Distinguishing Algal Blooms from Aquatic Vegetation in Chinese Lakes Using Sentinel 2 Image

(Supporting Information)

Data Preprocessing

Asynchronous images are widely used to identify algal blooms and aquatic vegetation in practice [1], but this method has two presumptions to identify algal blooms. First, although aquatic vegetation varies physically throughout the year, its distribution area is generally stable, in contrast to algal blooms, which float in the water, and their distribution constantly changes its shape and position due to, e.g., wind, light, and temperature. Therefore, when calculating the frequency of vegetation signals over a continuous period for the same pixel, the frequency of aquatic vegetation is usually higher than that of algal blooms. Second, as aquatic vegetation contributes to water purification through, e.g., nutrient uptake and sedimentation [2], the water quality in areas where aquatic vegetation grows is usually better without the occurrence of algal blooms, i.e., the spatial distributions of the two do not overlap. Based on these two assumptions, we can use the frequency index to distinguish between algal blooms and aquatic vegetation (Figure S1).

Step 1. Obtaining high-quality SA2 images.

Step 2. The images were processed by ArcGIS to Obtain remote sensing reflectance data of algal blooms with different outbreak intensities and aquatic vegetation for different areas of each of the five lakes.

Step 3. Constructing decision trees to extract algal blooms and aquatic vegetation.

Step 4: Calculate the VPF based on the phenological information.

Based on the climatic zoning and the growth of the aquatic vegetation, the period with the largest stable area of aquatic vegetation growth was selected as the VPF time range for the calculation. The formula for calculating VPF has been adjusted a little here. The difference between the VPF calculation method in this paper and the original formula is that the vegetation signal in this paper is based on the extraction results of multiple indices, replacing the vegetation signal extracted using only FAI in the VPF calculation formula proposed by Liu in 2015 [3]. The improved formula for calculating the VPF here is as follows:

$$VPF(j)=\frac{\sum_{i=1}^{n} Lv(j,i)}{n}$$

where VPF(j) is the vegetation presence frequency of pixel j in n images. Pixel j of image i enters the Lv layer when it is determined to be a vegetation signal by the decision tree and is assigned a value of 1. Otherwise, it is assigned a value of 0. n is the total number of images in the calculated time range.

Step 5: Delineating the VPF threshold based on the distribution of algal blooms and aquatic vegetation, removing low-frequency algal bloom distribution areas, and obtaining the aquatic vegetation boundaries.

Step 6: Clipping the algal blooms and aquatic vegetation areas extracted in Step 3 with the extracted aquatic vegetation boundaries. The area falling within the aquatic vegetation boundary is aquatic vegetation; otherwise, it is algal blooms.

The indices used in the mamuscript

(1) Normalized Difference Vegetation Index (NDVI)

The NDVI is sensitive to vegetation, effectively detects green plants, and calculates their biomass [4]. It is widely used in the study of national vegetation identification [5]. Algae contain chlorophyll and display similar vegetation features in their spectral characteristics. The NDVI is often used to distinguish algal blooms from water.

(2) Floating Algae Index (FAI)

The FAI is a MODIS-based index for identifying algal outbreaks proposed by Hu in 2009. The FAI has been tested and found suitable for the stable extraction of algal outbreak areas in the water column under relatively turbid conditions, as the FAI is insensitive to environmental variations, such as aerosol thickness and solar flares [6].

(3) Normalized Difference Water Index (NDWI)

The NDWI was proposed to identify water bodies using a normalized index of visible green light and near-infrared to eliminate soil and terrestrial vegetation features [7]. The NDWI has been modified to extract vegetation liquid water content using a combination of near-infrared and shortwave infrared [8]. as well as using a combination of the green band and shortwave infrared to extract water body information more consistently, effectively eliminating noise caused by the background, e.g., buildings. Due to continuous development and case studies, the NDWI has many band combinations, and this paper tested these to select the most efficient combination for our objectives.

