

饶品增, 蒋卫国, 王晓雅, 等. 基于 MODIS 数据的洪涝灾害分析研究——以 2017 年洞庭湖区洪水为例[J]. 灾害学, 2019, 34(1): 203–207. [RAO Pingzeng, JIANG Weiguo, WANG Xiaoya, et al. Flood disaster analysis based on MODIS data—Taking the flood in dongting lake area in 2017 as an example [J]. Journal of Catastrophology, 2019, 34(1): 203–207. doi: 10.3969/j.issn.1000-811X.2019.01.037.]

## 基于 MODIS 数据的洪涝灾害分析研究 ——以 2017 年洞庭湖区洪水为例\*

饶品增<sup>1,2</sup>, 蒋卫国<sup>1,2</sup>, 王晓雅<sup>1,2</sup>, 陈 坤<sup>3</sup>

(1. 北京师范大学 地理科学学部 环境演变与自然灾害教育部重点实验室, 北京 100875;

2. 环境遥感与数字城市北京市重点实验室北京师范大学 地理科学学部, 北京 100875;

3. 中国环境科学研究院土壤与固体废物研究所, 北京 100012)

**摘 要:** 洪涝灾害对人类的生产生活造成重要影响。本文根据提出的洪水水体提取流程, 选用时间分辨率较高的 MOD09A1 数据和高效稳定的随机森林方法提取洞庭湖区 2017 年的洪水淹没水体, 并获取洪水淹没区的农作物分布数据, 统计分析计算得本次洪水造成的粮食产量损失等。结果表明: ①随机森林方法相比传统的指数方法(NDVI、NDWI 和 MNDWI)和常用的机器学习方法(逻辑回归和支持向量机), 水体提取效果较好; ②多时期影像叠加获取的洪水淹没水体提取效果相比单期影像效果更好; ③此次洪水造成的水体淹没面积为 1 643.11 km<sup>2</sup>, 主要位于洞庭湖周边地区; ④洪水造成的农作物淹没面积为 503.46 km<sup>2</sup>, 作物损失量约为 32.37 万 t, 农作物导致的经济损失约为 14.44 亿元。

**关键词:** 洪涝灾害; 洞庭湖区; MOD09A1; 随机森林; 农作物

**中图分类号:** X43; P407 **文献标志码:** A **文章编号:** 1000-811X(2019)01-0203-05

doi: 10.3969/j.issn.1000-811X.2019.01.037

洪水灾害是发生最频繁和影响最严重的自然灾害之一。较高分辨率的光学遥感影像能应用于洪水水体淹没范围的监测和研究, 尤其是平原地区的湖泊和河流等<sup>[1]</sup>。目前, 常用于监测洪水的遥感数据包括 Landsat 和 MODIS 等。其中 Landsat 系列数据具有较高的空间分辨率(普遍为 30 m), 能提取较小范围的水体, 但由于时间间隔较大(16 d 及以上), 对周期短和强度大的洪水的监测常存在不足<sup>[2-3]</sup>。MODIS 数据具有较高的空间分辨率(250 m 和 500 m 等), 且其时间周期短, 能很好地监测洪水动态变化过程<sup>[4-5]</sup>。对于时间周期较长的影像如 Landsat 数据, 洪水时期通常只有一期或少数几期影像能满足要求, 有时甚至没有。但洪水期间常存在着巨大的水体淹没范围波动, 不同日期洪水的淹没范围完全不同。可见, 时间间隔较大的遥感影像所获取的洪水水体淹没结果的实时性较差, 且不全面。而对于时间周期短的遥感影像如 MODIS, 能很好的捕捉到洪水期间水体淹没的动态变化过程, 这有助于更准确地探究洪水所造成的灾害影响等<sup>[6]</sup>。理论上说, 洪水期淹没的最大范围减去正常水体覆盖即为洪水

淹没的有效范围。

遥感数据常用的水体提取方法主要包括两大类: 指数方法和监督/非监督分类的方法。指数方法通过利用 2 个及以上的波段构建一个新的指数, 通过设定阈值区分水体和非水体。指数方法几乎都是在正常水体情况下构建的, 在提取洪水范围时, 会受到很多因素的干扰, 包括水质和云等。监督/非监督分类方法不用设定阈值, 利用相关的数理统计方法和合适的分类器便可达到分类的目的, 当下热门的机器学习方法便属于此类方法, 典型的如支持向量机<sup>[7]</sup>、神经网络和随机森林等。该方法通过选取适量的样本点和足量的波段信息, 便可构建相应的分类模型并实现水体提取的目的。

