MDPI

Article

Monitoring and Forecasting Green Tide in the Yellow Sea Using Satellite Imagery

Shuwen Xu [1,2], Tan Yu [1,2,3,*], Jinmeng Xu [2], Xishan Pan [4,5], Weizeng Shao [2,3,6], Juncheng Zuo [2,3] and Yang Yu [7]

1   Donghai Laboratory, Zhoushan 316021, China
2   College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China
3   Key Laboratory of Marine Ecological Monitoring and Restoration Technologies (MNR),
    Shanghai 201206, China
4   Tidal Flat Research Center of State Oceanic Administration, Nanjing 210036, China
5   The Key Laboratory of Port, Waterway and Sedimentation Engineering of the Ministry of Transport,
    Nanjing 210029, China
6   National Satellite Ocean Application Service, Ministry of Natural Resources, Beijing 100081, China
7   Department of Marine Sciences and Biology, Qingdao University of Science and Technology,
    Qingdao 266042, China
*   Correspondence: tyu@shou.edu.cn; Tel.: +86-021-6190-0169

Abstract: This paper proposes a semi-automatic green tide extraction method based on the *NDVI* to extract Yellow Sea green tides from 2008 to 2022 using remote sensing (RS) images from multiple satellites: GF-1, Landsat 5 TM, Landsat 8 OLI_TIRS, HJ-1A/B, HY-1C, and MODIS. The results of the accuracy assessment based on three indicators: Precision, Recall, and F1-score, showed that our extraction method can be applied to the images of most satellites and different environments. We traced the source of the Yellow Sea green tide to Jiangsu Subei shoal and the southeastern Yellow Sea and earliest advanced the tracing time to early April. The Gompertz and Logistic growth curve models were selected to predict and monitor the extent and duration of the Yellow Sea green tide, and uncertainty for the predicted growth curve was estimated. The prediction for 2022 was that its start and dissipation dates were expected to be June 1 and August 15, respectively, and the accumulative cover area was expected to be approximately 1190.90–1191.21 km^2.

Keywords: Yellow Sea green tide; NDVI; multi-source RS; source tracing; growth curve; prediction

1. Introduction

Green tide, an ecological anomaly formed by the explosive growth and aggregation of macroalgae in the ocean, is common in coastal countries around the world [1,2]. The green tide in the Yellow Sea is one of the largest green tide events in the world; it has been occurring annually for 15 years, lasts for a long time every year, affects a wide area of the sea, and is difficult to clean up [3–7]. It has a significant negative impact on the local marine ecosystem and coastal economy, and it is relatively tolerant to high temperatures and light intensity [8], so it is highly persistent with explosive growth and has gradually become a regular marine ecological hazard. Therefore, there is an urgent need for effective dynamic monitoring of the green tide to guide prevention and control measures.

Compared with field surveys, satellite remote sensing (RS) has the advantages of a wide monitoring area, fast information acquisition, a short update period, timeliness, and strong economic benefits and is thus an effective means of real-time monitoring and control of green tide events [9,10]. The green tide index threshold method based on RS, which follows the principle that green algae have unique spectral characteristics in the visible and infrared bands, has been widely used to extract green tide information and is no longer limited by the type of satellite sensor. For example, various green tide index threshold methods based on band calculation, such as the normalized vegetation index [11], the

enhanced vegetation index (EVI, a MOD13 satellite-derived index), and the normalized difference algal index [12]. Moreover, there are other methods based on baseline subtraction that are less sensitive to changes in environmental and observation conditions, and can extract green tide information through thin clouds [9,13], such as baseline-based floating algae index [14], virtual-baseline drifting algae indices [15], and the RGB floating algae index [16]. All of these indices are widely used as they can clarify the differences between seawater and green tide features. Although the index calculation method and principle for green tide have been improved, optimization of the threshold selection method is still far from adequate; continuous refinement and improvement from a manual empirical judgment to a more automatic threshold selection are needed to improve the efficiency of green tide extraction and the dynamic monitoring capability of RS. Thus, one of the purposes of this study is to propose an effective, universal, and more automatic method for extracting green tide from RS images.

Since 2007, major green tide events have occurred in the Yellow Sea from March to August of each year. Since the first outbreak of this green tide, early prevention and control have been key concerns, and the tracing of the source of the Yellow Sea green tide is of utmost importance. RS satellite monitoring is the most common and effective means to trace the green tide. According to previous studies, the Subei shoal along the Jiangsu coast is considered to be the direct origin of the Yellow Sea green tide outbreak, as local geographical, aquaculture, and biological factors provide sufficient nutrients and a suitable growth environment for the green algae [1,3,4,13]. Further, morphological and phylogenetic analyses have shown that this green tide is homologous to the algal species in the purslane culture area of Jiangsu [17]. Consistently, RS satellite monitoring and field surveys have demonstrated that most of the initial discovery sites of the Yellow Sea green tide were also near the laver farms in the Subei shoal along the Jiangsu coast [1,3,4,13]. Therefore, evidence from various sources in previous studies has indicated that the Subei shoal along the Jiangsu coast is the source of the Yellow Sea green tide. At the same time, the current traceability of the Yellow Sea green tide source, whether by RS monitoring or field investigation, is limited to the Subei shoal. Since tracing and investigation of the source of the green tide remains an important first step toward its early prevention and control of green tide, so another purpose of this study is to use multi-source RS data to expand the monitoring range of the Yellow Sea green tide source, such as the southeastern Yellow Sea.

In terms of green tide prevention and control, in addition to tracing and early control of green tide sources, it is important to predict and assess subsequent green tide growth trends. The development of a green tide is dynamic, so an annual outbreak is not appropriate to define or assess only by some static indicators like the accumulative or the maximum cover area, which may lead to the wrong prediction and evaluation of the bloom scale of green tide [18]. Therefore, many numerical models have been introduced to predict the growth trends of green tides. For example, the physical–ecological model based on the FVCOM model is used to simulate the distribution and growth of green tides [19]. The LTRANS model, coupled with an ecological module, is constructed to predict the drift path, growth, and dissipation processes of green tide [20]. The pixel unmixing method was used to further estimate the coverage area of *U. prolifera*, so as to obtain a more accurate scale of the green tide [21]. Due to the limitations of the marine environment and resources, the growth curve of algal development is generally S-shaped [21,22]. Therefore, the growth curve of the green tide is also widely used to model and predict algal growth trends. To predict the growth scale and duration of the Yellow Sea green tide based on Gompertz and Logistic growth curve is also one of the main objectives of this study.

This work is organized as follows: Section 1—Introduction; Section 2—Description of the dataset; Section 3—Method, the methodology of the study is explained in detail; Section 4—Results, the results for different RS datasets are presented, the accuracy assessed, the green tide's growth curves calculated, forecast curve for 2022 estimated, tide's bloom times and accumulative/maximum cover area simulated; Section 5—Discussion, statistics

of source sites of the green tides in period from 2008 to 2022 is analyzed, and Section 6—Conclusions.

2. Description of the Dataset

2.1. Study Area

The Yellow Sea green tide has been breaking out for 15 years and has become a regular marine ecological hazard along the Chinese coast. The study area was the Yellow Sea (119°–126°E, 31°–38°N), as shown in Figure 1 (bathymetry data in Figure 1 are from ETOPO1). To validate the accuracy under different environments, we partitioned the Yellow Sea according to its water depth, as shown in Figure 1a according to the water depth of the Yellow Sea. Such partitioning mainly referred to Li et al., who argue that the normalized difference vegetation index (*NDVI*) will be more or less affected by muddy water, and shallow areas (seawater depth of <-35 m) would affect *NDVI* to some extent, and that environmental factors would cause uncertainty in the extraction results on satellite images [23]. As seen in Figure 1b, we chose the accuracy assessment areas A1–A3, B1–B3, and C1–C3 corresponding to the three partitions of A, B, and C.
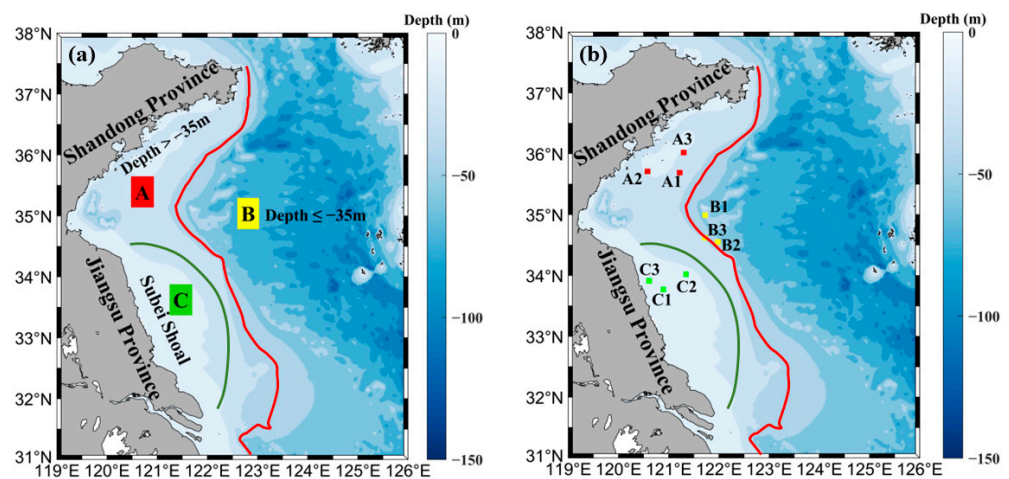

Figure 1. Study area and Location diagram of extraction regions in different environments for accuracy assessment ((**a**) Partition diagram, where the red line is the -35 m isobath and the area within the green line represents Zone C, which is the muddy water area. (**b**) Accuracy assessment areas A1–A3, B1–B3, and C1–C3 corresponding to the three partitions of A, B, and C).

2.2. RS Dataset

The main satellites from which RS data were obtained for this study were Gaofen-1 (GF-1), Landsat 5 Thematic Mapper (TM), Landsat 8 operational land imager_thermal infrared sensor (OLI_TIRS), Huanjing-1A/B (HJ-1A/B), HaiYang-1C (HY-1C), and moderate-resolution imaging spectroradiometer (MODIS).

As the first satellite in the space-based system of the major special project of China's high-resolution earth observation system, the GF-1 satellite goes beyond optical RS by combining high-temporal-resolution, multi-spectral, and wide coverage and is a leader among international civil optical RS satellites in terms of the comprehensive indices of resolution and width. We used the GF-1 Wide Field-of-View (WFV) satellite data product (http://36.112.130.153:7777/, accessed on 20 July 2022), which was launched in 2013. The spatial resolution of all bands is 16 m, with a temporal resolution of 4 days. We used RS image band 4 (*Red*) and band 3 (Near-Infrared, *NIR*).

Further, we used the Landsat 5 TM Level 1 (L1) and Landsat8 OLI_TIRS L1 products (https://earthexplorer.usgs.gov/, accessed on 20 July 2022). The Landsat 5 TM satellite has seven bands. The spatial resolution of bands 1–5 and 7 is 30 m, and that of band 6 is 120 m. The Landsat8 OLI_TIRS satellite has 11 bands. The spatial resolution of bands 1–7 and 9–11 is 30 m, and band 8 is a panchromatic band with a resolution of 15 m. These two satellites

provide global coverage every 16 days. We used RS image band 3 (*Red*) and band 4 (*NIR*) from Landsat 5 TM, and band 4 (*Red*) and band 5 (*NIR*) from Landsat 8 OLI_TIRS. We also used Landsat 8 OLI_TIRS L2 C2 products; this dataset is the atmospherically corrected surface reflectance from the Landsat 8 OLI/TIRS sensor for the identification of *U. prolifera* and Sargassum by the spectral shape.

