

## 基于遥感数据的太湖蓝藻水华信息识别方法\*

陈云<sup>1,2</sup>, 戴锦芳<sup>1</sup>

(1: 中国科学院南京地理与湖泊研究所湖泊与环境国家重点实验室, 南京 210008)

(2: 中国科学院研究生院, 北京 100049)

**摘要:** 2007年太湖大规模蓝藻暴发, 再次引起了人们对太湖环境的关注. 有效地提取蓝藻水华信息对分析蓝藻动态分布有重要意义, 而卫星遥感技术是进行太湖水质监测与保护的措施之一. 本文以2007年4月23日CBERS-02星CCD数据为主要的数据库, 以NDVI值为测试变量, 运用CART算法确定分割阈值, 从而通过构建决策树的方法识别蓝藻水华信息, 分析其蓝藻水华的提取结果, 取得了较好的效果. 文中还在GIS技术支持下, 提取了2007年5月17日MODIS影像中的蓝藻水华信息. 本次研究为以后开展长期的太湖蓝藻水华动态监测提供技术参考.

**关键词:** 太湖; 蓝藻水华; NDVI; 决策树

### Extraction methods of cyanobacteria bloom in Lake Taihu based on RS data

CHEN Yun<sup>1,2</sup> & DAI Jinfang<sup>1</sup>

(1: *State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P.R. China*)

(2: *Graduate School of Chinese Academy of Sciences, Beijing 100049, P.R. China*)

**Abstract:** This year blue algae blooms occurred largely in Lake Taihu, which caused people pay close attention to environment of Lake Taihu once again. It is very important to extract the information of cyanobacteria bloom-forming to analyze the distribution of cyanobacteria bloom and remote sensing is one of measures of monitoring and protecting water quality. In this paper, it takes the CBERS-02 CCD data as a main data source. The information of cyanobacteria bloom was extracted by the method of building decision tree, and with NDVI as the test variable, threshold value was determined based on CART. It obtains better effects according to analyzing the extraction result. The information of cyanobacteria bloom is also extracted from MODIS image of May 17th, 2007. The present study provides technical reference for long-term monitoring cyanobacteria bloom-forming in Lake Taihu.

**Keywords:** Lake Taihu; cyanobacteria bloom; NDVI; decision tree

太湖位于我国长江三角洲的核心区域, 湖泊水面面积为2338.1km<sup>2</sup>, 平均水深1.2m, 是我国第三大淡水湖. 太湖流域的主要行政单元分属江苏、浙江、上海二省一市. 流域内除包含上海市全境外, 还分布有杭州、苏州、无锡、常州、嘉兴、湖州等著名城市. 历史上, 太湖不仅在灌溉农田、水产生产、沟通航运方面发挥了重要作用, 而且是上海、无锡、苏州等周边城市的重要水源地. 因此, 保护好太湖的水资源尤为迫切与重要. 自20世纪90年代以来, 太湖的水环境形势越加严峻, 湖泊水体中氮、磷等营养物质含量过高, 加之夏秋季节湖水温度升高, 引起了蓝藻水华暴发, 给太湖地区的经济建设和人民生活带来了巨大的损失, 2007年4、5月间的太湖蓝藻暴发事件, 更加引起了人们对太湖水环境安全问题的高度关注.

蓝藻暴发时绿色的藻类生物体聚集于水体表面, 且藻类细胞在光合作用下快速繁殖, 藻类与陆地植被的光谱特征基本上类似, 蓝藻覆盖区的光谱特征则与湖泊水面有明显的差异<sup>[1-2]</sup>. 这种光谱特征被卫星

\* 江苏省测绘科研项目(JSCHKY200704)资助. 2007-09-05 收稿; 2007-11-06 收修改稿. 陈云, 女, 1984 生, 硕士研究生; E-mail: chen013@163.com.



的探测器记录下来,通过对卫星遥感数据的分析、处理,能够快速地获取湖区蓝藻水华暴发的信息,对分析蓝藻暴发的动态过程及其分布状况有重要意义<sup>[3]</sup>.

