

基于NDVI时间序列影像的农业种植结构提取

——以河北省邯郸市广平县为例

刘芮廷¹, 孙建伟², 吕正虎³, 任 帅⁴, 刘光辉¹, 赵 娴¹

¹河北工程大学水利水电学院, 河北 邯郸

²邯郸市水利局节约用水促进中心, 河北 邯郸

³河北省邢台水文勘测研究中心, 河北 邢台

⁴河北工程大学地球科学与工程学院, 河北 邯郸

收稿日期: 2023年10月1日; 录用日期: 2023年11月10日; 发布日期: 2023年11月20日

摘 要

农业种植结构提取是基础性环节, 针对原来以人工传统调查统计的农作物信息获取方法主观性较强且误差巨大难以满足信息时代背景下对农业的管理。近年来, 水资源匮乏不断制约着灌溉农业发展, 不同作物在全生育期所需的灌溉量存在巨大差异, 因此快速准确的了解种植结构对于节水农业有巨大效益。本文以广平县为研究区, 以哨兵2号遥感影像作为基础数据, 以人工采集和目视解译的样本点为标准, 采用随机森林算法对广平县的种植结构进行分类。结果表明: 分类结果的总体精度为89.95%, Kappa系数为0.88, 总体精度和Kappa系数均高于0.85, 说明各个农作物分类结果较为精准。经统计, 广平县内冬小麦-夏玉米、棉花、玉米、红薯、谷子和花生的种植面积分别为148.23 km²、9.34 km²、18.25 km²、2.31 km²、0.77 km²和3.98 km²。

关键词

种植结构提取, 哨兵2号, 遥感, 随机森林, 广平县

Extraction of Agricultural Planting Structure Based on NDVI Time Series Image

—A Case Study of Guangping County, Handan City, Hebei Province

Ruiting Liu¹, Jianwei Sun², Zhenghu Lv³, Shuai Ren⁴, Guanghui Liu¹, Xian Zhao¹

¹School of Water Conservancy and Hydroelectric Power, Hebei University of Engineering, Handan Hebei

²Handan Water Conservation Promotion Center, Handan Water Conservancy Bureau, Handan Hebei

³Hydrological Survey Research Center, Xingtai Hebei

⁴School of Earth Science and Engineering, Hebei University of Engineering, Handan Hebei

Received: Oct. 1st, 2023; accepted: Nov. 10th, 2023; published: Nov. 20th, 2023

文章引用: 刘芮廷, 孙建伟, 吕正虎, 任帅, 刘光辉, 赵娴. 基于 NDVI 时间序列影像的农业种植结构提取[J]. 地球科学前沿, 2023, 13(11): 1268-1275. DOI: 10.12677/ag.2023.1311120

Abstract

Extraction of agricultural planting structure is a fundamental process. The traditional manual survey and statistical methods for obtaining crop information were subjective and prone to significant errors, which could not meet the management requirements of agriculture in the context of the information era. In recent years, the scarcity of water resources has constantly hindered the development of irrigation agriculture. Different crops have significant differences in the amount of irrigation required throughout their growth period. Therefore, a rapid and accurate understanding of the planting structure is of great benefit to water-saving agriculture. In this study, Guangping County was selected as the research area, and Sentinel-2 satellite remote sensing images were used as the basic data. With the standard of manually collected and visually interpreted sample points, the random forest algorithm was employed to classify the planting structure in Guangping County. The results showed that the overall accuracy of the classification results was 89.95%, with a Kappa coefficient of 0.88. Both the overall accuracy and Kappa coefficient were higher than 0.85, indicating that the classification results of various crops were accurate. According to statistics, the planting areas of winter wheat-summer maize, cotton, maize, sweet potato, millet and peanut in Guangping County were 148.23 km², 9.34 km², 18.25 km², 2.31 km², 0.77 km² and 3.98 km² respectively.

