# Quantification of floating macroalgae blooms using the scaled algae index

Rodrigo A. Garcia,[1] Peter Fearns,[1] John K. Keesing,[2,3] and Dongyan Liu[3]

Received 18 June 2012; revised 9 October 2012; accepted 5 November 2012; published 18 January 2013.

[1] Quantifying the spatial coverage of floating macroalgae from satellite imagery, using methods such as the normalized difference vegetation index (NDVI) and the floating algae index (FAI), requires the use of a scene-wide threshold to isolate and then compute the number of floating macroalgae pixels. The problem faced is the sensitivity of the NDVI and, to a lesser extent, the FAI to radiance contributions from atmospheric aerosols and turbid water. Both these factors can vary significantly across a satellites' field-of-view generating irregular apparent reflectance of ocean and floating macroalgae pixels across an NDVI/FAI scene, leading to inaccuracies in spatial coverage estimates. We present a simple image processing algorithm, termed the scaled algae index (SAI), that removes any variability present in ocean and floating macroalgae pixels in NDVI or FAI imagery. The SAI does this by subtracting a given pixel's index by that of a local ocean pixel, effectively scaling ocean pixels to values near zero, and macroalgae pixels to positive values. The SAI algorithm has been tested on NDVI and FAI scenes of the 2008/2009 floating macroalgae blooms that occurred in the Yellow Sea, China. These SAI images show a major reduction in variability with scene-wide histograms being unimodal. Histogram analysis also indicates that sufficient contrast exists between ocean and floating macroalgae pixels to enable segmentation by a scene-wide threshold. A semiautomated threshold determination procedure is also presented, which together with the SAI algorithm can be used to compute accurate estimates of the spatial coverage of floating macroalgae.

**Citation:** Garcia, R., P. Fearns, J. K. Keesing, and D. Liu (2013), Quantification of floating macroalgae blooms using the scaled algae index, *J. Geophys. Res. Oceans*, *118*, 26–42, doi:10.1029/2012JC008292.

## 1. Introduction

[2] The use of satellite remote sensing for monitoring the initiation, growth, and decline of algal blooms [e.g., *Keesing et al.*, 2011] is crucial for effective coastal management. Information regarding the spatial coverage of algal blooms can help coastal resource managers make informed decisions that can potentially reduce any subsequent environmental damage [*Sneller et al.*, 2003], or help trace the source of the algal blooms [e.g., *Liu et al.*, 2009].

[3] The *Ulva prolifera* blooms over the Yellow Sea, China in 2008, 2009, and 2010 [*Hu and He*, 2008; *Liu et al.*, 2009; *Hu et al.*, 2010] were novel due to the size of the blooms ($>4000\,km^2$) [*Keesing et al.*, 2011] and their being caused by the expansion of coastal seaweed aquaculture [*Liu et al.*, 2009, 2010]. The blooms are initiated each spring when thousands of tons of waste *U. prolifera*, which fouls the culture rafts used to grow nori (*Porphyra yezoensis*) on extensive coastal sand flats in Jiangsu Province, are discarded into the sea. From there the organic waste floats offshore to the north and west, transported by wind-driven surface currents, growing rapidly in the warm nutrient rich waters of the Yellow Sea and eventually washing up on the shores of the Shandong Peninsula. The huge biomass of algae has resulted in significant economic losses to tourism and aquaculture [*Keesing et al.*, 2011]. In response to the first bloom in 2008, attempts were made to track the bloom using remote sensing methods and several satellite image algorithms were developed that both highlighted the algal bloom and enabled the quantification of its spatial coverage. The buoyant nature of *U. prolifera* and its close similarity to the reflectance spectrum of terrestrial vegetation has lead to the application of algorithms such as the normalized difference vegetation index (NDVI) [e.g., *Hu and He*, 2008], and more recently, the floating algae index (FAI) [*Hu*, 2009] to its remote detection in satellite imagery.

[4] Floating macroalgae can also be detected using satellite image algorithms that highlight chlorophyll fluorescence, such as $F_{sat}$—the chlorophyll fluorescence product from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor [*Behrenfeld et al.*, 2009]—and the maximum chlorophyll index (MCI) [*Gower et al.*, 1999, 2005, 2006]. These algorithms, however, also

[1]Curtin University, Remote Sensing and Satellite Research Group, Perth, Australia.
[2]CSIRO Wealth from Oceans Flagship, Marine and Atmospheric Research, Wembley, Australia.
[3]Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, Shandong, China.

Corresponding author: R. A. Garcia, Remote Sensing and Satellite Research Group, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. (rodrigo.garcia@postgrad.curtin.edu.au)


2169-9275/13/2012JC008292

highlight elevated levels of chlorophyll caused by phytoplankton blooms, with the MCI additionally detecting submerged aquatic vegetation (SAV) in shallow waters [*Gower et al*., 1999]. Additional information is therefore required to distinguish macroalgae from phytoplankton/SAV, making these algorithms unsuitable for simple quantification purposes. In contrast, algorithms such as the NDVI and FAI, applied to an oceanographic scene, will only detect species that exhibit a "red-edge" in their above-water reflectance spectrum.