近年来,一些学者在利用遥感技术分析蓝藻事件方面已进行了有益的探索.杨顶田等分析了利用遥感监测蓝藻的可行性,并对未来的研究趋势作了展望<sup>[4]</sup>.黄家柱等应用 TM 数据监测了 1998 年 8 月太湖水域的蓝藻暴发分布变化<sup>[5]</sup>,李旭文等利用图像处理技术,分析太湖梅梁湾湖区藻类暴发的原因<sup>[6]</sup>.本文以 2007 年 4 月和 5 月间太湖蓝藻大暴发为背景,选取太湖西北湖区(主要包括竺山湖和梅梁湾)为研究区域,以 CBERS-02 星 CCD 数据为主要信息源,结合使用 MODIS 数据,探讨了蓝藻水华的卫星遥感识别方法,旨在为以后开展长期的太湖蓝藻水华动态监测提供技术参考.

## 1 数据资料及预处理

### 1.1 数据资料

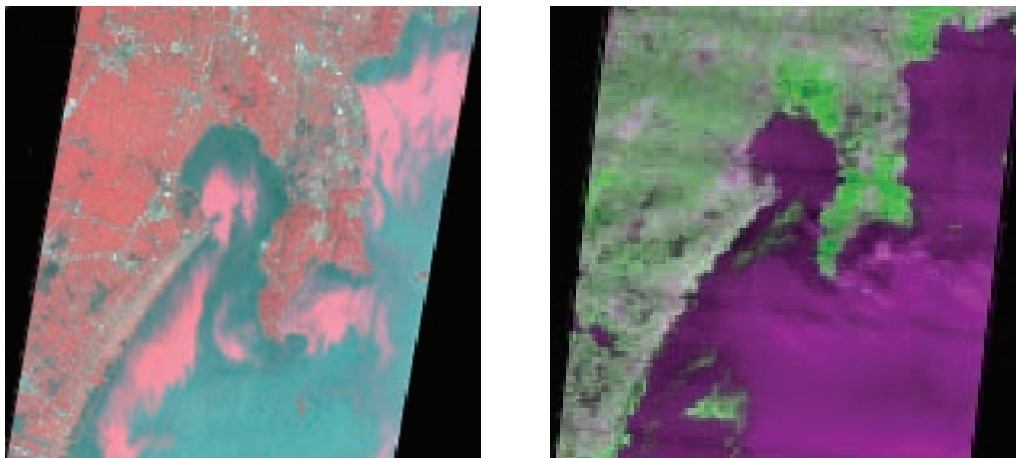
本研究应用 CBERS-02(中巴资源卫星 02 星)的 CCD 多光谱数据和美国的 EOS/MODIS(中分辨率成像光谱辐射计)数据. CBERS 是我国与巴西合作研制的传输型对地观测资源遥感卫星,携带 CCD 相机、红外多光谱扫描仪和宽视场成像仪三种遥感器,具有 11 个光谱波段,空间分辨率为 19.5–258m,适宜对多种目标的识别. MODIS 是搭载于美国 EOS 系列卫星之上的一个重要的遥感传感器,具有 36 个光谱波段,空间分辨率为 250–1000m,每 1–2d 观测地球一次. MODIS 数据时间分辨率高,有利于捕捉地面动态变化的信息<sup>[7]</sup>.

本研究中选用 2007 年 4 月 23 日的 CCD 多光谱数据,包括蓝、绿、红(波长分别为 0.45–0.52 $\mu\text{m}$ 、0.52–0.59 $\mu\text{m}$ 、0.63–0.69 $\mu\text{m}$ )三个可见光波段和一个近红外(波长为 0.77–0.89 $\mu\text{m}$ )波段,空间分辨率为 19.5m.选用的 MODIS 影像的空间分辨率为 250m,共两个波段(光谱范围分别为 0.62–0.67 $\mu\text{m}$  和 0.841–0.876 $\mu\text{m}$ )<sup>[8]</sup>,成像时间是 2007 年 5 月 17 日.

此外选用 1:100000 的太湖地区地形图作为影像几何纠正的依据.

### 1.2 数据预处理

MODIS 数据存在着 Bow-tie 现象,表现为相邻两个扫描行之间有部分数据相同,无法用几何校正方法去除这种现象,因此必须在几何校正之前加以处理.参照太湖地区 1:100000 地形图对 MODIS 影像和 CBERS 影像,进行了几何精校正,并对研究区影像进行截取(图 1).