Keywords

Crop Structure Extraction, Sentinel-2, Remote Sensing, Random Forest, Guangping County

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

农业种植结构是计算农业产值、需水量、粮食产量的基础数据,为了能够保障农业的稳定发展,国家必须制订适合的农业政策和经济发展规划,这就需要准确及时的了解农业的种植面积[1]。由于过去科学技术发展的限制,农作物种植信息是通过人力统计的方式实现,这种方式需要层层上报,往往耗时耗力,而且在准确性时效性方面存在很大缺陷[2]。随着遥感技术的兴起,为农作物种植面积及其空间分布提取提供了新思路, Jakubauskas [3]于 2002 年利用光谱信息的变化区分玉米大豆等作物,但分类结果欠佳。

之后,遥感解译工作逐步趋向成熟,同分辨率条件下,由单时相提取逐步变为多时相提取,分类精度有所提高。Wardlow B D [4]于 2008 年利用 NDVI 和 EVI 两种时间序列识别大面积农田,并比较其差异得到最佳区分结果图; 闫慧敏[5]等于 2005 年利用多时相遥感影像,对中国农田种植信息进行分类; Julien [6]等于 2011 年采用最大似然法提取西班牙部分区域种植结构的变化,分类效果较为理想。同年,国内的赵丽花[7]、潘耀忠[8]利用 EVI 时间序列进行分类识别,得到较好结果。张健康[9]于 2012 年运用多时相的遥感影像数据对黑龙江地区的棉花、玉米、小麦和蔬菜进行农业种植结构的提取,结果显示其总体分类精度达到了 91.3%,分类精度较高,能够较好的反映作物的分布情况,为该地区主要的种植结构调整提供科学依据。刘焕军[10]于 2015 年建立棉花的光谱信息变化时间序列图,实现对其产量的预测。

随着以美国国家航天局的 Landsat 系列等卫星发射和完善,其后随着中高分辨率卫星不断发射,尤其是以高分 6 号,哨兵 2 号等卫星为代表的开源卫星数据,使得分类精度再次突破。熊元康[11]等于 2018

年对天山北坡经济带的棉花、玉米、小麦利用随机森林分类器进行分类, 总体精度为 92.19%, Kappa 系数为 0.88; 刘嘉文[12]于 2020 年以县域为尺度使用哨兵 2 号影像数据对桦南县主要农作物玉米、大豆水稻进行空间种植结构提取研究, 总体分类精度 95.22%, Kappa 系数为 0.89; 2022 年, 田鑫[13]以甘肃张掖灌区为研究区, 将基于时序 NDVI 特征和光谱特征提取的玉米面积相比, 发现基于时序 NDVI 特征的提取的结果优于基于光谱特征提取的结果; 2023 年, 牛昱杰[14]采用哨兵 2 号遥感影像, 并基于多时间序列的归一化植被指数构建模型, 获取了 2018~2021 年甘肃省甘谷县冬小麦的分布情况, 提取的总体精度均在 90% 以上, 为当地冬小麦遥感种植监测提供借鉴和参考。

河北省处于华北平原西北部是我国重要的农业基地, 但人均水资源量仅为全国平均水平的 11%, 其中农业用水占比超过 70%, 且地下水占比的份额较大, 是典型的资源型缺水省份[15]。从上个世纪 70 年代到现在, 年均地下水超采量约为 59.7 亿 m^3 , 尤其是中东部的平原区已经成为超采最严重地区, 近三十年超采量达 1500 亿 m^3 , 随着地下水位的持续下降, 目前河北省地下水超采问题十分严重, 爆发了一系列连带问题, 上世纪六十年代仅邯郸市就有 13 条地裂发生, 对当地的居民生活产生极大危害, 并制约着社会经济的可持续发展[16]。邯郸市广平县作为河北省水资源超采区的农业县, 摸清该县的种植结构无疑为水资源管理、粮食宏观调控和促进农业可持续发展提供有效途径。

2. 数据及方法

2.1. 研究区概况

本文选择广平县为研究区, 该地区地处河北省南部, 邯郸市东部, 介于东经 $114^{\circ}84'$ 至 $115^{\circ}19'$, 北纬 $36^{\circ}37'$ 至 $36^{\circ}62'$ 之间, 全县境域长 31.4 km, 宽 27.0 km, 总面积达 320 km^2 。县政府驻地广平镇, 距省会石家庄 198 km, 距邯郸市中心城区 45 km。全县地势起伏不大, 自西南向东北逐渐走低, 平均坡度 1/3500, 海拔高度在 40~53.9 m 之间, 最大高差 13.9 m。广平地理位置如图 1 所示。

