结果的基础上, 利用 ArcGIS 平台半自动提取大量样本点以补充分类样本集的操作可行性。在基于两种深度模型和 400 个实地采样点的分类结果中选择总体精度较高的结果作为目标栅格, 运用 ArcGIS 平台中的“自动生成点”“多值提取至点工具”并对照各点的值与相应作物物候曲线进行人工筛选后, 获得各类地物和种植模式样本点共 4 000 个。按照 7:3 的比例将样本随机分为训练样本和精度验证点, 再次进行模型训练和结果预测。

本研究通过构建混淆矩阵来评价本研究分类算法的精度。评价指标包括用户精度 (Users' Accuracy, UA, %)、总精度 (Overall Accuracy, OA, %) 和 Kappa 系数。

## 3 结果与分析

### 3.1 精度对比分析

#### 3.1.1 总体精度对比分析

如表 1 所示, 在所有分类结果中, 基于半自动采样点 (4 000 个) 的 gcForest 模型的总体精度和 Kappa 系数最高, 总体精度达到 87.25%, 相比 DNN (400)、gcForest (400) 和 DNN (4 000) 的 78.33%、79.17%、83.83%, 分别提高了 8.92、8.08、3.42 个百分点; 同时, 无论基于 400 采样点或 4 000 半自动采样点, gcForest 都表现出更好的分类精度和 Kappa 系数, 表明该模型在复杂的作物种植模式分类提取的问题上有更强的优越性。

表 1 基于不同数量采样点的两种模型总体分类精度

Table 1 Overall classification accuracy of two models based on different number of sampling points

模型 Model	样点数 Number of samples	总体精度 Overall accuracy/%	Kappa 系数 Kappa coefficient
多粒度级联森林 Multi-grained cascade Forest (gcForest)	400	79.17	0.768
	4 000	87.25	0.856
深度神经网络 Deep Neural Network (DNN)	400	78.33	0.703
	4 000	83.83	0.820

注: 400 为野外采样点个数; 4 000 为半自动采样点个数。图 4 同。  
Note: The number of field sampling points are 400; The number of semi-automatic sampling points are 4 000. The same as Fig.4.



从总体上看,随着样本点个数的增加,两种深度模型的用户精度都有所提高。*gcForest* 的总体精度提高了 8.08 个百分点, DNN 的总体精度提高了 5.50 个百分点,表明半自动提取并扩充样本点数量的操作可行,且样本数量的增加会在一定程度上提高深度模型精度。

### 3.1.2 不同种植模式精度对比分析

由图 4 可知,基于 4 000 个采样点的 *gcForest* 模型总体用户精度都处于较高水平,表明该模型对于各作物及种植模式都有较好的识别能力。在所有待提取的类别中,相较 2 种单季的种植模式,2 个模型的 4 种轮作种植模式用户精度均整体较低。结合混淆矩阵热力图(图 5)可知,精度较低的 4 种轮作中,小麦-棉花大多与小麦-水稻错分,油菜-水稻也大多与油菜-棉花错分。这是由于棉花和水稻拥有相似的物候周期和物候曲线,区别的关键特征时期在 6—7 月,此时棉花与水稻的 NDVI 上升趋势和数值差异有所不同,棉花此时处于现蕾期,上升速度较缓慢,但后期增长较快,而水稻则相对匀速增长。但由于该趋势易受人为或气候影响,在面对实际问题时曲线更为复杂,趋势走向不明显且无法定量标准分析;小麦与油菜极少错分,原因在于 3 月中旬的油菜处在开花期,其 NDVI 远低于该时期的小麦,光谱特征较为明显。针对不同种植模式,*gcForest* 模型的用户精度均在 75% 以上,尽管在基于实地采样的小样本集也有较好的分类效果,而当样本集扩大时,其平均精度可达 88%,在单一特征,复杂种植模式的提取研究中展现了较高的识别能力。