洞庭湖由于其特殊的气候条件和地理位置, 是中国洪水灾害发生最频繁的地区之一。2017 年 6 月~7 月, 洞庭湖上游地区普降暴雨, 局部地区出现特大暴雨, 湘江、资水、沅水三条干流及洞庭湖区等地均出现了超警戒水位。7 月中旬, 洞庭湖的入湖洪量达 64 630 m<sup>3</sup>/s, 超过了 1998 年的

\* 收稿日期: 2018-05-27 修回日期: 2018-08-13

基金项目: 国家自然科学基金(41571077); 国家重点研发计划专项项目(2017YFB0504102); 水资源安全北京实验室项目  
第一作者简介: 饶品增(1990-), 男, 江西上饶人, 硕士研究生, 主要从事洪水灾害和降雨遥感方面的研究。

E-mail: raopinzen@mail.bnu.edu.cn

通讯作者: 蒋卫国(1976-), 男, 湖南衡阳人, 教授, 主要研究方向为洪水灾害及湿地生态遥感监测与评价研究。

E-mail: jiangweiguo@bnu.edu.cn

63 360 m<sup>3</sup>/s, 直逼 1954 年的 67 000 m<sup>3</sup>/s<sup>[8]</sup>, 造成了该地区前后一个多月严重的洪水灾害。目前针对该次洪水事件的相关研究较少, 且对洪水造成的灾害损失的评估存在欠缺。本文选取随机森林模型和多期 MOD09A1 产品数据获取到该次洪水造成的洪水淹没范围, 并统计淹没区内可能造成的农作物损失及相应的经济损失, 得到了较为可靠的结论, 这为相关部门和该地区的洪水灾害损失提供了一种较好的评估途径。

## 1 研究区概况

洞庭湖是中国第二大淡水湖泊, 仅次于鄱阳湖, 位于长江中游湖南省北部, 经纬度介于 28°30'N ~ 30°20'N 和 112°25'E ~ 113°15'E 之间, 北与长江相连, 南部地区为洞庭湖的水系来源, 流域面积达 26.28 万 km<sup>2</sup>, 主要涵盖了湖南省大部分地区及周边省份部分地区。流域内降水丰富, 且存在明显的季节性变化, 夏季降水频繁且强度大, 冬季则降水稀少。夏季丰水期时, 洞庭湖水面辽阔, 洪水灾害时有发生; 冬季枯水期时, 水面严重萎缩, 形成多草的滩地。本文选择洞庭湖及和湖周边低洼平原地区, 主要涵盖了湖南省的 15 个市县(图 1)。

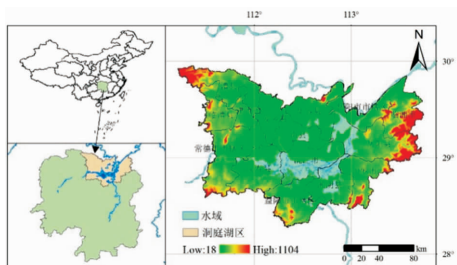


图 1 研究区地理位置

## 2 数据来源与研究方法

### 2.1 数据来源

#### 2.1.1 MODIS 数据

##### (1) MOD09A1

MOD09A1 是 Terra 卫星经过大气辐射和气溶胶校正后的 MODIS 地表反射率合成产品, 属于陆地 2 级产品, 其空间分辨率为 500 m, 时间分辨率为每 8 d, 主要涵盖的数据为 MODIS 的 1~7 波段反射率数据, 以及其他一些辅助数据, 包括太阳天顶角, 像元生产日期和波段质量数据等。本文从 USGS 官网 (<https://lpdaac.usgs.gov/>) 下载了 2017 年 2017001~2017361 天共 46 期的 MOD09A1 数据。数据格式为 HDF, 采用 MRT (MODIS Reprojection Tool) 工具进行批处理获取得到相应的 1~7 个波段反射率数据。

##### (2) MOD44W

MOD44W 最早是由 Salomon 等<sup>[9]</sup>在 2004 年生产的一套全球 250 m 分辨率的陆表水体数据, 目前该套产品已被 MODIS 官方收录在 MODIS 产品当