[5] The NDVI is typically defined as a ratio of the difference between the top-of-atmosphere (TOA) radiance (or reflectance) at red and near-infrared (NIR) wavelengths divided by their sum,

$$NDVI = \frac{(R_{\mathrm{NIR}} - R_{\mathrm{RED}})}{(R_{\mathrm{NIR}} + R_{\mathrm{RED}})} \tag{1}$$

[6] This algorithm utilizes the unique red-edge present in the reflectance spectrum of photosynethically active species to highlight vegetation from the surrounding environment [*Huete et al*., 2002]. The spectral red-edge of vegetation is indicative of low reflectance at red wavelengths due to chlorophyll absorption and high reflectance at NIR wavelengths due to "intra and interleaf scattering in the plant canopy" [*Tucker*, 1979]. This characteristic spectral signature causes vegetation to have an NDVI that ranges approximately between 0.2 and 0.9 (see the histogram analysis presented by *Huete et al*. [2002]).

[7] The use of the NDVI to detect floating macroalgae over the Yellow Sea, China, was initiated by *Hu and He* [2008]. Their investigation showed the ability of the MODIS NDVI algorithm, at 250 m spatial resolution, to: (a) exclusively highlight floating macroalgae from the surrounding clear and turbid waters of the Yellow Sea and (b) estimate the spatial extent of the algae bloom by applying a threshold, subjectively determined from histogram analysis, to the NDVI image. Note that floating macroalgae pixels tend to have a higher NDVI than the immediate surrounding water pixels.

[8] Despite the high selectivity of the NDVI, isolating floating macroalgae pixels using a scene-wide threshold tends to be problematic due to the irregular apparent brightness across a scene containing floating macroalgae and ocean pixels. These irregular distributions can arise from the sensitivity of the NDVI to atmospheric aerosols [*Kaufman and Tanre*, 1992] and turbid water, and the fact that MODIS's large field-of-view can potentially capture scenes with spatially varying atmospheric and water turbidity, particularly evident in the Yellow Sea region [*Hu et al*., 2007]. This causes the apparent "illumination" problem in NDVI imagery of oceanic waters, particularly over the Yellow Sea, where ocean pixels in some locations have a higher NDVI than floating macroalgae pixels at other locales. In such cases, using a scene-wide threshold will exclude an amount of algae pixels, which leads to underestimations in the spatial coverage of algae blooms. The amount of this underestimation is dependent on the severity of the NDVI "illumination" problem.

[9] Two methods have been published that attempt to overcome this limitation; the floating algae index (FAI) by *Hu* [2009] and an image-based processing scheme by *Shi and Wang* [2009]. The former is a new index exclusively for species that exhibit a red-edge in their above-water reflectance spectrum over an oceanographic setting. This index, analogous to the MCI, is a measure of the height of the NIR peak relative to a baseline value that is linearly interpolated from adjacent bands in the red and short wave infrared (SWIR) wavelengths.

$$FAI = R_{\mathrm{RC}}(NIR) - R'_{\mathrm{RC}}(NIR) \tag{2}$$

$$R'_{\mathrm{RC}}(NIR) = R_{\mathrm{RC}}(RED) + [R_{\mathrm{RC}}(SWIR) - R_{\mathrm{RC}}(RED)]\frac{\lambda(NIR) - \lambda(RED)}{\lambda(SWIR) - \lambda(RED)} \tag{3}$$

where, $R_{\mathrm{RC}}$ is the Rayleigh corrected top-of-atmosphere reflectance; $\lambda(SWIR)$, $\lambda(NIR)$, and $\lambda(RED)$ are the wavelengths of the SWIR, NIR, and red bands, respectively. The FAI shows significant improvement in reducing the variability observed in regions affected by Sun-glint or a hazy atmosphere. This is due to the nature of the algorithm where it is a measure of the height of the NIR signal relative to the red and SWIR signals. Note that atmospheric aerosols tend to dominate the SWIR radiance signal over ocean [*Wang*, 2007]; thus the inclusion of this band would cause partial atmospheric aerosol correction. *Hu* [2009] performed a sensitivity analysis of the FAI and NDVI to varying atmospheric conditions and showed that the FAI is still sensitive to scene variability—although to a lesser degree than the NDVI. One aspect that *Hu* [2009] ignored was the sensitivity of the FAI to turbid water, which in the Yellow Sea is an important aspect due to its shallow depth and continuous resuspension of very fine sediments of terrigenous origin [*Wang et al*., 2011].