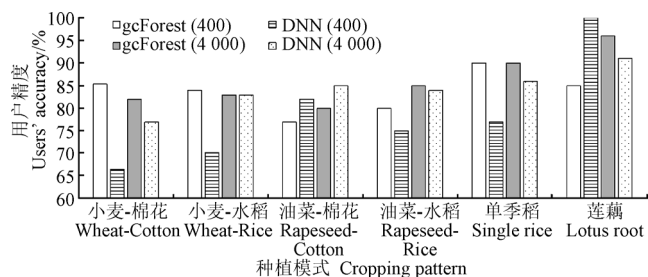


图 4 2 种模型基于不同采样点的用户精度

Fig.4 The users' accuracy in two models based on different sampling points

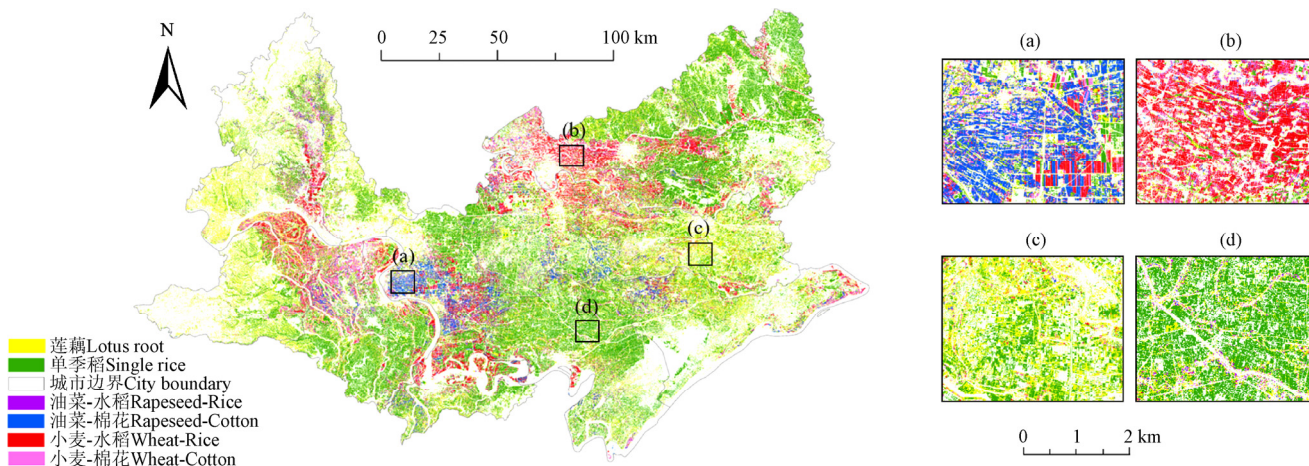
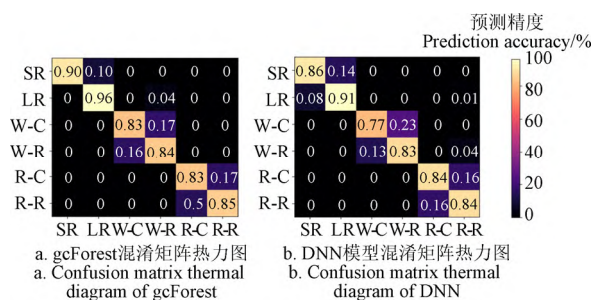


图 6 基于半自动采样点的 *gcForest* 模型分类结果

Fig.6 Classification results of *gcForest* model based on semi-automatic sampling points



注: SR: 单季稻; LR: 莲藕; W-C: 小麦-棉花; W-R: 小麦-水稻; R-C: 油菜-棉花; R-R: 油菜-水稻。  
Note: SR: Single rice; LR: Lotus root; W-C: Wheat-Cotton; W-R: Wheat-Rice; R-C: Rapeseed-Cotton; R-R: Rapeseed-Rice.

