

doi: 10.6046/gtzyyg.2020.04.29

引用格式: 王德军,姜琦刚,李远华,等. 基于 Sentinel - 2A/B 时序数据与随机森林算法的农耕区土地利用分类[J]. 国土资源遥感,2020,32(4):236 - 243. (Wang D J, Jiang Q G, Li Y H, et al. Land use classification of farming areas based on time series Sentinel - 2A/B data and random forest algorithm [J]. Remote Sensing for Land and Resources, 2020, 32(4): 236 - 243.)

## 基于 Sentinel - 2A/B 时序数据与随机森林算法的农耕区土地利用分类

王德军<sup>1</sup>, 姜琦刚<sup>2</sup>, 李远华<sup>2</sup>, 关海涛<sup>1</sup>, 赵鹏飞<sup>1</sup>, 习靖<sup>2</sup>

(1. 黑龙江省第五测绘地理信息工程院, 哈尔滨 150081;

2. 吉林大学地球探测科学与技术学院, 长春 130026)

**摘要:** 农耕区土地覆被信息是土地资源管理与规划的基础,在合理开发土地资源,调整土地利用结构以及土地动态监测等方面起着重要作用。由于农耕区土地类型复杂并且具有高异质性的特点,土地覆被信息提取的精度一直面临着挑战。因此,以 Sentinel - 2A/B 多光谱遥感数据作为数据源,首先构建归一化植被指数(normalized difference vegetation index, NDVI)时序数据集和缨帽-湿度分量(tasseled cap wetness, TCW)时序数据集;其次,利用 J - M (Jeffries - Matusita)距离对地物进行可分离性分析和挑选出 NDVI 和 TCW 最佳时序数据组合;最后,结合随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、最大似然分类(maximum likelihood classification, MLC)3 种分类算法以及利用单时相遥感数据对农耕区典型地物进行分类研究。研究表明:基于时间序列数据结合随机森林分类算法取得了较高的分类精度,其总体分类精度达到 88.87%,Kappa 系数达到 0.855 7,与利用单时相影像数据分类结果的精度相比分别提高了 10.05 个百分点和 0.209 3,这充分说明利用时序数据结合 RF 分类算法在农耕地区能够有效提高典型地物的分类精度。

**关键词:** 时间序列; 随机森林; 土地利用分类; 农耕区; Sentinel - 2A/B

**中图法分类号:** TP 79 **文献标志码:** A **文章编号:** 1001 - 070X(2020)04 - 0236 - 08

### 0 引言

土地利用分类研究是土地覆被信息提取的重要基础,在土地资源管理、动态监测以及环境保护等方面起到十分重要的作用<sup>[1]</sup>。近些年,遥感技术得到了快速的发展,尤其在农耕区能够快速获取作物类型和种植面积,已成为农耕区土地变化监测和作物产量估算的主要技术手段<sup>[2]</sup>。

目前,针对农耕区土地利用分类一直是遥感领域的研究热点。在以往的研究中,由于中高分辨率数据对植被具有较高的识别精度,继而被广泛用于分类研究。例如,林楠等<sup>[3]</sup>利用单时相资源一号 02C 遥感数据,结合最小二乘支持向量机(support vector machine, SVM)分类法,快速提取了农耕区土地信息;王月如等<sup>[4]</sup>以 Landsat8 OLI 为数据源,采用一种新的水体差异增强指数,获得研究区内

的富贵竹信息。但由于时间分辨率较低和气候因素的影响,很难获取植被关键时期的影像和构建时序数据集。为了解决时间分辨率低的问题,有学者便利用高时间分辨率数据进行土地利用分类研究。例如, Wardlow 等<sup>[5]</sup>结合 MODIS 数据,创建归一化植被指数(normalized difference vegetation index, NDVI)时间序列对农业区典型地物信息进行提取; Vintrou 等<sup>[6]</sup>采用 MOD13Q1 时间序列数据对马里南部农耕区进行分类研究;朱永森等<sup>[7]</sup>以 HJ - 1A/B 为数据源,创建 NDVI 和 PCI 时序数据,提取长株潭城市群的土地利用信息。以上虽然解决了时间分辨率较低的问题,但是空间分辨率的限制使得在随季节更替地物类型变化明显的农耕地区地物分类精度较低。为了具有高时空分辨率的特点,有学者采用多源遥感数据融合的方法。例如,张猛等<sup>[8]</sup>利用 MODIS 数据和 Landsat 数据,提取了洞庭湖区水稻的种植面积;郝鹏宇等<sup>[9]</sup>同样结合 MODIS 和 Landsat 数据,进而提

收稿日期: 2020 - 02 - 07; 修订日期: 2020 - 08 - 25

基金项目: 中国地质调查局项目“辽吉黑区自然资源更新调查”(编号: 3S2170124423)和中国地质调查局资金资助项目(编号: GFZX0404130302)共同资助。

