

## 基于决策树的 Landsat TM/ETM + 图像中太湖蓝藻水华信息提取\*

夏晓瑞<sup>1</sup>, 韦玉春<sup>1\*\*</sup>, 徐 宁<sup>2</sup>, 袁兆杰<sup>1</sup>, 王 沛<sup>1</sup>

(1: 南京师范大学虚拟地理环境教育部重点实验室, 南京 210023)

(2: 水利部海河水利委员会水资源保护科学研究所, 天津 300170)

**摘 要:** 以 9 期 Landsat TM/ETM + 影像为数据源, 基于 K-T 变换和归一化植被指数 (NDVI), 建立了湖泊蓝藻水华信息提取的决策树模型。基于大气顶面反射率图像, 选用 2005 年 10 月 17 日太湖图像进行了对比验证, 表明决策树模型比单波段阈值法、多波段阈值法 (RVI、DVI、NDVI) 能够更有效地提取蓝藻水华信息, 区分陆生植被、水生植物和水华, 省去了水体掩膜的过程。使用太湖 2002 年 10 月 25 日和 2011 年 7 月 22 日图像、巢湖 2005 年 8 月 12 日的图像, 验证决策树模型方法和工作流程的有效性。使用多期 TM 图像确定了阈值的取值范围, 其中, 亮度、绿度、NDVI 的下限值依次为 0.191、-0.007、-0.054, 湿度下限范围为 0.07 ~ 0.15; 亮度阈值上限范围为 0.3 ~ 0.7、绿度为 0.2 ~ 0.5、湿度为 0.1 ~ 0.3, 这些结果可作为湖泊蓝藻水华遥感监测的参考。

**关键词:** 蓝藻水华; 太湖; Landsat TM/ETM + 影像; K-T 变换; 决策树; 遥感信息提取

## Decision tree model of extracting blue-green algal blooms information based on Landsat TM/ETM + imagery in Lake Taihu

XIA Xiaorui<sup>1</sup>, WEI Yuchun<sup>1</sup>, XU Ning<sup>2</sup>, YUAN Zhaojie<sup>1</sup> & WANG Pei<sup>1</sup>

(1: *Key Lab of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, P. R. China*)

(2: *Research Institute of Water Resources Protection, Haihe Water Conservancy Committee, Tianjin 300170, P. R. China*)

**Abstract:** A decision tree model (KTNI) based on the K-T transformation and Normalized Difference Vegetation Index (NDVI) was developed to extract the blue-green algal blooms information in the lakes using 9 Landsat TM/ETM + images. This paper compared the KTNI with the single-band threshold value method and the multi-band threshold value methods (RVI, DVI, NDVI) in accuracy of extracting the blue-green algal blooms information, and then validated the KTNI using two observed images of the Lake Taihu and one image of the Lake Chaohu. The results showed that the KTNI can mask the water body automatically and separate the blue-green algal blooms from the land vegetation and aquatic plants effectively. Model thresholds and their ranges were determined based on the TM images; the lower limits of the *Brightness*, *Greenness* and *NDVI* were 0.191, -0.007 and -0.054, respectively, and *Wetness* ranged from 0.07 to 0.15. For the upper limits, the *Brightness* was from 0.3 to 0.7, *Greenness* was from 0.2 to 0.5, and *Wetness* was from 0.1 to 0.3. The decision tree model and threshold values can be used as the reference for remote sensing monitoring on the blue-green algal blooms in lakes.

**Keywords:** Blue-green algal blooms; Lake Taihu; Landsat TM/ETM + image; K-T transformation; decision tree; remote sensing information extraction

富营养化引起的蓝藻水华环境污染事件频频发生, 降低了水资源利用效能, 给水产养殖业、供水系统、旅游产业及人们的生命健康安全等带来了极大的危害。目前, 关于湖泊蓝藻水华的发生机理尚不明确, 难以清楚解释水华的发生规律<sup>[1]</sup>, 那么能够全面、及时、有效地获取蓝藻水华的发生发展信息和分布情况显得尤

\* 国家自然科学基金项目 (40771152) 和江苏省普通高校自然科学研究计划项目 (09KJA420001) 联合资助。2013 - 11 - 21 收稿; 2014 - 02 - 28 收修改稿。夏晓瑞 (1987 ~), 女, 硕士研究生; E-mail: henanxia@163.com.

\*\* 通信作者; E-mail: weiyuchun@njnu.edu.cn.

为重要. 因此遥感以其宏观、快速、成本低的特点, 对水环境进行动态监测并为水资源管理提供依据, 具有现实意义.

湖泊蓝藻水华提取的基础是被测水体、蓝藻水华与其他地物间的光谱差异. 水体光谱由水体中物质的吸收和散射性质决定. 水体中藻类有无及浓度不同造成水体颜色、透明度等的区别, 这些区别表现为光谱的差异, 在遥感影像上则为不同的色调、形态、纹理等. 从可见光到近红外, 正常水体的反射率逐步降低; 蓝藻水华在 560 nm 附近具有反射峰, 在近红外波段达到最高. 黄家柱等<sup>[2]</sup>利用 TM4(红)、TM3(绿)、TM2(蓝)假彩色合成图像, 发现太湖蓝藻区呈绯红色, 与周围深蓝色、蓝黑色湖水有明显区别. Kahru 等<sup>[3,4]</sup>利用绿波段对水华信息进行了提取. 考虑到水华与植被光谱特征的相似性, 大多数学者使用植被指数进行水华提取, 如: 比值植被指数 (RVI)<sup>[5,6]</sup>、差值植被指数 (DVI)<sup>[7]</sup>、归一化植被指数 (NDVI)<sup>[8-10]</sup>、增强型植被指数 (EVI)<sup>[11-12]</sup>. 有研究者根据光谱特征提出了新的藻华提取指数 (FAI)<sup>[13-14]</sup>.

