



文献 DOI:

10.11922/csdata.2018.0076.zh

数据 DOI:

10.11922/sciencedb.681

文献分类: 地球科学

收稿日期: 2018-11-16

开放同评: 2018-12-21

录用日期: 2019-03-25

发表日期: 2019-09-24

1988–2017 年黄山自然与文化遗产地森林扰动数据集

张少宇^{1,2}, 朱岚巍^{1*}, 霍艳辉^{1,3}, 杨旭⁴, 唐世林⁵

1. 中国科学院遥感与数字地球研究所, 数字地球重点实验室, 北京 100094

2. 东华理工大学, 南昌 330013

3. 华北理工大学, 河北唐山 063210

4. 清华同方威视技术股份有限公司, 北京 100094

5. 中国科学院南海海洋研究所广东省海洋遥感重点实验室, 广州 510220

摘要: 基于 Google Earth Engine (GEE) 云计算平台, 利用 1988–2017 年共计 30 期 Landsat 5、Landsat 7、Landsat 8 数据进行在线批量处理, 结合森林扰动提取算法, 提取 30 年的每年森林扰动变化; 利用 GEE 云平台及 ArcGIS 10.5 软件进行结果统计及精细制图, 采用流程化的处理方式, 得到长时间序列的黄山自然与文化遗产地每年的森林扰动数据集。本数据集可用于黄山遗产地森林时空变化分析、遗产地的要素监测、遗产地的可持续发展等研究, 能够清晰地反映黄山自然遗产地在长时间序列监测作用下的变化特征。

关键词: 自然遗产地; 森林扰动; GEE; 黄山; Landsat

数据库(集)基本信息简介

数据库(集)名称	1988–2017 年黄山自然与文化遗产地森林扰动数据集
数据作者	张少宇、朱岚巍、霍艳辉、杨旭、唐世林
数据通信作者	朱岚巍 (zhulw@radi.ac.cn)
数据时间范围	1988–2017年
地理区域	森林扰动数据集的范围为黄山自然与文化遗产地保护区, 东经 118°01'–118°17'、北纬30°01'–30°18', 包括遗产地的官方划定范围, 以及黄山管理委员会规划范围, 其中规划范围包含核心区。
空间分辨率	30 m
数据量	8.19 MB
数据格式	*.shp, *.tif, *.lyr, *.jpg, *.mxd
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/681
基金项目	中国科学院A类先导专项(XDA19030500); 海南省面上自然科学基金(sy17zm01132); 国家发展和改革委员会促进大数据发展重大工程(2016-999999-65-01-000696-01)。

* 论文通信作者

朱岚巍: zhulw@radi.ac.cn

数据库（集）组成

本数据集主要包括 1988–2017 年长时间序列 30 年的黄山自然遗产地的森林扰动结果数据。这些数据保存为一个压缩文件（1988–2017 年黄山自然与文化遗产地森林扰动数据集.zip），其中包含 4 个文件夹（地理要素、源数据、符号系统、制图）。

引言

世界遗产主要包括自然遗产、文化遗产、混合遗产 3 种，是由联合国教科文组织和世界遗产委员会发起并确认的世界人民公认的、目前无法替代的财富，是全人类最具有代表性和最具有普遍价值的文物古迹及自然景观。其中自然遗产包含奇特的自然景观、完整的生态系统、原始的生物保护圈，突出体现了大自然的鬼斧神工、动植物的良好生活家园、生物进化的显著特征等。因此世界遗产逐渐受到人类的广泛重视，越来越多的国家和人民加入保护自然遗产、保护世界遗产的行列当中。

黄山在 1990 年被评为世界自然与文化遗产，在 2015 年被评为世界地质公园，在 2018 年被列入世界生物圈保护区，因此黄山在中国乃至世界上都占据非常高的地位。本文利用遥感空间技术，在自然遗产的研究课题下针对中国的自然与文化遗产地—黄山进行长时间序列的监测研究，并利用云计算平台及森林扰动提取算法得到 1988–2017 年 30 年的森林扰动数据产品，并进行数据共享。相关的研究成果作为自然遗产的研究成果的一部分，反映了自然遗产的历史及现在的生存状况，同时也可作为相关学者研究其他自然遗产地提供相关参考。