第一作者: 王德军(1995 -),男,硕士,助理工程师,研究方向为遥感制图及其技术应用。Email: 360539842@qq.com。

取作物面积。上述融合方法虽然具备了高时空分辨率的特点,提高了分类精度,但是不同数据源的归一化问题依然很难解决。

随着各国大力发展遥感事业,中高分辨率的卫星数据突破了重访周期较长的困境,这为农耕地区土地利用分类提供了更多的数据源。欧洲航天局分别在2015年6月和2016年7月发射了 Sentinel-2A/B 2 颗卫星,每颗卫星都具有13个光谱波段,其空间分辨率最高可达10 m,2颗卫星协同工作可使重访周期缩短至5 d,这便可利用同一数据源构建高时空分辨率时序数据,对于土地利用分类研究具有重要意义<sup>[10-12]</sup>。而目前如何充分利用高时空分辨率的 Sentinel-2A/B 数据实现农耕地区的土地利用分类研究还很少。鉴于此,本研究选取吉林省大安市舍力镇农耕区为研究区,充分利用 Sentinel-2A/B 具有高时空分辨率的优势,分别构建 NDVI 指数和缨帽-湿度分量(tasseled cap wetness, TCW)时序数据,并结合随机森林(random forest, RF)、SVM 和最大似然分类(maximum likelihood classification, MLC)3种分类算法识别农耕区不同地物类型,为农耕地区的土地覆被信息提取提供有效的技术方法和途径<sup>[13-15]</sup>。

## 1 研究区概况及数据源

### 1.1 研究区概况

研究区选取吉林省西北部大安市舍力镇农耕区,地处松嫩平原腹地,坐标范围 N45°32′28.17″~45°36′15.20″,E123°20′7.23″~123°27′1.47″,如图1所示。该区为中温带季风气候,四季分明,年平均气温为4.3℃,年平均降雨量为413.7 mm。研究区内主要土地利用类型包括林地、草地、旱田(玉米)、水



图1 研究区 Sentinel-2A B4(R), B3(G), B2(B) 波段合成影像

Fig. 1 Image of Sentinel-2A B4(R), B3(G), B2(B) bands in the study area

体、建设用地、盐碱地及裸地。

### 1.2 数据源及其预处理

本研究选取 Sentinel-2A/B 影像共计12景,时相分别为2017年3—12月10景和2018年1—2月2景,时间跨度满1 a。影像数据为 Level-1C 级别,质量较好,清晰无云,无条带噪声且覆盖整个研究区,均由欧洲航天局官网下载。获取到的12景影像数据如表1所示。

表1 Sentinel-2A/B 影像数据列表

Tab. 1 Sentinel-2A/B images data list

编号	卫星传感器	获取日期	数据级别
1	Sentinel-2B	2018-01-12	Level-1C
2	Sentinel-2B	2018-02-18	Level-1C
3	Sentinel-2A	2017-03-13	Level-1C
4	Sentinel-2A	2017-04-02	Level-1C
5	Sentinel-2A	2017-05-12	Level-1C
6	Sentinel-2A	2017-06-28	Level-1C
7	Sentinel-2B	2017-07-16	Level-1C
8	Sentinel-2B	2017-08-22	Level-1C
9	Sentinel-2A	2017-09-09	Level-1C
10	Sentinel-2A	2017-10-19	Level-1C
11	Sentinel-2B	2017-11-20	Level-1C
12	Sentinel-2B	2017-12-20	Level-1C

下载到的 Sentinel-2A/B 原始影像数据为 Level-1C 级别,已经过系统几何精纠正处理,并且均方根误差(root mean square error, RMSE)控制在1个像元内,所以只需要对其进行大气校正,进而得到 Level-2A 级别数据产品。根据欧洲航天局发布的独立式遥感影像处理模块 Sen2Cor (Sentinel-2 Sen2Cor processor),对遥感影像进行大气校正。最后,利用研究区矢量边界裁剪影像数据。

根据研究区内部各种地类的分布情况,均匀选取各地类样本。以基于 GF-1 影像的2016年大安市自然资源调查成果数据和 Google Earth 高分辨率遥感影像数据作为本次研究的地面验证数据。

## 2 研究方法

本研究利用多时相 Sentinel-2A/B 影像数据,首先构建 NDVI 和 TCW 时序数据集,通过分析 NDVI 和 TCW 时间序列曲线寻找不同地物之间的差异特征,再计算 J-M 距离(Jeffries-Matusita)分析典型地物类型之间的可分离性;其次,确定研究区内部用于土地利用类型分类的最佳时序数据组合,分别采用 RF, SVM 和 MLC 3种分类器对其进行分类研究;最后,对比分析3种分类结果以及利用单一时相分类结果的精度,研究过程如图2所示。