图 5 基于 4 000 采样点的各模型混淆矩阵热力图

Fig.5 Confusion matrix thermal diagram of different models based on 4 000 sampling points

### 3.2 作物种植模式时空分布

对比各类模型方法,选择精度最高的基于半自动提取采样点的 *gcForest* 模型对整个江汉平原作物类型及种植模式进行分类制图,同时针对不同地区各自的主要种植模式进行了局部制图,得到图 6 的分布特征。对几大主要作物的空间分布特征进行分析,结合实地调查和研究,江汉平原作物种植模式及其影响因素主要包括以下几个方面:

1) 水稻是江汉平原最主要的农作物,尤其是单季稻,广泛分布在地块连续、土壤肥沃、靠近水系的地区。近年来,虾稻养殖技术的广泛推广和高经济效益的驱动,大量双季作物,尤其是双季水稻面积急剧下降,变为虾稻养殖模式;另外,随着城镇化加剧,农村青壮年劳动力转为非农就业类型比例加大,农业种植劳动力日趋老龄化,导致不少农户因为劳力不足,选择种植一季水稻,因此,单季稻的种植面积占主体并扩大。例如单季稻。

2) 小麦主要分布在平原的西部和北部地区,种植区域相对集中,在天门市西部、潜江市东北部、枝江东部和荆州市西北地区。平原北部的麦区主要是小麦-棉花为主的轮作模式,西部沿江地带则是小麦-棉花和小麦-水稻更加广泛。随着国家机耕政策的推广,江汉平原麦区的范围正在逐年扩大。

3) 油菜主要分布在江汉平原的中部, 包括仙桃西南部、江陵地区、潜江与监利交接地也有少量分布。多数为油菜-棉花、油菜-水稻轮作的模式, 以靠近长江周边区域种植更为集中, 这是因为油菜与棉花各个生长阶段对水的需求都很大, 更适宜种植在土壤湿润的地区。

## 4 讨论

由表1可知, 无论基于400或4 000个采样点, gcForest模型的总体精度均高于DNN; 各种种植模式的错分概率更小(图5), 针对中国复杂农田景观下多种种植模式的识别问题有更好的适用性。gcForest模型主要由级联森林和多粒度扫描两部分组成, 在模型中, 多粒度扫描提供了多尺度窗口提取特征, 级联森林使用多个随机森林集成学习, 输入的特征向量会输出成三维的类向量概率分布, 该结构使得模型所需要的参数较少, 更容易训练, 同时在小数据集上也有良好的性能<sup>[22]</sup>。相较于DNN模型复杂的内部结构和最优参数调整过程, gcForest以较少的数据和调参时间获得更好的分类效果。

大区域尺度下的农业精准监测一直都受到样本缺乏, 实地采集耗时耗力等问题的制约<sup>[25]</sup>。因此本研究探究了半自动提取大量样本点的途径。以实地采样点的模型分类结果为基础, 以ArcGIS平台的自动生成点为方法, 按照不同类别分别提取大量样本点。但由于初分类结果会有错分的可能, 生成的大量样本点仍要经过人为的筛选。在本研究中, 将生成的大量样本时序曲线与各种种植模式标准NDVI曲线对比, 人为筛选精准样本, 尽管仍无法实现全自动快速得到样本数据, 但该方法在一定程度上缓解了大尺度区域下农业样本信息收集困难的限制。由试验结果可以看到, 该方法有效提高了两种模型的分类精度, 说明该方法具有可行性。

中国南方农业区气候类型与农业景观类型复杂, 多云多雨的气候会造成光学影像在某时期缺失, 无法构建涵盖全部物候特征的时序数据集, 识别更多复杂种植模式<sup>[26]</sup>。针对上述问题, 一些学者已经提出将不受云影响的Sentinel-1号雷达数据与Sentinel-2号光学影像相融合, 并已开展了相关作物提取研究<sup>[27-30]</sup>。在未来, 高精度光学与雷达数据的融合应用于大面积作物及种植模式监测, 必然是时间序列精准农业制图的重要研究方向。

## 5 结论

本文以江汉平原为研究区, 基于GEE平台, 利用Sentinel-NDVI时序数据构建物候曲线并选择特征时期对研究区不同种植模式进行提取, 探究半自动提取大量样本点用于扩充样本集的途径, 对比不同样本点数量下gcForest和DNN两类深度模型分类效果与精度。得出结论如下:

1) 江汉平原种植模式主要以油菜-水稻、小麦-水稻、油菜-棉花、小麦-棉花4种轮作; 单季稻、莲藕2种单季种植结构为主, 其中水稻是种植面积最大的区域;

2) 从总体上看, 无论是基于400个实地采样点或4 000个半自动大量采样点, gcForest模型分类精度均

高于DNN, 基于半自动采样点的gcForest模型的总体精度和Kappa系数最高, 总体精度达到87.25%, 表明gcForest模型针对大尺度的复杂种植模式提取问题更加适用;

3) 水稻尤其是单季稻, 广泛分布在平原各地; 小麦-棉花和主要分布在平原北部, 西部沿江地带则是小麦-棉花和小麦-水稻更加广泛; 油菜-棉花、油菜-水稻以靠近长江周边区域种植更为集中。

本研究证实, 基于物候的开源Sentinel-2号时序数据在复杂农业景观区域的种植模式制图有较大潜力, 生成的高精度空间数据集填补了中国南方种植模式的空白, 可为中国南方地区的耕地轮作、农业可持续发展等相关问题的政策制定提供科学依据。

## 参考文献

- [1] Alvarado R, Toledo E. Environmental degradation and economic growth: Evidence for a developing country[J]. *Environment, Development and Sustainability*, 2017, 19(4): 1205-1218.
- [2] Li A, Wu J, Zhang X, et al. China's new rural "Separating Three Property Rights" land reform results in grassland degradation: Evidence from inner mongolia[J]. *Land Use Policy*, 2018, 71: 170-182.
- [3] Giltrap D L, Li C, Saggat S. Dndc: A process-based model of greenhouse gas fluxes from agricultural soils[J]. *Agriculture, Ecosystems & Environment*, 2010, 136(3/4): 292-300.
- [4] Lenz-Wiedemann V I S, Klar C W, Schneider K. Development and test of a crop growth model for application within a global change decision support system[J]. *Ecological Modelling*, 2010, 221(2): 314-329.
- [5] Lobell D B, Thau D, Seifert C, et al. A scalable satellite-based crop yield mapper[J]. *Remote Sensing of Environment*, 2015, 164: 324-333.
- [6] Waldhoff G, Lussem U, Bareth G. Multi-data approach for remote sensing-based regional crop rotation mapping: A case study for the rur catchment, germany[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 61: 55-69.
- [7] Boryan C G, Yang Z, Willis P. US geospatial crop frequency data layers[C]. *Proceedings of the 2014 the Third International Conference on Agro-Geoinformatics*. Beijing, China: IEEE, 2014.
- [8] Leteinturier B, Herman J L, De Longueville F, et al. Adaptation of a crop sequence indicator based on a land parcel management system[J]. *Agriculture, Ecosystems & Environment*, 2006, 112(4): 324-334.
- [9] Chen Y, Lu D, Moran E, et al. Mapping croplands, cropping patterns, and crop types using modis time-series data[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 69: 133-147.
- [10] 黄青, 唐华俊, 周清波, 等. 东北地区主要作物种植结构遥感提取及长势监测[J]. *农业工程学报*, 2010, 26(9):