中, 而且有了第 6 代产品。该套数据覆盖了全球主要的湖泊和较大型河流, 其水体分类效果较好已被许多研究人员证实, 但整体上略偏大<sup>[10]</sup>。MOD44W 产品主要提供了粗略的水体淹没数据, 其中水体值为 1, 陆地值为 0, 空值或者边界值为 253。本文使用的 MOD44W 数据为第 5 代的产品数据, 用作水体样本点选取的先验知识。

#### 2.1.2 农作物分布数据

本文获取了陈军等<sup>[11]</sup>生产的 2010 年全球 30 m 的土地利用数据, 该数据提供了 10 种地表类型, 包括水体、湿地、人工表面、林地、草地、耕地、灌丛、冰雪、苔原和裸地等。本文选取了洞庭湖区的土地利用数据, 涵盖其中的 8 种土地利用类型, 其中占比较大的包括耕地、水体、湿地和林地等(图 2)。

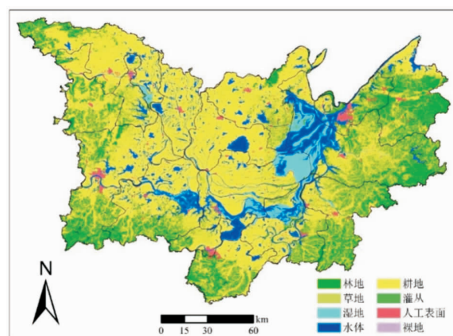


图 2 洞庭湖区土地利用数据

### 2.2 研究方法

本文使用的方法主要包括两部分: 水体提取方法和洪水淹没范围确定方法。采用机器学习的方法做监督分类, 主要需要考虑三个问题: ①水体和非水体样本点的选择; ②解释变量的获取; ③机器学习方法的高效性。本文通过多种方法比较, 包括传统的指数方法和目前应用较广泛的逻辑回归、支持向量机和随机森林方法等, 最终选择了随机森林作为洪水水体提取的分类方法。另外, 对于洪水水体淹没范围的获取, 本文采用了空间叠置和统计的方法, 构建了洪水淹没水体获取的整体流程。

#### 2.2.1 随机森林模型

随机森林 (Random Forest, RF) 是由 Leo Breiman 等<sup>[12]</sup>于 2001 年提出来的。它是基于传统的决策树原理改进成的集成分类器, 决策树是单一的树, 而随机森林则是由多棵决策树组合成分类回归树 (Classification And Regression Tree, CART) 构成。在训练过程中, 对于每一棵树来说, 他们之间都没有关联, 选择从已知提供的总训练集中有放回地随机选取样本用作训练, 其训练样本都不完全相同。训练结果通过某种判决条件来进行判断, 好的结果保留, 差的去除, 这使得样本被使用的频率完全不同, 其中有效样本被多次、高效利用, 无效样本的使用率降低, 使得其噪音影响也随之下降。基于此原理, 相比于单一的决策树分类方法, 随机森林由于具有多棵树和重复采样的特点在分类过程中表现出更稳健, 泛化能

力更好的特点。

根据 Leo Breiman 的文章和其他一些相关材料, 随机森林的过程可以定义为: 随机森林是由一系列树状分类器组成的一个集成分类器。它主要通过一个边际函数来进行收敛, 公式如下:

$$mg(X, Y) = \text{av}_k(I(h_k(X) = Y)) - \max_{j \neq Y} \text{av}_k(I(h_k(X) = j)) \quad (1)$$

式中:  $\text{av}_k$  表示取平均;  $I(\cdot)$  是一个指示函数, 显然,  $I(h_k(X) = Y)$  和  $I(h_k(X) = j)$  等于 0 或者 1, 前者代表了投票正确的概率, 后者则代表了其它分类(分类的目标变量应该大于等于 2)投票错误的概率, 两者取平均相减, 当差值  $mg(X, Y)$  越大, 则表明分类正确的置信度越高。