1 数据采集和处理方法

1.1 数据来源

本研究使用的云计算在线处理平台，具有高效、快速、方便等多个特点，解决了数据年份过多、研究区过大、数据下载繁琐等众多问题。在线获取的原始数据集是美国地质勘探局（<http://glovis.usgs.gov/>）提供的 Landsat 产品数据集。Landsat 具有较高的分辨率（30 m）、较长的时间跨度、开放式使用等众多特点^[1]，对长时间序列监测研究具有显著优势。区域尺度的植被变化监测研究，主要使用的是 Landsat 影像数据^[2-4]。通过人工预览选取 30 年中较好的数据产品（表 1），并进行相应的预处理得到可在线使用的 1988–2017 年的基础数据集。

表 1 黄山自然遗产地遥感影像数据列表

序号	成像日期	卫星	传感器	轨道号
1	1988-06-03	Landsat 5	TM	120-039
2	1989-11-29	Landsat 5	TM	120-039
3	1990-10-15	Landsat 5	TM	120-039
4	1991-11-03	Landsat 5	TM	120-039
5	1992-11-21	Landsat 5	TM	120-039
6	1993-12-26	Landsat 5	TM	120-039
7	1994-09-24	Landsat 5	TM	120-039
8	1995-10-13	Landsat 5	TM	120-039

序号	成像日期	卫星	传感器	轨道号
9	1996-04-22	Landsat 5	TM	120-039
10	1997-11-03	Landsat 5	TM	120-039
11	1998-04-28	Landsat 5	TM	120-039
12	1999-12-27	Landsat 5	TM	120-039
13	2000-10-10	Landsat 5	TM	120-039
14	2001-02-15	Landsat 5	TM	120-039
15	2002-09-30	Landsat 5	TM	120-039
16	2003-02-05	Landsat 5	TM	120-039
17	2004-04-20	Landsat 7	ETM	120-039
18	2005-03-06	Landsat 7	ETM	120-039
19	2006-05-20	Landsat 5	TM	120-039
20	2007-01-07	Landsat 7	ETM	120-039
21	2008-02-27	Landsat 7	ETM	120-039
22	2009-03-09	Landsat 5	TM	120-039
23	2010-03-28	Landsat 5	TM	120-039
24	2011-05-18	Landsat 5	TM	120-039
25	2012-04-26	Landsat 7	ETM	120-039
26	2013-04-07	Landsat 8	OLI	120-039
27	2014-10-25	Landsat 7	ETM	120-039
28	2015-02-06	Landsat 8	OLI	120-039
29	2016-03-28	Landsat 8	OLI	120-039
30	2017-10-09	Landsat 8	OLI	120-039

1.2 数据处理方法

近几十年来，许多遥感变化检测方法被提出、应用、对比分析和评价^[2]。同时也出现了很多监测森林扰动的指数，如 MODIS 全球扰动指数（MGDI）、基于缨帽变化的扰动指数（DI）、归一化湿度指数（NDMI）、综合森林特征指数（IFZ）、归一化植被指数（NDVI）、归一化燃烧比指数（NBR）等^[3]。但是每个指数具有不同的特点，针对不同的扰动类型具有不同的监测能力，同时众多指数对于不同的研究区也会出现一定的差异性。本文参考众多文献进行相关总结分析，选取 DI 指数作为研究方法的基础，并采取因地制宜的方式进行改进，对黄山自然遗产地进行监测。

根据森林扰动监测算法原理，基于森林扰动 DI 算法加以改进，DI 指数在根据对影像波段的进行缨帽变化后，提取其主要的分量进行数学运算，以获取主要的森林扰动信息，其改进后 DI 算法的主要原理如下^[5]：

$$\text{森林扰动指数 } DI = B_t - (G_t + W_t) \quad (1)$$

$$\text{标准化后的亮度 } B_t = (B - B_\mu) / B_\sigma \quad (2)$$

$$\text{标准化后的绿度 } G_t = (G - G_\mu) / G_\sigma \quad (3)$$

$$\text{标准化后的湿度 } W_t = (W - W_\mu) / W_\sigma \quad (4)$$

$$\text{森林扰动指数平均值 } DI_\mu = B_\mu - (G_\mu + W_\mu) \quad (5)$$

$$\text{扰动像元平均值 } DI_m = (DI_{\mu 1} + DI_{\mu 2} + \dots + DI_{\mu n}) / n \quad \text{其中 } 1 \cdots n \text{ 为年份} \quad (6)$$

式(1)中 B_t 、 G_t 、 W_t 分别为标准化后的亮度、绿度、和湿度。式(2)中 B 为缨帽变化后第一个分量波段代表亮度, B_μ 表示第一分量波段平均值, B_σ 为第一分量波段方差。式(3)、式(4)同理分别对应缨帽变化后第二分量和第三分量, 代表绿度和湿度以及其平均值和方差。式(5)中 DI_μ 代表森林扰动指数平均值, 根据其亮度平均值、绿度平均值以及湿度平均值计算而来; 式(6)中 DI_m 表示由多年份森林扰动指数平均值计算而来, n 为时间年份。