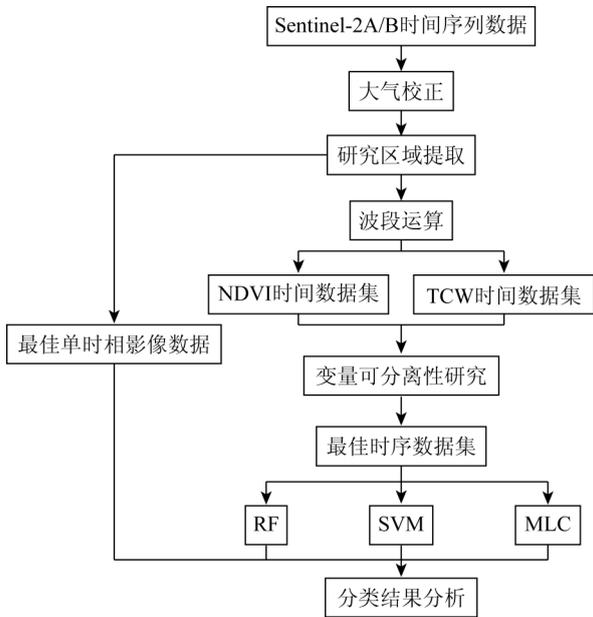


图 2 土地利用分类流程

Fig. 2 Technology flowchart for the land use classification

### 2.1 NDVI 时序数据集的构建

NDVI 是红光波段和近红外波段的数学计算,能够有效区分绿色植被信息和土壤背景信息,目前是应用最为广泛的指数之一。NDVI 数值位于  $[-1, 1]$  之间,正值代表有植被覆盖,并随着植被覆盖度增加而增加<sup>[16]</sup>。NDVI 时间序列曲线可以表达植被的物候信息。因此,本研究利用 ENVI5.3 软件对 Sentinel-2A/B 的近红外和红光波段进行波段运算,即

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}, \quad (1)$$

式中  $\rho_{nir}$  和  $\rho_{red}$  分别为近红外和红光波段的反射率值。根据式(1)得到农耕地不同地物类型的 NDVI 时序集,如图 3 所示。

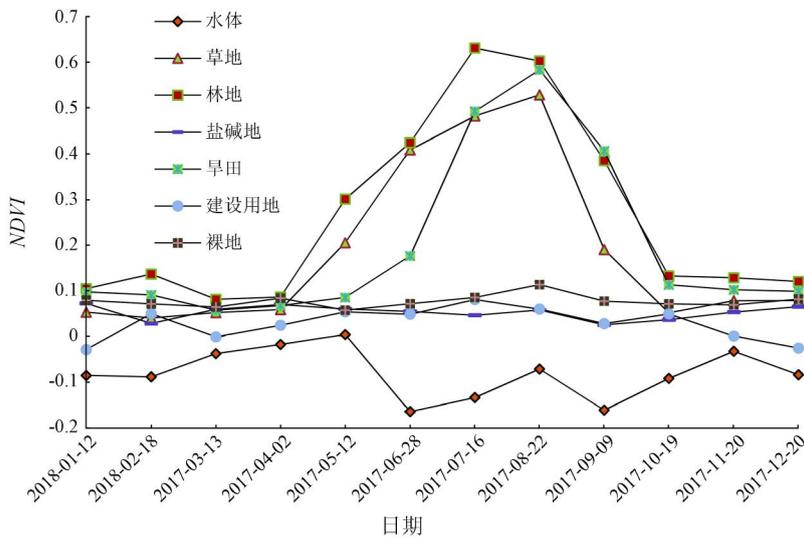


图 3 典型地物 NDVI 时间序列曲线

Fig. 3 NDVI time series curve of typical features

通过典型地物的 NDVI 时间序列曲线可以发现:林地的 NDVI 值从 5 月逐渐上升,7 月达到峰值,并且 NDVI 值高于其他地物,这主要是因为林地在 5—7 月正处于旺盛生长阶段,植被生长较快,冠幅大,易于遮盖其他植被;草地各期的 NDVI 值略低于林地,但整体趋势与林地保持一致,5 月 NDVI 值逐渐上升,并在 8 月达到峰值;根据我国东北地区主要农作物的物候特征,旱田等作物主要集中在 4—5 月开始播种,6 月出苗,此时 NDVI 值略有增加,但是整体 NDVI 值低于草地,7—9 月为快速生长期,NDVI 值持续增加,超过草地 NDVI 值,并在 8 月份达到峰值,10 月份完成收割,其 NDVI 值急剧下降;此外,水体的 NDVI 值均小于 0,而其他地物类型的 NDVI 值均大于 0,可见水体的差异性特征显

著;建设用地、裸地和盐碱地 3 类地物的 NDVI 值变化不明显,曲线形状特征相似,利用 NDVI 时间序列区分该 3 类地物存在困难。因此,根据林地、草地、旱田和水体 NDVI 值的变化情况,构建的 NDVI 时间序列数据集可以作为区分林地、草地、旱田和水体的有效分类特征,而想要区分出建设用地、裸地和盐碱地,必须对 Sentinel-2A/B 时序数据集进一步处理。

### 2.2 TCW 时序数据集的构建

建设用地、裸地和盐碱地 3 种地物类型的湿度会随着季节的变化而不同,因此可尝试构建湿度分量的时序数据集来区分此 3 类地物。对 12 景 Sentinel-2A/B 影像进行缨帽变换,变换后可将原始影像的 10 个波段,转换为亮度、绿度和湿度 3 个分量。其中亮度分量(tasseled cap brightness, TCB)代表整

体的反射率值,绿度分量 (tasseled cap greenness, TCG)代表了植被覆盖情况,湿度分量 TCW 则代表土壤湿度情况。依据传感器的差异,缨帽变换的计

$$TCW = 0.150 9\rho_2 + 0.197 3\rho_3 + 0.327 9\rho_4 + 0.340 6\rho_8 + 0.711 2\rho_{11} + 0.457 2\rho_{12} \quad (2)$$

式中  $\rho_i$  为 Sentinel-2A/B 影像中第  $i$  波段反射率

算公式也不同,本文对 Sentinel-2A/B 影像进行缨帽变换,计算公式为<sup>[17-18]</sup>:

值。图 4 为构建的 TCW 时间序列数据集曲线。

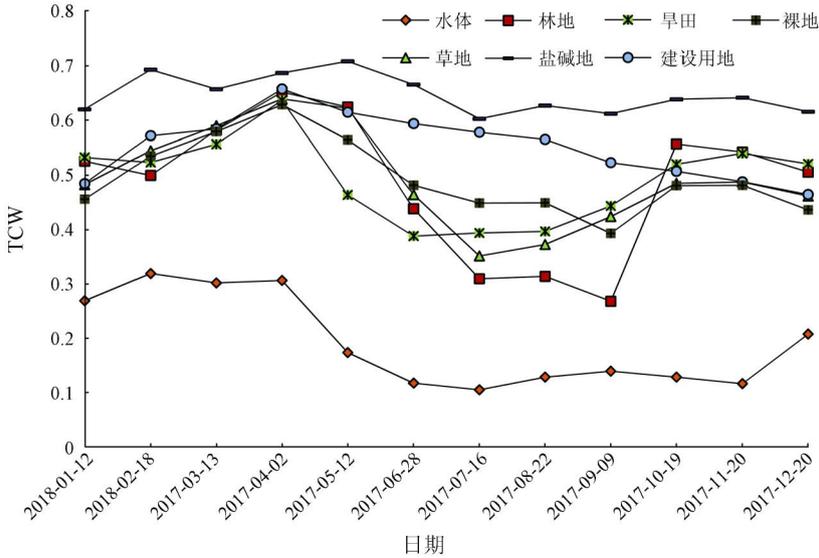


图 4 典型地物 TCW 时间序列曲线  
Fig. 4 TCW time series curve of typical features

从构建的典型地物 TCW 时间序列曲线可以发现:各类型地物的曲线形态和反射率值都有明显的差异,并且随着地物含水量的增加 TCW 值越低;盐碱地的 TCW 值高于其他地物类型,这主要是因为盐碱地土壤蓄水能力差,含水量较低;建设用地 1—4 月随着积雪融化,TCW 值逐渐升高,在 4 月达到峰值,5—10 月陆续出现降雨,TCW 值又逐渐降低,介于盐碱地和其他地类之间;裸地具有较好的储水能力,在 6—8 月期间因为降雨作用,其 TCW 值低于建设用地,但高于林地、草地和旱田,可见利用 TCW 可以有效地区分裸地和植被。因此,在 NDVI 时序数据集中较难区分的建设用地、裸地和盐碱地,在 TCW 时序数据集中却较容易区分。

### 2.3 地物可分离性分析及选择最佳时序数据组合

本文利用 J-M 距离方法,进行地物可分离性分析。J-M 距离是基于条件概率理论的光谱可分性指标,通过计算某一特征 2 类样本间的距离,衡量 2 类样本间的可分离度,因此被认为是可分离性的最佳判定指标<sup>[19]</sup>。基于某一特征 2 类样本间 J-M 距离的计算公式为:

$$J = 2(1 - e^{-B}) \quad (3)$$

$$B = \frac{1}{8} (m_1 - m_2)^2 \frac{2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2} \ln \left( \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1\sigma_2} \right) \quad (4)$$

式中:  $B$  为基于某一特征 2 种地类的距离;  $m_i$  为某

类特征的均值;  $\sigma_i$  为某类特征的标准差, ( $i = 1, 2$ )。  $J$  的取值为  $[0, 2]$ , 其值大小代表 2 种地类间的可分离程度。当  $J = 0$  时,表明 2 种地类在所选特征上无法区分,当  $J = 2$  时,表明 2 种地类在所选特征上完全分离<sup>[20]</sup>。

本文以 2016 年大安市自然资源调查成果数据和 Google Earth 高分辨率遥感影像数据作为后期验证数据。根据典型地物的生长周期规律和物候特征,首先选择 NDVI 和 TCW 特征明显的月份进行组合,再逐次增加不同月份数据,分别计算不同月份数据组合下的 J-M 距离,如表 2 所示。通过表 2 可知:典型地物之间的可分离程度随着不同时相数据组合会存在很大差异;9 种不同的组合方式均能很好地区分旱地、林地和盐碱地;当利用 3 个时相数据时,旱地与草地、盐碱地与建设用地和建设用地与裸地区分效果不佳, J-M 距离均未超过 1.9;随着不同时相数据个数的增加,可分性效果越来越好,尤其是盐碱地与建设用地的可分性明显增加;当不同时相数据个数为 7 及以上时,典型地物之间的可分性不再随着不同时相个数的增加而增大,如图 5 所示。因此,根据典型地物的 NDVI 和 TCW 时间序列曲线特征,选取 4—10 月 7 个时相的 NDVI 和 TCW 共计 14 个波段作为此次农耕区土地利用分类研究的最佳时序组合数据。

表 2 Sentinel-2A/B 不同时间序列数据组合的 6 种典型地物间的 J-M 距离

Tab. 2 J-M distance of six typical features under different time series combination of Sentinel-2A/B data