4 月 23 日的 CBERS CCD 多光谱数据(4、3、2 合成)

5 月 17 日的 MODIS 影像图(1、2、1 合成)

图 1 2007 年太湖西北湖区的遥感影像

Fig.1 The remote sensing images of northwest area in Lake Taihu, 2007

## 2 研究方法

### 2.1 基于 CBERS CCD 数据的蓝藻水华识别方法研究

2.1.1 归一化差异植被指数 NDVI 的计算 NDVI(Normalized Difference Vegetation Index)是目前应用效果较好的植被指数方法之一<sup>[9-10]</sup>, 蓝藻水华的光谱特征与植被的光谱特征相似, 所以根据 NDVI 能够有效地提取蓝藻水华的相关信息. NDVI 被定义为红光波段和近红外波段两波段的归一化比值<sup>[11]</sup>. 即,

$$NDVI = (RED - NIR) / (RED + NIR) \quad (1)$$

式中: NIR 和 RED 分别代表着近红外波段和红光波段的反射率, 对 CBERS-02 星的 CCD 影像而言, 分别代表着第 4 和第 3 波段. 通过对 4 月 23 日的太湖地区 CBERS 数据的 NDVI 计算, 可得到 NDVI 植被指数(图 2).

2.1.2 蓝藻水华光谱特征分析及训练样本的选取 分析研究区内地物的光谱特征, 在 4、3、2 波段合成的假彩色合成影像中, 蓝藻富集区呈绯红色, 与周围蓝色湖泊水面有明显的区别, 蓝藻聚集区中蓝藻量较少, 在合成影像中表现为淡粉色. 在 NDVI 图像中, 蓝藻富集区的 NDVI 值较大, 表现为亮白色, 蓝藻聚集区的 NDVI 的值相对较小, 呈现为灰白色, 与周围黑色、灰色水体有明显的差别. 结合 4、3、2 波段合成的影像和 NDVI 图像, 将湖区内地物分为三类: 蓝藻富集区、蓝藻聚集区和湖泊水面, 并分别进行采样, 各获取 250 个训练样本.

2.1.3 构建决策树 在遥感分类中, 决策树是制定分类规则的有效方法之一, 它是一个类似流程图的树型结构, 树的最高层节点为根节点, 是所有数据的集合; 树的每个内部节点代表着对某一属性的最佳测试, 其分支代表测试结果; 树的每个叶节点代表分类类别, 构建决策树的关键是确定叶节点测试变量和分割阈值<sup>[12-13]</sup>. 本文采用 Breiman 1984 年提出的一种决策树构建算法——CART(Classification and regression Tree)算法, CART 算法采用经济学中的基尼系数(Gini Index)作为选择最佳测试变量和分割阈值的准则<sup>[14]</sup>, 其数学定义如下:

$$GiniIndex = 1 - \sum_j p^2(j/h),$$

其中,  $P(j/h) = \frac{n_j(h)}{n(h)}$ ,  $\sum_j p(j/h) = 1$  (2)

式中,  $P(j/h)$ 是从训练样本集中随机抽取一个样本, 当某一测试变量值为  $h$  时属于第  $j$  类的概率;  $n_j(h)$ 为训练样本中测试变量值为  $h$  时属于第  $j$  类的样本个数;  $n(h)$ 为训练样本中测试变量值为  $h$  的样本个数;  $j$  为类别个数<sup>[15]</sup>. CART 算法具有如下的特点: 严格无参数, 对输入数据没有任何统计分布的假设要求; 能够清楚地指出变量对于分类的重要性, 选择出与分类相关的变量, 方法实现简单, 运行速度快; 结构清晰, 容易理解等<sup>[16]</sup>.