扰动像元以及恢复像元判别:

θ 为阈值, 阈值调整首先参考 DI_m 平均值的大小, 一般往大于 DI_m 的方向选择。连续三年 DI 值大于 θ , 即定义为扰动像元; 即 DI_n 、 DI_{n+1} 、 DI_{n+2} 均大于 θ , 则 DI_n 为第 n 年扰动值, 此时 DI_n 对应的该像元为扰动像元。 DI_n 、 DI_{n+1} 、 DI_{n+2} 中 DI_n 为扰动像元, DI_{n+1} 、 DI_{n+2} 均小于 θ , 则 DI_n 对应的该像元为恢复像元。本次实验通过计算 DI_m 值, 以及参考持久森林样本 DI 值, 综合考虑后本实验初次 θ 阈值选取为 4.0, 以步长为 0.5, 分别进行多次实验; 对检测结果的范围大小、变化程度进行比较, 经过多次试验后确定阈值。

本文的研究路线如图 1, 利用 Google Earth Engine 云计算在线处理平台, 先进行原始数据的大气校正、GEE 中 Landsat.simpleCloudScore 函数及 image.updateMask 函数进行去云掩膜; 然后进行监督分类、计算 NDVI 滤波; 然后结合 Landsat TOA 数据产品进行缨帽变换、扰动提取、结果导出等。以上工作大部分可依靠编程完成。在数据平台导出数据后, 然后通过 ArcGIS 软件进行处理, 同时利用 2017 年哨兵数据、Google Earth 高分辨率影像、实地验证进行结果分析和结果验证, 最终生成 1988–2017 年共 30 年逐年森林扰动产品数据集。