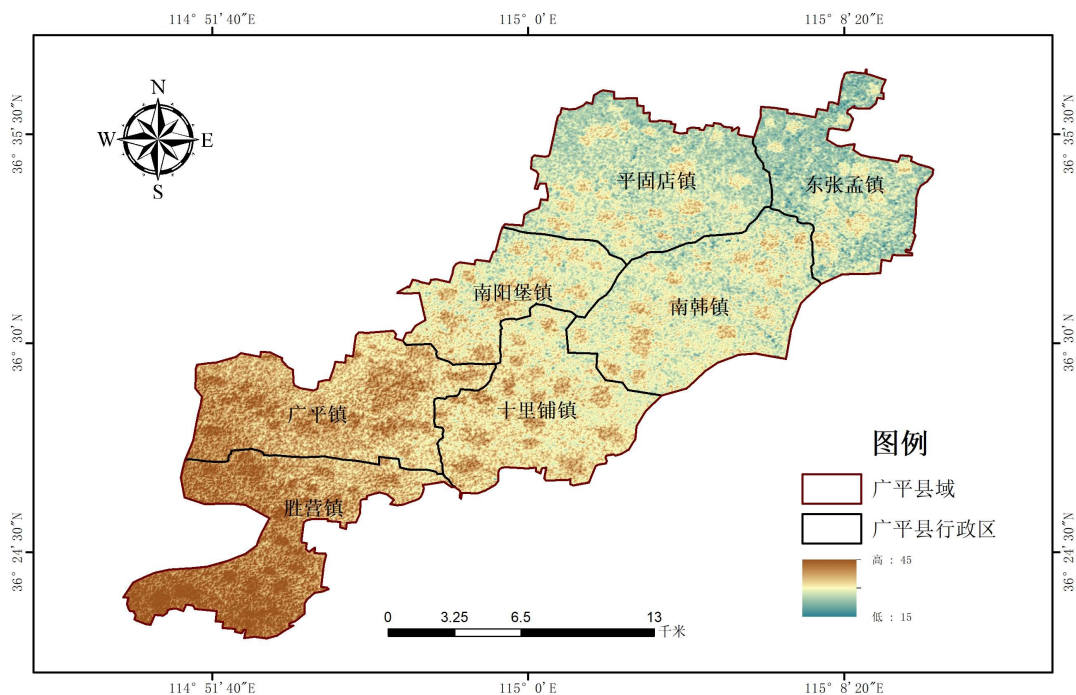


Figure 1. Geographical location map of Guangping County
图 1. 广平县地理位置图

2.2. 数据获取

本文选取哨兵 2 号(Sentinel-2)卫星作为原始数据, 该卫星源于欧洲委员会和欧洲航天局共同倡议的哥白尼计划, 由 Sentinel-2A 和 2B 两颗相同的卫星组成, 其光学任务主要是对于陆地和沿海地区进行各种自然资源灾害的监测, 涵盖可见光、近红外线与短波红外线的 13 个波段, 空间分辨率为 10~60 m, 时间分辨率为 5 天。本文采用空间分辨率为 10 m 的 B2、B3、B4、B8 波段进行后续分析, 共选取从 2019 年 9 月至 2020 年 10 月的 27 景 L1C 级无云影像数据。

2.3. 方法

2.3.1. 典型作物选取

根据当地调查的物候情况, 以及种植面积的大小, 选取冬小麦、夏玉米、棉花、花生、谷子、红薯共 6 种农作物作为本次种植结构提取的主要作物, 其中由于冬小麦和夏玉米轮作情况较为普遍, 故把本次遥感提取的农作物分为冬小麦-夏玉米、棉花、玉米、红薯、谷子、花生。

2.3.2. 图像预处理与 NDVI 计算

对 27 景原始遥感影像进行影像预处理操作。由于哨兵 2 号 L1C 级数据已正射校正和几何精校正, 所以影像预处理主要包括辐射定标、大气校正、影像裁剪三个部分。