基于 4 月 23 日的 CBERS CCD 影像, 首先对太湖水域以

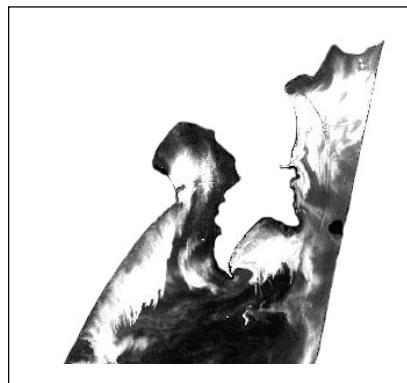


图 2 2007 年 4 月 23 日的 NDVI 植被指数  
Fig.2 The NDVI of April 23, 2007

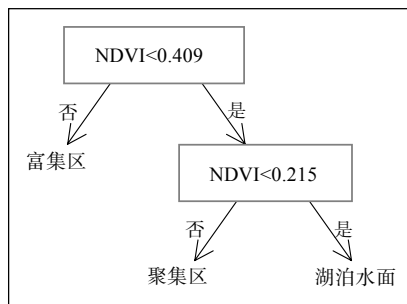


图 3 基于 CART 算法构建的决策树  
Fig.3 The decision tree based on CART

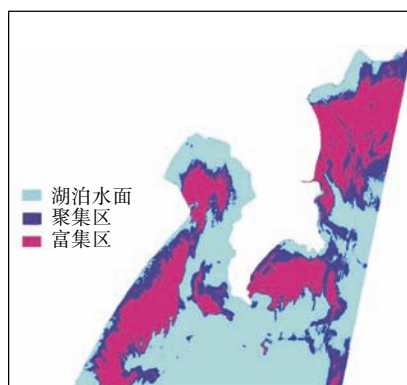


图 4 用决策树提取的蓝藻水华分布图  
(根据 2007 年 4 月 23 日太湖西北部的  
CBERS 数据提取)

Fig.4 The extraction result of cyanobacteria bloom-forming by decision tree

外的影像进行掩膜。由以上分析可知, NDVI 可以区分影像中三类地物, 选取 NDVI 为测试变量, 根据所选取的训练样本, 采用 CART 算法进行计算分析, 确定了 0.2152 和 0.4098 为分割阈值, 得到决策树(图 3), 当  $NDVI > 0.4098$  时, 判属于蓝藻富集区, 当  $0.2152 < NDVI \leq 0.4098$  时, 判属于蓝藻聚集区, 当  $NDVI \leq 0.2152$  时, 则判属于湖泊水面。将决策树规则运用到图像分类中, 提取蓝藻水华分布信息(图 4), 图中红色部分为蓝藻富集区, 紫色部分为蓝藻聚集区, 蓝色部分则为湖泊水面。

2.1.4 精度评价 通过叠置分析, 认为蓝藻水华分布图与原始影像有很好一致性。在实验区内随机选取检验样本(107 个), 建立混淆矩阵进行精度检验, 计算其生产者精度、用户精度以及 kappa 系数(表 1)。对分类结果进行精度评价, 其中蓝藻富集区分类精度为 90.48%, 蓝藻聚集区分类精度为 83.78%。验证表明, 根据 NDVI 值, 运用 CART 算法计算确定分割阈值, 能有效地提取蓝藻水华分布信息。

表 1 基于决策树的蓝藻分类精度评价

Tab.1 Accuracy assessment on the classification based on decision tree

| 类型       | 蓝藻富集区 | 蓝藻聚集区 | 湖泊水面  | 总和  | 使用精度(%) |
|----------|-------|-------|-------|-----|---------|
| 蓝藻富集区    | 38    | 1     | 0     | 39  | 97.44   |
| 蓝藻聚集区    | 4     | 31    | 2     | 37  | 83.78   |
| 湖泊水面     | 0     | 5     | 39    | 44  | 88.64   |
| 总和       | 42    | 37    | 41    | 120 |         |
| 生产者精度(%) | 90.48 | 83.78 | 95.12 |     |         |
| 用户精度(%)  |       |       |       | 90  |         |
| Kappa    |       |       |       |     | 0.7747  |

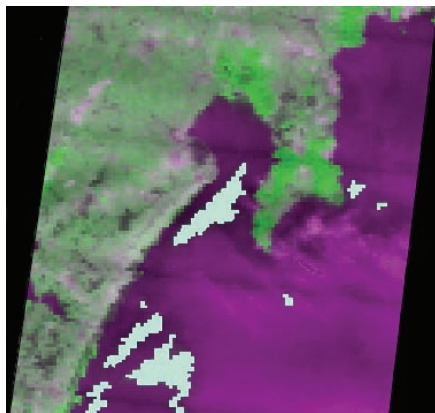


图 5 5 月 17 日的蓝藻水华提取结果

Fig.5 The extraction result of cyanobacteria bloom-forming on May 17

## 2.2 基于 MODIS 数据的蓝藻水华区识别

依据太湖及其周边地理环境在遥感影像的波谱特征、时相特征、空间特征等成像机理以及所掌握的各种地学规律, 可以获得蓝藻水华的识别特征<sup>[17-18]</sup>。在按 1、2、1 波段合成方案合成的 MODIS 卫星图像上, 湖泊水面呈紫色, 蓝藻水华为绿色, 其光谱特征与湖泊水面有明显的差异, 通过不同时相的遥感图像对比分析和地理相关分析, 能够较容易地提取太湖蓝藻水华区的相关信息。

本文在 GIS 技术的支持下<sup>[19]</sup>, 根据蓝藻水华的光谱特征, 5 月 17 日两个时相 MODIS 影像进行识别处理, 确定了蓝藻水华分布范围, 蓝藻提取结果图中的淡蓝色部分为蓝藻水华区(图 5)。