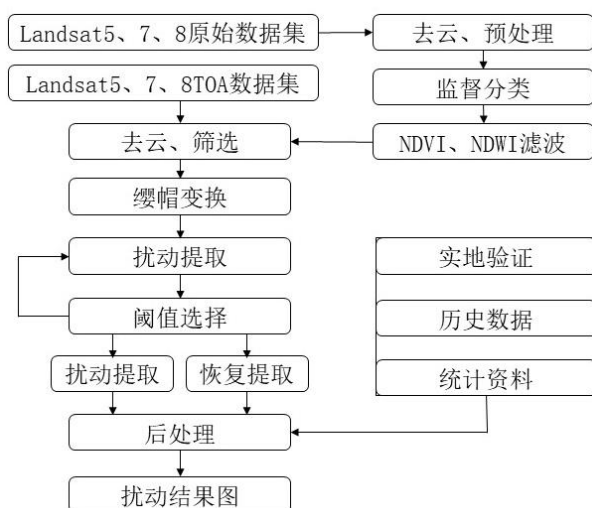


图 1 森林扰动研究技术路线

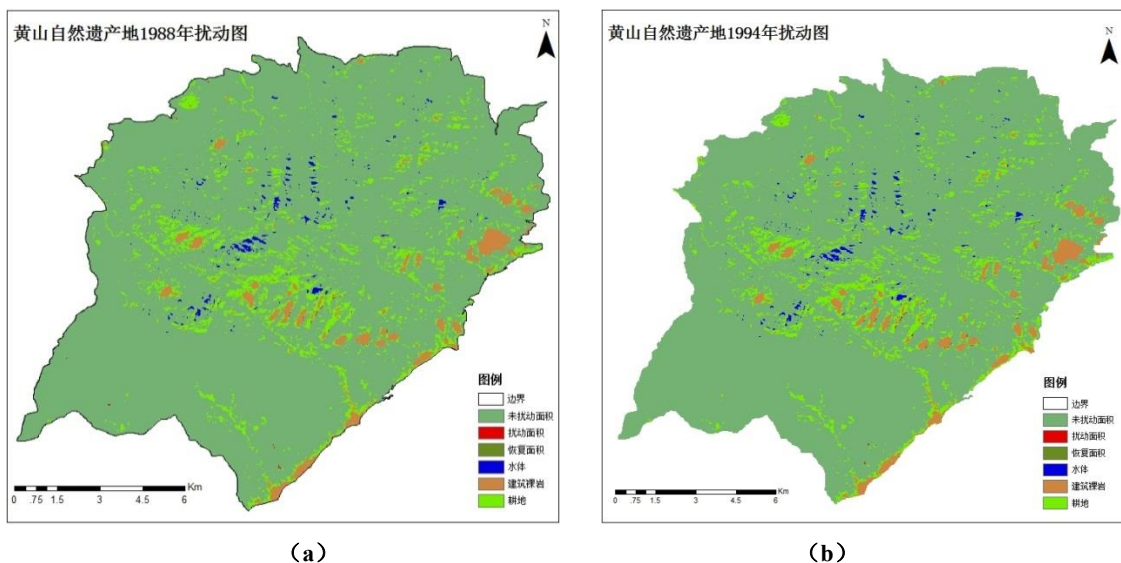
2 数据样本描述

2.1 数据组成

本数据集主要包括 1988–2017 年长时间序列 30 年的黄山自然遗产地的森林扰动结果数据, 这些数据保存为一个压缩文件。压缩文件中包含 4 个文件夹(地理要素、源数据、符号系统、制图), 总数据量为 8.19 MB。地理要素文件夹中包含建筑、水体、边界、索道、公路、步道对应的矢量数据, 文件为 shp 格式, 可直接进行查看编辑; 此边界为相关课题边界研究人员参考多方资料确定的黄山自然遗产地的边界范围; 其他要素为本研究工作人员搜集数据, 结合地形图, 规划图矢量化而成。源数据文件夹中每个文件夹单独存储每一年的对应的 tif 栅格数据文件, 共有 6 个属性值, 分别为未扰动森林、扰动森林、恢复森林、建筑裸岩、耕地和水体, 采用的 UTM-WGS84 坐标系; 符号系统文件夹中为 ArcGIS 软件中属性值中的符号系统, 此为模板方便赋值使用, 格式为 lyr 可直接使用或编辑; mxd 格式文件为 ArcGIS 软件格式文件, 包含所有的源文件信息可直观进行查看; 制图文件夹中为 1988–2017 年 30 年的制图成果, 格式为 jpg。

2.2 数据样本

图 2 为黄山自然遗产地 1988–2017 年中 6 期的森林扰动监测产品。通过 GEE 平台中的监督分类的方式, 选取足够多的样本进行分类, 主要分为森林、水体、建筑、裸岩、耕地; 然后通过 NDVI 滤波再次筛选森林像元。森林扰动信息的提取是对森林像元做变化监测, 根据连续 3 年的处理像元值的比较来判断是否为扰动像元或恢复像元。使用的方法是 GEE 平台中 Join 函数方法, 查看当前年份前后年份的像元值是否大于 0, 只要存在连续 3 年大于 0, 均定义为扰动像元。例如, 起始年份 1988 年只查看当前年份及后一年数据即 1989 年像元值是否大于 0, 以及终止年份只查看当前年份及左年份即 2016 年像元值是否大于 0, 以此定义是否为扰动像元。森林扰动点位主要集中在旅游路线的主干道两侧、山顶的宾馆附近、水系附近、较高海拔、以及森林与裸岩的交界处。主要的扰动类型为森林–裸岩、森林–建筑、森林–土地以及植被的自然变迁。例如, 2006 年黄山自然遗产地的保护面积为 160.6 平方公里, 缓冲区面积为 490 平方公里, 2006 年的森林扰动面积接近 7.6 万平方米, 恢复面积为 6.8 万平方米, 扰动面积占据总覆盖面积的 0.0473%, 恢复面积占据总覆盖面积的 0.0423%。