- 218-223.
- Huang Qing, Tang Huajun, Zhou Qingbo, et al. Remote sensing extraction and growth monitoring of planting structure of main crops in Northeast China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(9): 218-223. (in Chinese with English abstract)
- [11] 潘力, 夏浩铭, 王瑞萌, 等. 基于 Google Earth Engine 的淮河流域越冬作物种植面积制图[J]. 农业工程学报, 2021, 37(18): 211-218.
- Pan Li, Xia Haoming, Wang Ruimeng, et al. Mapping of overwintering crop planting area in Huaihe River Basin based on Google Earth Engine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 211-218. (in Chinese with English abstract)
- [12] Liu L, Xiao X, Qin Y, et al. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and google earth engine[J]. Remote Sensing of Environment, 2020, 239: 111624.
- [13] Belgiu M, Csillik O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis[J]. Remote Sensing of Environment, 2018, 204: 509-523.
- [14] Yan Y, Ryu Y. Exploring google street view with deep learning for crop type mapping[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 278-296.
- [15] Paulo F, Zhao Z, Jithin M, et al. Distinguishing volunteer corn from soybean at seedling stage using images and machine learning[J]. Smart Agriculture, 2020, 2(3): 61-74.
- [16] Sun Z, Di L, Fang H, et al. Deep learning classification for crop types in North Dakota[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2200-2213.
- [17] Elbeltagi A, Aslam M R, Malik A, et al. The impact of climate changes on the water footprint of wheat and maize production in the Nile Delta, Egypt[J]. Science of the Total Environment, 2020, 743: 140770.
- [18] Bermúdez J D, Achancaray P, Sanches I D, et al. Evaluation of recurrent neural networks for crop recognition from multitemporal remote sensing images[C]//Proceedings of the Anais do XXVII Congresso Brasileiro de Cartografia. Brasile: Conselho Brasileiro de Oftalmologia, 2017.
- [19] Gong L, Yu M, Jiang S, et al. Deep learning based prediction on greenhouse crop yield combined Tcn and Rnn[J]. Sensors, 2021, 21(13): 4537.
- [20] Hixson P K. Language stimulation group: Habilitation model[J]. Annals of Otolaryngology, Rhinology & Laryngology, 1980, 89(Supp.5): 175-178.
- [21] 樊东东, 李强子, 王红岩, 等. 通过训练样本采样处理改善小宗作物遥感识别精度[J]. 遥感学报, 2019, 23(4): 730-742.
- Fan Dongdong, Li Qiangzi, Wang Hongyan, et al. Improving the accuracy of remote sensing identification of small crops by sampling training samples[J]. Journal of Remote Sensing, 2019, 23(4): 730-742. (in Chinese with English abstract)
- [22] Zhou Z H, Feng J. Deep forest: Towards an alternative to deep neural networks[EB/OL]. 28 Feb. 2017 [25 Jan. 2022]. <https://arxiv.org/abs/1702.08835v1>.
- [23] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [24] Maimaitijiang M, Sagan V, Sidike P, et al. Soybean yield prediction from uav using multimodal data fusion and deep learning[J]. Remote Sensing of Environment, 2020, 237: 111599.
- [25] Tsiligiridis T, Ainali K. Remote sensing big agridata for food availability[C]//Proceedings of the 2018 International Conference on Image and Video Processing, and Artificial Intelligence International Society for Optics and Photonics. Shanghai, China: International Society for Optics and Photonics, 2018.
- [26] 赵红伟, 陈仲新, 姜浩, 等. 基于 Sentinel-1A 影像和一维 CNN 的中国南方生长季早期作物种类识别[J]. 农业工程学报, 2020, 36(3): 169-177.
- Zhao Hongwei, Chen Zhongxin, Jiang Hao, et al. Early growing stage crop species identification in southern China based on Sentinel-1A time series imagery and one-dimensional CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(3): 169-177. (in Chinese with English abstract)
- [27] Chen Y, Hou J, Huang C, et al. Mapping maize area in heterogeneous agricultural landscape with multi-temporal Sentinel-1 and Sentinel-2 images based on random forest[J]. Remote Sensing, 2021, 13(15): 2988.
- [28] D'andrimont R, Taymans M, Lemoine G, et al. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and Sentinel-2 time series[J]. Remote Sensing of Environment, 2020, 239: 111660.
- [29] 郭交, 朱琳, 靳标. 基于 Sentinel-1 和 Sentinel-2 数据融合的农作物分类[J]. 农业机械学报, 2018, 49(4): 192-198.
- Guo Jiao, Zhu Lin, Jin Biao. Crop classification based on Sentinel-1 and Sentinel-2 data fusion[J]. Journal of Agricultural Machinery, 2018, 49(4): 192-198. (in Chinese with English abstract)
- [30] 李恒凯, 王利娟, 肖松松. 基于多源数据的南方丘陵山地土地利用随机森林分类[J]. 农业工程学报, 2021, 37(7): 244-251.
- Li Hengkai, Wang Lijuan, Xiao Songsong. Random forest classification of land use in hilly and mountainous areas of southern China using multi-source remote sensing data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(7): 244-251. (in Chinese with English abstract)