蓝藻水华的日常监测流程一般是先利用湖体矢量边界将水域掩膜出来<sup>[15]</sup>, 去除水生植物和陆地信息, 以避免对提取结果的干扰. 但是矢量边界的裁剪误差会影响水华提取的精度, 更不利于自动化提取的进行. 另外, 常用的单波段、多波段提取蓝藻水华的算法, 对不同传感器图像采用的阈值不同, 该阈值受人为因素影响较大, 且提取结果中易存在湖泊水华与水生植物的混淆. 因此, 以上问题都不便于蓝藻水华的高效高精度提取. 本文基于太湖 Landsat TM/ETM+ 图像, 构建了一个湖泊蓝藻水华提取的决策树模型, 使用太湖和巢湖不同时期的图像进行了验证, 并对提取阈值范围进行了讨论.

## 1 数据与方法

### 1.1 数据源

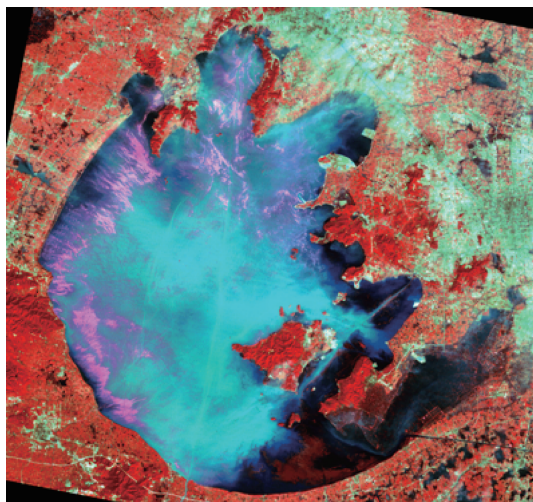


图 1 2005 年 10 月 17 日太湖 TM(4,3,2) 合成图像

Fig. 1 TM(4,3,2) composite image of Lake Taihu on October 17, 2005

大多数的卫星遥感传感器能够识别和监测蓝藻水华的空间分布和暴发范围<sup>[16-17]</sup>, Landsat TM/ETM+ 卫星影像在保证完全覆盖水体区域的同时, 还能提供湖泊水华的细节信息. 考虑到天气状况和水华的季节分布, 本文选取了 9 期 Landsat 图像, 其中太湖 6 期, 日期分别为 1998 年 7 月 26 日、2002 年 10 月 25 日、2004 年 7 月 26 日、2005 年 10 月 17 日、2011 年 7 月 22 日、2012 年 5 月 5 日; 巢湖 3 期, 日期分别为 2005 年 8 月 12 日、2007 年 10 月 5 日、2010 年 6 月 15 日. 图像数据来自国际科学数据服务平台, 采用 UTM 投影. 图像经过研究区裁剪和辐射定标后, 计算大气顶面反射率 (TOA) 图像.

在上述图像中, 太湖 2005 年 10 月 17 日的 TM 图像水华特征明显, 水生植物丰富, 且在标准假彩色影像上 (图 1), 水华、水体、水生植物差异明显. 水华呈绯红色絮状, 水体为蓝色或蓝黑色, 水生植物分布在太湖的东部, 呈暗红色. 因此选择该日期图像进行水华提取方法的对比研究.

### 1.2 湖泊蓝藻水华的提取

在遥感信息提取和分类上, 决策树方法通过不同的决策级别设定阈值, 克服了单指标阈值的缺点, 具有更好的适应性. 因此, 基于决策树思想, 本文建立湖泊蓝藻水华提取的决策树模型 (简称为 KTN1, 图 2). 模型的结点是 K-T 变换 (公式 (1)) 的 3 个分量和 NDVI (公式 (3)), 涉及到 7 个阈值, 分别用  $a_1$ 、 $a_2$ 、 $b_1$ 、 $b_2$ 、 $c_1$ 、 $c_2$ 、 $d_1$  表示. 为了方便规范提取阈值的取值范围, 对 TM/ETM+ 图像统一使用 Landsat 的 K-T 变换矩阵, 基本公式为:

$$U = R \cdot X + r \quad (1)$$

使用的系数矩阵为:

$$R = \begin{bmatrix} 0.2909 & 0.2493 & 0.4806 & 0.5568 & 0.4438 & 0.1706 \\ -0.2728 & -0.2174 & -0.5568 & 0.7221 & 0.0733 & -0.1648 \\ 0.1446 & 0.1761 & 0.3322 & 0.3396 & -0.6210 & -0.4186 \\ 0.8461 & -0.0731 & -0.4640 & -0.0032 & -0.0492 & 0.0119 \end{bmatrix} \quad (2)$$

式中,  $X = (x_1, x_2, x_3, x_4, x_5, x_7)$  是 TM/ETM + 图像的第 1、2、3、4、5、7 波段的像元值,  $r$  为常数项 10.3695、-0.7310、-3.3828 和 0.7879.