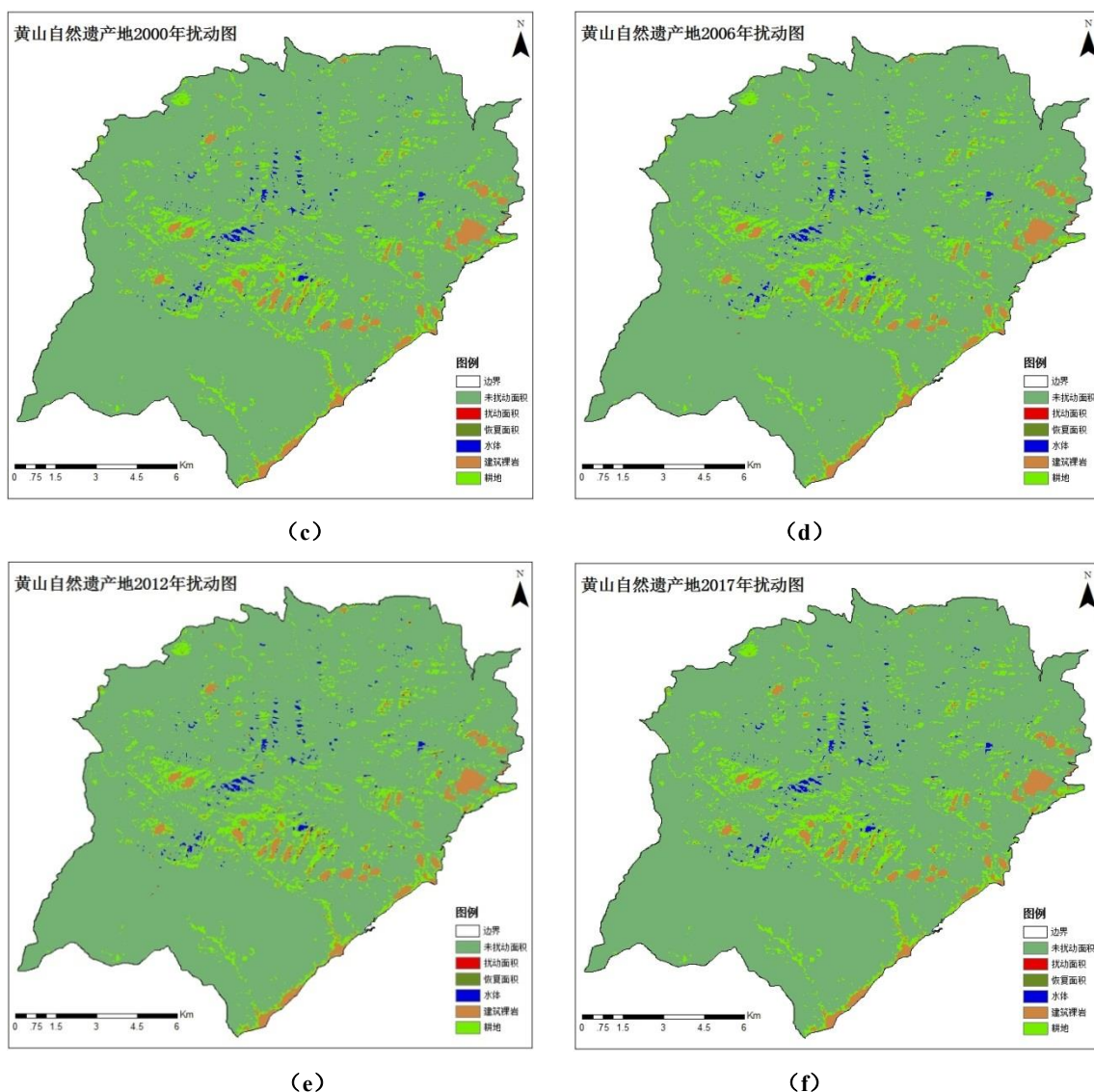


图2 黄山自然与文化遗产地 6 期森林扰动图

3 数据质量控制和评估

3.1 数据结果分析

根据结果统计可得 1988–2017 年 30 年中每年扰动面积和恢复面积见图 3。根据 30 年的森林扰动趋势可看出，早期及 2011 年前后具有大面积的扰动，2008 年前后出现大面积扰动，随后未发生较好的恢复。最近 10 年虽一直频繁出现不同程度的扰动，但整体占据比例很小，仍在自然变化的可接受范围之内。在每次扰动之后均有一定幅度的面积恢复，分析可知主要以森林自然变化为主。选取每年的具体扰动点位进行比较来看，大多扰动点位出现在低等植物与高等植物、裸岩交界之处；由于数据质量的限制，不得不采用每年份相对较好的影像，因此就满足不了同一月份进行比较；同时由于黄山季节变化也会出现很大的差异，数据选取问题给提取精度带来影响，造成一定程度的扰动信息误提取，导致提取精度一般；但鉴于此方法高效、快捷、可同时进行长时间序列分析，因此具有一定的普遍适用性和可扩展性，可优选其使用数据并用于其他地区研究。