The environmental and disaster monitoring and prediction small satellite constellation A/B (HJ-1A/B) was launched in September 2008. The HJ-1A satellite is equipped with a CCD camera and a hyperspectral imager, while the HJ-1B satellite is equipped with a CCD camera and an infrared camera (IRS). We used the HJ-1A/B satellite CCD products of band 3 (*Red*) and band 4 (*NIR*) (http://36.112.130.153:7777/, accessed on 20 July 2022), which have a spatial resolution of 30 m and a temporal resolution of 4 days.

The HY-1C satellite project provides a series of observations in the morning and afternoon, which improves the global coverage ability, enlarges the coverage width of the coastal zone imager, and improves spatial resolution to meet practical needs. We used the L1C product of the HY-1C satellite coastal zone imaging instrument (CZI), which has a spatial resolution of <50 m and a temporal resolution of 3 days (https://osdds.nsoas.org.cn/, accessed on 20 July 2022). Specifically, we used RS image band 1 (*Red*) and band 2 (*NIR*). We also used HY-1C L2A products; this dataset is the rayleigh corrected reflectance from the HY-1C sensor for the identification of *U. prolifera* and Sargassum by the spectral shape.

The MODIS sensor was launched as part of the NASA Earth Observation System series of satellites (https://ladsweb.modaps.eosdis.nasa.gov/, accessed on 20 July 2022). We used the MYD02 QKM and MOD02 QKM products, including the RS images in bands 1 (*Red*) and 2 (*NIR*), which are corrected earth observation data by calibrated and geolocated and with a spatial resolution of 250 m. The temporal resolution of it is nearly 1 day.

3. Method

3.1. RS Data Preprocessing

The ENVI software was used to perform geometric correction was for MODIS, Landsat 8, HY-1C, and HJ-1 images. GF-1 comes with a Rational Ploynomial Coefficient (RPC) file, so we used ENVI for RPC orthographic correction.

3.2. Semi-Automatic Green Tide Extraction Method

3.2.1. Green Tide Extraction Index

The optimized semi-automated method for green tide extraction proposed in this paper was based on the *NDVI*, which in turn is based on the principle that vegetation is more reflective in the *NIR* band: the higher the vegetation cover, the lower the reflection in the *Red* band and the higher the reflection in the *NIR* band [7,24,25]. The *NDVI* reinforces the difference in vegetation reflectance between the *Red* and *NIR* bands and is defined as in Equation (1), where *Red* and *NIR* are the digital number values of the *Red* and *NIR* bands, respectively. In addition, before calculating *NDVI*, we performed geometric correction.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

As the green tide tends to cover water bodies with high aggregation, the water bodies covered by green tide also show high reflectance in the *NIR* band, so the *NDVI* can also be used to monitor the scale and distribution of the green tide. The *Red* and *NIR* bands, which are mainly used in the *NDVI*, are common to most visible RS satellites, so the use of the *NDVI* can improve the universality of the method.

3.2.2. Land Mask

As the spectral characteristics of a green tide and vegetation are similar [15], in this study, land masking was performed on RS images to prevent terrestrial vegetation from interfering with threshold selection.

We used shoreline information and the land vertices of RS images (i.e., the vertex of the pentacle in Figure 2) for flexible masking based on the principle of the region of interest (ROI) extraction; i.e., we selected polygon vertices. Finally, a landmask was generated with a buffer of 20 grids extending from the shoreline to the outer sea. For low-resolution RS images, i.e., those from MODIS, the shoreline information of each image was obtained by using prior known shoreline information in the geographic information system and the latitude and longitude coordinates of RS images for point-to-point distance calculation and matching the point with the shortest distance between prior known shoreline information and RS images. For high-resolution RS images (i.e., those from GF-1, Landsat 5 TM, Landsat 8 OLI_TIRS, HJ-1A/B, and HY-1C), which are larger, the shoreline matching method is more computationally intensive, and the masking efficiency is lower; therefore, to improve the masking efficiency, the shoreline was first extracted directly using our proposed semi-automatic shoreline extraction algorithm [26], and then the next step of masking was performed.
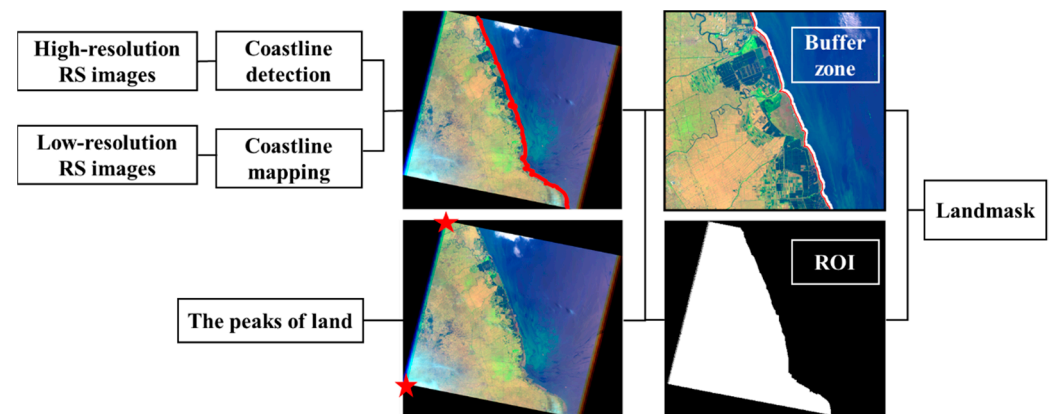

Figure 2. Terrestrial masking process.

3.2.3. Threshold Selection

After removing the influence of terrestrial vegetation, selecting a suitable threshold range to binarize the RS images to separate seawater from the green tide is necessary. The *NDVI* values of RS images were counted using a frequency distribution histogram [27,28], and the statistical principle was to divide the interval of *NDVI* values into k disjoint intervals ($[x_1, x_2), [x_2, x_3), \ldots, [x_{k-1}, x_k]$), divide the sample observations into each interval one by one, and calculate the frequency ni of the samples falling in each interval: the frequency $f = ni/n$, where n is the total number of samples, and the frequency/interval length is given as $[f_i(k) = f(k)/(x_k, -x_{k-1})]$. In this study, we chose $k = 100$ to obtain the frequency histogram of *NDVI* values (Figure 3). Then, using the statistical results $[x_1, f_i(1)], [x_2, f_i(2)], \ldots, [x_k, f_i(k)]$ from Equation (2), the parameters a_1, a_2, b_1, b_2, c_1, and c_2 of the second-order Gaussian fit function were calculated to obtain the final second-order Gaussian fit curve, as shown in Figure 3.

$$H(x) = a_1 \cdot e^{\frac{-(x-b_1)^2}{c_1^2}} + a_2 \cdot e^{\frac{-(x-b_2)^2}{c_2^2}} \tag{2}$$

For vegetation, $-1 \leq NDVI \leq 1$ [29]. Negative values indicate ground coverage by features such as clouds, water, and snow; 0 indicates the presence of features such as rocks and bare soil, whose *NIR* and *Red* are approximately equal; and positive values indicate green tide coverage and increase with increasing coverage [30]. When the green tide coverage of the water body is incomplete, an area with a small amount of green tide distribution may also show negative values [23]; therefore, we chose the first inflection point after the maximum distribution point of the *NDVI* value as the lower limit of the threshold (i.e., the green dot in Figure 3) and 1 as the upper limit of the threshold (elimination of anomalous values) for the discrimination of green tide; i.e., the *NDVI* value less than the

lower limit of the threshold was judged as seawater (blue part in Figure 3), and a value greater than the lower limit of the threshold but less than 1 was indicative of the green tide (green part in Figure 3).
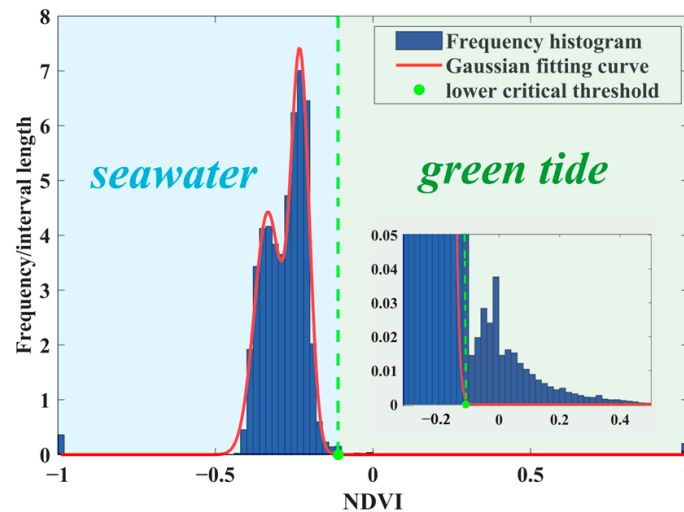

Figure 3. *NDVI* frequency histogram and second-order Gaussian fit curve (The data in the figure are from Landsat 8 (Date: 23 June 2021; Worldwide Reference System: path 119 and row 35).

3.3. Semi-Automated Green Tide Extraction

After selecting a suitable threshold range, the RS image was binarized by threshold segmentation, as seen in Figure 4a, where the white value was 1, representing the green tide, and the black value represented the seawater and the land part that has been masked, separating the seawater and the green tide. As there may also be interfering information, such as ships and thick clouds that cannot be separated by thresholding, and manual ROI is always used to select training samples or feature regions [31,32], so the green tide region was artificially boxed out using an ROI (as seen in Figure 4a, the red polygon boxed area (ROI region)) which was combined with the RS image for further judgment, and finally, the green tide information was obtained with the white value represented by 1 in the ROI region in Figure 4a, and the extraction result of the green tide was shown in Figure 4b. ROI could be used not only to pick out the green tide information but also to remove confounding factors.
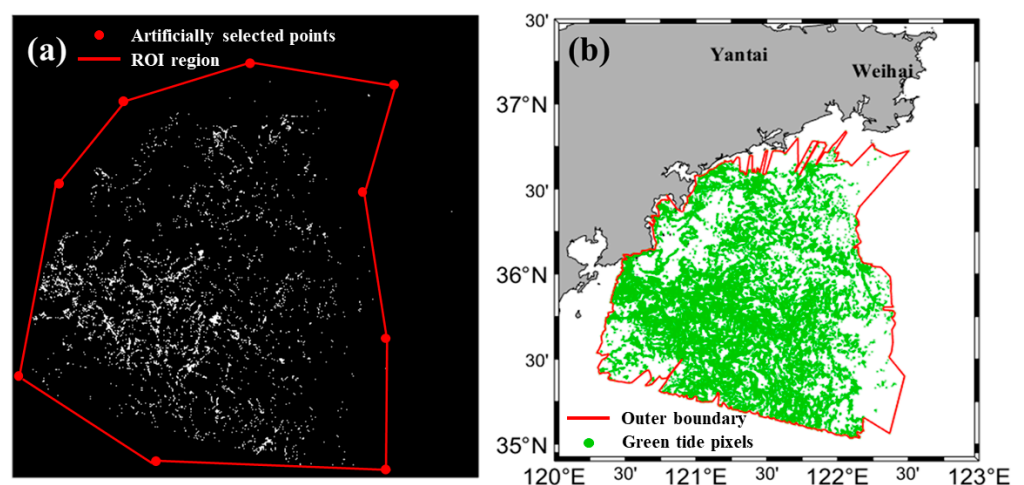

Figure 4. Threshold segmentation separating the seawater and green tide binarized images, (**a**) is the binarized green tide information and (**b**) is the extraction result of green tide, the outer boundary red line is indicated to the distribution of green tide. The data were extracted from Landsat 8 (Date: 23 June 2021; Worldwide Reference System: path 119 and row 35).