## Extraction of cropping patterns in Jiangnan Plain based on GEE and Sentinel-NDVI time series data

Zhang Zijing<sup>1</sup>, Hua Li<sup>1\*</sup>, Zheng Xuan<sup>2</sup>, Li Jialin<sup>1</sup>

(1.School of Resource Environment, Huazhong Agricultural University, Wuhan 430000, China;

2.School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518000, China)

**Abstract:** Rapid and accurate extraction of cropping patterns is of great significance for regional resource capacity evaluation, green and sustainable agricultural development and national food security. However, there are few spatial datasets with the precision, breadth of coverage, or sufficient information as required in the mapping of complex crop types and rotation patterns. Furthermore, the majority of existing research on crop extraction focuses on extracting crops either from high-resolution images in small areas or from low-to-medium resolution images in large areas, thus missing complex and dynamic cropping patterns. To make up this shortage, a phenology-based crop type and the mapping technique of cropping patterns were proposed based on the GEE platform and Sentinel-2 time series imagery for high-accuracy crop mapping of large areas. The Jiangnan Plain, an important grain-producing region in south China, was studied. The method for semi-automatic extraction of a large number of samples was explored, and the classification accuracy of multi-grained cascade Forest (gcForest) and that of the Deep Neural Network (DNN) were compared to identify and map the types and cropping patterns of rice, wheat, rapeseed, cotton and lotus root in Jiangnan Plain. The results showed that: 1) the main crops in the Jiangnan Plain include rice growing in a single season, lotus root, wheat, rapeseed, and cotton. The findings also highlight the major single-season cropping structures, including single rice and lotus root, and the major rotation patterns, consisting of wheat-rice, wheat-cotton, rapeseed-rice, and rapeseed-cotton; 2) the formation and distribution of different crop types and crop rotation patterns are driven by multiple factors, such as climate, topography, socioeconomics, and farmers' subjective wishes: rice is the most widely planted crop in the Jiangnan Plain, especially single rice. In recent years, as the urbanization is intensified and the labor force in agricultural planting is aging, many farmers choose to plant single-season rice due to the lack of labor. Therefore, the planting area of single-season rice dominates and expands; wheat is mainly distributed in the western and northern regions of the plain, and its planting areas are relatively concentrated. Areas sown to wheat in the northern part of the plain mainly carry out wheat-soybean and wheat-rice-based crop rotation patterns, while wheat-cotton and wheat-rice patterns are more frequently seen in the western areas along the river. With the promotion of the national policy on farming with machine, areas sown to wheat in Jiangnan Plain are expanding year by year; rapeseed is mainly distributed in the middle of the Plain, mostly in the modes of rapeseed-cotton and rapeseed-rice rotation, for rapeseed and cotton have a large demand for water at each growth stage, and are more suitable for planting in areas with moist soil; 3) the gcForest model based on 4 000 semi-automatic sampling points has the highest overall accuracy, which can reach 87.25%. The classification accuracy of the two models based on 4 000 sampling points is 8.08 and 5.5 percentage points higher than that based on 400 field sampling points respectively. This semi-automatic extraction approach of sampling points can effectively improve the classification accuracy. Besides, whether based on 400 or 4 000 sample points, the classification accuracy of the gcForest model is higher than that of the DNN, indicating that gcForest has an edge when extracting planting patterns in complex agricultural landscapes in south China. In conclusion, the results demonstrate that the phenology-based open-source Sentinel-2 sequential data can effectively support the mapping of planting patterns in large and complex agricultural areas. Thus, the mapping of crop planting patterns in south China presented in this study provides a scientific basis for formulating policies related to crop rotation and sustainable agricultural development.

**Keywords:** models; time series data; depth model; semi-automatic extraction; planting pattern