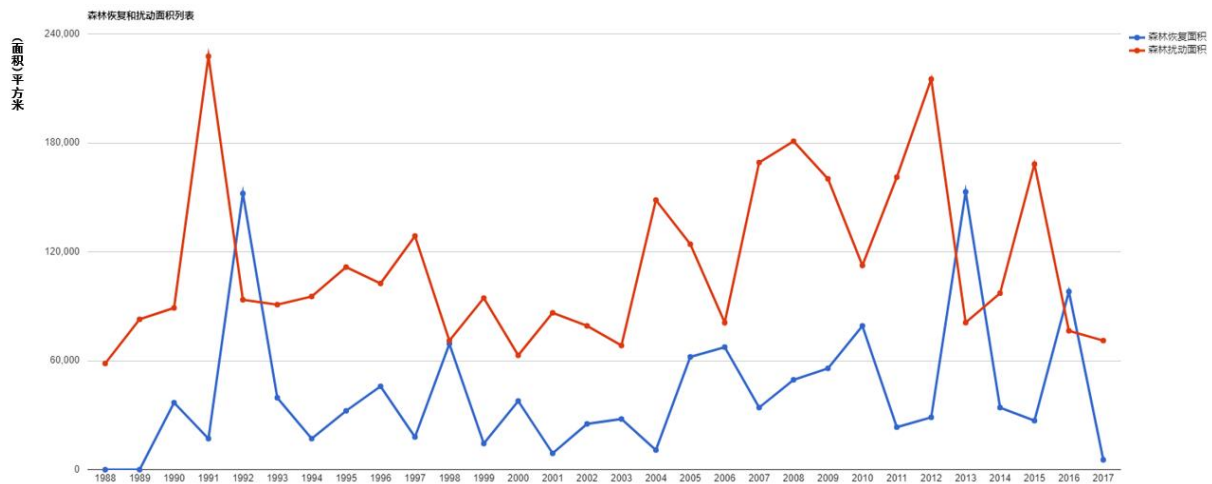


图 3 森林恢复和扰动面积列表

3.2 精度验证方式

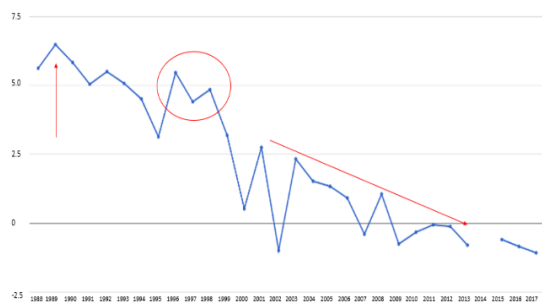
采用辅助数据，2017 年哨兵数据、Landsat 原始数据、高分影像 Google Earth 数据进行目视解译。采用目视解译的方法，根据随机分层抽样原理^[6]；由于 Google 高分影像限制，选取 2008–2017 年共 10 年每年的扰动点，每层随机抽样 30 个样本；未扰动森林、非森林每层随机抽取样本 100 个，共 500 个像元点。根据精度评价混淆矩阵，计算制图精度、用户精度、总体精度、误检率、漏检率、Kappa 系数加以评价。总体精度为 80.66%，Kappa 系数为 0.77，见表 2。

表 2 精度混淆矩阵及评价表

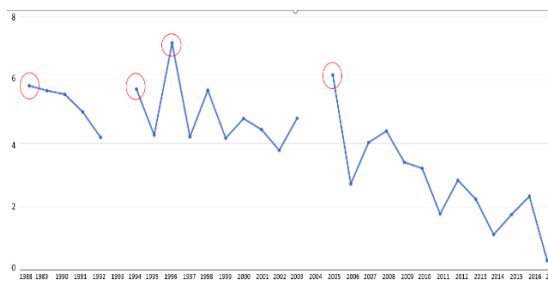
预测值 真实值	持续森林	非森林	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	总计	用户精度	误检率%
持续森林	94	6											100.0	94.00	6.00
非森林	3	97											100.0	97.00	3.00
2008	3	2	25										30.00	83.33	16.67
2009	8			22									30.00	73.33	26.67
2010	1	1	1		26				1				30.00	86.66	13.34
2011	8	1				21							30.00	70.00	30.00
2012	5	1			1		22	1					30.00	73.33	26.67
2013	6	1						22	1				30.00	73.33	26.67
2014	5								25				30.00	83.33	16.67
2015	1							1		28			30.00	93.33	6.67
2016		1									29		30.00	96.67	3.33
2017	1								1	3	3	22	30.00	73.33	26.67
总计	135	110	26	22	27	21	22	24	28	31	32	22	500.0		
制图精度	69.63	88.18	96.15	100.0	96.29	100.0	100.0	91.67	89.29	90.32	90.63	100.0		总体	kappa

预测值 真实值	持续森林	非森林	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	总计	用户精度	误检率%
														%	
漏检率%	30.37	11.82	3.80	0.00	3.70	0.00	0.00	2.90	1.71	9.70	9.40	0.00		80.66	0.77