[10] Removing the variability displayed in the NDVI or FAI requires the use of atmospheric and turbid water correction algorithms. Atmospheric correction for ocean products generally uses several combinations of NIR bands to compute the aerosol reflectance in the visible wavelengths [*Gordon and Wang*, 1994; *Wang*, 2007]. This procedure assumes a "black ocean" at NIR wavelengths—i.e., water absorbs all solar downwelling irradiance at these wavelengths and consequently no photons are backscattered through the water-air interface. In such cases the radiance signal detected by the satellite remote sensor is solely caused by aerosols and Rayleigh scattering. In turbid waters, this assumption generally becomes invalid because a high load of suspended solids leads to increased back-scattering in the water column and thus a non-negligible water-leaving radiance contribution in the NIR wavelengths. *Wang* [2007] developed an atmospheric correction algorithm for such cases that uses two SWIR wavelengths to determine the aerosol reflectance at the NIR and visible wavelengths. This method assumes that turbid water backscatters negligible amounts of SWIR radiation to the satellite sensor, and would consequently appear "black". Unfortunately this assumption does not generally hold for the coastal region of the Yellow Sea as illustrated in Figure 1, which shows a transect of the TOA Rayleigh corrected reflectance of the 2130 nm MODIS band across the Yellow Sea (31.51°N/119.70°E to 31.51°N/123°E) for several dates ranging between 2008 to 2011. These dates coincide with days that showed floating macroalgae in MODIS NDVI imagery.

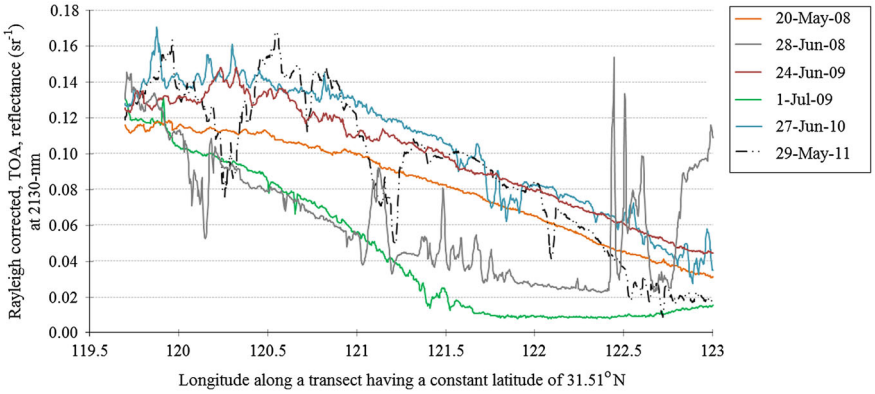


**Figure 1.** A horizontal transect, starting near the coast at 31.51°N/119.70°E and ending at 31.51°N/123°E, of Rayleigh-corrected TOA reflectance ($sr^{-1}$) of the 2130 nm MODIS band.

[11] Figure 1 shows that the reflectance at 2130 nm is highest nearest to the coast (left-hand extent of the figure) and decreases by an average of 30% at a longitude of 122.7°E (right-hand extent of the figure). Inspection of numerous true color MODIS images for this region suggests the increased reflectance nearer to the coast is due to increased water turbidity and enhanced atmospheric optical thickness (haze), most likely having drifted east off the mainland. The radiance contribution from such turbid water is not negligible, and using the SWIR atmospheric correction algorithm [*Wang*, 2007] would lead to an overestimation of the aerosol reflectance in the coastal regions, and thus overcorrections of the reflectance of the visible and NIR wavelengths. Furthermore, atmospheric correction using the MODIS SWIR bands will lead to large uncertainties owing to their low signal-to-noise ratios [*Wang*, 2007]. These two factors would generate both inaccuracies and low precision in the estimated spatial coverage of an algal bloom. An image processing technique that removes variability caused by atmosphere/water turbidity would circumvent these two issues (low precision and accuracy) that the *Wang* [2007] atmospheric correction faces for scenes of the Yellow Sea, and thus is more suitable.

[12] Recently, *Shi and Wang* [2009] developed an image-based method that isolates floating macroalgae from spatially irregular background reflectance. This procedure utilizes the Normalized Difference Algae Index (NDAI), an algorithm equivalent to the NDVI except that Rayleigh-corrected TOA reflectances are used. This imaging procedure is summarized in the following steps: (a) the median value of the NDAI scene is computed and used as a threshold on the NDAI image, where pixels with NDAI ≥ threshold are classed as "algae". This generates a binary image consisting of "algae" and "nonalgae" pixels; (b) a refinement kernel is then applied to the binary image, in which a $10 \times 10$ pixel region is created and centered on a given "algae" pixel. The NDAI values of all "nonalgae" pixels within this $10 \times 10$ pixel region are averaged and subtracted from the NDAI of the central pixel. This is equivalent to subtracting the average ocean value immediately surrounding an "algae" pixel. Step (b) generates a difference image where, according to *Shi and Wang* [2009], pixels with values greater than 0.05 are macroalgae-contaminated pixels.

