

# 基于随机森林模型提取干旱区人工绿洲耕地

## -以阿拉尔绿洲区为例

王静

新疆水利水电科学研究院

DOI:10.32629/eep.v3i5.823

**[摘要]** 本文以干旱区阿拉尔2013年-2018年的Landsat8为主要数据来源,基于随机森林模型对干旱区绿洲植被信息进行提取。结果表明:总体分类精度为87%,满足干旱区绿洲提取要求。随机森林模型可以很好地提取干旱区绿洲植被,为干旱区提取绿洲植被提供了依据。

**[关键词]** 干旱区; 绿洲提取; 随机森林模型

### 引言

人工绿洲耕地提取是大众尺度遥感数据分析和相关数据分析的前提和基础,在生态学、地理学及新兴的生态水文学、水文土壤等学科中都有广泛应用。从成本和时效性角度考虑,区域绿洲耕地覆盖状况,绝大部分都基于遥感技术获取。而基于遥感技术获取植被信息的分类精度,对于陆面过程研究结果至关重要。遥感技术在区域上实现了实时快速的植被动态监测<sup>[1-2]</sup>,为相关领域的研究提供了技术支撑。随着遥感技术的发展,影像的空间分辨率得到了很大的提高,由此对影像的处理技术有了更高的要求。目前地物分类技术正在发展和探讨中,如侍昊等利用随机森林模型对太湖水生植被进行遥感信息提取<sup>[3]</sup>,赵艺淞采用随机森林模型提取了深圳市某区域不透水面<sup>[4]</sup>,随机森林模型对不同地区目标要素提取取得了很好的效果。但是在干旱区,植被信息提取研究仍然较少,这是由于干旱区的绿洲植被分布特点、气候等因素的限制,在进行绿洲植被信息提取时,部分绿洲植被不易被计算机区分。然而干旱区植被空间分布信息获取对开展区域生态恢复和环境管理有着重要的理论和实践意义,开展这些工作的前提即是干旱区的绿洲植被信息提取。文中以干旱区阿拉尔为研究区,利用随机森林模型对遥感影像进行信息提取试验,以期为干旱区植被信息精确提取提供科学的方法以及科学依据。

### 1 研究区概况

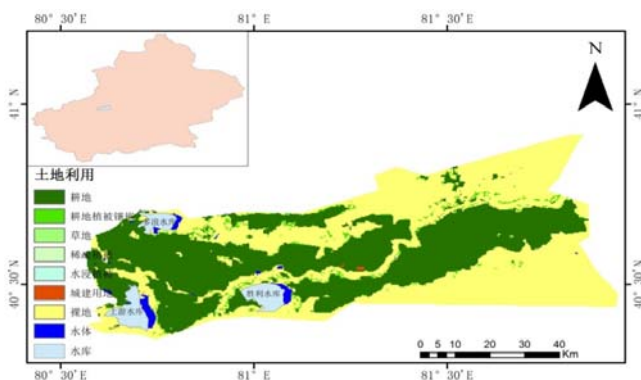


图1 研究区位置

Fig.1 The study area

研究区位于新疆南部阿拉尔地区,包含塔里木灌区及周边荒漠区,地势由西北向东南倾斜,坡降1/200~1/300,海拔高程1041~1000m,

地理坐标为东经80° 30' ~81° 58', 北纬40° 22' ~40° 57', 总面积为4161km<sup>2</sup>,区域包括阿拉尔市(九团)、七团、八团、十团、十一团、十二团、十三团、十四团、十五团、十六团。研究区域分布示意图如图(1)所示。

### 2 数据与方法

#### 2.1 数据

##### 2.1.1 地理信息数据

选用的地理信息数据包括:阿拉尔地区边界及DEM,矢量边界来自于<http://www.sbsm.gov.cn>。DEM数据来自于“国际科学数据服务平台”<http://datamirror.csdb.cn/index.jsp>网站,空间分辨率为90m×90m。

##### 2.1.2 遥感数据

本文选用Landsat8 2013年-2018年数据(尽量选择云量少于20%),采用ENVI进行预处理(包括大气校正、辐射定标等等步骤)为后期数据分类以及进行各种实验做好准备工作。