辐射定标: 将遥感影像传感器纪录的无量纲 DN 值转换成具有实际物理意义的表现辐射亮度的过程, 其意义在于尽可能地除去由传感器自身造成的影响, 减少遥感影像与地面物体辐射亮度之间的差异, 使信息更加准确, 为农作物种植结构提取、面积估算提供精准的数据支撑。计算公式如下[17]:

$$Le = Gain \times DN + offset \quad (1)$$

式中: Le 为影像的辐射亮度值, $Gain$ 表示传感器定标系数的增益值, DN 为影像的数字量化值, $offset$ 为定标系数的偏移值。

大气校正: 太阳辐射与物体表面相互作用时, 需要透过大气层, 而大气中的气溶胶、二氧化碳、臭氧等会吸收部分反射光, 同时伴随着少量大气散射现象, 使得成像过程中农作物反射光谱信息出现了偏差, 存在一定程度上的辐射变形, 消除以上干扰的过程通常称为大气校正。

图像裁剪: 图像裁剪是按照某一规定的边界进行裁剪, 目的是去除研究区范围以外的多余部分, 以免操作过程中带来的巨大工作量。

对预处理之后的 27 景影像使用归一化植被指数 NDVI 作为分类特征指数[18]。它由 Rouse 等提出, 在许多研究中得到应用[19][20]。该指数通过建立两波段间的数学运算关系, 表达了红光波段与近红外波段即植物吸收强烈的波段与植被反射强烈的波段之间的关系, 其计算公式如下[2]:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (2)$$

式中: ρ_{NIR} 代表近红外波段的反射率; ρ_{Red} 代表红波段的反射率。NDVI 能够很好地反映植被生长状况, 取值大小与农作物生长茂盛程度相关, 一般作物取值在 0~1 之间, 约接近 1 作物生长越茂盛。

2.3.3. 遥感影像分类算法

本研究主要采用监督分类方法中的随机森林算法对种植结构进行识别。基本思路是在收集到的原始样本中, 随机抽取一定量样本组成决策树, 这些决策树具有样本和特征空间, 抽取之后把样本放回原始样本中, 使得每棵决策树内部都可能存在相同的样本和特征, 实现每棵决策树都有对应的随机且独立数据集, 以此增加算法的随机性, 提高模型的泛化能力, 之后将所有的决策树组合, 对未知个体进行识别,

最终对所有决策树分类的结果进行对比, 选取个体的最优分类作为最终的输出结果。

一般抽取每个样本集的原则是不超过总样本数的三分之二, 相对于每次抽中的样本而言, 约有三分之一的样本未被抽中, 利用这一部分未被抽中的样本估计训练的内部误差[2]:

$$Q_i = \frac{M_i}{S} < \frac{2}{3} \quad i=1,2,3 \quad (3)$$

式中: $i=1,2,3 \dots n$ 为抽取样本的次数, Q_i 表示抽中样本占总样本的百分比, M_i 表示每次抽取的样本数, S 为原始样本数。然后根据样本建立各自对应的分类决策树, 并将所有的决策树合并为一个森林, 基于每棵决策树对未知个体所属类别进行判断, 统计每一类作物在所有树的识别结果中所出现的频率, 出现最多的作物类别就认定为该个体最终的分类型结果[2]:

$$p(x) = \sum_{i=1}^K P_i(x) \quad (4)$$

式中: K 为随机森林中树的个数, x 为类别, P_i 表示在决策树中该农作物类型在分类过程中所出现的比重。

2.3.4. 样本采集

本文采用实地勘测的方式采集样本点并使用目视解译法样本点进行扩充, 对典型作物进行提取以及精度验证。采用亚米级手持 GPS (ICE980RTK) 进行研究区种植类型野外样本调查, 仪器坐标定义为 WGS-84 坐标系, 测量值采用地理坐标, 测量时间为 2020 年 7 月 3 日~2020 年 7 月 17 日, 询问当地农民 2020 年上半年种植冬小麦的情况之后再采集样本点, 每种类型各采样 20 个, 共采集 120 个样本。根据收集到的 120 个野外样本的特征信息采用目视解译法在地图上进行扩充, 共得到 22,570 个样本点, 其中 70% 作为训练样本, 30% 作为验证样本。