3.4. Accuracy Assessment Method

　　　Green tide extraction based on RS is a pixel classification problem in image semantic segmentation, where the green tide elements are defined as positive samples (true value) and the background information as negative samples. The true value was the pixel that we manually interpreted as the green tide based on RS false-color images. For GF-1, we chose band 3 (*Red*), band 2 (*Green*), and band 1 (*Blue*), which correspond to R, G, and B, respectively. For Landsat, we chose band 4 (*Red*), band 5 (*NIR*), and band 4 (*Green*), respectively. For HJ-1 and HY-1C, we both chose band 3 (Red), band 4 (*NIR*), and band 2 (*Green*), respectively. For MODIS, we chose band 1 (*Red*), band 2 (*NIR*), and band 1, respectively. Moreover, the green tide true value pixels were shown as green ones in the image.

　　　The extraction results of all pixels can be classified into four categories (Table 1): true positive (TP) represents the number of correctly classified green tide pixels, true negative (TN) represents the number of correctly classified background pixels, false positive (FP) represents the number of background pixels misclassified as green tide pixels, and false negative (FN) represents the number of green tide pixels misclassified as background pixels. The true value was the pixel that we manually interpreted as the green tide; classification results for TP, FP, FN, and TN were shown in Figure 5 to descript the selection definition of true values. Green tide pixels are usually green in RS false-color images. The circular marks were green tide data points extracted by our method. The green circle represented TP, the pixels that were green in the image and determined by our method to be green tide pixels. The red circle represented FP that was not green in the image but also determined by our method to be green tide pixels. The orange pentagram was green in the image, but it was determined as a background pixel by our method, i.e., FN. TN were the pixels that were not marked, which means these were correctly classified as background pixels.

Table 1. Schedule of precision evaluation formulas.

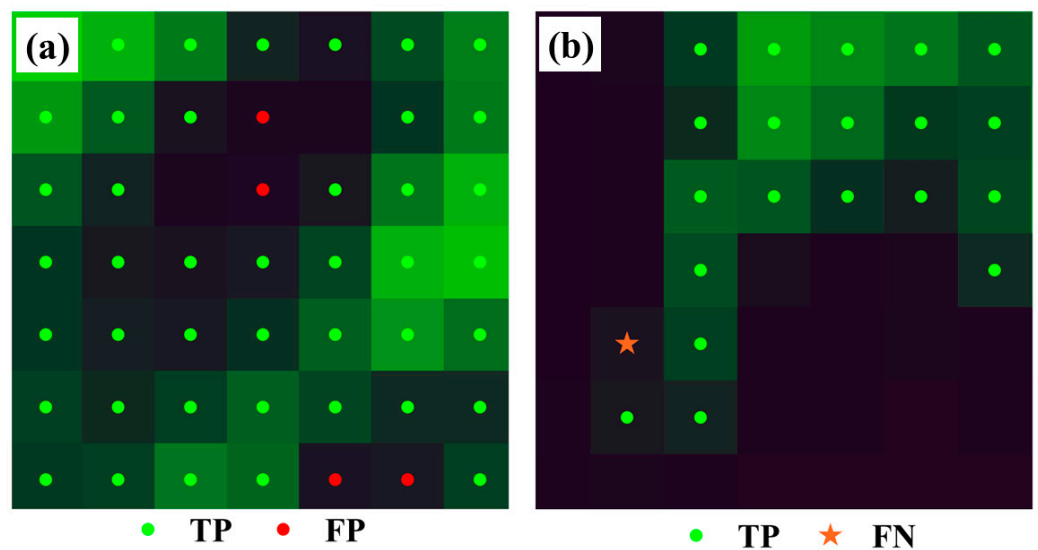|	**Predicted Value = 1**	**Predicted Value = 0**
True value = 1	TP	FN
True value = 0	FP	TN

Figure 5. Classification results for TP, FP, FN, and TN, ((**a**) is classification results for TP and FP, where the green solid circle represents TP, and the red solid circle represents FP. (**b**) is classification results for TP and FN, where the green solid circle represents TP, and the orange pentagram represents FN).

To evaluate the accuracy of our method in extracting green tides from different RS data sources, three commonly used judgment indicators were selected: Precision, Recall, and F1-score. Precision was defined as the percentage of extracted green tide elements (all elements with a predicted value of 1) that were accurate, i.e., Precision = TP/(TP + FP). That was, Precision focused on the ability of green tide pixels to be correctly classified, which was biased to evaluate the percentage of the number of background pixels misclassified as green tide pixels (FP). Recall was defined as the percentage of green tide elements (all elements with a true value of 1) correctly extracted from the RS images, i.e., Recall = TP/(TP + FN). That was, Recall focused on the ability of green tide pixels not to be missed, which was biased to evaluate the percentage of the number of green tide pixels misclassified as background pixels (FN). F1-score was defined as the summed average of Precision and Recall, which was the balance point between the two, i.e., F1-score = 2 (Precision × Recall)/(Precision + Recall). F1-score can comprehensively evaluate the ability of each data to find positive pixels (green tide pixels). Therefore, F1-score was taken as the final indicator to measure the accuracy; the higher the value of F1-score, the better the accuracy of the extraction effect.

The annual average coefficient of determination, R^2, as shown in Equation (3), was calculated to verify the fitting effect of the Gompertz model and the Logistic model.

$$R^2 = 1 - \frac{\sum (f_i - y_i)^2}{\sum (f_i - \overline{y})^2} \tag{3}$$

where f is the fitted value of the growth curve, y is the accumulative cover area (ACA) of the green tide for the corresponding date we input and the $\overline{y}$ is the mean of the input data.

3.5. Area Consistency Validation Method

The combination of multi-source RS images has been widely used for real-time monitoring of the Yellow Sea green tides, but the classification results of different images in the same period and area are usually different. In order to use multi-source RS data uniformly for green tide extraction and study its growth curve prediction, it is necessary to evaluate the consistency from extraction results of different remote sensing data sources. However, due to the fast drift speed of green tide, which can reach roughly 1–5 km/day [33], multi-source RS data will have at least a few hours' intervals, which is enough to produce differences in green tide drift and deformation. Therefore, it is difficult to conduct pixel-level consistency validation, such as Kappa. So, we chose to validate the consistency by comparing the area of the results, as shown in Equation (4).

$$Acc = \left(1 - \left|\frac{S_i - S_j}{S_j}\right|\right) \times 100\% \tag{4}$$

where Acc is the area consistency coefficient of green tide extraction results among each RS satellite data. S_i and S_j are the reference data and the sample data, respectively.

3.6. Approximate Nonlinear Method

The approximate nonlinear method (see Table 2) was used to fit the Gompertz and Logistic curve models separately.

Table 2. Approximate nonlinear method fitting formula.

Curve Model	Gompertz	Logistic
Formula	$y = ae^{-be^{-cx}}$	$y = \frac{a}{1+be^{-cx}}$
Straightening of the curves	$\ln(\ln \frac{y}{a}) = -\ln b + cx$ $y' = \ln(\ln \frac{y}{a}), bb = \ln b$ $y' = bb + cx$	$\ln(\frac{a-y}{y}) = \ln b - cx$ $y' = \ln \frac{a-y}{y}, bb = \ln b$ $y' = bb - cx$
The constant term a	substitute $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ into $\ln a = \frac{\ln y_1 \ln y_2 - \ln y_3^2}{\ln y_1 + \ln y_2 - 2\ln y_3}$	substitute $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ into $a = \frac{2y_1 y_2 y_3 - y_3^2 y_1 - y_3^2 y_2}{y_1 y_2 - y_3^2}$

3.7. Uncertainty Estimate Method

The fitting process is only a statistical correlation, not a deterministic functional relationship, and only the regression value f of the growth curve can be obtained. While the actual value y will be affected by other random factors and must fluctuate around the regression value, and the uncertainty of such fluctuations can be measured by the residual standard deviation S, the formula was as follows:

$$S = \sqrt{\frac{\sum\limits_{i=1}^{n}(y_i - f_i)^2}{n-2}} \tag{5}$$

where f was the fitting regression value of the growth curve, and y was the accumulative coverage area of the actually extracted green tide.

Usually, the fluctuation of the actual value from the regression value f follows a normal distribution. So, for a fixed x value, the y value is distributed symmetrically around the regression value f. The probability of a y value occurring is related to the S. The smaller the S, the more centralized the data, the higher the accuracy of the estimate of y. Therefore, in the process of the approximate nonlinear method, the S was taken as the precision mark of fitting, and two parallel lines are drawn near the regression line $y' = cx + bb$, as shown in Equation (6).

$$\begin{cases} y'_1 = cx + bb - 2s \\ y'_2 = cx + bb + 2s \end{cases} \tag{6}$$

Then, the corresponding curve formula can be obtained according to the approximate nonlinear method. In all possible y values, about 95% of data points will fall in these two curves.

3.8. Growth Curve Area Proportion Calculation Method

As shown in Figure 6, the area and the percentage points corresponding to the start and dissipation times of the green tide, i.e., the statistic P related to the proportion of integral area at the corresponding time (d), was calculated according to the definite integral of the Gaussian fitting result of the first derivative, as shown in Equation (7):

$$P(x_\alpha) = \frac{s_\alpha}{s_{all}} = \frac{\int_{x_1}^{x_\alpha} y'dx}{\int_{x_1}^{x_2} y'dx} \qquad 0 \leq P \leq 1 \tag{7}$$

where x_α is the start and dissipation time of the green tide bulletin; x_1, x_2 are day 1 and day 365, respectively; and S_α, S_{all} are the integral areas corresponding to day x_α and the whole year, respectively. When x_α is the start time of the green tide, $P(x_\alpha)$ is the starting percentile, defined as P_{start}; when x_α is the end time of the green tide, $P(x_\alpha)$ is the dissipation percentile, defined as $P_{dissipation}$.
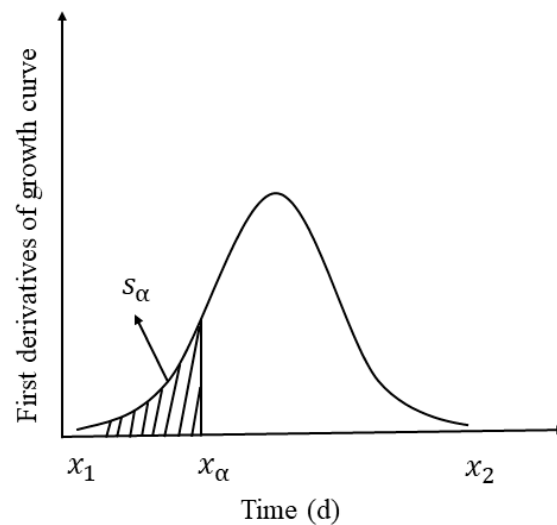
Figure 6. Schematic diagram of percentile points.

3.9. Kurtosis Coefficient

The kurtosis coefficient was calculated as in Equation (8):

$$g_2 = \sqrt{\frac{n}{24}} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y'_i - \overline{y'}}{ss} \right)^4 - 3 \right] \tag{8}$$

$$ss = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y'_i - \overline{y'})^2} \tag{9}$$

where y'_i is the value of the first derivative of the input data; $\overline{y}$ is the mean value of y'_i; and ss is the standard deviation (see Equation (6)). g_2 is dimensionless.