[13] The refinement kernel in step (b) requires at least two "nonalgae" pixels to be present in a given $10 \times 10$ pixel region in order to compute their average NDAI. Meeting this requirement is problematic because the binary image, generated in step (a), typically contains large "algae" regions that do not have any "nonalgae" pixels (Figure 2b). The refinement kernel cannot be computed in these large "algae" regions, which is a significant issue because blooms typically occur in these regions. Figure 2a shows the NDAI image, captured by MODIS (Aqua) over the Yellow Sea, China, on 20 June 2008. The scene-wide median NDAI value of this image was –0.147 and was used as a threshold to generate the binary image of Figure 2b. Note that land and cloud pixels were excluded from the calculation of the scene-wide median NDAI value. It should be noted that *Shi and Wang* [2009] did not specify: (a) whether land and/or clouds were included or excluded when determining the scene-wide median value; (b) why a $10 \times 10$ pixel region was chosen; or (c) the effect (if any) this $10 \times 10$ "box size" has on the resultant estimates of the spatial coverage of floating macroalgae.

[14] In this paper, we propose an improved semiautomated image-based algorithm, which we have termed the scaled algae index (SAI). This algorithm uses an image kernel to scale a given pixel in an NDVI or FAI image by its local ocean index value to generate a scaled image. The algae pixels in this scaled image can then be segmented with the application of a global threshold. The SAI has previously been used to estimate the spatial coverage of floating macroalgae in the Yellow Sea, China [*Keesing et al.*, 2011]. Since then several improvements have been made to the SAI algorithm, most notably the development of a robust threshold method and an additional scaling routine designed to produce more accurate spatial coverages of algae. This new threshold method does, at the current stage, require minor operator assistance for a given scene and as such the SAI is not an entirely automated procedure. *Keesing et al.* [2011] alluded to the fact that the spatial coverage of algae is sensitive to both the threshold and the size of the processing kernel. In this paper, we investigate in detail the effect these two variables have, and we also state an optimum range of kernel sizes that would likely output the most accurate spatial coverage of algae. The following sections describe the SAI algorithm, its characteristics and segmentation of algae pixels and the optimum kernel and threshold range.

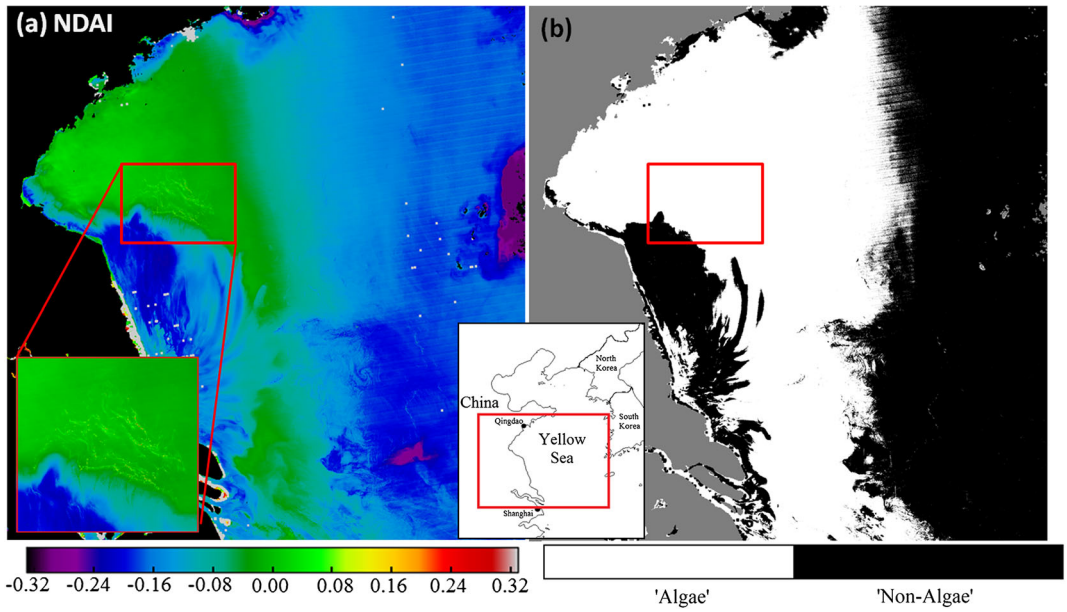


**Figure 2.** (a) NDAI image of the Yellow Sea, on 20 June 2008, bounded by 119°E/31°N and 126°E/37°N. The scene-wide median NDAI value for this image is –0.147. (b) The application of the median NDAI value as the threshold—pixels with values greater or equal to the threshold are classified as algae and are shown in white, those less than the threshold are "nonalgae" and displayed in black. Land pixels are shown as grey, and the region highlighted in red in Figures 2a and 2b shows the location of the algae bloom for this date.

## 2. Methodology

[15] The SAI method was developed using NDVI imagery generated from the 250 m spatial resolution MODIS' red (645 nm) and NIR (859 nm) bands captured over an area of the Yellow Sea, China, enclosed by 119°E/31°N and 126°E/37°N. Prior to the computation of the NDVI, these bands were processed from raw radiance counts (MODIS L0 files were downloaded from http://oceancolor.gsfc.nasa.gov) to Rayleigh-corrected TOA reflectance using SeaDAS 6.1 and then mapped to 250 m spatial resolution using cylindrical equidistant projection. The SAI algorithm was also applied to FAI imagery generated from MODIS' red, NIR, and SWIR (1240 nm) bands. Note that the two former bands have a spatial resolution of 250 m while the latter has 500 m resolution. Before computing the FAI, these three bands were processed to Rayleigh-corrected TOA reflectance as specified by *Hu* [2009], and then mapped to 250 m spatial resolution using a cylindrical equidistant projection.