图像经过 K-T 变换后,产生的前 3 个分量依次是亮度 (*Brightness*)、绿度 (*Greenness*) 和湿度 (*Wetness*). 在本文的模型中,使用亮度来区分水体与陆地. 绿度分量反映了绿色生物量的特征,因此用来区分一般水体与水华、陆生植被. 湿度分量是可见光和近红外与较长的红外 (第 5、7 波段) 的差值,第 5 和 7 波段对湿度比较敏感,可用来区分水华、水生植物和陆地信息.

模型中的 *NDVI* 用来进一步剔除水华周围的水体,提高水华信息的提取精度.

$$NDVI = \frac{\rho_{nir} - \rho_r}{\rho_{nir} + \rho_r} \quad (3)$$

式中,  $\rho_{nir}$  是近红外波段的大气顶面反射率,  $\rho_r$  是红光波段的大气顶面反射率.

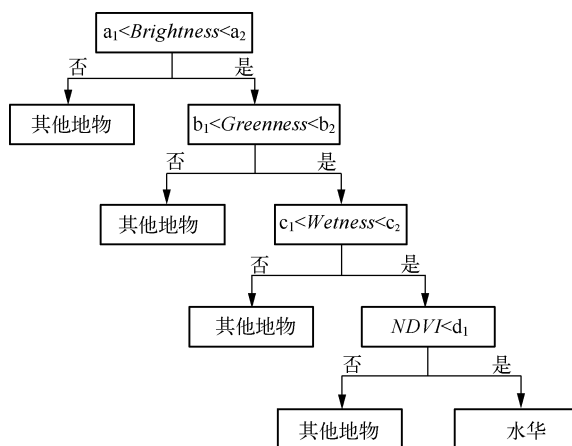


图 2 湖泊蓝藻水华提取的决策树模型

Fig. 2 The decision tree model of blue-green algal blooms information extraction

按照水华、一般水体、水生植物、陆生植被、建设用地 5 类地物选取典型样点,分别统计 *Brightness*、*Greenness*、*Wetness* 和 *NDVI* 的值,然后通过比较确定决策树模型的结点阈值. 使用  $B_i$  表示第  $i$  波段的图像值. 图像处理在 ENVI 4.8 中进行.

## 2 结果与讨论

### 2.1 KTNi 的提取结果

对 2005 年 10 月 17 日的太湖 TM 图像进行 K-T 变换和 *NDVI* 计算,然后统计典型地物 K-T 变换的前 3 个分量和 *NDVI* 的最大值、最小值、均值和标准差以便获得决策树结点的阈值. 统计结果见表 1.

*Brightness* 分量:建筑用地的亮度较高,陆生植被和水生植物的一部分小于蓝藻水华,因此亮度阈值区间设置为 (0.261, 0.647). 这样可以剔除大部分的建设用地和小部分的水生植物、水体及陆生植被.

*Greenness* 分量:蓝藻水华的值域分布较大,与其他地物有一定的差别,小于 -0.025 的包括大部分水体和一些水生植物. 因此,绿度阈值为 -0.025. 主要剔除大部分水体和剩余的建设用地.

表 1 2005 年 10 月 17 日图像中典型地物的决策树结点的统计特征

Tab. 1 TOA statistical characteristics of typical objects in the decision tree nodes, October 17, 2005

结点	统计量	水华	水生植物	一般水体	陆生植被	建设用地
<i>Brightness</i>	最小值	0.261	0.248	0.313	0.252	0.298
	最大值	0.647	0.422	0.357	0.504	0.876
	均值	0.468	0.295	0.329	0.369	0.471
	方差	0.050	0.026	0.008	0.047	0.078
<i>Greenness</i>	最小值	-0.025	-0.048	-0.091	0.031	-0.046
	最大值	0.428	0.110	-0.024	0.245	0.157
	均值	0.076	0.013	-0.078	0.121	0.015
	方差	0.076	0.018	0.005	0.037	0.028
<i>Wetness</i>	最小值	0.141	0.067	0.145	-0.084	-0.337
	最大值	0.230	0.142	0.177	0.084	0.137
	均值	0.174	0.116	0.162	0.034	0.013
	方差	0.015	0.011	0.004	0.025	0.049
<i>NDVI</i>	最小值	-0.040	-0.136	-0.323	0.218	-0.128
	最大值	0.591	0.335	-0.137	0.576	0.435
	均值	0.377	0.117	-0.252	0.423	0.076
	方差	0.103	0.062	0.028	0.065	0.067

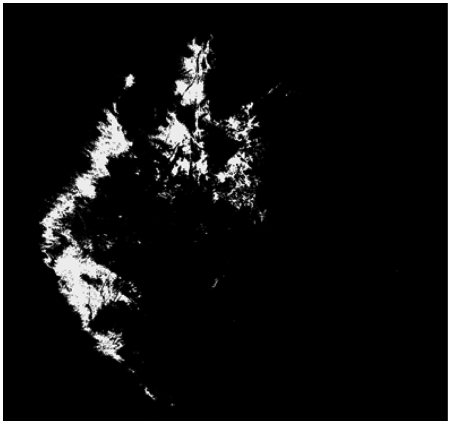


图 3 2005 年 10 月 17 日 TM 图像的水华提取结果

Fig. 3 Blue-green algal blooms information in TM image, October 17, 2005

*Wetness* 分量:从表 1 可以看出,蓝藻水华与水生植物的差别较大,小于 0.142 的包括水生植物、陆生植被和建设用地的全部.而蓝藻水华的湿度值大于 0.141,由于其湿度值大于 0.142 的部分占 90% 以上,因此该分量设置大于 0.142 的阈值可剔除水生植物.