3.10. Identification of U. prolifera and Sargassum by RS

We usually use the spectral properties of macroalgae, which have a distinct reflection peak (700–800 nm) in the near-infrared band, to separate them from seawater. Although this reflection peak separates *U. prolifera* and Sargassum from seawater, the presence of *U. prolifera* can be demonstrated by the other reflection peak appears near the green band (550–570 nm), a spectral feature absents from Sargassum [7,34,35]. The RS images usually contain the spectral information of the above bands (Landsat 5 has spectral bands 1–6: 485, 569, 660, 840, 1676, 2223 nm; GF-1 has spectral bands 1–4: 485, 555, 676, 774 nm; HY-1C has spectral bands 1–4: 460, 560, 650, 825 nm). Therefore, the spectral characteristics of different bands of RS images can distinguish *U. prolifera* and Sargassum, taking GF-1 as an example.

As shown in Figure 7, the black points were the remote sensing reflectance of GF-1 RS data in different bands, and the solid black line connects all these points. The black dotted line connects the mean reflectance of band 1 (*Red* band) and band 3 (*Blue* band), and the formula of the line was fitted by $y = ax + b$. The green line segment represents the difference between $y2$ and $y1$ ($d = y2 - y1$), where $y2$ is the mean reflectance of band 2 (Green band) of each RS datum and $y1$ is the y coordinate of the same x value (band 2) on the black dotted line ($y = ax + b$). If $d > 0$, it meant that there was a reflection peak near the green band, which means this pixel was *U. prolifera*; otherwise, if $d <= 0$, there was no reflection peak which means this pixel may be Sargassum.
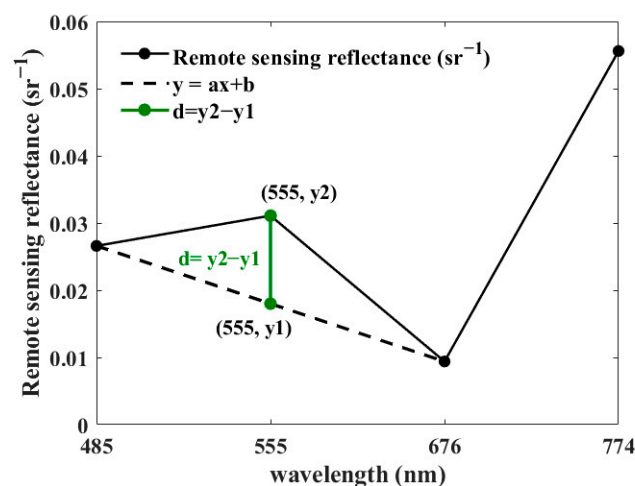
Figure 7. Spectral information of one *U. prolifera* pixel from GF-1 image (6 April 2021).

4. Results

4.1. Semi-Automated Green Tide Extraction

4.1.1. Characteristics of Different RS Data

To compare the extraction results of different satellite data sources, we matched the contemporaneous satellite images of each data source and selected the same area for comparison to ensure that the objective conditions, such as cloud volume, weather conditions, imaging effect, and green tide coverage, of the satellite images, were as similar as possible. Due to the different revisit periods of the satellites, it was difficult to match the data exactly to the same day, so we tried to select a set of data with relatively close dates, and the results of the comparison are shown in Table 3. Even though the selected data for GF-1 and HJ-1 did not match the RS image on the same day as the other data, both of them are the main sources of data for the Yellow Sea Green Tide traceability, so we still added them to the comparison. The region selection schematic diagram and extraction effect of each dataset are shown in Figure 8.

Table 3. Selection regions of the same area size from different data sources.

Satellite	Time of Guardian Film	Resolution (m)	Number of Grids	Corresponding Area (km^2)
GF-1	20 June 2021	16	562 × 562	80.86
Landsat	23 June 2021	30	300 × 300	81.00
HJ-1	22 June 2021	30	300 × 300	81.00
HY-1C	23 June 2021	50	180 × 180	81.00
MODIS	23 June 2021	250	36 × 36	81.00

In terms of the ability of the RS image to display the green tide information, the high-resolution RS images (Figure 8b,d,f,h) presented the green tide information more completely, regardless of whether the green tide was large or sporadically distributed, whereas the low-resolution RS images (Figure 8j) cannot present the sporadic green tide information. In terms of the effect of the extraction method, the green tide information presented by each data source was extracted successfully, although the extraction results were, as expected, significantly better for the high-resolution RS images (Figure 8c,e,g,i) than for the low-resolution images (Figure 8k), especially when the early green tide bloom was sporadically distributed [36]. To sum up, thanks to the advantage of the presentation of sporadically distributed green tides, the high-resolution RS images had absolute advantages over low-resolution images for the traceability of early green tide sources.
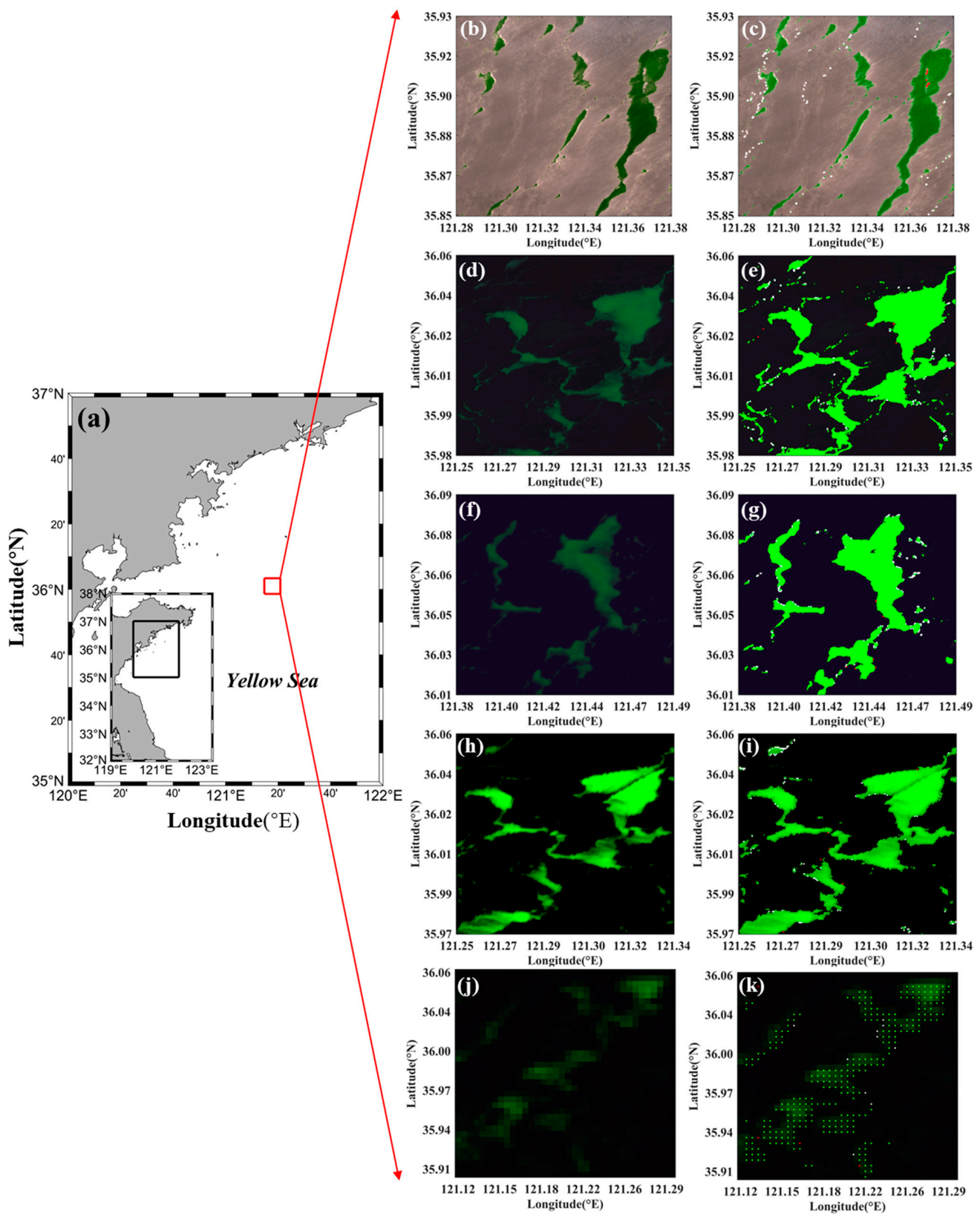
Figure 8. Schematic diagram of the location of the satellite extraction area. (**a**) is the selected approximate area; (**b,d,f,h,j**) are false-color images of GF-1, Landsat, HJ-1, HY-1C, and MODIS, respectively (band combination: GF-1: band 3-2-1; Landsat: band 4-5-3; HJ-1 and HY-1C: 3-4-2; MODIS:1-2-1); and (**c,e,g,i,k**) are the images' respective green tide extraction effects (the red point represents FP, the white point represents FN, and the green point represents TP).

4.1.2. Accuracy Assessment

Accuracy Assessment for Different RS Data

We assessed the accuracy of the different RS data in Table 3. As shown in Table 4, the Precision values of different satellites (GF-1, Landsat, HJ-1, HY-1C, and MODIS) were all up to 98.0%, the Recall values of different satellites were all higher than 96.0%, and all of the F1-score values can reach 97.0%. The average Precision, Recall, and F1-score values of all datasets were 99.6%, 98.3%, and 99.0%, respectively. Regardless of the other indicators (Precision and Recall), or the aggregative indicator F1-score used to evaluate the extraction ability of green tide, all of them were up to 97.0%. In conclusion, our method worked well for both high- and low-resolution RS images for green tide extraction, indicating that our extraction method can be applied to the images of most satellites.

Table 4. Accuracy evaluation of the extraction effect.

Satellite	Precision $\frac{TP}{TP+FP}$	Recall $\frac{TP}{TP+FN}$	F1-Score $2 \times \frac{Precision \times Recall}{Precision + Recal}$
GF-1	99.8%	99.3%	99.6%
Landsat	99.9%	99.0%	99.5%
HJ-1	99.9%	98.5%	99.2%
HY-1C	99.9%	98.3%	99.1%
MODIS	98.6%	96.6%	97.6%
Mean	99.6%	98.3%	99.0%

Accuracy Assessment for Different Environments

We uniformly selected Landsat RS images with high spatial resolution, a large amount of data, and comprehensive coverage of the study area as datasets for accuracy assessment in different environments. Detailed information is shown in Table 5. In addition to the comparison of extraction accuracy in different water depth environments, we also included a group covered by thin clouds (C3).

Table 5. Detailed information on RS images for accuracy assessment in different environments.