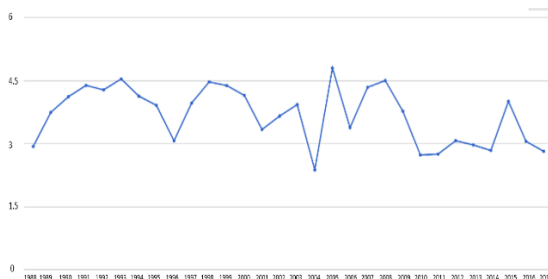
通过实地验证的方式，共选取 4 种典型样本点（图 4），验证点位分别为 a（118°09'42.27"E、30°04'00.78"N），b（118°10'38.27"E、30°4'30.89"N），c（118°14'15.68"E、30°8'32.67"N），d（118°9'32.237"E、30°8'36.161"N）。图 4a 为早期扰动，现已经完全恢复，森林密集程度走势较好，左侧为该样本点 DI 值的走势图，由图可知当地森林覆盖良好；图 4b 为早期多次扰动，近几年逐渐恢复，由实地情况可知为较新树种覆盖良好。图 4c 为持续未扰动森林，一直处于较好的状态，由图可知此样本相比图 4a、4b 两样本明显有更好的覆盖效果。图 4d 为近几年发生扰动但没有完全恢复样本，经过现场查看可知此处多为植被生长变化造成的自然扰动，由于是裸岩与植被交界处，受到季节植被变化影响，容易出现监测扰动值持续较高，由此可见此处的覆盖度较差。



(a)



(b)



(c)

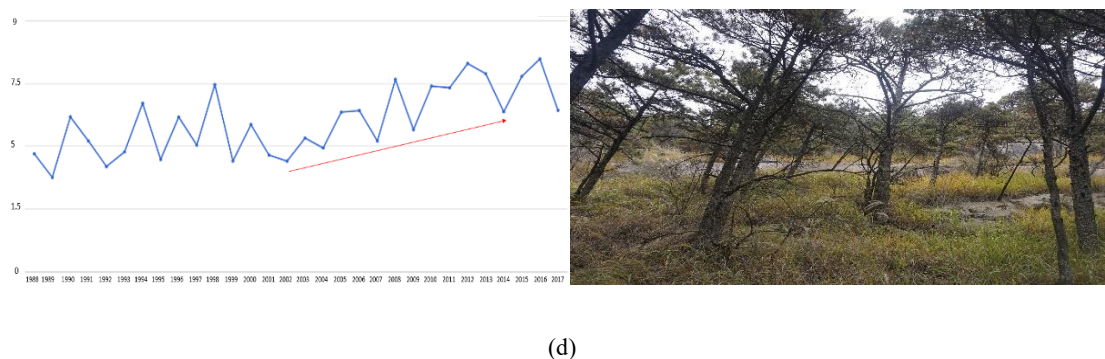


图 4 4 种典型样本点

4 数据使用方法和建议

数据使用的 tif 格式、shp 格式的文件、lyr 格式文件、mxd 格式文件，便于研究工作者查看和进行编辑。本文采用 ArcGIS 软件进行编辑绘制，同时数据集中也包含了本文所设置的图层模板，便于更好的查看和制图。本数据集具有很强的代表性，可以为众多研究学者研究森林变化、植被变化、自然遗产地监测提供一定程度的参考。同时本研究数据集可直接用于黄山遗产地保护、规划、可持续发展等研究。

致 谢

非常有幸能够参与到自然遗产的保护研究工作当中，能够有机会致力于遗产的保护和发展；同时感谢杨瑞霞副研究员在验证工作中给予指导，感谢杜文杰、王普、陈亚亚等同学在研究过程中提供了很多帮助。

数据作者分工职责

张少宇（1994—），男，河南省驻马店人，硕士研究生，研究方向为“一带一路”自然遗产地时空变化遥感监测。主要承担工作：处理程序的编写、数据处理工作、结果分析、数据制图等工作。

朱岚巍（1980—），女，内蒙古自治区赤峰人，副研究员，研究方向为数字自然遗产研究、微波遥感研究、遥感应用研究等。主要承担工作：实地验证、数据下载、统计资料、相关资料的收集等工作。

霍艳辉（1994—），男，河北省邢台市人，硕士研究生，研究方向为“一带一路”自然要素遥感监测、多源数据珊瑚礁提取研究等。主要承担工作：验证思路设计、内业验证、实地验证、数据整理等工作。

杨旭（1978—），男，辽宁营口人，程序设计及算法研究，研究方向遥感国土资源监测。主要承担工作：程序设计，方案设计，数据分析等。

唐世林（1981—），男，湖南邵阳市人，博士研究生，研究方向为海洋资源遥感。主要承担工作：野外数据验证辅助工作，野外验证方案设计等。

参考文献

[1] 汤冬梅, 樊辉, 张瑶. Landsat 时序变化检测综述[J]. 地球信息科学学报, 2017, 19(08): 1069-1079.