[16] The SAI algorithm is an image processing technique in which the high variability of an image is removed. The algorithm does this by subtracting a local ocean index value from each pixel in an NDVI or FAI image using the following kernel: an odd-numbered pixel square region (e.g., $3 \times 3$, $15 \times 15$, $21 \times 21$, etc.) is created and centered on a given pixel of an NDVI or FAI image. The index value of the median pixel within the square region (henceforth referred to as a kernel) is computed and subtracted from the index value of the central pixel. This kernel is iterated through all the pixels in an image, except for cloud and land pixels, to generate the SAI. Note that (a) cloud and land pixels were excluded from the median calculation, and (b) the size of the SAI kernel is a highly influential factor in computing the spatial coverage of floating macroalgae, and as such kernel sizes ranging from $3 \times 3$ to $101 \times 101$ were used to evaluate the appropriate kernel size. The reason for using the median pixel rather than the mean when calculating the local ocean index value is that the former is less affected by the presence of algae pixels having high index values. This is particularly important when the SAI kernel is passed through a region heavily populated by floating macroalgae pixels. Throughout this text, the notation $SAI_{INDEX}$(kernel size) will be used, for example, $SAI_{NDVI}(13 \times 13)$ states that the SAI algorithm at a kernel size of $13 \times 13$ was implemented on an NDVI image.

[17] The spatial coverage of floating macroalgae is subsequently computed in two steps: (1) segmentation of a given SAI image using a threshold, where pixels with a greater SAI are classified as "algae", and those less than or equal to the threshold are classed as "nonalgae", and (2) scaling the SAI of a given "algae" pixel to determine its relative coverage of a 250 m × 250 m area (the spatial resolution of a given pixel). As illustrated in the following equation, this scaling is bounded by the threshold value, which represents 0% coverage, and the maximum SAI of the "algae" pixels in the image, which is assumed to represent 100% coverage.

$$Rel.Coverage = \left( \frac{SAI_{\text{Algae Pixel}} - SAI_{\text{Threshold}}}{\max SAI_{\text{Algae Pixel}} - SAI_{\text{Threshold}}} \right) \times (0.25\ \text{km})^2 \quad (4)$$

[18] Summation of the relative coverage of all the algae pixels in an SAI image produces the spatial coverage of algae. In this paper a simple threshold method termed the exclusion method is presented, which determines the SAI value of a suitable threshold by using the frequency

distribution of ocean pixels in SAI imagery. For a given SAI image, these frequency distributions were obtained from the manual selection of multiple ocean regions that did not appear (by "eye") to have algae or cloud pixels. The exclusion method computes the threshold by finding an SAI whose value is greater than a specified proportion of ocean pixels from the selected ocean regions. Thus, for example an exclusion threshold of 99.90% refers to that SAI value, which is greater than 99.90% of the SAI ocean values in the selected ocean regions. The proportions of ocean pixels we have chosen to exclude throughout this paper ranges from 99.50% to 99.99%. Note that an exclusion threshold of 100% was not used, due to the possibility of outliers, which may result in the exclusion of floating macroalgae pixels.

## 3. Results

### 3.1. Characteristics of SAI Imagery

[19] The limitation of using NDVI and FAI imagery for quantifying the spatial coverage of algal blooms is illustrated in the histogram analysis of such MODIS-derived images of the Yellow Sea on 28 June 2008 (Figure 3). During this date the algal bloom, shown as red-yellow pixels, was located in two regions (displayed as regions A and B in Figure 3) that experienced different atmospheric/water conditions. Histogram comparisons (Figures 3b and 3d) between these two regions in the NDVI and FAI show that: (1) histograms for regions A and B are mono-modal, each with a significant leading tail. Analysis has shown that algae pixels constitute this leading tail feature while the ocean pixels form the dominant histogram peak; (2) the NDVI and FAI of ocean and algae pixels in region A are lower than those in region B; and (3) the majority of the algae pixels in region A have lower NDVI and FAI than the ocean pixels in region B. This latter detrimental characteristic demonstrates that a global threshold (such as "High-Threshold" shown in Figures 3b and 3d) would exclude algae pixels from region A and thus lead to underestimations in the spatial coverage of algae.

[20] This bimodal image histogram characteristic displayed by the FAI and NDVI is absent in SAI imagery derived from both these indexes. Figures 4a and 4c shows the $SAI_{FAI}(25 \times 25)$ and $SAI_{NDVI}(25 \times 25)$ images of the same scene and date as in Figure 3. Note that Figures 4b and 4d present the histogram comparison between regions A and B highlighted in the SAI images.