Sentinel-2A/B 数据组合方式	旱地-林地	旱地-草地	旱地-盐碱地	盐碱地-建设用地	建设用地-裸地
6 7 8	1.997 8	1.882 7	1.999 9	1.552 7	1.861 4
5 6 7 8	1.999 9	1.964 1	1.999 9	1.761 5	1.971 6
4 5 6 7 8	1.999 9	1.989 9	1.999 9	1.883 4	1.992 1
4 5 6 7 8 9	1.999 9	1.998 6	1.999 9	1.966 7	1.997 0
4 5 6 7 8 9 10	2.000 0	1.999 1	1.999 9	1.994 3	1.999 0
3 4 5 6 7 8 9 10	2.000 0	1.999 5	2.000 0	1.999 7	1.999 8
2 3 4 5 6 7 8 9 10	2.000 0	1.999 7	2.000 0	1.999 9	1.999 9
1 2 3 4 5 6 7 8 9 10 11	2.000 0	1.999 9	2.000 0	1.999 9	1.999 9
1 2 3 4 5 6 7 8 9 10 11 12	2.000 0	1.999 9	2.000 0	1.999 9	1.999 9

注：Sentinel-2A/B 数据组合包括每月的 NDVI 和 TCW, Sentinel-2A/B 数据组合方式为不同月份组合。

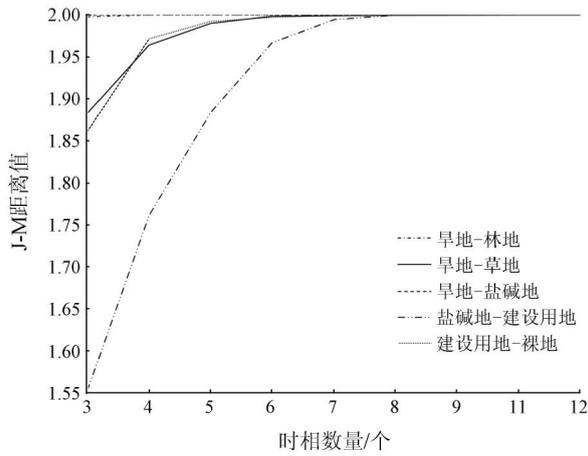


图 5 不同时相数据个数的典型地物之间 J-M 距离变化曲线

Fig. 5 J-M distance variation curve of typical features with different time phase data

### 2.4 基于 RF 的遥感分类与精度评价

在高维特征影像数据中,为了提高地物的分类精度,选择合适的分类器极为重要。2001 年 Breiman<sup>[21]</sup> 提出了 RF 算法,该算法是由多个 CART 决策树组合而成,可以有效地解决因单一决策树造成的过拟合和欠拟合问题。RF 分类器中任意 2 棵决策树都是相互独立的,当有新的测试样本输入时,每棵决策树都会对其进行分类,最后采用投票法得出分类结果<sup>[22]</sup>。RF 分类器需要调节的参数分别是

树的个数  $N$  和随机选择特征变量的个数  $m$ ,设置合适的参数,可有效提高分类精度<sup>[23-24]</sup>。通过将多组参数进行分类对比分析,确定生成决策树的个数为 50,随机选择特征变量个数为 4 时,RF 分类器的分类效果最佳。结合挑选出的 4—10 月 7 个时相的 14 个特征变量,对研究区内部林地、草地、旱田、水体、建设用地、盐碱地以及裸地进行信息提取。

为了与 RF 分类器进行对比分析,再选取 SVM 和 MLC 2 种分类器结合相同的时序数据,分别对影像进行分类。最后通过混淆矩阵获取的总体精度、Kappa 系数、生产者精度和用户精度对 3 种分类器的分类结果进行定量评价。

## 3 结果与分析

### 3.1 典型地物分类结果

基于上述 J-M 距离分析挑选出的最佳时序数据,结合 RF 分类器,得到研究区内典型地物分类结果(图 6(b))。此外,选取农作物生长茂盛的 8 月影像数据作为单一时相数据源,同样采用 RF 分类器,得到分类结果(图 6(c))。为进一步评估基于时序数据 RF 分类器的分类性能,分别与 SVM 分类结果(图 6(d))和 MLC 分类结果(图 6(e))进行对比分析。