- [2] HEALEY S P , COHEN W B , SPIES T A , et al. The Relative Impact of Harvest and Fire upon Landscape-Level Dynamics of Older Forests: Lessons from the Northwest Forest Plan[J]. *Ecosystems*, 2008, 11(7): 1106-1119.
- [3] 杨辰, 沈润平. 森林扰动遥感监测研究进展[J]. *国土资源遥感*, 2015, 27(01): 1-8.
- [4] 张连华, 庞勇, 岳彩荣, 等. 基于缨帽变换的景洪市时间序列 Landsat 影像森林扰动自动识别方法研究[J]. *林业调查规划*, 2013, 38(02):6-12, 19.
- [5] HEALEY S P , COHEN W B , YANG Z Q, et al. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection[J]. *Remote Sensing of Environment*, 2005, 97(3): 301-310.
- [6] 张宇欣, 倪静, 杨存建, 等. 基于扰动指数的岷江上游森林扰动时间序列研究——以理县为例[J]. *浙江林业科技*, 2017, 37(03): 48-53.

论文引用格式

张少宇, 朱岚巍, 霍艳辉. 1988–2017 年黄山自然与文化遗产地森林扰动数据集[J/OL]. *中国科学数据*, 2019, 4(3). (2019-04-23). DOI: 10.11922/csdata.2018.0076.zh.

数据引用格式

张少宇, 朱岚巍, 霍艳辉. 1988–2017 年黄山自然与文化遗产地森林扰动数据集[DB/OL]. *Science Data Bank*, 2018. (2018-11-16). DOI: 10.11922/sciencedb.681.

A dataset of forest disturbance on the site of Huangshan Natural and Cultural Heritage, 1988 – 2017

Zhang Shaoyu^{1,2}, Zhu Lanwei^{1*}, Huo Yanhui^{1,3}, Yang Xu⁴, Tang Shilin⁵

1. Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, P. R. China

2. East China University of Technology, Nanchang 330013, P. R. China

3. North China University of Science and Technology, Tangshan 063210, P. R. China

4. Tsinghua Nuctech Company Limited, Beijing 100094, P.R. China

5. Guangdong Key Laboratory of Ocean Remote Sensing, South China Sea Institute of Oceanology Chinese Academy of Sciences, Guangzhou 510220, P.R. China

*Email: zhulw@radi.ac.cn

Abstract: On the cloud computing platform Google Earth Engine (GEE), we performed batch processing on Landsat 5, Landsat 7, Landsat 8 data for the Huangshan Natural and Cultural Heritage site over the past 30 years in 1988 – 2017, and used forest disturbance extraction algorithm to extract annual variations of forest disturbance during this period. Then GEE cloud platform and ArcGIS10.5 software were used for fine mapping and result statistics. This dataset can be used for analyzing the temporal and spatial changes of

forests, monitoring the heritage elements and studying the sustainable development of the Huangshan heritage site. It clearly reflects the changing characteristics of the site under long-term sequence monitoring.

Keywords: natural heritage; forest disturbance; GEE; Huangshan; Landsat

Dataset Profile

Title	A dataset of forest disturbance on the site of Huangshan Natural and Cultural Heritage, 1988 – 2017
Data corresponding author	Zhu Lanwei (zhulw@radi.ac.cn)
Data authors	Zhang Shaoyu, Zhu Lanwei, Huo Yanhui, Yang Xu, Tang Shilin
Time range	1988 – 2017
Geographical scope	The forest disturbance dataset covers the scope of the Huangshan Natural and Cultural Heritage Site (E 118°01'—118°17', N 30°01'—30°18'). It includes the officially delineated area of the site and the area under planning by the Huangshan Management Committee, the latter of which designates the core area.
Spatial resolution	30 m
Data volume	8.19 MB
Data format	*.shp, *.tif, *.lyr, *.jpg, *.mxd
Data service system	< http://www.sciencedb.cn/dataSet/handle/681 >
Sources of funding	Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19030500); General Programs of Natural Science Foundation of Hainan Province (sy17zm01132); Major Program for Big Data Development of the National Development and Reform Commission (2016-999999-65-01-000696-01).
Dataset composition	This dataset mainly includes data of the forest disturbance results across long-term sequences of Huangshan Natural Heritage Site from 1988 to 2017. The dataset is saved as a compressed file (A dataset of forest disturbance on the site of Huangshan Natural and Cultural Heritage_1988–2017.zip), comprised of four folders (namely, geographical element, source data, symbol system, mapping).