### 2.2.2 洪水水体提取流程

对于一次洪水过程, 其造成的水体淹没范围应为洪水时期的最大淹没范围去掉洪水前的淹没范围。水体提取流程包含 6 个步骤: 第 1 步, 获取 MOD09A1 数据并预处理, 通过相应波段生产 NDVI、NDWI 和 MNDWI 三个指数, 和 MOD09A1 本身含的 7 个波段一共构建 10 个特征变量; 第 2 步, 利用 MOD44W 人工采集水体和非水体样本点, 总数约 1 000 个左右, 水体和非水体均约 500 个; 第 3 步, 基于拥有的特征变量和样本点, 利用随机森林分类器进行分类; 第 4 步, 初步分类结果通过人工目视解译去除云和云阴影等噪声; 第 5 步, 区分洪水期和非洪水期水体影像, 将洪水期水体影像进行叠加得到洪水期最大淹没面积, 背景水体影像则采用非洪水期的最大一期水体影像; 第 6 步, 利用背景水体掩膜洪水期最大淹没水体, 获得洪水淹没水体范围, 其它部分还包括正常水体和非水体像元(图 3)。

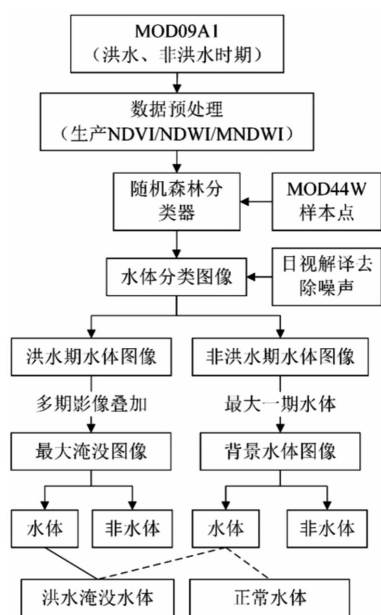


图3 洪水水体提取流程图

### 2.2.3 其他方法

除随机森林方法外, 本文还利用了 NDVI、NDWI 和 MNDWI 等三个指数提取水体。另外, 为了验证随机森林方法相比其它类似的机器学习类

的监督分类方法效果更好, 本文还尝试了应用逻辑回归(Logistic Regression, LR)和支持向量机(Support Vector Machine, SVM)提取研究区的水体。逻辑回归方法为改进的多元线性回归方法, 被广泛应用于分类和回归, 其中应用较多的研究领域为地质灾害易发性评价等<sup>[13-14]</sup>。支持向量机是非线性分类器, 提出于 1995 年, 目前被广泛应用于数据分析、地物分类和模式识别等。在遥感地表分类方面, 其效果也得到了广泛的认可<sup>[15]</sup>。

## 3 评价结果分析

### 3.1 水体提取方法比较分析

选取洪水时期含云量较少的一期影像(2017185 天), 采用三个指数(NDVI、NDWI 和 MNDWI)和三种机器学习方法(LR、SVM 和 RF)提取水体, 不做去噪声处理, 并和 2、5、3 波段假彩色影像进行比较(图 4)。其中, 通过多次尝试, NDVI、NDWI 和 MNDWI 选择的最优阈值分别为 0.1、-0.1 和 -0.1。对比图 4 中效果, 彩色图中的蓝色圆圈为薄云覆盖区域, NDVI、NDWI 和 MNDWI 指数存在将较大部分的云错分为水体, 尤其是 MNDWI, 云几乎都被识别为水体。红色圆圈里面的水体在 LR 和 SVM 方法中均存在较明显的漏分现象, 尤其是 LR, 大量水体像元未被识别。RF 在错分和漏分两个方面表现均较稳定, 相比其它五种方法较为精确和稳定。因此, 本文选择 RF 作为最终的洪水水体提取方法。

根据假彩色影像, 目视解译其中的水体部分, 并与随机森林的水体提取结果进行叠加比较。结果显示, 水体重叠率为 95.65%, 这可看作为该影像的水体提取精度。

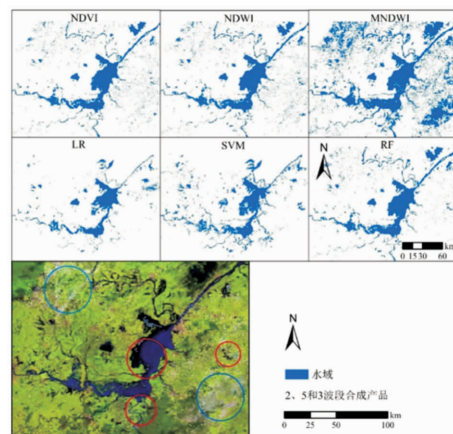


图4 2017185 天 MOD09A1 遥感影像水体提取结果

为了确定非洪水时期的水体提取效果, 本文选择洪水前的一期影像(2017169 天)进行精度验证。该影像本身也含有部分云, 本文利用 MOD09A1 的波段质量数据, 去除云层, 并手动去除噪声和修复部分水体, 得到最终如图 5b 中的水体分布结果。目视解译图 5a 中的水体部分, 将两者进行叠加, 统计得水体像元的重合率为 95.53%。可见, 随机森林方法能很好地应用于研