#### 2.2 方法

随机森林基本原理:随机森林算法是由很多没有关联的决策树随机组合而成的一个森林集合,变量(列)和数据(行)被随机抽取进行训练,然后形成决策树,最终汇总决策树的结果。具体的过程为:采用Bootstrap从原始数据集N中有放回的抽取k个训练样本,并构建k棵决策树,每次抽取剩余的数据组成k个测试数据,假设共有m个变量,每一个节点训练时抽取w个变量,然后比较w个变量中分类重要性最强的变量,这个变量的阈值是通过检查每个节点确定。最后将生成的很多棵决策树组合成随机森林,并使用此随机森林对新的数据集进行判别或者分类,最终的结果取决于决策树分类器的投票情况。随机森林分类器对参数不敏感,训练速度较快,更适合多分类问题。

### 3 结果与讨论

#### 3.1 基于耕地的绿洲识别

选取Landsat8数据,根据研究发现,植被在生长季在影像上表现的更突出,所以耕地植被提取选在2013年-2018年6月-8月云量小于20%的影像,并进行预处理(大气校正、辐射定标等),采用随机森林的提取方法提取并得到如图2所示。

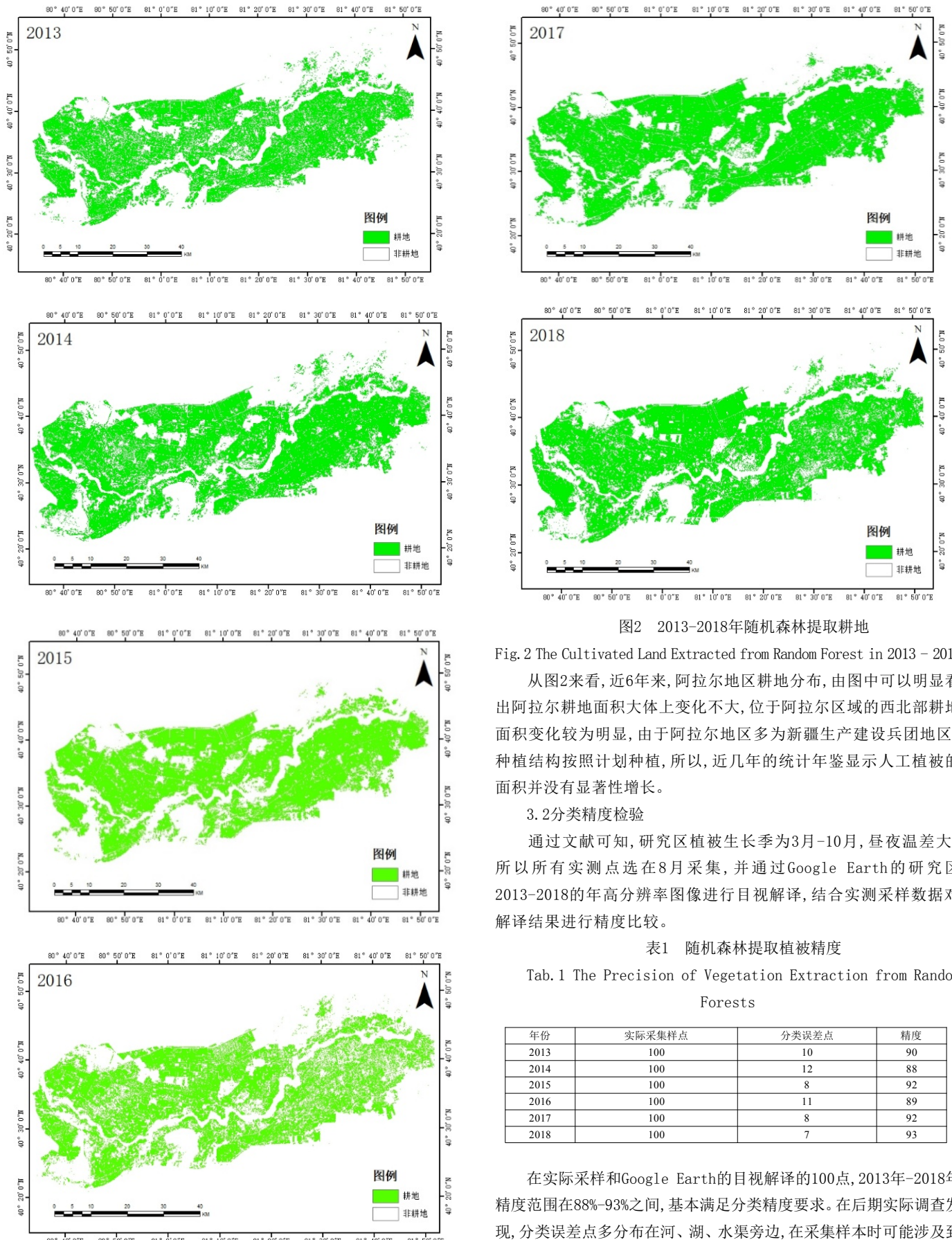


图2 2013-2018年随机森林提取耕地

Fig. 2 The Cultivated Land Extracted from Random Forest in 2013 - 2018

从图2来看,近6年来,阿拉尔地区耕地分布,由图中可以明显看出阿拉尔耕地面积大体上变化不大,位于阿拉尔区域的西北部耕地面积变化较为明显,由于阿拉尔地区多为新疆生产建设兵团地区,种植结构按照计划种植,所以,近几年的统计年鉴显示人工植被的面积并没有显著性增长。