[21] The histograms for regions A and B of the $SAI_{NDVI}$ (Figure 4d) and $SAI_{FAI}$ (Figure 4b) images overlap, such that both these distributions have the same mean and median. Moreover, the algae pixels from regions A and B have the same spread of values, contrary to the NDVI and FAI (Figure 3) where the majority of algae pixels in region A had lower indexes than the ocean pixels in region B. The ocean pixels in regions A and B of the $SAI_{NDVI}$ $(25 \times 25)$ image have been normalized to a narrow spread

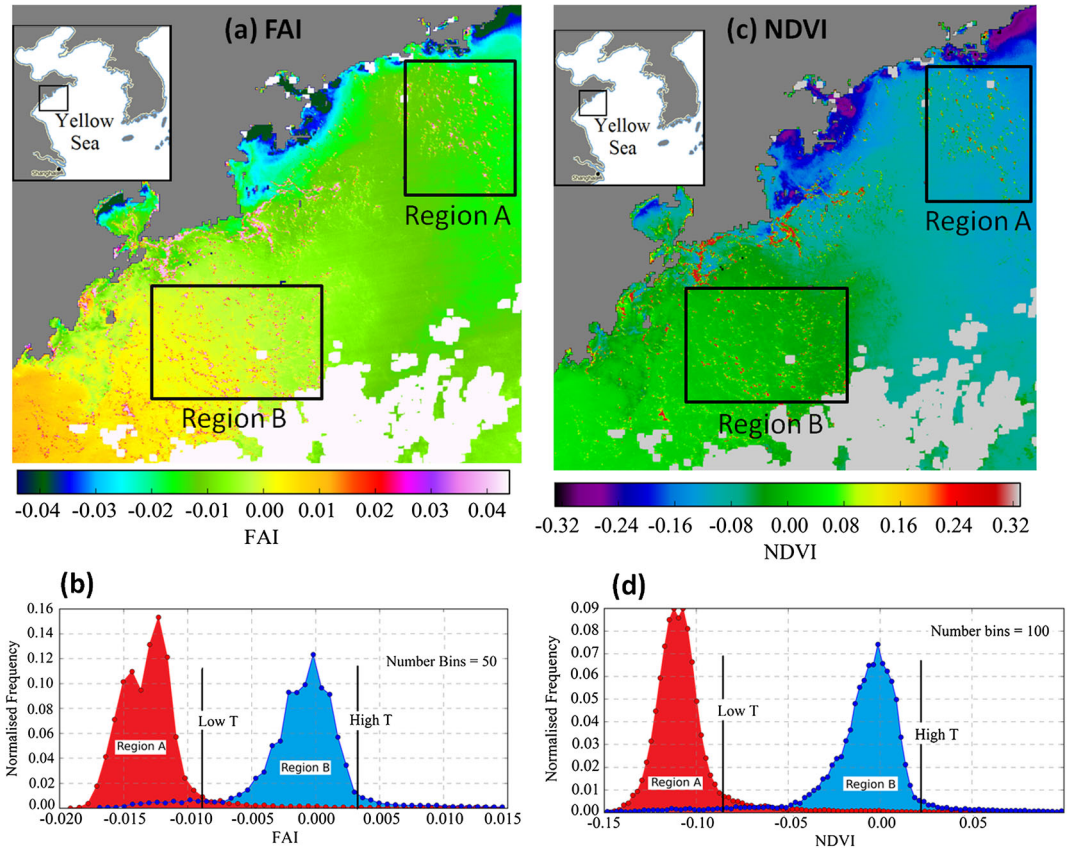


**Figure 3.** MODIS, aqua, (a) derived FAI, and (c) NDVI image of the Yellow Sea, 28 June 2008. The algae bloom is observed as slicks while land and cloud are shown as gray and white pixels, respectively. (b, d) Histogram for regions A and B in the FAI and NDVI image. Note that "Low T" and "High T" stands for "Low Threshold" and "High Threshold", respectively.