Zone	Serial Number	Corresponding RS Image	Location of the Images
A (Depth of < -35 m)	A1	Landsat 8 (4 June 2020)	121.17°–121.27°E, 35.65°–35.73°N
	A2	Landsat 9 (25 June 2022)	120.53°–120.63°E, 35.67°–35.75°N
	A3	Landsat 8 (23 June 2021)	121.25°–121.35°E, 35.98°–36.06°N
B (Depth of ≥ -35 m)	B1	Landsat 8 (23 June 2021)	121.68°–121.77°E, 34.95°–35.03°N
	B2	Landsat 8 (23 June 2021)	121.93°–122.03°E, 34.51°–34.59°N
	B3	Landsat 8 (2 June 2019)	121.67°–121.77°E, 34.58°–34.66°N
C (Muddy water area)	C1	Landsat 8 (22 May 2021)	120.85°–120.94°E, 33.73°–33.81°N
	C2	Landsat 8 (23 June 2021)	121.30°–121.39°E, 33.98°–34.06°N
	C3 (Think cloud)	Landsat 8 (24 May 2016)	120.56°–120.66°E, 33.87°–33.95°N

As shown in Figure 9, the minimum number of pixels needed for the accuracy assessment was around 10^4, and the optimal number of pixels was around $10^5 \sim 10^6$. So the total number of pixels in Table 5 was nearly on the order of 10^5.
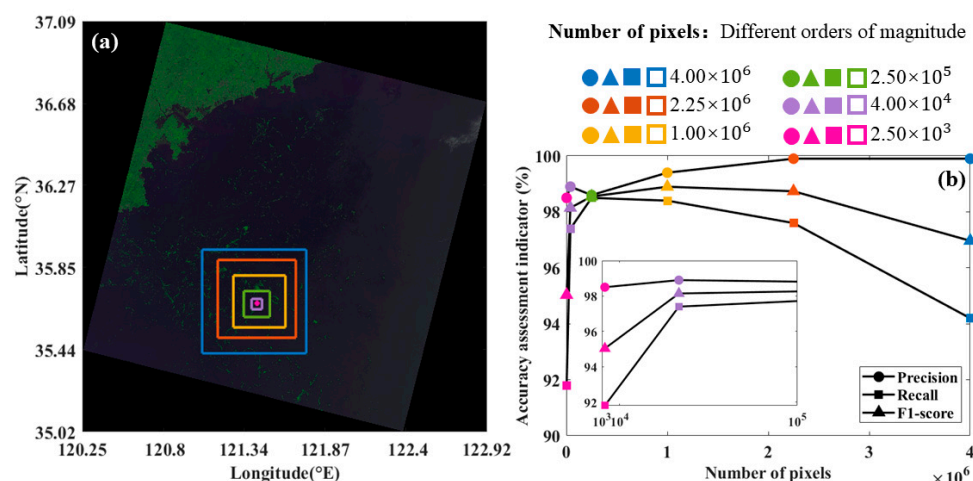
Figure 9. Accuracy assessment of extraction results with different orders of magnitude of pixel numbers (RS data: Landsat 8, 23 June 2021, Worldwide Reference System: path 119 and row 35). (**a**) is a schematic diagram of extraction regions of different orders of magnitude pixel numbers and (**b**) is the corresponding accuracy of different orders of magnitude of pixel numbers.

The accuracy assessment results of the extraction effect in different environments are shown in Table 6. The Precision was not much different across different environments and was as high as 99.0%. This was because the *NDVI* value of the main background pixel (seawater pixel) was mostly negative, and the threshold value we automatically selected was generally positive. Therefore, it was rare to misclassify the background pixel as a green tide pixel (FP). The Recall varied little under different water depths, but it was significantly higher than that in muddy water areas (Zone C) and under thin cloud cover (C3).

Table 6. Accuracy assessment results of the extraction effect in different environments.

Zone	Serial Number	Precision $\frac{TP}{TP+FP}$	Recall $\frac{TP}{TP+FN}$	F1-Score $2\times\frac{Precision\times Recall}{Precision+Recal}$	Coverage Rate $\frac{TP}{300\times300}$
A (Depth of <-35 m)	A1	99.6%	90.0%	94.6%	7.79%
	A2	99.8%	96.5%	98.1%	10.4%
	A3	99.9%	99.0%	99.5%	18.9%
B (Depth of ≥-35 m)	B1	99.9%	98.3%	99.1%	13.1%
	B2	99.9%	98.4%	99.1%	29.6%
	B3	99.9%	94.6%	97.2%	8.59%
C (Muddy water area)	C1	99.9%	84.2%	91.4%	5.70%
	C2	99.9%	93.4%	96.6%	12.6%
	C3 (Think cloud)	99.8%	69.5%	82.0%	5.67%

There were still some differences in accuracy results within the same zone with a similar environment, such as A1 and A2. We think this was related to the proportion and distribution of green tide in the selected area and the amount of sporadic floating green algae. When the amount of green tide in a one-pixel grid was small (the case of sporadic floating green algae pixel), it could easily be misclassified as a background pixel (FN). Therefore, we used the Coverage rate to roughly reflect the distribution of green tide in the selected area; the larger the value, the higher the quantity and aggregation degree of green tide in the area. As we can see in Table 6, the higher the value, the higher the accuracy.

In general, the results of the comprehensive evaluation index F1-score were all above 90%, except C3, which is covered by thin clouds.

4.2. Area Consistency Validation Results

According to Equation (3) in Section 3.5, the area consistency coefficient (Acc) of green tide extraction results among remote sensing satellite data in Table 3 was calculated, and the results were shown in Table 7. The Acc between Landsat and HY-1C was the best, reaching 96%. The Acc between MODIS, Landsat, and HY-1C was also between 85% and 90%. However, GF-1 and HJ-1 had relatively low Acc with the other three, which was because the other data were on the same day, and the time was relatively closer, so the drift and deformation of the green tide were relatively small, while GF-1 and HJ-1 differed by 3 and 1 days, respectively, resulting in lower Acc of the area. The difference between GF-1 and HJ-1 was 2 days, and the Acc between them reached 88–89%. The consistency validation results show that the green tide extraction results from multi-source RS data were generally consistent, and multi-source RS data could be used to monitor green tide more accurately and comprehensively in real-time.

Table 7. Area consistency coefficient of green tide extraction results.

Satellite	GF-1	Landsat	HJ-1	HY-1C	MODIS
GF-1	100%	67.6%	89.7%	65.4%	59.2%
Landsat	52.1%	100%	67.2%	96.7%	87.5%
HJ-1	88.6%	75.3%	100%	72.8%	65.9%
HY-1C	47.1%	96.6%	62.7%	100%	90.5%
MODIS	31.0%	85.7%	48.3%	89.5%	100%

4.3. Green Tide Growth Curve

We extracted single-day data on the Yellow Sea green tide from 2008 to 2021 based on cloud-free/low-cloud multi-source RS images to effectively monitor the green tide in real time and analyze its development each year in terms of the annual green tide bloom, development duration, and impact range and route. We chose the commonly used growth curve models—Gompertz [37] and Logistic [38]—to fit the relationship between the ACA and the time of the green tide.

The cover area of the green tide (2008–2021) was extracted from RS images, and an approximate nonlinear method (see Table 2) was used to fit the two curve models separately (Figure 10). The input was the ACA extracted from remote sensing images without cloud or with thin cloud from 2008 to 2021. The output was the formula of Gompertz and the Logistic curve fitted by the approximate nonlinear method described in Table 2. The R^2 (see Equation (3)) was calculated to verify the fitting effect and it was espectively labeled in Figure 10. The mean R^2 of Gompertz model was 0.9740 and was slightly better than that of the Logistic model (0.9720).

The two curves fit similarly to the annual data of green tide, so the best of the two models was chosen to fit the growth curve trends of the green tide cover area for each year. Combined with the approximate nonlinear method, the data from more than three scenes extracted by real-time monitoring could fit the growth curve of the current year and predict the general bloom situation, providing a reference for the next steps toward green tide prevention and control. The larger the amount of basic data, the higher the accuracy of the fitted growth curve.

For any type of optical sensor, *NDVI* is more or less affected by atmospheric aerosol, muddy water, and shallow water layer. Similarly, the change of environmental factors in satellite images will lead to the uncertainty of extraction results. In addition, the ACA depends on surface currents and the number of images. Therefore, an uncertainty for the growth curve was estimated, as shown in Figure 10.
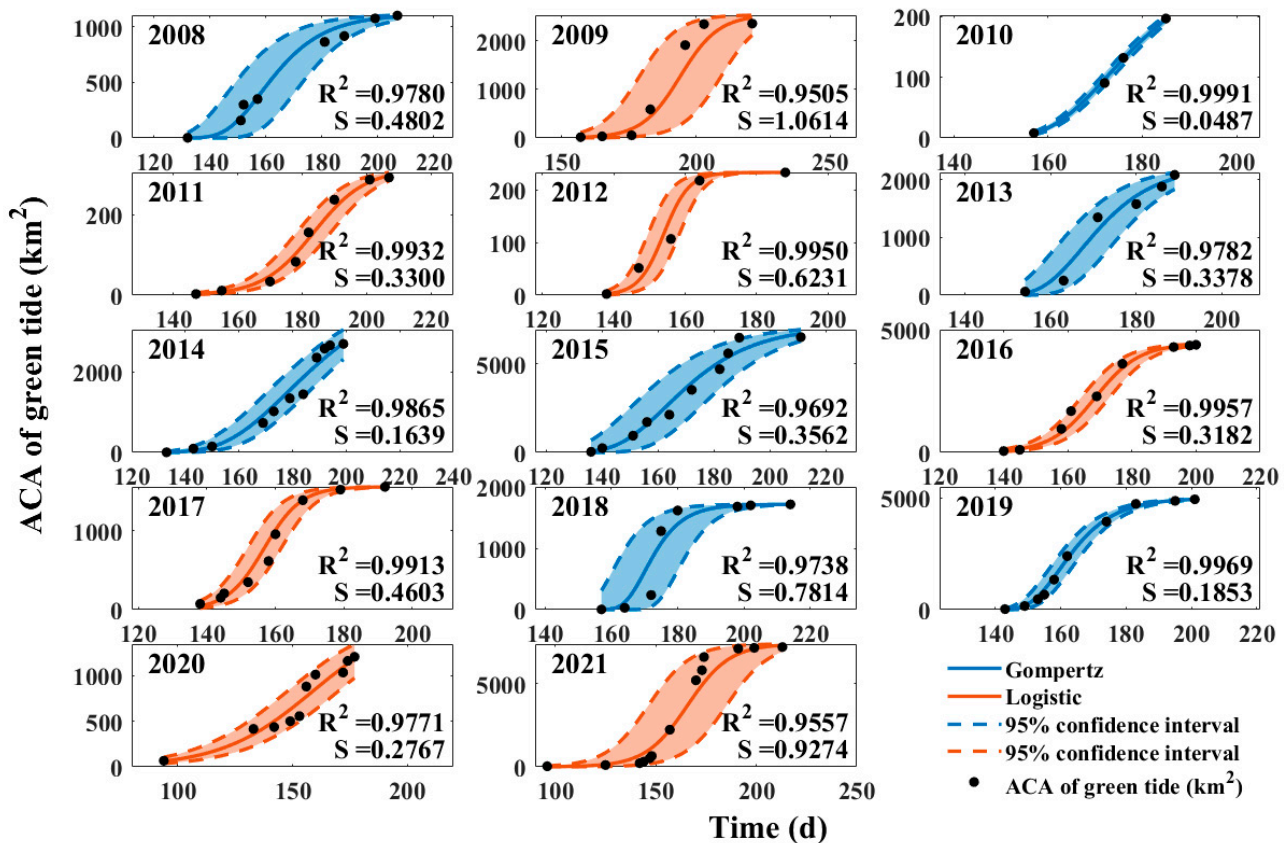
Figure 10. Growth curve fitting effect with 95% confidence interval.