究区的水体提取。

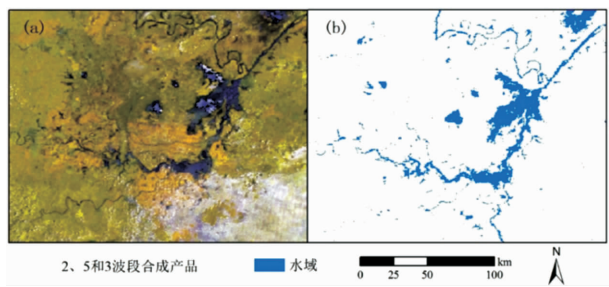


图5 2017169天MOD09A1遥感影像水体提取结果(RF)

3.2 洪水水体变化分析

据调查和分析,6-7月洪水造成的大范围淹没水体主要发生在6月底和7月,2017169天可看作为洪水前的背景水体,2017177-2017209天的5期MOD09A1影像可看作为洪水淹没水体,结果见图6。

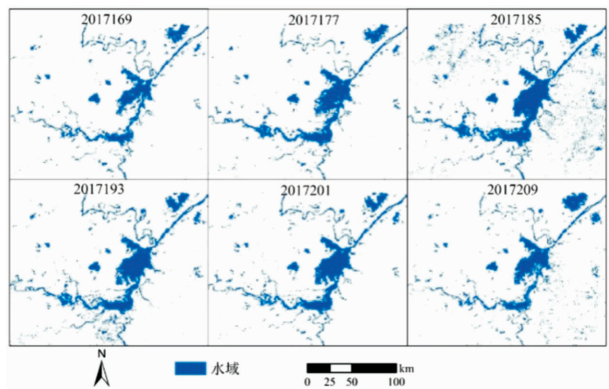


图6 2017年洞庭湖区洪水前后淹没范围变化

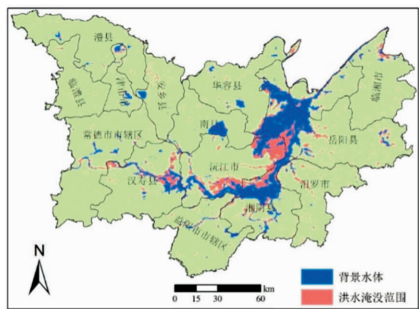


图7 2017年洞庭湖区洪水淹没范围

根据2.2.2中的洪水水体提取流程和方法,叠加2017177-2017209天的5期水体结果影像得本次洪水淹没的最大水体面积,该结果被认为是本次洪水的水体淹没范围,背景水体选择洪水前2017169天的水体提取结果影像,洪水淹没范围见图7。经统计,洪水时期最大的水体淹没面积为3569.8 km<sup>2</sup>,去掉背景水体的面积1926.7 km<sup>2</sup>,实际造成的水体淹没面积为1643.1 km<sup>2</sup>,淹没面积相当于正常水体面积的85.28%。受淹地区主要位于洞庭湖周边,包括东洞庭湖南侧、南洞庭湖北侧和西洞庭湖西侧等地。按行政区域划分,受淹面积最大的县级单元为沅江市,离洞庭湖水域较远的县市受淹面积相对较小,如临澧县等。

3.3 农作物损失评估

叠加洪水淹没范围图和土地利用图,统计得本次洪水淹没的各种土地利用类型面积(表1)。损失面积最大的为耕地,面积达503.46 km<sup>2</sup>。

表1 洪水淹没区各土地利用类型面积

类型	总面积/km <sup>2</sup>	洪水淹没面积/km <sup>2</sup>
耕地	15080.19	503.46
林地	5899.39	165.67
草地	1117.23	93.98
湿地	849.56	362.98
水体	3184.35	490.87
人工表面	516.16	26.16