2.3.5. 精度评价

使用验证样本对分类结果进行评价, 通过混淆矩阵进行计算分析, 获取提取结果的精度。混淆矩阵基于像元尺度进行计算, 通过对比真实像元与影像对应像元位置的分类结果, 利用相应指标将结果表示出来, 主要的精度评价指标分为总体分类精度 OA、Kappa 系数、用户精度和制图精度。总体分类精度是指每一类验证样本所对应的分类结果的正确个数与所有实验验证样本个数之间的比值; Kappa 系数是对一致性程度的评价判断指标, 大于 0.8 表示分类结果较好。具体计算公式如下[21]:

$$P_{OA} = \frac{\sum_{i=0}^n x_i}{N} \quad (5)$$

$$Kappa = \frac{P_{\text{总体}} - \frac{\sum_{i=0}^n a_i \times b_i}{N^2}}{1 - \frac{\sum_{i=0}^n a_i \times b_i}{N^2}} \quad (6)$$

式中: P_{OA} 为总体分类精度; x_i 为地物中被正确分类的个数; a_i 为 n 类地物中的真实样本像元数; b_i 为样本中被归为 n 类地物的像元数; n 为分类结果总类别数; N 为样本像元数。总体分类精度表示样本于分类结果的相似程度, Kappa 系数的计算结果通常处于 0~1 之间。小于 0.6 则表示分类效果较差; 位于 0.6~0.8 之间, 分类精度较高; 大于 0.8 表示分类结果与真实地物具有高度一致性。