3.2分类精度检验

通过文献可知,研究区植被生长季为3月-10月,昼夜温差大,所以所有实测点选在8月采集,并通过Google Earth的研究区2013-2018的年高分辨率图像进行目视解译,结合实测采样数据对解译结果进行精度比较。

表1 随机森林提取植被精度

Tab.1 The Precision of Vegetation Extraction from Random Forests

年份	实际采集样点	分类误差点	精度
2013	100	10	90
2014	100	12	88
2015	100	8	92
2016	100	11	89
2017	100	8	92
2018	100	7	93

在实际采样和Google Earth的目视解译的100点,2013年-2018年精度范围在88%-93%之间,基本满足分类精度要求。在后期实际调查发现,分类误差点多分布在河、湖、水渠旁边,在采集样本时可能涉及到



经纬度定位不准等问题,可能是造成分类不准的原因。

研究区阿拉尔包含兵团农一师7团-14团,通过查阅农一师阿拉尔年鉴可以得到各团耕地面积,将各团面积相加得到整体面积,可以得到面积精度。

根据已有农一师阿拉尔年鉴[7](2013-2015),将2013年、2014年、2015年植被面积统计出来,与用随机森林的方法对比,得到精度验证。

表2 随机森林植被提取面积精度

Tab. 2 The Precision of Random Forest Vegetation Extraction Area

年份	统计数据	提取数据	精度
2013	1650	1846	89.38%
2014	1940	2018	96.13%
2015	1895	1911	99.16%