通过上述处理后,还有水华的周围仍然混淆有部分水体,可使用 *NDVI* 来进一步区分.一般水体的 *NDVI* 值小于 -0.137,绝大多数的蓝藻水华像元值则大于 -0.040,因此阈值定为 -0.040,即大于 -0.040 的 *NDVI* 为蓝藻水华.

使用上述阈值提取的蓝藻水华结果见图 3.对照图 1 中的影像,决策树模型提取结果完全去除了水生植物;在没有湖区边界裁剪的情况下可剔除绝大部分陆地信息;提取的蓝藻水华分布与假彩色影像相比基本一致.从技术流程来看,决策树方法简单明了,层层剥离,容易控制对地物信息的筛选.

2.2 提取结果比较

提取 Landsat 图像中蓝藻水华的单一指数,主要有单波段 (*B4*) 阈值法和 *RVI*、*DVI*、*NDVI* 等指数.选择 2005 年 10 月 17 日图像中蓝藻水华、一般水体、水生植物的区域进行统计,样点分布见图 4.统计结果表明,这些提取指数值在不同地物间有明显的交叉,提取结果中地物混淆明显(图 5).由于陆地信息对水华提取的干扰很大,使用这些方法前,必须先进行湖泊水体的掩膜处理.

通过水华区域统计确定提取水华的阈值,在这些指数中,*B4* 波段为 (0.145, 0.695).

对于比值、差值或归一化植被指数来提取水华,根据图像特征,确定的阈值如下:

$$RVI > 0.800 \tag{4}$$

$$DVI > -0.100 \tag{5}$$

$$NDVI > -0.040 \quad (6)$$

上述指数不能完全提取水华,且多将东湖的水生植物混淆为水华.部分浑浊水体近红外波段反射率相对较高<sup>[18-19]</sup>,容易与蓝藻水华光谱混淆.因此,单波段容易扩大或者减小蓝藻水华信息.对于  $RVI$ 、 $DVI$  和  $NDVI$  指数,高度透明的水体在近红外波段的反射率较低,蓝藻水华易受周围水体的影响导致差值小于 0 或比值小于 1,一般要下调提取阈值;对于悬浮物浓度高的水体,近红外反射率较高,与蓝藻水华的光谱特征相似,使得差值大于 0 和比值大于 1,易出现蓝藻水华的误提现象,导致水华面积高估.

图 6 是不同方法对太湖漫山岛附近水华提取效果的比较.  $TM(4,3,2)$  假彩色合成原始影像中,深红色部分为有植被覆盖的小岛,绕在小岛边缘的浅红色部分为水华,蓝色为水体(图 6a).与之相对比,只有决策树模型能有效地剔除陆相信息,完整保留水华(图 6b),近红外单波段剔除了小岛中部的一部分(图 6d),其他方法则无能为力(图 6c、e、f).

图 7 是不同方法对太湖南部水生植物分布区内水华的提取结果.假彩色合成图像中,红色斑点为水生植物,蓝色部分为水体(图 7a).不同方法在剔除水生植物能力上的结果表明,决策树模型表现出良好的效果,只有零星的水生植物信息未被剔除(图 7b);其他方法则保留了大部分的水生植物信息,而这些信息混淆在水华中,影响了水华提取的精度(图 7c、d、e、f).

为了进一步比较各方法提取结果的差异,以标准假彩色合成影像目视解译出来的水华区域作为标准,进行精度评价.虽然目视解译靠人工干预,操作麻烦,但其提取结果具有较好的可信度.将提取结果与目视解译结果进行叠加分析,以验证水华的位置和数量信息.

表 2 是提取精度的比较结果.其中,决策树方法和其他方法的水华提取像元个数为  $a$ ;目视解译水华的像元个数为  $a_0$ ,作为参考的真值;对于任一提取方法, $a$  与  $a_0$  相交的部分为该方法正确提取的部分,记为  $b$ ;  $a$  中不相交的部分为错提部分,记为  $c$ ;  $a_0$  中不相交的部分为漏提部分,记为  $d$ .为了保证在同一水平上比较各方法的提取精度,以  $a_0$  为基数进行计算,则蓝藻水华的正确提取精度为  $R = b/a_0 \times 100\%$ ,漏提精度是  $M = d/a_0 \times 100\%$ ,错提精度是  $W = c/a_0 \times 100\%$ .由于  $d = a_0 - b$ ,所以  $R + M = 1$ ,那么  $W$  在这里做不同方法间的相对比较.