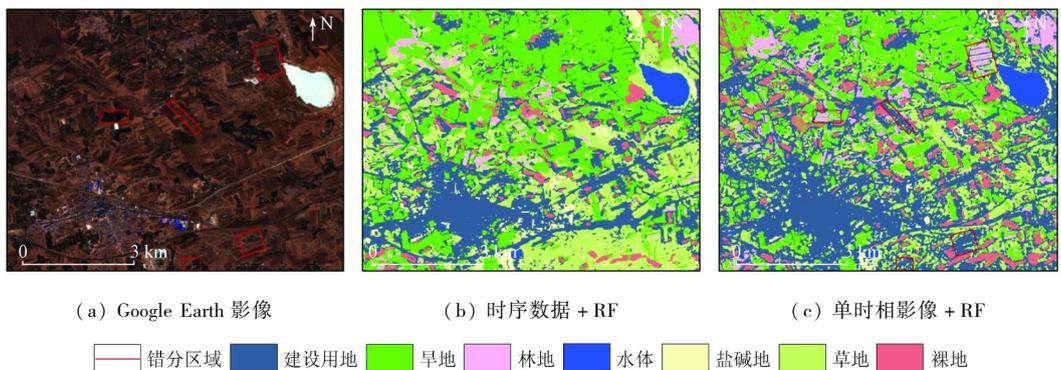


图 6-1 分类结果对比

Fig. 6-1 Comparison of classification results

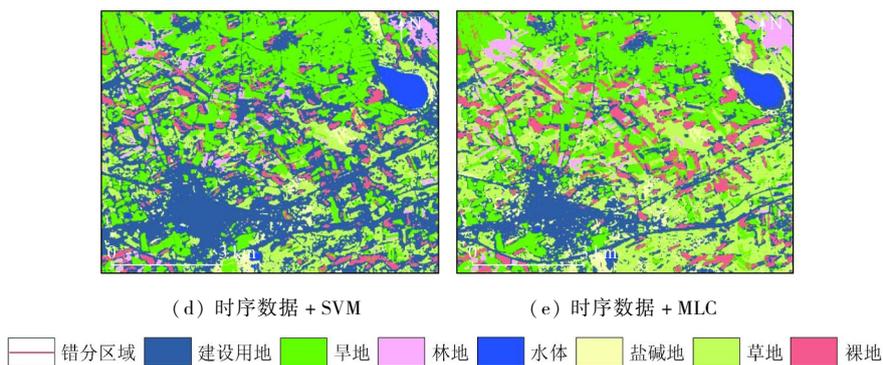


图 6-2 分类结果对比

Fig. 6-2 Comparison of classification results

从图 6(c) 观察到林地与旱地、盐碱地与建设用地以及耕地与建设用地之间存在严重的错分现象, 分类结果较为破碎, “椒盐”现象也比较明显, 因此利用单时相数据区分光谱相近的 2 类地物时, 存在一定难度, 分类结果的准确性也较低。在充分考虑地物的时序变化趋势基础上, 加入时序数据能够有效地解决同谱异物的问题。从视觉上观察, 图 6(b) 和 (d) 都能够很好地区分林地与旱地、盐碱地与建设用地, 并且其分类结果与各地物类别的分布范围基本保持一致, “椒盐”现象也得到改善; 图 6(e) 显

示盐碱地与建设用地区分效果明显, 但是依然存在将部分水体错分成建设用地的现象。

### 3.2 分类精度评价与分析

结合基于 GF-1 影像的 2016 年大安市自然资源调查成果数据和 Google Earth 高分辨率遥感影像数据, 在研究区内部随机选择 500 个地面验证点, 均匀分布在研究区范围内, 其中水体 7 个、草地 94 个、林地 18 个、盐碱地 4 个、旱地 232 个、建设用地 102 个、裸地 43 个, 并应用混淆矩阵方法对不同分类结果进行精度评价, 其分类精度指标如表 3 所示。

表 3 分类精度指标对比

Tab. 3 Comparison of classification accuracy index

类别	时序数据 + RF		时序数据 + SVM		时序数据 + MLC		单时相数据 + RF	
	生产者精度/%	用户精度/%	生产者精度/%	用户精度/%	生产者精度/%	用户精度/%	生产者精度/%	用户精度/%
水体	85.66	89.49	87.52	83.46	82.21	83.60	88.91	78.00
草地	96.89	82.64	93.72	88.91	76.26	72.25	73.91	73.12
林地	90.21	93.79	89.83	91.02	83.53	88.96	62.27	67.06
盐碱地	91.08	90.08	92.41	87.69	90.32	85.98	83.33	99.87
旱地	96.01	91.77	95.87	89.09	82.67	87.89	74.07	79.20
建设用地	86.75	86.86	84.45	87.42	84.38	82.66	81.95	69.26
裸地	89.63	90.48	86.91	85.04	83.56	87.88	82.94	85.10
Kappa 系数	0.855 7		0.802 3		0.783 2		0.646 4	
总体精度/%	88.87		87.51		84.26		78.82	

从表 3 中可以看出, 基于单时相数据分类的总体精度为 78.82%, Kappa 系数为 0.646 4, 而利用时序数据结合 3 种不同的分类方法获得的分类结果均优于基于单时相的分类结果, 其中利用时序数据结合 RF 分类法的分类效果最佳, 总体分类精度达到 88.87%, Kappa 系数达到 0.855 7。将时序数据结合 RF 分类法与单时相数据结合 RF 分类法的分类结果进行对比分析, 发现除水体外各地物类型的生产者精度均有所提高, 其中草地、林地和旱地的生产者精度提高幅度最大, 分别提高了 22.98、27.94 和 21.94 百分点, 而盐碱地、建设用地和裸地的生产者精度提高幅度较小, 这充分说明时序数据能更好地区分具有物候规律的地物, 能够客观真实地反映各地物类型之间的差异。此外, RF 分类法与 SVM 分

类法和 MLC 分类法相比, RF 分类法的结果总体精度最高, 并且具有分类速度快, 处理时间短, 能够有效地处理高维度数据等优点。以上结果说明, 基于时序数据结合 RF 分类法在农耕地区能够有效地区分典型地物, 其分类结果精度较高, 在农耕地区具有很好的适用性。

## 4 结论

1) Sentinel-2A/B 影像数据具有 3 个波宽很窄的“红边”波段, 这在农耕区土地利用信息提取过程中具有优越性, 尤其对具有物候特征的草地、林地和旱地地物类型更为敏感, 因此本研究挑选最佳 NDVI 和 TCW 组合的时序数据, 可有效提取出农耕区土地