从面积可以看出提取面积略高于实际面积,相差不大,所提面积值得参考,统计数据的面面积变化趋势与提取数据的面面积变化趋势一致。

### 3.3 人工绿洲时空变化分析

由上节可知精度随机森林算法值得信任,研究区一共有农一师7团-14团,研究将7团-14团分别提出来比较面积变化。

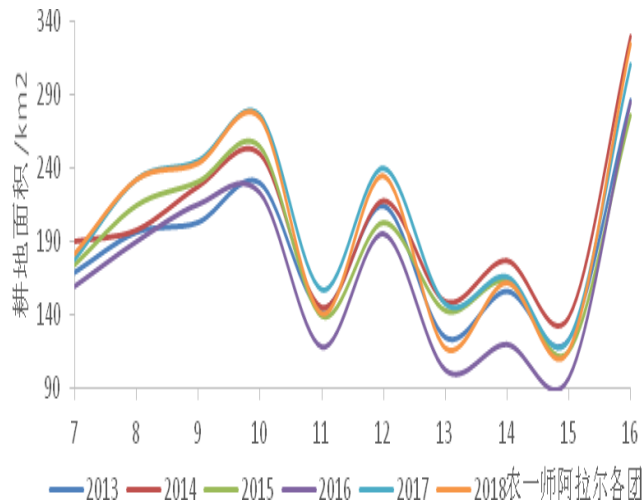


图3 2013年-2018年7团-14团耕地面积

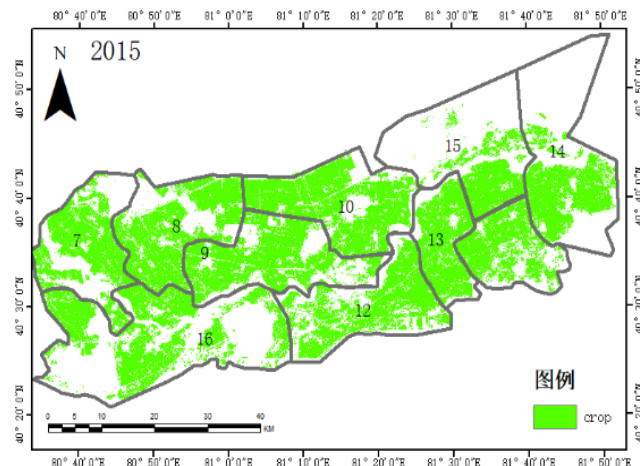
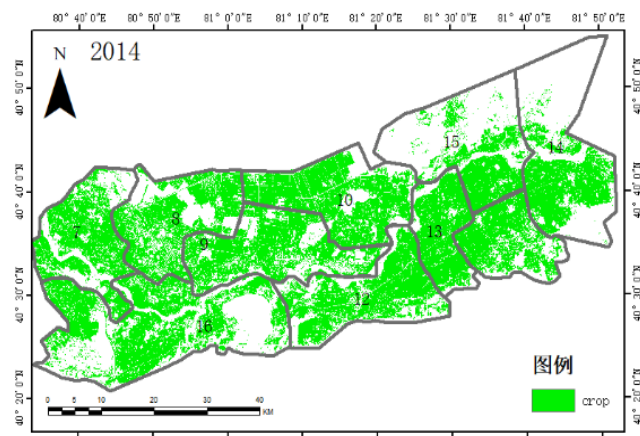
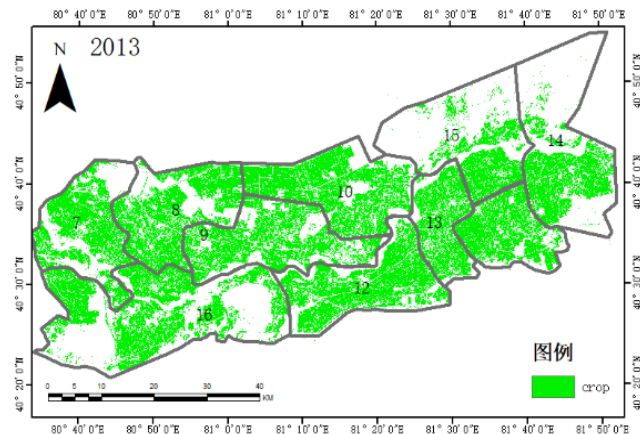
Fig. 3 Cultivated Land Area of 7th and 14th Regimes from 2013 to 2018

由图3可知,2013年-2018年各团的人工植被面积,较为稳定。主要原因,此区域为生产建设兵团区域,各团有计划的进行植被的面积分配,所以近几年植被面积趋于稳定。

通过对各团面积的提取分析,可以得到的其面积值并通过了精度检验。因此可以分析其各团的面积变化。

由图4可知,阿拉尔地区部分团随着时间的变化绿洲面积一直持续动态上升。如下所示:2013年7团耕地面积为168.96km<sup>2</sup>,相较于2013年耕地面积,7团2014年-2018年耕地面积上升了12.3%、3%、-5.41%、

5.12%、7.33%;2013年8团耕地面积为196.27km<sup>2</sup>,相较于2013年耕地面积,8团2014年-2018年耕地面积上升了0.57%、9.3%、-3.06%、18.29%、18.2%;2013年9团耕地面积为203.04km<sup>2</sup>,相较于2013年耕地面积,9团2014年-2018年耕地面积上升了11.9%、13.64%、6.24%、20.64%、19.73%;2013年10团耕地面积为229.71km<sup>2</sup>,相较于2013年耕地面积,10团2014年-2018年耕地面积上升了8.55%、10.81%、-2.74%、20.17%、19.31%;2013年16团耕地面积为156.07km<sup>2</sup>,相较于2013年耕地面积,16团2014年-2018年耕地面积上升了15.28%、-3.51%、-0.24%、8.62%、13.43%。