结果表明,不同方法提取的蓝藻水华差异较大.其中,决策树模型的提取精度最高,正确提取精度达到 97.77%;漏提精度和错提精度最低,分别是 2.23% 和 3.78%,说明漏提和错提的程度最小.  $NDVI$  的提取精度次之,正确提取精度为 89.50%.  $RVI$  和  $DVI$  提取的像元数较多.  $RVI$  的错提和漏提信息最多,分别为 16.34% 和 22.19% (表 2).单一提取指数只有一个变量来控制提取结果,并且各类地物的参数值存在交集,使得在阈值选择上难以判断.阈值过于狭窄容易造成漏提现象;阈值范围扩大会增加其他类被错提为水华的信息.决策树模型综合了多个变量的优点,较易剔除其他地类,尤其是陆生植被和水生植物的干扰,弥补

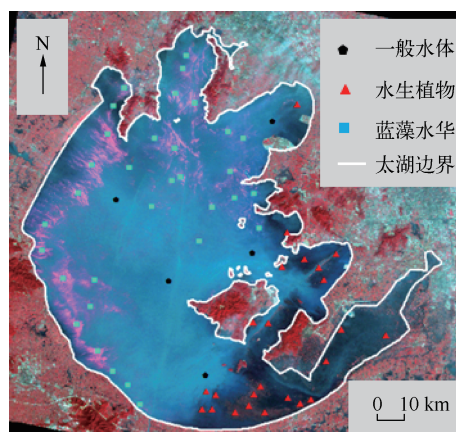


图 4 样点分布

Fig. 4 Distribution of sample sites

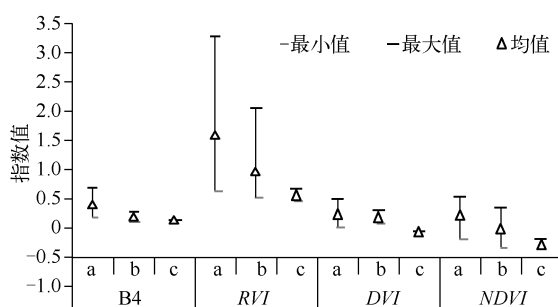


图 5 蓝藻水华、水生植物和一般水体单一指数的值域 (a 为蓝藻水华, b 为水生植物, c 为一般水体)

Fig. 5 Value ranges of the single index from blue-green algal blooms, aquatic plants and water body

了单一指数的不足,降低了错提和漏提率.

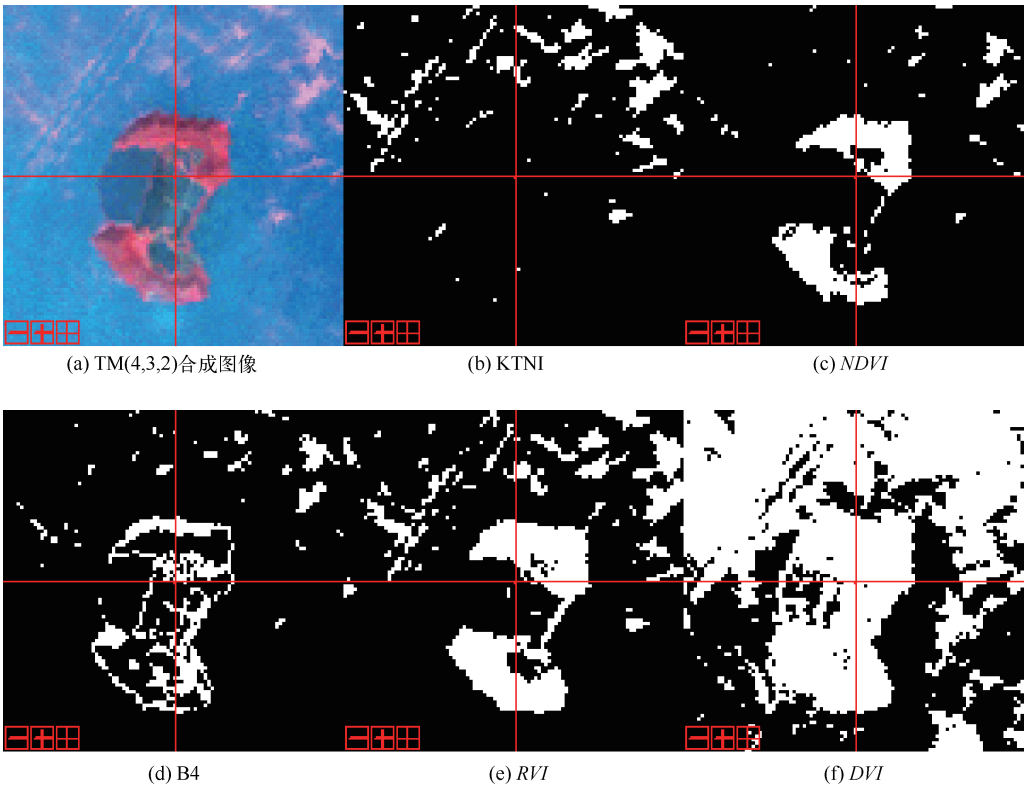


图 6 不同方法对 Landsat TM 图像中太湖漫山岛周围水华的提取效果  
Fig. 6 Blue-green algal blooms information around Manshan Island in Lake Taihu  
in Landsat TM image extracted by different methods