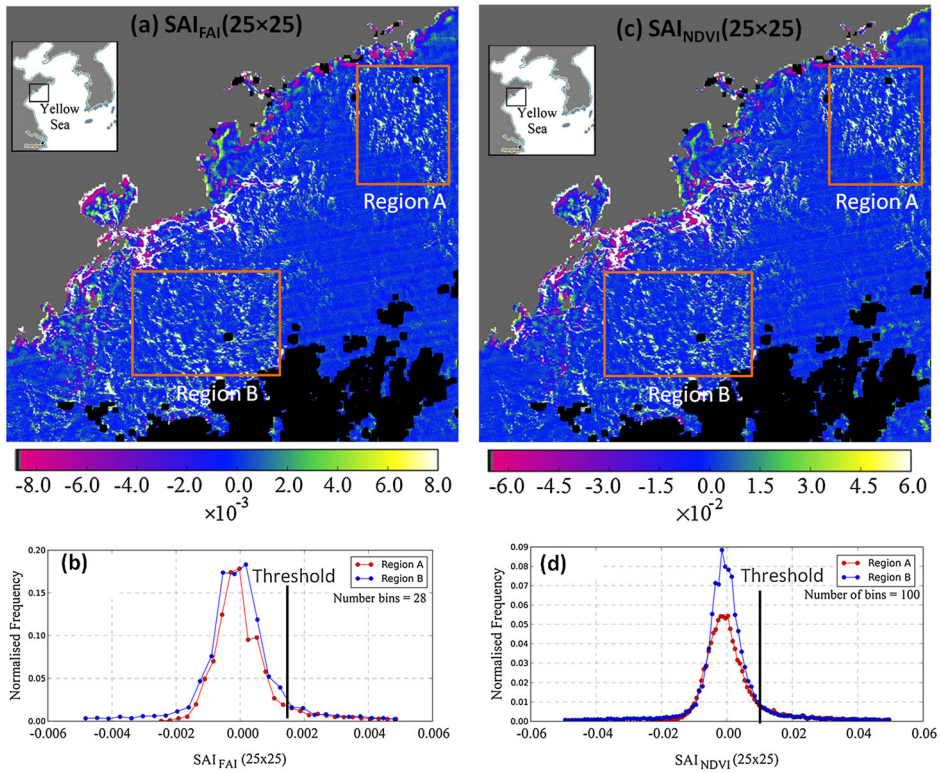


**Figure 4.** SAI derived from (a) FAI and (c) NDVI using a 25 × 25 kernel size, captured on the 28 June 2008 over the Yellow Sea. Land and cloud pixels are masked and shown as grey and black pixels respectively. (b, d) Histogram comparison between regions A (red) and B (blue) in the $SAI_{NDVI}$ and $SAI_{FAI}$ image, respectively.

of indexes ranging from –0.06 to approximately 0.02. The algae pixels, alternatively, form the leading tails of both distributions and range from approximately 0.02 to 0.07. This same characteristic is observed for the histogram comparison of regions A and B of the $SAI_{FAI}(25 \times 25)$ image.

[22] This transformation to a homogeneous image is further demonstrated by the scene-wide histogram comparison between the NDVI and its respective $SAI_{NDVI}(25 \times 25)$ image (Figure 5). The three scene-wide NDVI histograms presented show that the ocean pixels are highly variable forming either negatively skewed (Figure 5a), monomodal (Figure 5b) or multimodal (Figure 5c) distributions. Application of the SAI algorithm, at a kernel size of 25 × 25 to these NDVI images generates histograms that are monomodal, and thus demonstrates the normalizing effect of the SAI algorithm to highly variable scenes.

[23] The size of the image processing kernel has a significant effect on the resultant SAI image. Not all odd-numbered kernel sizes generate the desired, highly uniform SAI image. This is evident by the increase in the intravariability of an SAI image as the kernel size is increased. Figure 6a shows the intravariability expressed as the standard error of the means (SEMs) obtained from 10 randomly selected ocean regions, each having 10,201 pixels. Note that the locations of these 10 ocean regions are different for the different dates analyzed.

[24] Figure 6a highlights the increase in the spread of the 10 means as the kernel size is increased. At kernel sizes ranging from $3 \times 3$ to $21 \times 21$, the SEMs is less than $2.5 \times 10^{-4}$ indicating that the 10 random ocean regions (and hence all the ocean pixels) have very similar SAI values. Note, in an ideal image the background pixels, i.e., ocean pixels in this case, would have a single index value. In such cases, the mean of any given background region would be the same regardless of its location, which would result in an SEM of zero. This intuitively suggests that using the smallest possible kernel size ($3 \times 3$) would be ideal due to the highly uniform image that it would generate. However, investigations have shown the presence of a highly detrimental phenomena (discussed later) at low kernel sizes, which will negate their use. Figure 6a also shows that the SEMs for the different dates increases steadily until a kernel size of $45 \times 45$, past which the SEMs increases nonlinearly, indicating that the SAI images are becoming more variable (less uniform) with increasing kernel size.

[25] Coincident with this increase in the intravariability is the approximate linear increase in the average Euclidean distance (km) between the median and central pixel from 257 m at a kernel size of $3 \times 3$ to 8.51 km at $101 \times 101$ (Figure 6b). Physically this states that the median pixel, on average, progressively increases its distance from the central pixel. As these two pixels become more spatially separated, they would begin to experience different atmospheric/water turbidity radiance contributions, leading to greater differences in their NDVI or FAI. This effectively causes the increase in the intravariability observed in Figure 6a.