After determining the growth curve of the green tide cover area versus time, the duration of development of the green tide was estimated, and the approximate start and dissipation times of the green tide were projected. We found that the first-order derivative of the growth curve, i.e., the rate of increase in the green tide cover area, conformed to the normal distribution, so a Gaussian fit to the first-order derivative was performed to obtain the curve equation y' (see Figure 11).
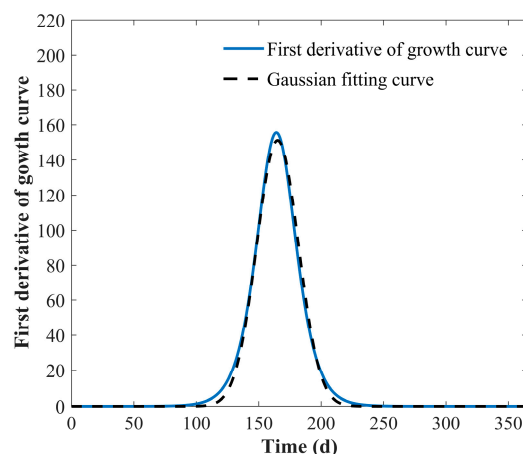
Figure 11. First derivatives of growth curve and its Gaussian fitting curve.

We calculated the proportion area of the first-order derivative of the growth curve corresponding to the start and dissipation times of the green tide according to the growth curve area proportion calculation method described in Section 3.8. The start and dissipation times of the green tide event were determined according to the China Marine Disaster

Bulletin and the Jiangsu Marine Disaster Bulletin, as shown in Table 8. For periods that were not accurate to days, the 15th day was considered as the middle, the 20th day as mid-late, and the 25th day as late.

Table 8. Start and dissipation times of the green tide event.

Year	Start Time	Starting Percentile	Dissipation Time	Dissipation Percentile
2008	May ★	0.832%	August ★	97.64%
2009	24 March ★	0.000%	Late August ★	90.38%
2010	20 April ★	0.000%	Mid-August ★	100.0%
2011	27 May ★	0.000%	21 August ★	100.0%
2012	Late March ★	0.000%	30 August ★	100.0%
2013	Mid to late March ★	0.000%	Mid-August ★	99.98%
2014	30 April ⊼	0.115%	8 September ⊼	99.60%
2015	Mid-April ★	3.934%	14 August ⊼	99.99%
2016	Mid-April ★	1.773%	25 August ⊼	100.0%
2017	Mid-April ★	0.000%	24 July ⊼	100.0%
2018	4 May ⊼	0.000%	11 August ⊼	100.0%
2019	Mid to late April ★	0.000%	1 August ⊼	100.0%
2020	19 March ⊼	4.767%	20 July ⊼	88.92%
2021	Mid-April ★	1.085%	Late August ★	97.28%

Note: ★ indicates data from the China Marine Disaster Bulletin; ⊼ indicates data from the Jiangsu Provincial Marine Disaster Bulletin.

We found that the start percentile $\left(P_{start}\right)$ and dissipation percentile $\left(P_{dissipation}\right)$ were related to the flatness of the Gaussian fit curve y' of the first-order derivative of the growth curve; therefore, we established the relationship between the start and dissipation percentiles and the kurtosis coefficient g_2 of y' to better infer the start and dissipation times of the green tide. First, the kurtosis coefficient of the first derivative y' of the growth curve was calculated as in Equation, and the input data were the growth curve.

The relationships between the kurtosis coefficient of Gaussian fit curve y' of the first derivative of the growth curve and the start and dissipation percentiles were established separately and found to be roughly exponential. According to the principle of the least squares method, the curves of the start and dissipation percentiles and the kurtosis coefficients were fitted, as shown in Figure 12. By combining these relationships with the growth curve fitted using the approximate nonlinear method, it was possible to predict the start and dissipation times of the Yellow Sea green tide and the extent of the green tide bloom based on the predicted ACA of the green tide. These predictions provide a reference basis for future work on green tide management.
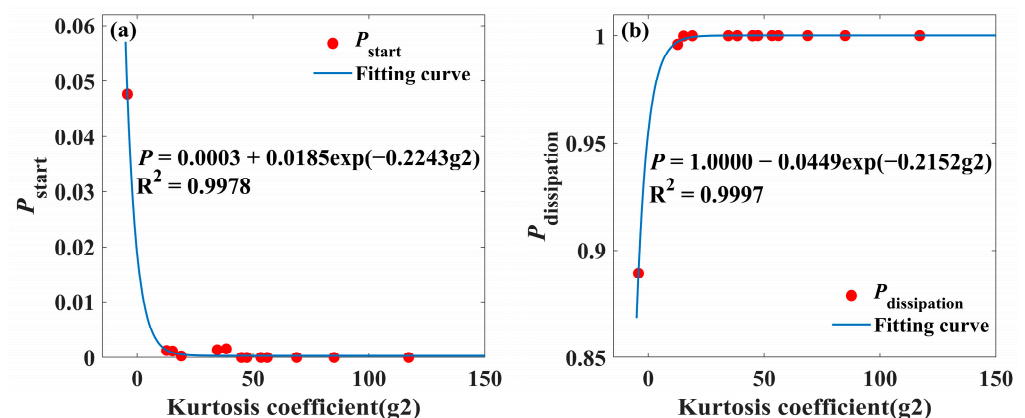

Figure 12. Curvilinear relationships between the start and dissipation percentiles and the kurtosis coefficients ((**a**) is curvilinear relationships between the start percentiles and the kurtosis coefficients and (**b**) is curvilinear relationships between the dissipation percentiles and the kurtosis coefficients).

4.4. Green Tide Trend Forecast for 2022

Data for predicting the 2022 green tide were extracted from multi-source RS images. The available data were for four scenes, which were combined with the approximate nonlinear method to fit the predicted growth curve of this green tide for 2022. According to the prediction results (Figure 13), the bloom time was expected to be from 1 June to 15 August 2022, the duration was expected to be 73 days, and the ACA was expected to be approximately 1190.90–1191.21 km^2. Compared with the green tide data from recent years, the predicted bloom duration and ACA in 2022 were respectively shorter and smaller, suggesting that this would be a smaller green tide bloom event in the Yellow Sea than those seen in recent years (see Figure 14).
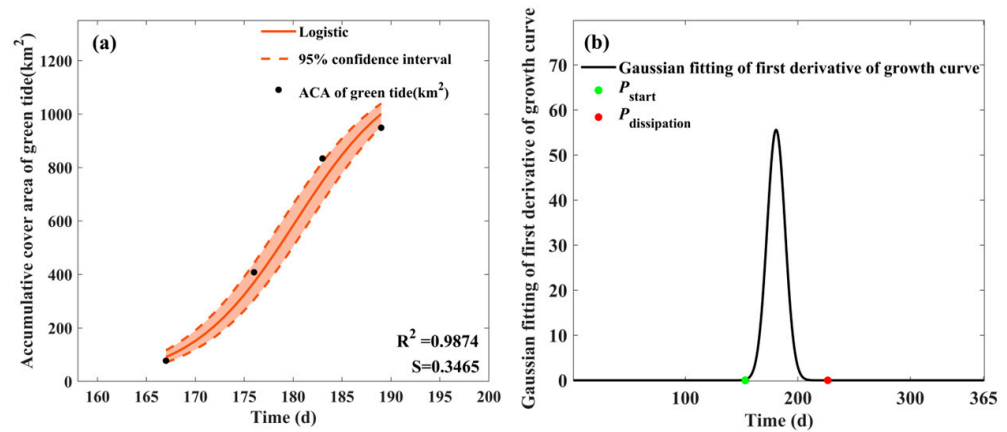
Figure 13. Forecasted start and dissipation times of the Yellow Sea green tide for 2022 ((**a**) is Logistic growth curve fitting effect of the ACA of the Yellow Sea green tide (2022) and (**b**) is the forecast results of start and dissipation times of the Yellow Sea green tide for 2022).
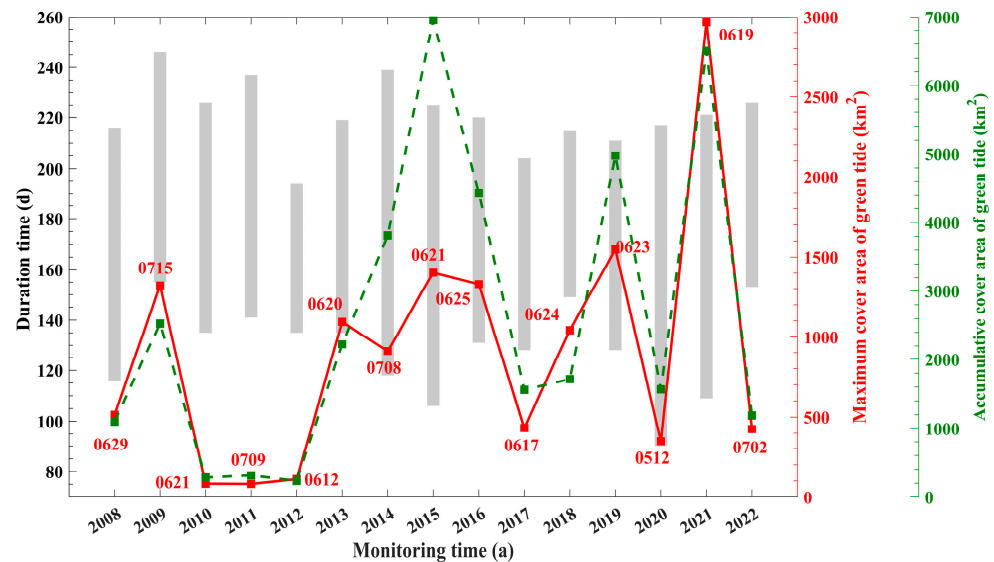
Figure 14. Results of simulation of the bloom duration and accumulative and maximum cover area of the Yellow Sea green tide from 2008 to 2022.

For the true condition of 2022, according to the Ministry of Natural Resources of the People's Republic of China, the 2022 pre-salvage operation of the Yellow Sea green tide began on May 19 and finished around mid-August in coastal cities such as Yantai. In addition, the bloom was much smaller than last year. The actual duration and bloom condition were roughly in line with our predictions.

In fact, there have been a lot of artificial measures for the Yellow Sea green tide in recent years, such as artificial salvage, which is bound to have an impact on prediction.

However, we believe that these human factors may affect the accuracy of the results over a short period but will not affect the general direction of the forecast. One was that the main function of the prediction model was to get a rough estimate of the size of the green tide blooms and to know its bloom duration and inflection point. The second was that monitoring was a dynamic process; for example, we did not know whether this RS image we selected was before or after artificial salvage. In addition, salvage and defense measures are mainly inshore, so the impact on the center of the sea is limited.

However, the positive impact of human intervention is also one of the hot topics of attention, and we will pay further attention to this content. In addition to human factors, climate factors can also impact the green tide. Therefore, we also tried a predicted model considering climate factors, but the effect of the model was not ideal, and the stability was not good, so we did not adopt it.

5. Discussion

5.1. Green Tide Traceability

Due to the influences of temporal and spatial resolution, cloud, rain, fog, and other weather factors, the earliest time range for green tide monitoring by RS is generally from mid-late May to early June [25,39], and the traceable range is generally limited to the Subei shoal along the Jiangsu coast. To obtain an earlier time range for monitoring and a wider range of green tide traceability locations, the newly proposed green tide extraction algorithm was used to extract data on the Yellow Sea green tide from 2008 to 2022 through multi-source high-resolution RS images, and the source of the Yellow Sea green tide was successfully traced. The main data of the high-resolution RS satellites used in this study and their in-orbit operation times are shown in Table 9. HY-1C and GF-1 advanced the source tracing time to early April (as shown in Table 9), and the overall tracing time of the multi-source satellite images was advanced to early May. We used multi-source high-resolution RS data to trace the source of the Yellow Sea green tide as far as possible, with a spatial resolution between 16 m and 50 m. This resolution range can monitor part of sporadic floating *U. prolifera*, as shown in Figure 8. Even if we cannot monitor the very early process of the green tide, at least we will be able to monitor the green tide when it starts to take shape.