表 2 湖泊蓝藻水华提取精度比较

Tab. 2 Accuracy comparison of blue-green algal blooms distribution in the lake

编号	方法	提取的水华像元数( $a$ )	正确提取精度( $R$ )	漏提精度( $M$ )	错提精度( $W$ )
1	NDVI	344487	89.50%	10.50%	12.44%
2	RVI	357701	83.66%	16.34%	22.19%
3	DVI	359123	87.77%	12.23%	18.50%
4	B4 单波段	345218	88.61%	11.39%	13.54%
5	决策树	343169	97.77%	2.23%	3.78%
	目视解译( $a_0$ )	337899			

2.3 KTNI 的验证

使用太湖 2002 年 10 月 25 日、巢湖 2005 年 8 月 12 日的 TM 影像及太湖 2011 年 7 月 22 日有云覆盖的 ETM+ 图像,根据图像特征选取典型地物样本统计阈值,进行提取模型的验证,表明决策树模型可以将影像中的水华信息较完整地提取出来,提取的水华像元个数分别为 6617、8993、47894. 陆地部分也被很好地分离出去,个别零星像元一般存在于小的坑塘和河流的入湖位置,与水华的实际分布相符合(图 8).



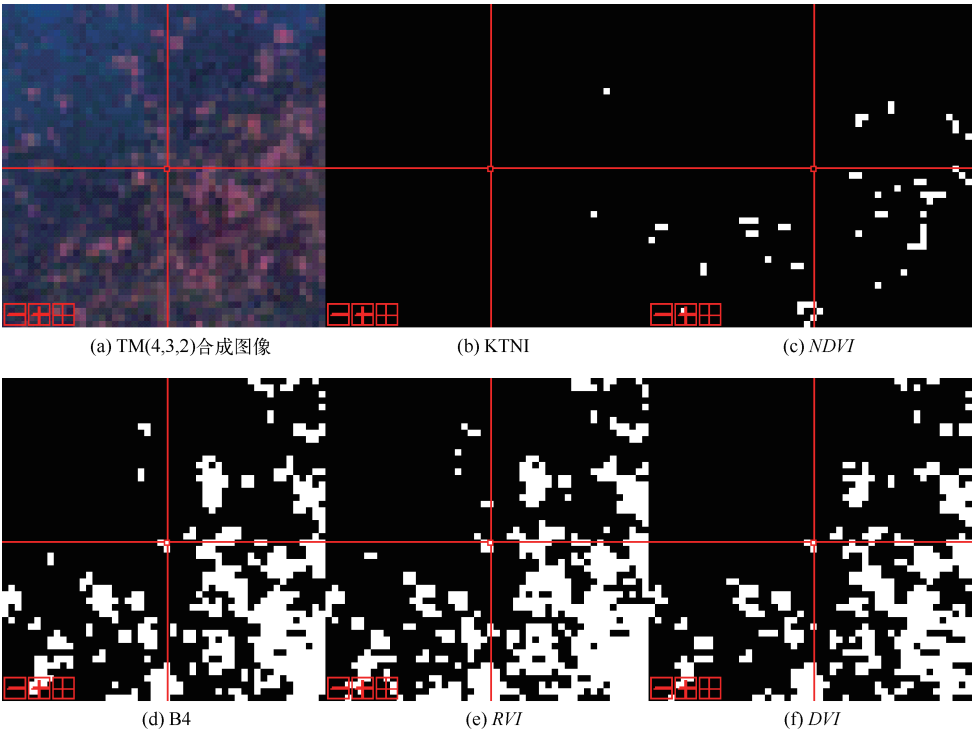


图 7 不同方法对 Landsat TM 图像中太湖南部水生植物的剔除效果

Fig.7 Aquatic plants information in the south of Lake Taihu in Landsat TM image removed by different methods

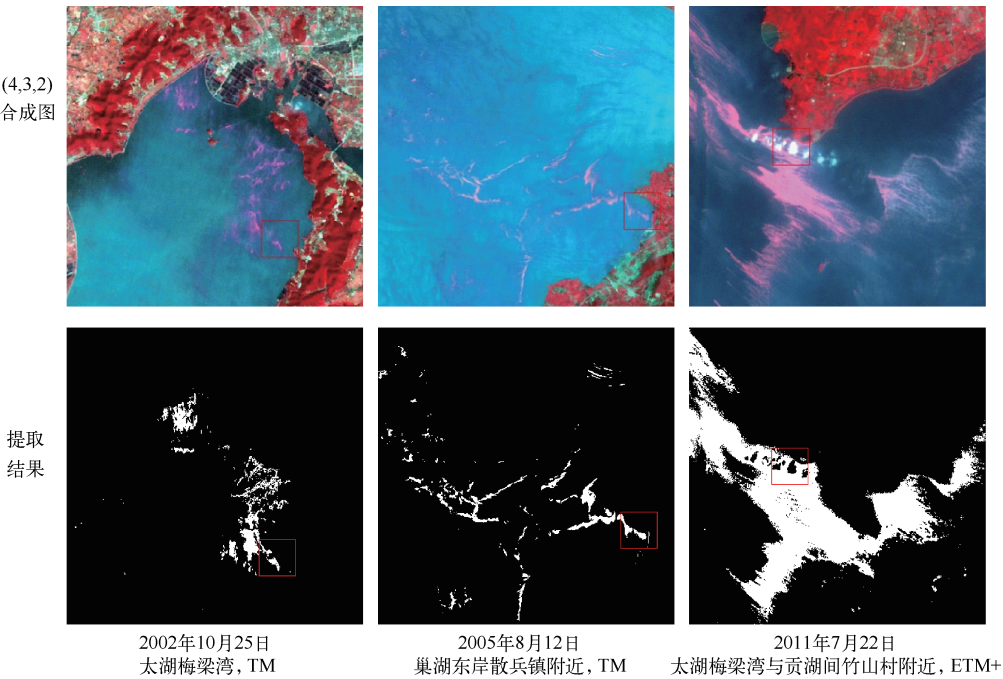


图 8 不同日期图像中 KTNi 提取的蓝藻水华分布

Fig. 8 Blue-green algal blooms information in images on different dates extracted by different methods