覆被信息。

2) 本研究选取最佳的 Sentinel - 2A/B 时间序列组合数据, 分别采用 RF 分类法、SVM 分类法和 MLC 分类法, 以及利用单时相影像数据分别对大安市舍力镇农耕地土地覆被信息进行提取, 发现基于时间序列数据结合 RF 分类法取得了较高的分类精度和准确度, 其总体分类精度达到 88.87%, Kappa 系数达到 0.855 7。

3) 将时序数据结合 RF 分类法与单时相数据结合 RF 分类法对比分析发现, 利用时序数据结合 RF 分类法分类结果的总体精度和 Kappa 系数较后者分别提高了 10.05 个百分点和 0.209 3, 其中草地、林地和旱地的生产者精度提高幅度最大, 分别为 22.98, 27.94 和 21.94 百分点, 这充分说明时序数据能更好地区分具有物候规律的地物, 能够客观真实地反映各地物类型之间的差异。

## 参考文献 (References):

- [1] 张 静, 张 翔, 田 龙, 等. 西北旱区遥感影像分类的支持向量机法[J]. 测绘科学, 2017, 42(1): 49 - 52, 58.  
Zhang J, Zhang X, Tian L, et al. The support vector machine method for RS images' classification in northwest arid area[J]. Science of Surveying and Mapping, 2017, 42(1): 49 - 52, 58.
- [2] 马 玥, 姜琦刚, 孟治国, 等. 基于随机森林算法的农耕地土地利用分类研究[J]. 农业机械学报, 2016, 47(1): 297 - 303.  
Ma Y, Jiang Q G, Meng Z G, et al. Classification of land use in farming area based on random forest algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(1): 297 - 303.
- [3] 林 楠, 姜琦刚, 杨佳佳, 等. 基于资源一号 02C 高分辨率数据的农业区土地利用分类[J]. 农业机械学报, 2015, 46(1): 278 - 284.  
Lin N, Jiang Q G, Yang J J, et al. Classifications of agricultural land use based on high - spatial ZY1 - 02C remote sensing images [J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 278 - 284.
- [4] 王月如, 韩鹏鹏, 关舒婧, 等. 基于 Landsat8 OLI 数据的富贵竹种植区域信息提取[J]. 国土资源遥感, 2019, 31(1): 133 - 140. doi:10.6046/gtzyyg.2019.01.08.  
Wang Y R, Han P P, Guan S J, et al. Information extraction of Dra-caena sanderiana planting area based on Landsat8 OLI data[J]. Remote Sensing for Land and Resources, 2019, 31(1): 133 - 140. doi:10.6046/gtzyyg.2019.01.08.
- [5] Wardlow B D, Egbert S L. Large - area crop mapping using time - series MODIS 250 m NDVI data: An assessment for the U. S. Central Great Plains [J]. Remote Sensing of Environment, 2008, 112(3): 1096 - 1116.
- [6] Vintrou E, Desbrosse A, Bégué A, et al. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products[J]. International Journal of Applied Earth Observation and Geoinformation, 2012, 14(1): 83 - 93.
- [7] 朱永森, 曾永年, 张 猛. 基于 HJ 卫星数据与面向对象分类的土地利用/覆盖信息提取[J]. 农业工程学报, 2017, 33(14): 266 - 273.  
Zhu Y S, Zeng Y N, Zhang M. Extract of land use/cover information based on HJ satellites data and object - oriented classification [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(14): 266 - 273.
- [8] 张 猛, 曾永年. 基于多时相 Landsat 数据融合的洞庭湖区水稻面积提取[J]. 农业工程学报, 2015, 31(13): 178 - 185.  
Zhang M, Zeng Y N. Mapping paddy fields of Dongting Lake area by fusing Landsat and MODIS data [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(13): 178 - 185.
- [9] 郝鹏宇, 牛 铮, 王 力, 等. 基于历史时序植被指数库的多源数据作物面积自动提取方法[J]. 农业工程学报, 2012, 28(23): 123 - 131, 297.  
Hao P Y, Niu Z, Wang L, et al. Multi - source automatic crop pattern mapping based on historical vegetation index profiles [J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(23): 123 - 131, 297.
- [10] 龚 燃, 哨兵 - 2A 光学成像卫星发射升空[J]. 国际太空, 2015(8): 36 - 40.  
Gong R. Sentinel - 2A optical imaging satellite launched [J]. Space International, 2015(8): 36 - 40.
- [11] 范唯唯. Sentinel - 2B 卫星发射成功[J]. 空间科学学报, 2017, 37(4): 371 - 372.  
Fan W W. Sentinel - 2B satellite launched successfully [J]. Chinese Journal of Space Science, 2017, 37(4): 371 - 372.
- [12] 吴 静, 吕玉娜, 李纯斌, 等. 基于多时相 Sentinel - 2A 的县域农作物分类[J]. 农业机械学报, 2019, 50(9): 194 - 200.  
Wu J, Lyu Y N, Li C B, et al. Fine classification of county crops based on multi - temporal images of Sentinel - 2A [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(9): 194 - 200.
- [13] Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24 - 31.
- [14] 赵理君, 唐 婷. 典型遥感数据分类方法的适用性分析——以遥感图像场景分类为例[J]. 遥感学报, 2016, 20(2): 157 - 171.  
Zhao L J, Tang P. Scalability analysis of typical remote sensing data classification methods: A case of remote sensing image scene [J]. Journal of Remote Sensing, 2016, 20(2): 157 - 171.
- [15] 喻小倩, 刘 娜, 李 红, 等. 基于线性光谱混合分解和最大似然分类相结合的土地覆被分类——以红寺堡灌区为例[J]. 国土资源遥感, 2010, 22(1): 96 - 100. doi:10.6046/gtzyyg.2010.01.08.  
Yu X Q, Liu N, Li H, et al. Land cover classification based on linear spectral mixture decomposition combined with maximum likelihood classification: A case study of Hongsipu Irrigation Area [J]. Remote Sensing for Land and Resources, 2010, 22(1): 96 - 100. doi:10.6046/gtzyyg.2010.01.08.
- [16] 程红芳, 章文波, 陈 锋. 植被覆盖度遥感估算方法研究进展[J]. 国土资源遥感, 2008, 20(1): 13 - 18. doi:10.6046/gtzyyg.2008.01.02.  
Cheng H F, Zhang W B, Chen F. Advances in researches on application of remote sensing method to estimating vegetation coverage