## 3 结论

基于 CBERS CCD 和 MODIS 遥感影像特征的分析研究, 采用相应的方法提取影像中的蓝藻水华信息, 可有效地捕捉藻类的分布特征。CBERS CCD 影像光谱信息丰富, NDVI 是应用效果较好的植被指数之一, 常用于陆地植被的提取, 藻类与陆地植被的光谱特征相似, 所以可以根据 NDVI 提取蓝藻水华区信息。试验也表明: 选取 NDVI 值为测试变量, 利用 CART 算法分析计算分割阈值, 构建决策树有效地把蓝藻水华富集区、蓝藻聚集区和水体区分开来。文中所使用 MODIS 影像只有两个波段, 且空间分辨率较低, 所以在 GIS 技术的支持下, 就能够容易地提取蓝藻水华区的信息。

本文中主要研究了从多源遥感影像中提取蓝藻水华区信息的方法, 从蓝藻提取的结果图中可以直观地看出不同时期蓝藻水华分布变化, 为以后进行长时期的太湖蓝藻水华动态监测提供技术参考。

#### 4 参考文献

- [1] 秦伯强, 胡维平. 太湖水环境演化过程与机理. 北京: 科学出版社, 2004: 1-8.
- [2] 王海君, 李云梅. 用 ASTER 数据监测梅梁湖湖区蓝藻分布的方法. 南京师范大学报, 2005, **28**(1): 103-106.
- [3] Vincent RK, Qin XM. Phycocyanin detection from Landsat TM Data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment*, 2004, **89**: 381-392.
- [4] 杨顶田, 潘德炉. 蓝藻卫星遥感研究进展. 国土资源遥感, 2006, (4): 1-5.
- [5] 黄家柱, 赵锐. 卫星遥感监测太湖水域蓝藻暴发. 遥感信息, 1999, **10**(4): 43-44.
- [6] 李旭文. 太湖梅梁湖湾生物量的遥感估算. 国土资源遥感, 1995, (2): 23-28.
- [7] 唐俊梅, 张树文. 基于 MODIS 数据的宏观土地利用/土地覆盖监测研究. 遥感技术与应用, 2002, **17**(2): 104-107.
- [8] 黄家浩, 万幼川. MODIS 的特性及其应用. 地理空间科学, 2003, **2**(4): 31-37.
- [9] 郑元润, 周广胜. 基于 NDVI 的中国天然森林植被净第一性生产力模型. 植物生态学报, 2000, **24**(1): 9-12.
- [10] Huete A, Didan K, Miura T. Overview of the radiometric and biophysical performance of the MODIS vegetation indexes. *Remote Sensing of Environment*, 2002, (83): 195-213.
- [11] Liu HQ, Huete A. A feedback based modification of the NDVI to Minimize Canopy Background and Atmosphere Noise. *IEEE Trans Geosci Remote Sensing*, 1995, **33**: 457-465.
- [12] 李俊杰, 何隆华等. 基于遥感影像纹理信息的湖泊围网养殖区提取. 湖泊科学, 2006, **18**(4): 337-342.
- [13] 赵萍, 傅云飞, 郑刘根等. 基于分类回归树分析的遥感影像土地利用/覆被分类研究. 遥感学报, 2005, **9**(6): 708-716.
- [14] 赵萍. 基于知识的江南典型区土地利用/覆被分类研究[学位论文]. 南京: 南京大学, 2003.
- [15] Breiman L, Friedman JH, Olshen RA *et al.* Classification and regression tree. Monterey, California, U.S.A. Wadsworth International Group, 1984: 1-358.
- [16] Friell MA, Brodley CE. Decision tree classification of land cover from Remotely Sensed Data. *Remote Sensing of Environment*, 1997, **61**: 399-409.
- [17] 陈华芳, 王金亮等. 山地高原地区 TM 影像水体信息提取方法比较. 遥感技术与应用, 2004, (6): 479-484.
- [18] Yang DT, Pan DL, Zhang XY. Detection of algal bloom with in situ and MODIS in Lake Taihu, China. *Proceedings of SPIE*, 2005, **5977**: 431-439.
- [19] 张冬梅, 袁清. 一种辅助草原遥感影像目视解译方法的探讨——影像与基础图件半透明叠加法. 遥感技术与应用, 2006, **4**(6): 21-28.