3. 结果与讨论

本文通过随机森林模型分类法得到的 6 种典型作物种植结构分布如图 2 所示。

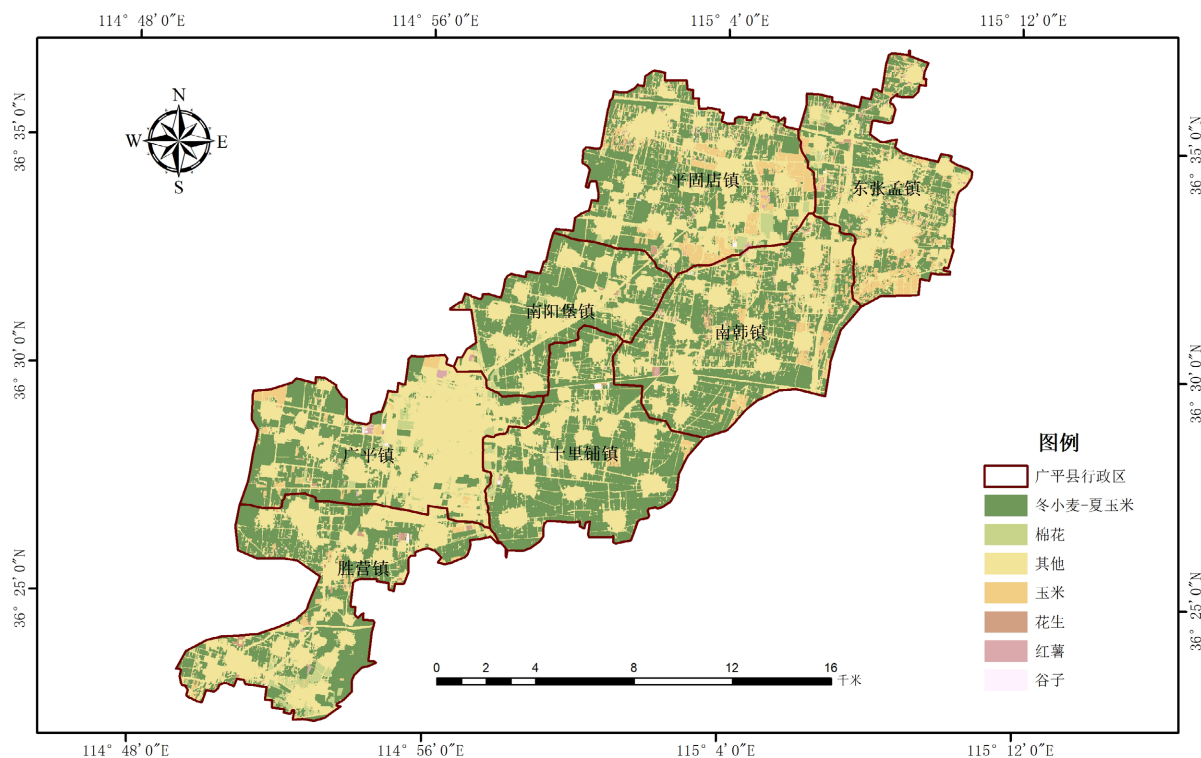


Figure 2. Spatial distribution map of six typical crops in Guangping County

图 2. 广平县六种典型作物的空间分布图

从图中可以看出, 该县冬小麦-夏玉米和夏玉米种植面积占比超过 70%, 其余农作物零星分布。冬小麦-夏玉米的种植分布范围最广, 其中南阳堡镇、十里铺镇、南韩镇的种植比例较高。玉米和棉花在县域北部及南部分布较多。玉米绝大部分分布在平固店镇、东张孟镇; 棉花主要分布在平固店镇南部、十里铺镇西部和胜营镇南部。花生、红薯和谷子的面积较为分散, 在各镇均有少量种植。经统计, 6 种典型作物面积如表 1 所示。

Table 1. Area of various crops

表 1. 各类作物的面积

作物	面积/km ²	作物	面积/km ²
冬小麦-夏玉米	148.23	红薯	2.31
棉花	9.34	花生	3.98
玉米	18.25	谷子	0.77

如表所示, 广平县内冬小麦-夏玉米、棉花、玉米、红薯、谷子和花生的种植面积分别为 148.23 km²、9.34 km²、18.25 km²、2.31 km²、0.77 km² 和 3.98 km²。其中, 种植比例分别为 81.05%、5.11%、9.98%、1.26%、0.42% 和 2.18%, 这与《2021 年邯郸统计年鉴》上面的数据较为吻合, 说明提取结果较为合理。分类作物混淆矩阵如表 2 所示。

由表可知, 分类结果的总体精度为 89.95%, Kappa 系数为 0.88, 总体精度和 Kappa 系数均高于 0.85, 说明总体分类质量较精准。冬小麦-夏玉米、玉米和棉花制图精度都高于 90%, 冬小麦-夏玉米更是高

达 93.01%，说明这几种农作物的类别错分现象极低，分类结果基本正确。而制度精度稍低的红薯、谷子和花生由于种植面积较小验证样本总数少，导致的错分一个点对制图精度的干扰极大，但整体的用户精度都高于 80%，说明其他类别之间相互错分的几率极低。

Table 2. Classification of crops classification confusion matrix

表 2. 分类作物分类混淆矩阵

		用户精度/制图精度(%)					总体精度(%)/ Kappa 系数
冬小麦 - 夏玉米	棉花	玉米	红薯	谷子	花生	其他	
94.95/ 93.01	93.21/ 91.50	91.73/ 92.78	90.3/ 89.85	88.92/ 89.29	88.58/ 87.70	81.97/ 85.29	89.95/ 0.88

4. 结论

利用随机森林与哨兵 2 号遥感影像成功获取了 2020 年研究区内 6 种典型作物的空间分布图。与传统的方法相比，在提高精度的同时增加了作物的种类。冬小麦 - 夏玉米、棉花、玉米、红薯、谷子和花生提取之后的制图精度分别为 94.95%、93.21%、91.76%、90.30%、88.92%、88.58%；用户精度分别为 93.01%、91.5%、92.78%、89.85%、89.29%、87.70%，总体精度达到了 89.95%，Kappa 系数为 0.88。各精度均满足要求，说明基于随机分类算法进行种植结构分类稳定性高，值得在生产实践中进行应用。经统计，冬小麦 - 夏玉米与玉米的面积最大，约占整个研究区总面积的 70% 左右，其余农作物呈零星分布。经统计，冬小麦 - 夏玉米、玉米、红薯、谷子、棉花和花生的种植面积分别为 148.23 km²、9.34 km²、18.25 km²、2.31 km²、0.77 km²、3.98 km² 和 24.4503 km²。