土地利用数据没有给出明确的耕地类型,据先验知识和调查,洞庭湖区耕地,尤其是洞庭湖周边的耕地基本都为水田。因此,受淹耕地均可看作为水田。由于阳历6月底和7月为洞庭湖水稻成熟和收割季节,本次洪水很大程度上会摧毁该地区的水稻作物。通过查阅2016年湖南省经济社会发展报告,最近一年(2015年)湖南省稻谷单位面积产量为6429 kg/hm<sup>2</sup><sup>[16]</sup>,假设所有农田水稻均被破坏,根据此计算得该次洪水造成的粮食产量损失为32.37万t。以2017年12月湖南省大米均价每市斤2.23元<sup>[17]</sup>计算,得农作物受损导致的经济损失约为14.44亿元。

统计洞庭湖区各个市县受损作物面积、作物损失量和经济损失(图8)。以县级单元统计,受损最严重的为沅江市,受损作物面积达到11.95 km<sup>2</sup>,对应的作物损失量和经济损失分别为7.68万t和17.13亿元;以地市级单元统计,受损最严重的为岳阳市,受损作物面积达到20.76 km<sup>2</sup>,对应的作物损失量和经济损失分别为5.95万t和13.35亿元。

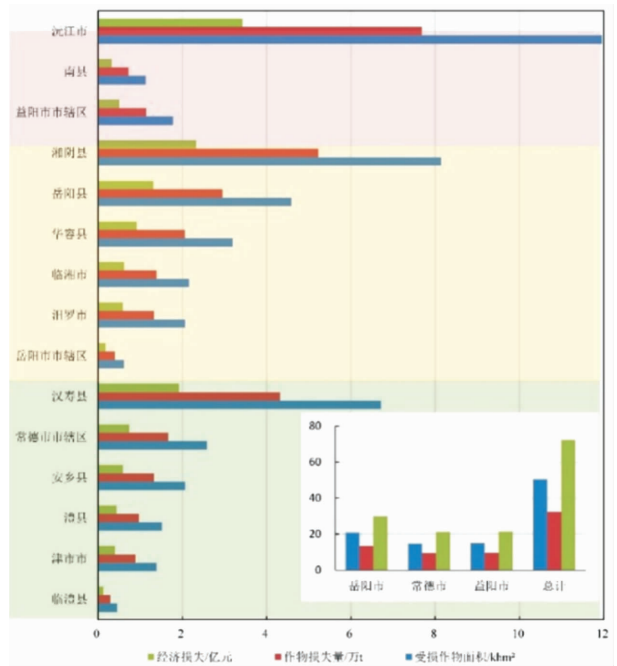


图8 洞庭湖区各市县洪水受灾结果图

## 4 结论与讨论

不同于前人利用单期影像反映洪水水体, 本文提出了基于多期影像叠加的方式获取洪水水体淹没范围。此外, 本文对比了多种水体提取方法, 最终确认了随机森林方法的高效性和鲁棒性。最后, 基于获取的洪水水体淹没范围统计了本次洪水可能造成的农作物损失等。结果表明:

(1) 随机森林方法相比传统的指数方法 (NDVI、NDWI 和 MNDWI) 和常用的机器学习方法 (逻辑回归和随机森林) 具有较好的水体提取效果, 在研究区的水体提取精度约为 95.59%。

(2) 根据洪水水体提取流程, 本次洪水淹没面积为 1 643.1 km<sup>2</sup>, 淹没面积相当于正常水体面积的 85.28%, 受淹较严重的区域主要位于东洞庭湖南侧、南洞庭湖北侧和西洞庭湖西侧等地。

(3) 本次洪水造成的农作物淹没面积为 503.46 km<sup>2</sup>, 基于此估算的作物损失量约为 32.37 万 t, 导致的农作物经济损失约为 14.44 亿元。其中受灾最严重的地级市为岳阳市, 最严重的县级单元为沅江市。

## 参考文献:

- [1] 李香颜, 陈怀亮, 李有. 洪水灾害卫星遥感监测与评估研究综述 [J]. 中国农业气象, 2009, 30(1): 102–108.
- [2] WANG Y, Colby J D, Mulcahy K A. An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data [J]. International Journal of Remote Sensing, 2002, 23(18): 3681–3696.
- [3] 饶品增, 蒋卫国, 贾凯, 等. 2013 年黑龙江省洪水对植被影响评估 [J]. 灾害学, 2017, 32(4): 195–201.
- [4] 许超, 蒋卫国, 万立冬, 等. 基于 MODIS 时间序列数据的洞

庭湖区洪水淹没频率研究 [J]. 灾害学, 2016, 31(1): 96–101.