太湖蓝藻水华多发生在梅梁湾区域,东太湖的水生植物区很少见<sup>[20-21]</sup>. 上述太湖的两期验证影像都成功地去除了水生植物信息. 2011 年 7 月 22 日假彩色图像上红框内的高亮度厚云也被剔除. 巢湖湖区无水生植物,提取过程较简单. 太湖 2002 年和 2011 年两期图像截取了梅梁湾区域,巢湖 2005 年的图像截取了东岸. 对比假彩色图和蓝藻水华提取二值图可以看出,3 期数据的提取结果中陆地信息已被剔除,且无论蓝藻水华成片或零散,提取结果的形状大小和分布均与假彩色影像一致(图 8).

2.4 KTNi 提取阈值的确定

蓝藻水华提取阈值影响着提取方法的效率和提取结果的准确性,因此有必要作进一步分析. 为了较全面地揭示蓝藻水华提取阈值的特征,使用 5—10 月有蓝藻水华分布的典型图像,共计 12 景进行了蓝藻水华的提取,观察其阈值分布情况. 其中,鉴于 ETM+ 图像中的条带会影响阈值选取,使用能覆盖阈值变化区间的太湖 TM 图像进行本文阈值的确定,日期分别为 1998 年 7 月 10 日、2002 年 10 月 25 日、2004 年 7 月 26 日、2005 年 10 月 17 日. 这 4 期图像中的水华分布状况不同,其中 2004 年 7 月 26 日影像中水生植物较多. 2005 年 10 月 17 日水华分布于西太湖的大部分区域,水生植物分布在东太湖并以东南部较多. 决策树模型阈值统计特征见表 3.

表 3 蓝藻水华决策树模型结点阈值的统计特征

Tab. 3 TOA statistical characteristics of node threshold in the decision tree model of blue-green algal blooms

影像日期		Brightness		Greenness		Wetness		NDVI
		下限	上限	下限	上限	下限	上限	下限
2005 年 10 月 17 日		0.261	0.648	-0.025	0.428	0.141	0.230	-0.040
2004 年 7 月 26 日		0.205	0.471	-0.027	0.219	0.110	0.151	-0.100
2002 年 10 月 25 日		0.173	0.312	0.023	0.202	0.085	0.118	-0.050
1998 年 7 月 10 日		0.126	0.394	0.003	0.403	0.071	0.132	-0.026
统计量	均值	0.191	0.456	-0.007	0.313	0.102	0.157	-0.054
	方差	0.003	0.021	0.0006	0.014	0.001	0.003	0.001

由于各分量阈值下限的方差均远远小于上限,这里用以上 4 期影像的下限均值作为决策树模型阈值的下限,所以 *Brightness*、*Greenness*、*NDVI* 的下限值依次设定为:0.191、-0.007、-0.054. 在 *Wetness* 分量上,由于水生植物与水华典型区的统计值可能有少量相交部分,为了完全剔除水生植物,各期图像选用水生植物典型区湿度的最大值为该分量阈值下限,参考 4 期图像 *Wetness* 的下限最小值 0.071 和最大值 0.141 (表 3),将模型的湿度下限范围定为 0.07~0.15.

K-T 变换分量的上限随水华发生的情况而变化,如水华在假彩色影像上色彩较浓,其值偏高,反之偏低. 根据图像的统计结果,*Brightness* 的上限定在 0.3~0.7 之间、*Greenness* 为 0.2~0.5、*Wetness* 为 0.1~0.3.

选择 2007 年 10 月 5 日巢湖的 TM 图像及 2010 年 6 月 15 日巢湖、2012 年 5 月 5 日太湖的 ETM+ 图像,验证了阈值的适用性. 2007 年 10 月 5 日巢湖 TM 图像的提取阈值,即  $a_1$ 、 $a_2$ 、 $b_1$ 、 $b_2$ 、 $c_1$ 、 $c_2$ 、 $d_1$  依次为 0.165、0.350、-0.100、0.200、0.070、0.290、-0.200; 2010 年 6 月 15 日巢湖的 ETM+ 图像阈值分别为 0.285、0.386、-0.007、0.175、0.117、0.159、-0.051; 2012 年 5 月 5 日太湖的 ETM+ 图像阈值分别为 0.273、0.440、-0.007、0.268、0.117、0.140、-0.051. 上述阈值都在本文设定的参考阈值范围内.

这些结果表明:对于新的图像,参考阈值可用做决策树模型阈值的初始值. 以此为基础,通过进一步的调整可提取湖泊蓝藻水华,这样可以省去采集样本、阈值统计的步骤. 另外,模型本身较大地压抑了非蓝藻水华信息,节约了湖体掩膜的时间,从而提高了工作效率.