Table 9. In-orbit operation times of high-resolution RS satellites.

High-Resolution RS Data Source	In-Orbit Operation Time
Landsat5	March 1984–June 2013
HJ-1	September 2008–Present
Landsat8	February 2013–Present
GF-1	April 2013–Present
HY-1C	September 2018–Present

After extracting the earliest distribution data of the Yellow Sea green tide in all years (2008–2022), the central location was defined as the source location of the green tide, as shown in Figure 15a. The source locations were mainly clustered in two regions, and the k-means clustering algorithm was used to classify the central locations. The k-means clustering is a method of cluster analysis that is aimed at partitioning the data into k clusters, where each datum belongs to the cluster with the nearest mean [40]. This algorithm is widely used in spatial clustering analysis [41]. The source locations were classified into two categories using the k-means clustering algorithm (as shown in Figure 15b). Based on the tracing results, a source was also found in the southeastern Yellow Sea (blue box), in addition to the Subei shoal along the Jiangsu coast (the most commonly attributed source; red box). The sources of the Yellow Sea green tide were further clarified, which may guide the early prevention, control, and targeted management of the Yellow Sea green tide.
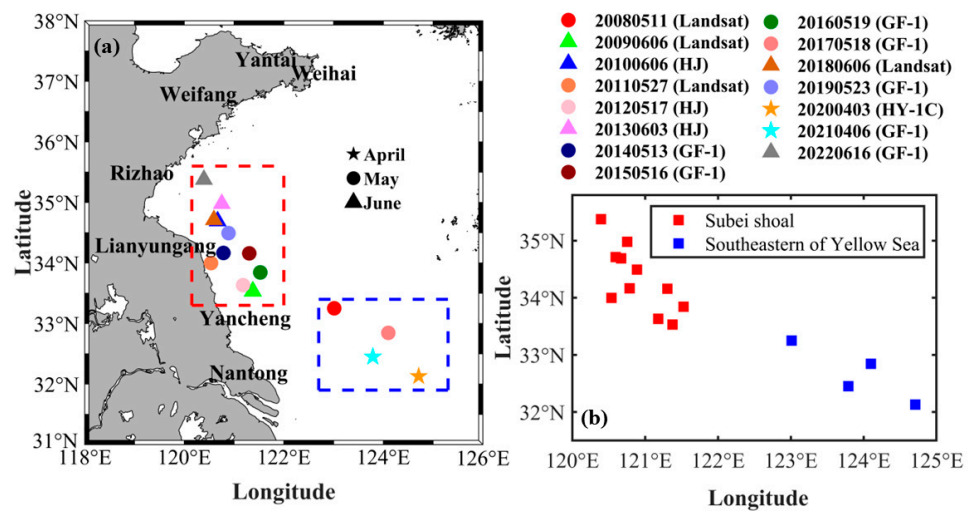
Figure 15. Source site statistics 2008–2022. (**a**) shows the source sites of green tide blooms over the years, and (**b**) shows the results of k-means clustering, which divided the green tide source into two clusters: the Subei shoal (red box) and the southeastern Yellow Sea (blue box).

5.2. Green Tide Source Identification

In addition to *U. prolifera* green tide, Sargassum gold tide also occurred frequently in the Yellow Sea, especially in the southeastern Yellow Sea [42,43]. Therefore, it cannot be ruled out that the green tide source found in the Southeastern Yellow Sea is Sargassum.

According to the identification method described in Section 3.6, we counted the spectral reflectance of the extracted green tide pixels and calculated the mean value of each band. As shown in Figure 16a,c,d shows that 99% of *d* (the green part of the pie chart) was positive, indicating that the majority of our extracted green tide pixels matched this spectral property of *U. prolifera* with a reflection peak near the green band. The reflection peak in Figure 16b was lower than in Figure 16b on the solid black line was not very pronounced, so we selected the pixels that $d > 0$ and drew as a red solid line to demonstrate the presence of the reflection peak near the green band, and thus prove the existence of *U. prolifera*. The pixel proportion of $d > 0$ was about 50% (the green part of the pie chart). The reflection peak near green band (550–570 nm) in Figure 16b is not as obvious as the other ones (Figure 16a,c,d), which was mainly affected by Sargassum.
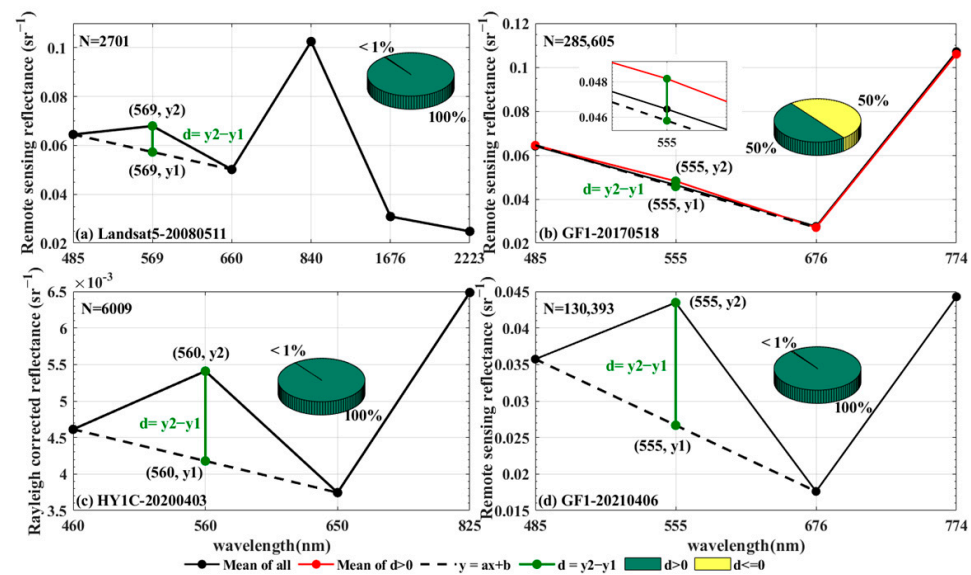

Figure 16. Spectral information of extracted green tide pixels for the source found in the southeastern Yellow Sea (blue box in Figure 6) for 2008, 2017, 2020, and 2021 (where N is the number of pixels).

Therefore, the green tide source of the southeastern Yellow Sea (blue box in Figure 15) was sufficiently possible to exist and should be taken seriously.

As seen in Figure 14, 2015, 2019, and 2021 were the three years with the largest green tide blooms in the Yellow Sea, and it was found that the green tide generally reached its maximum cover area in mid-to-late June to early July, which was consistent with previous studies [44–46]. In addition, the duration of the three years was relatively long, consistent with Zheng et al., 2022.

6. Conclusions

For many years, the repeated green tide blooms in the Yellow Sea have not only posed a great threat to the coastal ecology but also caused huge economic losses along the Yellow Sea coast. Therefore, there is an urgent need to predict future blooms of the Yellow Sea green tide to guide appropriate measures to prevent and control these events. In this paper, we propose a semi-automated method for extracting data on the Yellow Sea green tide based on the *NDVI*, which greatly improved the efficiency of green tide extraction and enabled analysis of the events and monitoring of the green tide in real-time.

In this study, we first compared the presentation of the green tide from various RS data sources (GF-1, Landsat 5 TM, Landsat 8 OLI_TIRS, HJ-1A/B, HY-1C, and MODIS). High-resolution RS images were shown to present the sporadic floating green algae more completely than low-resolution images, with the resolution being highly positively correlated with the completeness of representation of the details. To evaluate the accuracy of the newly proposed semi-automated green tide extraction method, Precision, Recall, and the composite index F1-score were used as indicators. For different satellites, the mean Precision, Recall, and F1-score were all over 98.0%, which indicated that the universality of this method is good. For different environments, the Recall varied little under different water depths, but it was significantly higher than that in muddy water areas (Zone C) and under thin cloud cover (C3). For the results of Precision, there was not much difference between different environments, and it was as high as 99.0%. F1-score were all above 90%, except when the data were covered by thin clouds. Overall, our method can be well adapted to different environments. To obtain earlier and more comprehensive green tide traceability results and promote the early prevention, control, and targeted management of the Yellow Sea green tide, we further advanced the tracing time to early May by using the combination of multi-source RS images, except for HY-1C, for which the tracing time was advanced to early April. Using *k*-means clustering analysis, the multi-year traceability results (i.e., the central location of the early-stage green tide) were divided into two categories: the Subei shoal along the Jiangsu coast and the southeastern Yellow Sea.

To analyze and predict the future development of green tides, single-day data on the Yellow Sea green tide were extracted from 2008 to 2022, the ACA data of the green tide over time were fitted using Gompertz and Logistic growth curves, and the relationships between the start and dissipation percentiles and the kurtosis coefficients were obtained. Using the approximate nonlinear method and the extraction results from at least the three Rs images, the green tide growth curve for 2022 was fitted, and the approximate bloom extent was successfully predicted. Such predictions can provide a reference basis for green tide prevention and control measures. The green tide prediction results showed that compared with recent years, the predicted bloom duration in 2022 would be shorter and have a smaller ACA. That prediction was in line with the actual duration and bloom condition in 2022.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: [http://36.112.130.153:7777/, accessed on 20 July 2022; https://earthexplorer.usgs.gov/, accessed on 20 July 2022; https://osdds.nsoas.org.cn/, accessed on 20 July 2022; https://ladsweb.modaps.eosdis.nasa.gov/, accessed on 20 July 2022].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Extraction result	Data determined as green tide pixels from RS images
Accumulative cover area	The accumulation cover area of extraction result
f	The *NDVI* statistic sample frequency
H	The equation for the second-order Gaussian fitted curve of *NDVI* values
TP	True positive (TP) indicates the number of correctly classified green tide pixels
FN	False negative (FN) indicates the number of green tide pixels that are misclassified as background
FP	False positive (FP) represents the number of background pixels that are misclassified as green tide pixels
TN	True negative (TN) indicates the number of correctly classified background pixels
Precision	The percentage of extracted green tide elements (all elements with a predicted value of 1) that are accurate
Recall	The percentage of green tide elements (all elements with a true value of 1) extracted from the Defender
F1-score	Summed average of Precision and Recall
Acc	The area consistency coefficient
Coverage rate	The roughly distribution of green tide in the selected area
$y1$	The *y* coordinates of the same *x* value point (band 2) on the black dotted line ($y = ax + b$).
$y2$	The mean reflectance of band 2 (Green band) of each RS data
d	The difference between $y2$ and $y1$ ($d = y2 - y1$)
y'	Gaussian fitted curve formula for the first-order derivative of the growth curve
R^2	Fitting coefficient of determination of the growth curve
P	The percentile corresponding to the time of start and dissipation, i.e., the statistic related to the percentage of area integrated with time (days)
S_α	The integral area of the first derivative of the growth curve for day x_α
S_{all}	The integral area of the first derivative of the growth curve for the whole year
$g2$	The kurtosis coefficient of y'
ss	The standard deviation of y'