- [5] 孔凡明, 蒋卫国, 李京, 等. 基于 MODIS 的 2011 年泰国洪涝受灾信息提取与分析 [J]. 灾害学, 2013, 28(2): 95–99.
- [6] Klein I, Gessner U, Dietz A, et al. Detection of inland water bodies with high temporal resolution – assessing dynamic threshold approaches [C]. Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE, 2016: 7647–7650.
- [7] 段秋亚, 孟令奎, 樊志伟, 等. GF-1 卫星影像水体信息提取方法的适用性研究 [J]. 国土资源遥感, 2015, 27(4): 79–84.
- [8] 郝家勇. 今年的洪水为啥来得猛来得快? [N]. 岳阳日报, 2017-07-09(003).
- [9] Salomon J, Hodges J C F, Friedl M, et al. Global land – water mask derived from MODIS Nadir BRDF – adjusted reflectances (NBAR) and the MODIS land cover algorithm [C]//Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International. IEEE, 2004.
- [10] Pham – Duc B, Prigent C, Aires F, et al. Comparisons of Global Terrestrial Surface Water Datasets over 15 Years [J]. Journal of Hydrometeorology, 2017, 18(4): 993–1007.
- [11] CHEN J, CHEN J, LIAO A, et al. Global land cover mapping at 30 m resolution: A POK – based operational approach [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 103: 7–27.
- [12] Breiman, Leo. Random forests [J]. Machine learning, 2001, 45(1): 5–32.
- [13] 邢秋菊, 赵纯勇, 高克昌, 等. 基于 GIS 的滑坡危险性逻辑回归评价研究 [J]. 地理与地理信息科学, 2004, 20(3): 49–51.
- [14] JIANG W, RAO P, CAO R, et al. Comparative evaluation of geological disaster susceptibility using multi – regression methods and spatial accuracy validation [J]. Journal of Geographical Sciences, 2017, 27(4): 439–462.
- [15] 和晓风. 基于 GF-1 遥感湿地类型提取研究 [D]. 长沙: 中南林业科技大学, 2017.
- [16] 张世平. 湖南省经济社会发展报告 [M]//宁吉喆, 中国发展报告, 北京, 中国统计出版社, 2016: 423–431.
- [17] 湖南省价格成本调查队. 湖南稻谷价格略降生猪价格小幅上涨——2017 年 11 月湖南价格指数点评 [J]. 价格理论与实践, 2017(12): 166.

## Flood Disaster Analysis Based on MODIS Data

### ——Taking the Flood in Dongting Lake Area in 2017 as an Example

RAO Pingzeng<sup>1,2</sup>, JIANG Weiguo<sup>1,2</sup>, WANG Xiaoya<sup>1,2</sup> and CHEN Kun<sup>3</sup>

(1. Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; 2. Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; 3. Chinese Research Academy of Environmental Sciences, Institute of soil and solid waste, Beijing 100012, China)

**Abstract:** Flood disasters have an important impact on human life and production. According to the flood water extraction process in this study, MOD09A1 data with high temporal resolution and highly efficient and stable random forest method were used to extract the flooded water in Dongting Lake District in 2017. What's more, the spatial distribution data of the crops in the flooded area was obtained, and Losses of food production are calculated for flooded area. The results show that: (1) Random forest classifier is better than the traditional exponential methods (NDVI, NDWI, and MNDWI) and the commonly used machine learning methods (logical regression and support vector machines) for water extraction; (2) the effect of flood submerged water extraction obtained by image overlay in more than 2 periods is better than that of single – stage image; (3) the flooded water area is 1, 641.11 km<sup>2</sup>, mainly located in the surrounding area of Dongting Lake; (4) the flooded crop area is 503.46 km<sup>2</sup>, and the crop loss is about 32.37 million tons and the economic loss caused by crops is about 1.444 billion Yuan.

**Key words:** flood disasters; Dongting Lake; MOD09A1; random forest; crop