3 结论

基于 2005 年 10 月 17 日太湖的大气顶面反射率图像,本文将 K-T 变换和 *NDVI* 植被指数相组合构建了湖泊蓝藻水华的决策树提取模型,通过增大地物像元值的差异,初步解决了水华与水生植物的混淆问题. 与



常用的太湖蓝藻水华提取的方法相比,本模型剔除了陆地的影响,省去了湖体掩膜的过程,更便于蓝藻水华的快速提取。经太湖和巢湖的典型图像验证,表明模型具有较好的适用性、速度快、提取精度高。本文初步确定了模型的阈值范围,即亮度、绿度和 *NDVI* 的下限值依次设定为:0.191、-0.007 和 -0.054,湿度下限范围设定为 0.07~0.15;亮度的阈值上限范围为 0.3~0.7、绿度为 0.2~0.5、湿度为 0.1~0.3。巢湖和太湖的 TM/ETM + 图像验证了阈值的适用性,表明可以在该阈值范围基础上通过微调快速提取蓝藻水华信息,为蓝藻水华的自动提取奠定了基础。

本文的湖泊蓝藻水华提取方法只能用于 Landsat 图像,如何针对其他传感器的图像特征构建决策树还需进一步研究。在湖泊蓝藻水华的遥感信息提取中,如何快速区分蓝藻水华、一般水体和水生植物,建立提取算法并实现自动化提取,如何针对一般水体、蓝藻水华和水生植物的混合特征,通过混合像元分解来提高提取精度,也还需开展更多的研究工作。

#### 4 参考文献

- [1] 马健荣,邓建明,秦伯强等. 湖泊蓝藻水华发生机理研究进展. 生态学报,2013,**33**(10):3020-3030.
- [2] 黄家柱,赵 锐. 卫星遥感监测太湖水域蓝藻暴发. 遥感信息,1999,(4):43-44.
- [3] Kahru M, Mitchell BG, Diaz A. Using MODIS medium-resolution bands to monitor harmful algal blooms. *Remote Sensing of the Coastal Oceanic Environment*, 2005, **5885**: 1-6.
- [4] Kahru M, Savchuk OP, Elmgren R. Satellite measurements of cyanobacterial bloom frequency in the Baltic Sea: interannual and spatial variability. *Marine Ecology Progress Series*, 2007, **343**: 15-23.
- [5] 马荣华,孔繁翔,段洪涛等. 基于卫星遥感的太湖蓝藻水华的时空分布规律认识. 湖泊科学,2008,**20**(6):687-694.
- [6] 周立国,冯学智,王春红等. 太湖蓝藻水华的 MODIS 卫星监测. 湖泊科学,2008,**20**(2):203-207.
- [7] 段洪涛,张寿选,张渊智. 太湖蓝藻水华遥感监测. 湖泊科学,2008,**20**(2):145-152.
- [8] 鲁韦坤,谢国清,余领翔等. MODIS 遥感监测滇池蓝藻水华分布. 气象科技,2009,**37**(5):618-620.
- [9] 张 红,黄 勇. 基于 NOAA/AVHRR 卫星资料的巢湖水华规律分析. 中国环境科学,2009,**29**(7):727-732.
- [10] 成 功,高泽润. 基于 ETM + 的洞庭湖蓝藻遥感监测研究. 中国环境科学学会学术年会论文集,2012:553-559.
- [11] 徐京萍,张 柏,李 方等. 基于 MODIS 数据的太湖藻华水体识别模式. 湖泊科学,2008,**20**(2):191-195.
- [12] 刘 健,王文成. 基于 ASAR 和 MODIS 遥感数据的蓝藻水华提取. 安徽农业科学,2013,(2):889-892.
- [13] Hu CM. A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, 2009, **113**: 2118-2129.
- [14] Hu CM, Lee ZP, Ma RH *et al.* Moderate resolution imaging spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research: Oceans (1978-2012)*, 2010, **115**(C4). doi: 10.1029/2009JC005511.
- [15] 李亚春,孙佳丽,谢志清等. 基于 MODIS 植被指数的太湖蓝藻信息提取方法研究. 气象科学,2011,**31**(6):737-741.
- [16] Sun LY, Chen YZ, Wang XQ *et al.* Comparison of multi-sensor data application in algal bloom detection. *International Congress on Image and Signal Processing*, 2010, **3**: 2144-2148.
- [17] Klemas V. Remote sensing of algal blooms: An overview with case studies. *Journal of Coastal Research*, 2012, **28**(1A): 34-43.
- [18] Doxaran D, Froidefond JM, Lavender S *et al.* Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations. *Remote Sensing of Environment*, 2002, **81**(1): 149-161.
- [19] Doron M, Bélanger S, Doxaran D *et al.* Spectral variation in the near-infrared ocean reflectance. *Remote Sensing of Environment*, 2011, **115**: 1617-1631.
- [20] Chen Y, Qin B, Teubner K *et al.* Long-term dynamics of phytoplankton assemblages: *Microcystis*-domination in Lake Taihu, a large shallow lake in China. *Journal of Plankton Research*, 2003, **25**(4):445-453.
- [21] 朱广伟. 太湖富营养化现状及原因分析. 湖泊科学,2008,**20**(1):21-26.