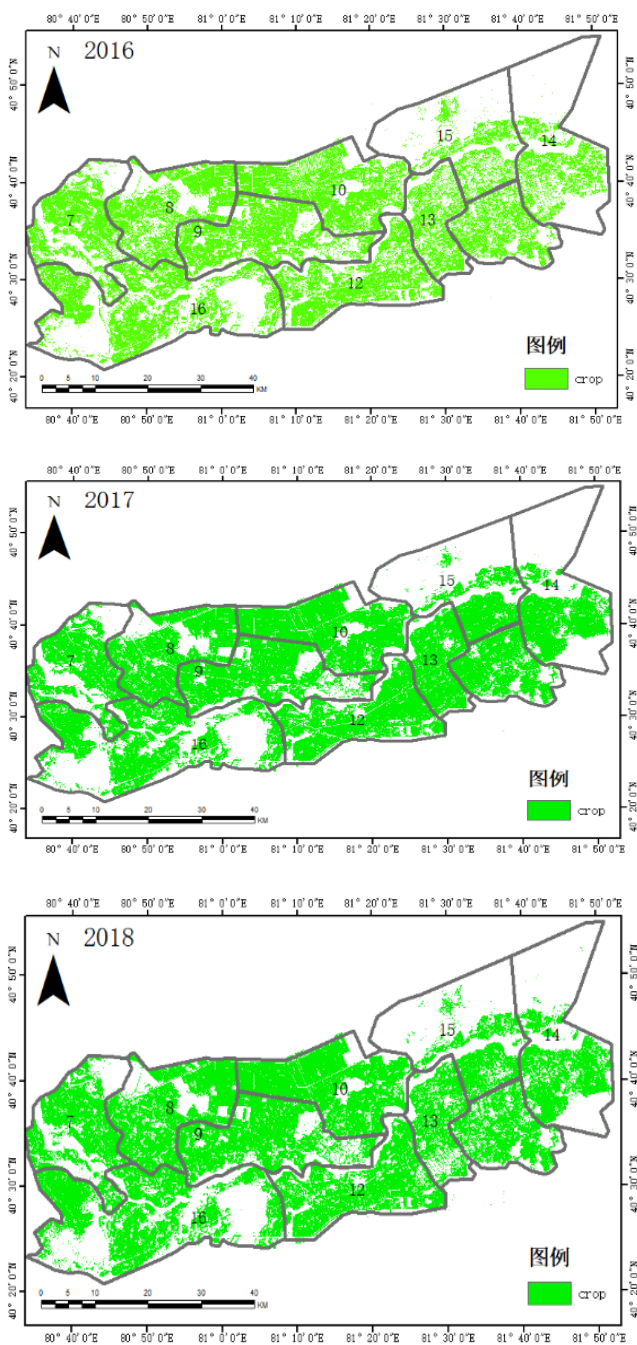


图4 2013年-2018年研究区7团-14团人工植被面积变化

Fig. 4 The Changes in Artificial Vegetation Area of 7th and 14th Regimes in the Research Area from 2013 to 2018

由图4中可以看出7、8、9、10、16团主要位于研究区的东面,东面位于三河汇合口处,阿拉尔地区有3个水库在7、8、16团团界中,因为水源的充足,所以此地区的人工植被面积一直在持续增长。

由图4可知,部分团随着时间的变化绿洲面积一直维持动态变化不明显上升。如下所示:2013年11团耕地面积为144km<sup>2</sup>,相较于2013年耕地面积,11团2014年-2018年耕地面积上升了0.42%、-3.52%、-17.81%、8.91%、-2.33%;2013年12团耕地面积为144km<sup>2</sup>,相较于2013年耕地面积,12团2014年-2018年耕地面积上升了1.47%、-5.38%、-8.78%、11.93%、9.31%;2013年13团耕地面积为124.72km<sup>2</sup>,相较于2013年耕地面积,13团2014年-2018年耕地面积上升了19.15%、14.39%、-17.61%、17.98%、-6.13%;2013年14团耕地面积为156.07km<sup>2</sup>,相较于2013年耕地面积,14团2014年-2018年耕地面积上升了13.07%、4.28%、-23.01%、6.11%、3.58%;2013年15团耕地面积为156.07km<sup>2</sup>,相较于2013年耕地面积,15团2014年-2018年耕地面积上升了11.95%、-6.78%、-21.16%、-0.55%、-6.31%。

#### 4 结论

随机森林模型能够准确地提取人工绿洲。通过使用随机森林的方法对2013年-2018年西北干旱地区阿拉尔人工绿洲耕地提取,能够准确地得到阿拉尔区域各团耕地面积的多少,研究发现阿拉尔耕地面积随着时间的变化在波动递增,这与当地的农业发展种植结构有巨大的相关关系。

#### [参考文献]

- [1]贾丽斯,谢积鉴,张雪玲.针对植被覆盖的遥感监测技术国内专利现状研究[J].科技展望,2017,027(018):267-268.
- [2]张航,仲波,洪友堂,等.近20多年来赣州地区稀土矿区遥感动态监测[J].遥感技术与应用,2015,30(2):376-382.
- [3]侍昊,李旭文,牛志春,等.基于随机森林模型的大湖水生植被遥感信息提取%Remote sensing information extraction of aquatic vegetation in Lake Taihu based on Random Forest Model[J].湖泊科学,2016,028(003):635-644.
- [4]赵艺淞,杨昆,王保云,等.随机森林在城市不透水面提取中的应用研究[J].云南师范大学学报:自然科学版,2017,(37):78.

#### 作者简介:

王静(1993--),女,新疆乌鲁木齐人,汉族,硕士研究生,主要从事水文水资源研究。