- [J]. Remote Sensing for Land and Resources, 2008, 20(1): 13 - 18. doi: 10.6046/gtzyyg. 2008. 01. 02.
- [17] 罗开盛, 陶福祿. 融合面向对象与缨帽变换的湿地覆被类别遥感提取方法[J]. 农业工程学报, 2017, 33(3): 198 - 203.  
Luo K S, Tao F L. Method for wetland type extraction using remote sensing combining object-oriented and tasseled cap transformation [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(3): 198 - 203.
- [18] Witke S, Yu X, Karjalainen M, et al. Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest [J]. International Journal of Applied Earth Observation and Geoinformation, 2019, 76: 167 - 178.
- [19] Sun Z P, Shen W M, Wei B, et al. Object-oriented land cover classification using HJ-1 remote sensing imagery [J]. Science China Earth Sciences, 2010, 53(1): 34 - 44.
- [20] 王长耀, 刘正军, 颜春燕. 成像光谱数据特征选择及小麦品种识别实验研究[J]. 遥感学报, 2006, 10(2): 249 - 255.  
Wang C Y, Liu Z J, Yan C Y. A experimental study on imaging spectrometer data feature selection and wheat type identification [J]. Journal of Remote Sensing, 2006, 10(2): 249 - 255.
- [21] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1): 5 - 32.
- [22] Adam E, Mutanga O. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2009, 64(6): 612 - 620.
- [23] 刘舒, 姜琦刚, 马玥, 等. 基于多目标遗传随机森林特征选择的面向对象湿地分类[J]. 农业机械学报, 2017, 48(1): 119 - 127.  
Liu S, Jiang Q G, Ma Y, et al. Object-oriented wetland classification based on hybrid feature selection method combining with Relief F, multi-objective genetic algorithm and random forest [J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(1): 119 - 127.
- [24] 何云, 黄翀, 李贺, 等. 基于 Sentinel-2A 影像特征优选的随机森林土地覆盖分类[J]. 资源科学, 2019, 41(5): 992 - 1001.  
He Y, Huang C, Li H, et al. Land-cover classification of random forest based on Sentinel-2A image feature optimization [J]. Resources Science, 2019, 41(5): 992 - 1001.

## Land use classification of farming areas based on time series Sentinel-2A/B data and random forest algorithm

WANG Dejun<sup>1</sup>, JIANG Qigang<sup>2</sup>, LI Yuanhua<sup>2</sup>, GUAN Haitao<sup>1</sup>, ZHAO Pengfei<sup>1</sup>, XI Jing<sup>2</sup>

(1. The Fifth Surveying Mapping and Geographic Information Engineering Institute of Heilongjiang Province, Harbin 150081, China; 2. College of Geo-Exploration Science and Technology, Jilin University, Changchun 130026, China)

**Abstract:** Land cover information in farming areas is the basis of land resource management and planning, which plays an important role in the rational development of land resources, adjustment of land use structure, and dynamic monitoring of land. Due to the complex land types and high heterogeneity in farming areas, the accuracy of land cover information extraction has been facing challenges. Therefore, this study used Sentinel-2A/B remote sensing data as the data source. Firstly, a normalized difference vegetation index (NDVI) time series data set and tasseled cap wetness (TCW) time series data set were constructed; Secondly, the J-M (Jeffries-Matusita) distance was used to analyze the separability of the surface features and select the best time series data combination of NDVI and TCW; Finally, combined with random forest (RF), support vector machine (SVM), maximum likelihood classification (MLC) and single phase remote sensing data, the classification of typical features in farming areas was studied, and the accuracy of classification results was evaluated and compared. The research results show that the classification accuracy of the time series data combined with the random forest classification algorithm is relatively high. The overall classification accuracy reaches 88.87%, and the Kappa coefficient reaches 0.8557, which improves the classification accuracy by 10.05 percentage points and 0.2093 respectively compared with that of the single remote sensing data. This fully demonstrates that the combination of time series data and random forest classification algorithm can effectively improve the classification accuracy of typical features in farming areas.

**Keywords:** time series; random forest; land use classification; farming area; Sentinel-2A/B

(责任编辑: 张仙)