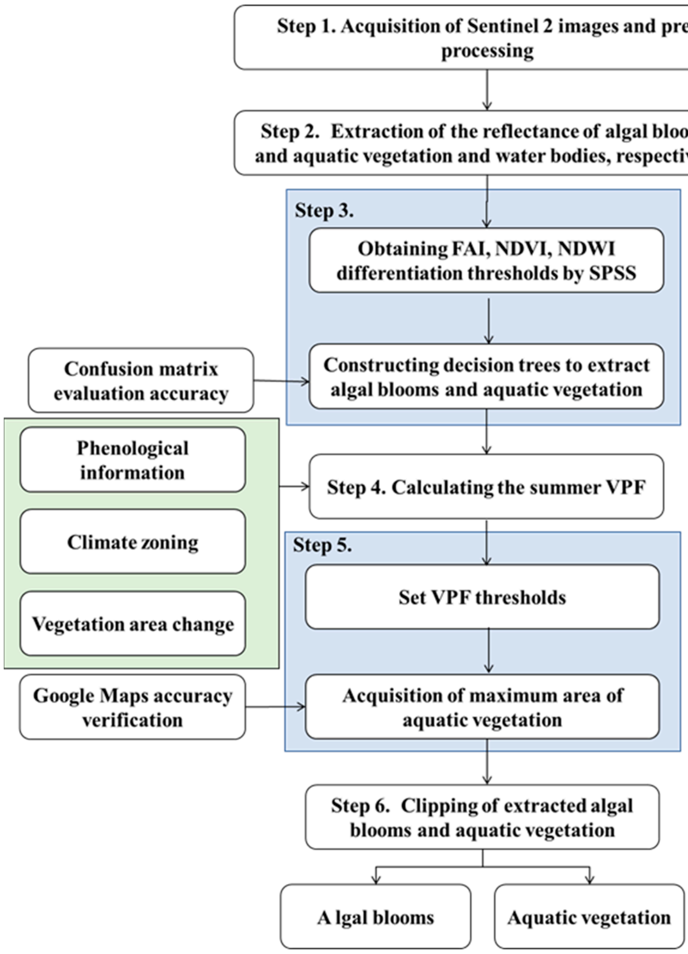
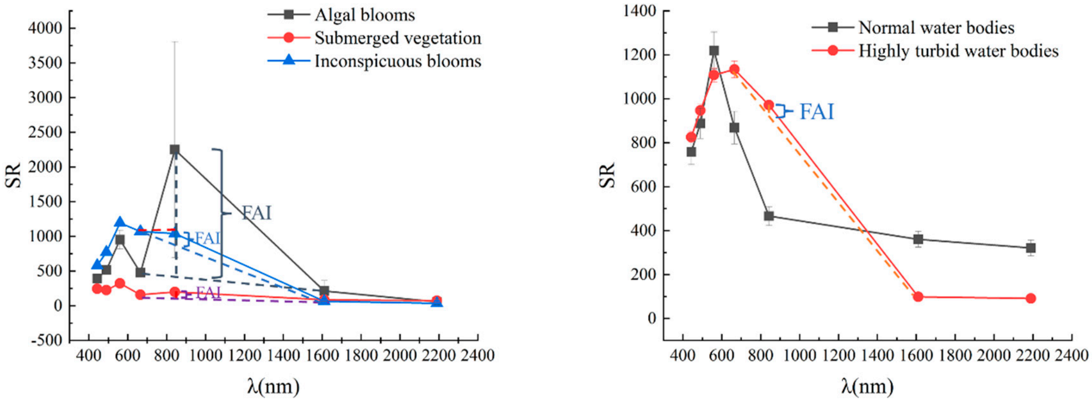


Figure S1. Flow chart for identifying algal blooms and aquatic vegetation in Sentinel 2 images.



(a) (b)

Figure S2. Reflectance spectral curves for the five features, figure (**a**) water group, black curve for normal water, red curve for highly turbid water; figure (**b**) vegetation group, black curve for normal algal blooms, blue curve for inconspicuous algal blooms (marginal areas of algal blooms) and red curve for submerged vegetation.

Table S1. Basic information on seven lakes covered in the article.

Lake	Long (E)	Lat (N)	Area (km^2)	Average Water Depth (m)	Max. Water Depth (m)	Water Storage Capacity ($*10^8 m^3$)	Climate Zone
Lake Hulun	116°58′-117°48′	48°43′-49°20′	2,339	5.7	8	138.5	temperate climate
Lake Hongze	118°10′-118°52′	33°06′-33°40′	1,805	1.77	4.37	27.9	southern temperate climate
Lake Chaohu	117°00′-118°17′	30°34′-32°12′	573	2.69	3.77	20.7	subtropical climate
Lake Taihu	119°52′-120°36′	30°55′-31°32′	2,338	1.9	2.6	51.4	subtropical climate
Lake Dianchi	102°36′-103°40′	24°40′-25°02′	330	2.93	5.87	11.69	subtropical climate
Taipingchi Reservoir	124°31′-124°58′	43°55′-44°55′	25	2.5	6	2.01	temperate climate
Lake Chenghai	100°38′-100°41′	16°27′-26°38′	77.22	25.7	35.1	19.87	subtropical climate

Table S2. Temporal distribution of downloaded Sentinel 2 images of the five lakes used for modelling.

	HLH	HZH	CH	TH	DC
Jan	0	4	0	1	5
Feb	0	1	2	2	1
Mar	0	0	0	1	3
Apr	0	2	1	4	1
May	7	1	5	2	3
June	6	0	2	1	1
Jul	6	3	1	1	1
Aug	6	2	4	3	0
Sep	8	2	3	1	0
Oct	4	1	1	3	1
Nov	0	3	2	3	1
Dec	0	2	1	1	2
Total	37	21	22	23	19

References

1. Song, K., Fang, C., Jacinthe, P.-A., et al. Climatic versus Anthropogenic Controls of Decadal Trends (1983-2017) in Algal Blooms in Lakes and Reservoirs across China. Environmental Science & Technology (2021) 55, 2929-2938.
2. Dong, B., Qin, B., Gao, G., et al. Submerged macrophyte communities and the controlling factors in large, shallow Lake Taihu (China): Sediment distribution and water depth. Journal of Great Lakes Research (2014) 40, 646-655.
3. Liu, X., Zhang, Y., Shi, K., et al. Mapping Aquatic Vegetation in a Large, Shallow Eutrophic Lake: A Frequency-Based Approach Using Multiple Years of MODIS Data. Remote Sensing (2015) 7, 10295-10320.
4. Huete, A., Didan, K., Miura, T., et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment (2002) 83, 195-213.
5. Shi, Y., Jin, N., Ma, X.L., et al. Attribution of climate and human activities to vegetation change in China using machine learning techniques. Agricultural and Forest Meteorology (2020) 294.
6. Hu, C. A novel ocean color index to detect floating algae in the global oceans. Remote Sensing of Environment (2009) 113, 2118-2129.
7. McFeeters, S.K. The use of the normalized difference water index (NDWI) in the delineation of open water features. International Journal of Remote Sensing (1996) 17, 1425-1432.
8. Gao, B.C. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment (1996) 58, 257-266.