参考文献

- [1] 赵彦茜. 河北平原粮食生产格局演变及水资源-产量效应研究[D]: [硕士学位论文]. 石家庄: 河北地质大学, 2018.
- [2] 王睿. 基于光谱和纹理特征综合的农作物种植结构提取方法研究[D]: [硕士学位论文]. 兰州: 兰州交通大学, 2019.
- [3] Jakubauskas, M.E., Legates, D.R. and Kastens, J.H. (2002) Crop Identification Using Harmonic Analysis of Time-Series AVHRR NDVI Data. *Computers and Electronics in Agriculture*, **37**, 127-139. [https://doi.org/10.1016/S0168-1699\(02\)00116-3](https://doi.org/10.1016/S0168-1699(02)00116-3)
- [4] Wardlaw, B.D. and Egbert, S.L. (2008) Large-Area Crop Mapping Using Time-Series MODIS 250m NDVI Data: An Assessment for the U.S. Central Great Plains. *Remote Sensing of Environment*, **112**, 1096-1116. <https://doi.org/10.1016/j.rse.2007.07.019>
- [5] 闫慧敏, 曹明奎, 刘纪远, 等. 基于多时相遥感信息的中国农业种植制度空间格局研究[J]. 农业工程学报, 2005, 21(4): 85-90.
- [6] Julien, Y., Sobrino, J.A. and Jiménez-Muñoz, J.C. (2011) Land Use Classification from Multitemporal Landsat Imagery Using the Yearly Land Cover Dynamics (YLCD) Method. *International Journal of Applied Earth Observation and Geoinformation*, **13**, 711-720. <https://doi.org/10.1016/j.jag.2011.05.008>
- [7] 赵丽花, 李卫国, 杜培军. 基于多时相 HJ 卫星的冬小麦面积提取[J]. 遥感信息, 2011(2): 41-45, 50.
- [8] 潘耀忠, 李乐, 张锦水, 等. 基于典型物候特征的 MODIS-EVI 时间序列数据农作物种植面积提取方法——小区冬小麦实验研究[J]. 遥感学报, 2011, 15(3): 578-594.
- [9] 张健康, 程彦培, 张发旺, 等. 基于多时相遥感影像的作物种植信息提取[J]. 农业工程学报, 2012, 28(2): 134-141.
- [10] 刘焕军, 孟令华, 张新乐, 等. 基于时间序列 Landsat 影像的棉花估产模型[J]. 农业工程学报, 2015, 31(17): 215-220.

-
- [11] 熊元康, 张清凌. 基于 NDVI 时间序列影像的天山北坡经济带农业种植结构提取[J]. 干旱区地理, 2019, 42(5): 1105-1114.
- [12] 刘嘉文. 基于 Sentinel-2 数据的农作物空间种植结构提取研究[J]. 测绘与空间地理信息, 2022, 45(11): 62-64.
- [13] 田鑫, 何海, 金双彦, 等. 基于遥感影像的张掖灌区作物种植结构提取研究[J]. 中国农村水利水电, 2022(8): 206-212, 217.
- [14] 牛昱杰, 杨永明. 基于哨兵2号时序影像的冬小麦空间分布研究[J]. 山东农业大学学报(自然科学版), 2023, 54(3): 352-359.
- [15] 许月卿. 京津以南河北平原地下水位下降驱动因子的定量评估[J]. 地理科学进展, 2003, 22(5): 490-498.
- [16] 张兆吉, 雒国中, 王昭, 等. 华北平原地下水资源可持续利用研究[J]. 资源科学, 2009, 31(3): 355-360.
- [17] 司凯凯. 基于卫星遥感的棉花种植区域提取方法研究[D]: [硕士学位论文]. 石河子: 石河子大学, 2023.
- [18] Sun, L., Gao, F., Xie, D., *et al.* (2021) Reconstructing Daily 30m NDVI over Complex Agricultural Landscapes Using a Crop Reference Curve Approach. *Remote Sensing of Environment*, **253**, Article ID: 112156. <https://doi.org/10.1016/j.rse.2020.112156>
- [19] 张雅芳, 郭英, 沈彦俊, 等. 华北平原种植结构变化对农业需水的影响[J]. 中国生态农业学报(中英文), 2020, 28(1): 8-16.
- [20] 白燕英, 高聚林, 张宝林. 基于 Landsat8 影像时间序列 NDVI 的作物种植结构提取[J]. 干旱区地理, 2019, 42(4): 893-901.
- [21] 李宾, 荆华, 张殷钦, 等. 多样种植区冬小麦 RS 影像精确提取方法研究[J]. 海河水利, 2023(7): 65-69, 128.