References

1. Liu, D.; Keesing, J.K.; He, P.; Wang, Z.; Shi, Y.; Wang, Y. The world's largest macroalgal bloom in the Yellow Sea, China: Formation and implications. *Estuar. Coast. Shelf Sci.* **2013**, *129*, 2–10. [CrossRef]
2. Ye, N.-H.; Zhang, X.-W.; Mao, Y.-Z.; Liang, C.-W.; Xu, D.; Zou, J.; Zhuang, Z.-M.; Wang, Q.-Y. 'Green tides' are overwhelming the coastline of our blue planet: Taking the world's largest example. *Ecol. Res.* **2011**, *26*, 477–485. [CrossRef]
3. Qiao, F.; Wang, G.; Lü, X.; Dai, D. Drift characteristics of green macroalgae in the Yellow Sea in 2008 and 2010. *Chin. Sci. Bull.* **2011**, *56*, 2236–2242. [CrossRef]
4. Li, Y.; Song, W.; Xiao, J.; Wang, Z.; Fu, M.; Zhu, M.; Li, R.; Zhang, X.; Wang, X. Tempo-spatial distribution and species diversity of green algae micro-propagules in the Yellow Sea during the large-scale green tide development. *Harmful Algae* **2014**, *39*, 40–47. [CrossRef]
5. Zhou, M.-J.; Liu, D.-Y.; Anderson, D.M.; Valiela, I. Introduction to the Special Issue on green tides in the Yellow Sea. *Estuar. Coast. Shelf Sci.* **2015**, *163*, 3–8. [CrossRef]
6. Zhang, Y.; He, P.; Li, H.; Li, G.; Liu, J.; Jiao, F.; Zhang, J.; Huo, Y.; Shi, X.; Su, R. Ulva prolifera green-tide outbreaks and their environmental impact in the Yellow Sea, China. *Natl. Sci. Rev.* **2019**, *6*, 825–838. [CrossRef]
7. Cao, H.; Han, L. Drift path of green tide and the impact of typhoon "Chan-hom" in the Chinese Yellow Sea based on GOCI images in 2015. *Ecol. Inform.* **2020**, *60*, 101156. [CrossRef]
8. Cui, J.; Zhang, J.; Huo, Y.; Zhou, L.; Wu, Q.; Chen, L.; Yu, K.; He, P. Adaptability of free-floating green tide algae in the Yellow Sea to variable temperature and light intensity. *Mar. Pollut. Bull.* **2015**, *101*, 660–666. [CrossRef]
9. Hu, C.; Li, D.; Chen, C.; Ge, J.; Muller-Karger, F.E.; Liu, J.; Yu, F.; He, M.X. On the recurrent Ulva prolifera blooms in the Yellow Sea and East China Sea. *J. Geophys. Res. Ocean.* **2010**, *115*, C05017. [CrossRef]
10. Hu, L.; Hu, C.; Ming-Xia, H. Remote estimation of biomass of Ulva prolifera macroalgae in the Yellow Sea. *Remote Sens. Environ.* **2017**, *192*, 217–227. [CrossRef]
11. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
12. Shi, W.; Wang, M. Green macroalgae blooms in the Yellow Sea during the spring and summer of 2008. *J. Geophys. Res. Ocean.* **2009**, *114*, C12010. [CrossRef]
13. Xu, Q.; Zhang, H.; Cheng, Y.; Zhang, S.; Zhang, W. Monitoring and tracking the green tide in the Yellow Sea with satellite imagery and trajectory model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 5172–5181. [CrossRef]
14. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [CrossRef]
15. Xing, Q.; Hu, C. Mapping macroalgal blooms in the Yellow Sea and East China Sea using HJ-1 and Landsat data: Application of a virtual baseline reflectance height technique. *Remote Sens. Environ.* **2016**, *178*, 113–126. [CrossRef]
16. Jiang, X.; Gao, M.; Gao, Z. A novel index to detect green-tide using UAV-based RGB imagery. *Estuar. Coast. Shelf Sci.* **2020**, *245*, 106943. [CrossRef]
17. Liu, D.; Keesing, J.K.; Dong, Z.; Zhen, Y.; Di, B.; Shi, Y.; Fearns, P.; Shi, P. Recurrence of the world's largest green-tide in 2009 in Yellow Sea, China: Porphyra yezoensis aquaculture rafts confirmed as nursery for macroalgal blooms. *Mar. Pollut. Bull.* **2010**, *60*, 1423–1432. [CrossRef]
18. Yuan, C.; Xiao, J.; Zhang, X.; Zhou, J.; Wang, Z. A new assessment of the algal biomass of green tide in the Yellow Sea. *Mar. Pollut. Bull.* **2022**, *174*, 113253. [CrossRef]
19. Zhou, F.; Ge, J.; Liu, D.; Ding, P.; Chen, C.; Wei, X. The Lagrangian-based floating macroalgal growth and drift model (FMGDM v1. 0): Application to the Yellow Sea green tide. *Geosci. Model Dev.* **2021**, *14*, 6049–6070. [CrossRef]
20. Wang, S.; Zhao, L.; Wang, Y.; Zhang, H.; Li, F.; Zhang, Y. Distribution characteristics of green tides and its impact on environment in the Yellow Sea. *Mar. Environ. Res.* **2022**, *181*, 105756. [CrossRef]
21. Jimenez-Perez, M.; Sánchez-Castillo, P.; Romera, O.; Fernandez-Moreno, D.; Pérez-Martınez, C. Growth and nutrient removal in free and immobilized planktonic green algae isolated from pig manure. *Enzym. Microb. Technol.* **2004**, *34*, 392–398. [CrossRef]
22. Banks, H.; Collins, E.; Flores, K.; Pershad, P.; Stemkovski, M.; Stephenson, L. Statistical error model comparison for logistic growth of green algae (*Raphidocelis subcapitata*). *Appl. Math. Lett.* **2017**, *64*, 213–222. [CrossRef]
23. Li, D.; Gao, Z.; Song, D. Analysis of environmental factors affecting the large-scale long-term sequence of green tide outbreaks in the Yellow Sea. *Estuar. Coast. Shelf Sci.* **2021**, *260*, 107504. [CrossRef]
24. Cui, T.-W.; Zhang, J.; Sun, L.-E.; Jia, Y.-J.; Zhao, W.; Wang, Z.-L.; Meng, J.-M. Satellite monitoring of massive green macroalgae bloom (GMB): Imaging ability comparison of multi-source data and drifting velocity estimation. *Int. J. Remote Sens.* **2012**, *33*, 5513–5527. [CrossRef]
25. Cao, Y.; Wu, Y.; Fang, Z.; Cui, X.; Liang, J.; Song, X. Spatiotemporal patterns and morphological characteristics of Ulva prolifera distribution in the Yellow Sea, China in 2016–2018. *Remote Sens.* **2019**, *11*, 445. [CrossRef]
26. Yu, T.; Xu, S.; Tao, B.; Shao, W. Coastline detection using optical and synthetic aperture radar images. *Adv. Space Res.* **2022**, *70*, 70–84. [CrossRef]
27. Anbazhagan, S.; Paramasivam, C. Statistical correlation between land surface temperature (LST) and vegetation index (NDVI) using multi-temporal landsat TM data. *Int. J. Adv. Earth Sci. Eng.* **2016**, *5*, 333–346.

28. Torres-Sánchez, J.; López-Granados, F.; Pena, J.M. An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Comput. Electron. Agric.* **2015**, *114*, 43–52. [CrossRef]

29. Myneni, R.B.; Hall, F.G.; Sellers, P.J.; Marshak, A.L. The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 481–486. [CrossRef]

30. Bhandari, A.; Kumar, A.; Singh, G. Feature extraction using Normalized Difference Vegetation Index (NDVI): A case study of Jabalpur city. *Procedia Technol.* **2012**, *6*, 612–621. [CrossRef]

31. Zhang, G.; Wu, M.; Wei, J.; He, Y.; Niu, L.; Li, H.; Xu, G. Adaptive threshold model in google earth engine: A case study of Ulva prolifera extraction in the south yellow sea, China. *Remote Sens.* **2021**, *13*, 3240. [CrossRef]

32. Huang, X.; Weng, C.; Lu, Q.; Feng, T.; Zhang, L. Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas. *Remote Sens.* **2015**, *7*, 16024–16044. [CrossRef]

33. Li, D.; Gao, Z.; Song, D.; Shang, W.; Jiang, X. Characteristics and influence of green tide drift and dissipation in Shandong Rongcheng coastal water based on remote sensing. *Estuar. Coast. Shelf Sci.* **2019**, *227*, 106335. [CrossRef]

34. Sun, D.; Chen, Y.; Wang, S.; Zhang, H.; Qiu, Z.; Mao, Z.; He, Y. Using Landsat 8 OLI data to differentiate Sargassum and Ulva prolifera blooms in the South Yellow Sea. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *98*, 102302. [CrossRef]

35. Qi, L.; Hu, C. To what extent can Ulva and Sargassum be detected and separated in satellite imagery? *Harmful Algae* **2021**, *103*, 102001. [CrossRef]

36. Huo, Y.; Zhang, J.; Chen, L.; Hu, M.; Yu, K.; Chen, Q.; He, Q.; He, P. Green algae blooms caused by Ulva prolifera in the southern Yellow Sea: Identification of the original bloom location and evaluation of biological processes occurring during the early northward floating period. *Limnol. Oceanogr.* **2013**, *58*, 2206–2218. [CrossRef]

37. Winsor, C.P. The Gompertz curve as a growth curve. *Proc. Natl. Acad. Sci. USA* **1932**, *18*, 1. [CrossRef]

38. Nelder, J. The fitting of a generalization of the logistic curve. *Biometrics* **1961**, *17*, 89–110. [CrossRef]

39. Liu, X.; Li, Y.; Wang, Z.; Zhang, Q.; Cai, X. Cruise observation of Ulva prolifera bloom in the southern Yellow Sea, China. *Estuar. Coast. Shelf Sci.* **2015**, *163*, 17–22. [CrossRef]

40. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [CrossRef]

41. Zhang, Z.; Zhang, J.; Xue, H. Improved K-means clustering algorithm. *Image Signal Process.* **2008**, *5*, 169–172.

42. Zhuang, M.; Liu, J.; Ding, X.; He, J.; Zhao, S.; Wu, L.; Gao, S.; Zhao, C.; Liu, D.; Zhang, J. Sargassum blooms in the East China Sea and Yellow Sea: Formation and management. *Mar. Pollut. Bull.* **2021**, *162*, 111845. [CrossRef] [PubMed]

43. Xing, Q.; Guo, R.; Wu, L.; An, D.; Cong, M.; Qin, S.; Li, X. High-resolution satellite observations of a new hazard of golden tides caused by floating Sargassum in winter in the Yellow Sea. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1815–1819. [CrossRef]

44. Li, D.; Gao, Z.; Xu, F. Research on the dissipation of green tide and its influencing factors in the Yellow Sea based on Google Earth Engine. *Mar. Pollut. Bull.* **2021**, *172*, 112801. [CrossRef]

45. Qi, L.; Hu, C.; Xing, Q.; Shang, S. Long-term trend of Ulva prolifera blooms in the western Yellow Sea. *Harmful Algae* **2016**, *58*, 35–44. [CrossRef]

46. Zheng, L.; Wu, M.; Cui, Y.; Tian, L.; Yang, P.; Zhao, L.; Xue, M.; Liu, J. What causes the great green tide disaster in the South Yellow Sea of China in 2021? *Ecol. Indic.* **2022**, *140*, 108988. [CrossRef]