[26] Although the creation of a homogeneous background is important in estimating the coverage of floating macroalgae, there are two important factors to consider: (1) the size of the kernel used in generating SAI imagery and (2) the

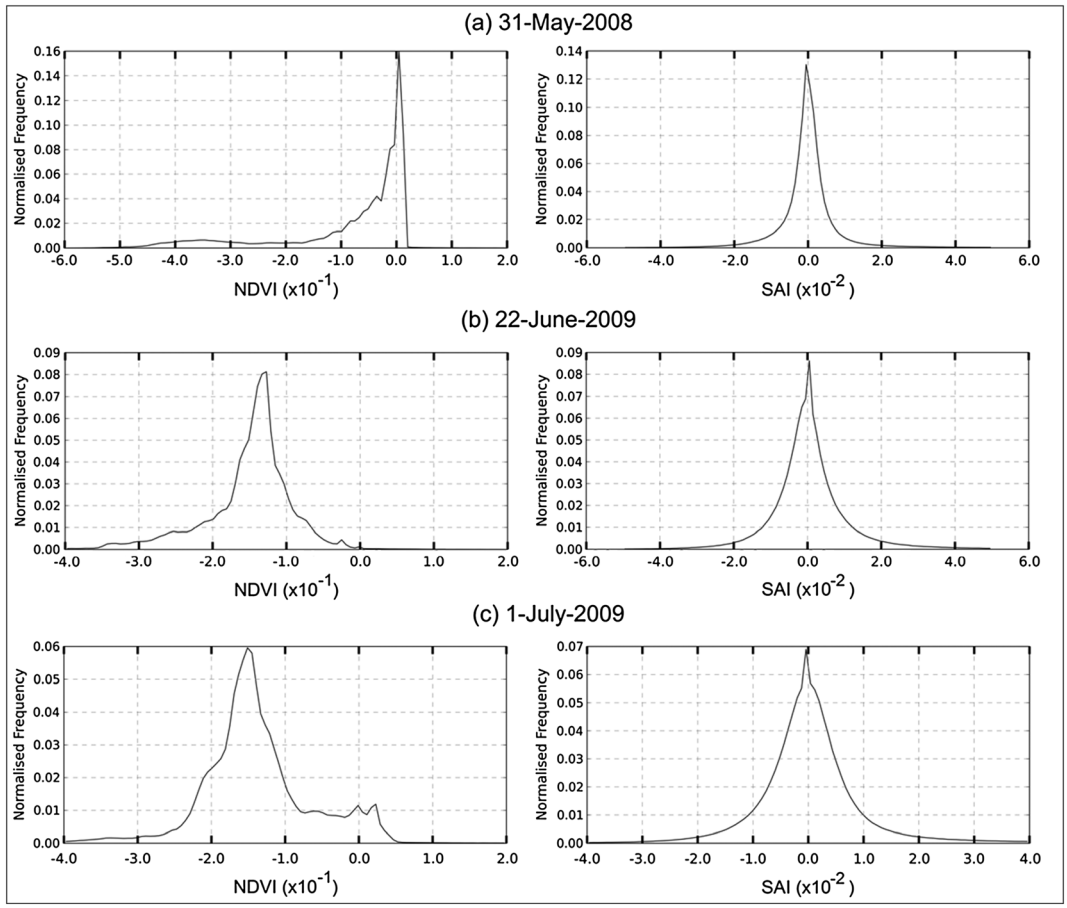


**Figure 5.** Scene-wide histogram comparisons between NDVI (left panels) and $SAI_{NDVI}(25 \times 25)$ imagery (right panels) for (a) 31 May 2008, (b) 22 June 2009, and (c) 1 July 2009. Note that floating macroalgae were observed in these images, and that land and cloud pixels were excluded from the histogram analysis.

threshold used to segment algae pixels. The subsequent sections present results detailing the effect of the kernel size and the validation of the exclusion threshold method.

### 3.2. Image Segmentation of SAI

[27] The exclusion method starts with the manual selection of ocean regions from SAI imagery, as demonstrated in Figure 7a, which displays the $SAI_{NDVI}(25 \times 25)$ image of the Yellow Sea on 22 June 2009. In this figure, four regions are highlighted: region A, comprising both ocean and algae pixels, and ocean regions 1, 2, and 3. The histograms of these regions, that is, of region A and of the combined ocean regions 1, 2, and 3 are compared in Figure 7b.

[28] The most notable feature in the histogram comparison of Figure 7b is that the ocean distribution approaches a normalized frequency of zero at an SAI of 0.025 compared to 0.07 for the distribution containing algae pixels. Under the assumption that the ocean pixels in region A follow a similar distribution to those in ocean regions 1, 2, and 3, the threshold would then be located where the leading tail of the ocean distribution approaches a frequency of zero, which for the above scene is 0.025. This location is equivalent to an exclusion threshold of 99.90%, that is, the ocean histogram (red curve in Figure 7b) is analyzed to determine the SAI value that is greater than 99.90% of all SAI ocean values. The sensitivity of the spatial coverage of algae to changes in threshold value is an important consideration and will be covered in later sections.

[29] The assumption that the selected ocean pixels, for threshold determination, have a similar SAI distribution to those occurring within the algae bloom is in some respects important. Investigations have shown that erroneous estimates of the spatial coverage of algae are obtained when this assumption becomes untrue. This is illustrated in Figure 8, which shows the effect of using variable, nonrepresentative, ocean regions on the algae coverage profile and exclusion threshold (99.90%) profile. We define the algae coverage profile as the spatial coverage of algae determined from an SAI image using a set of kernel sizes ranging from $3 \times 3$ to $101 \times 101$ for a given exclusion threshold. Similarly, the exclusion threshold profile is the SAI of an exclusion threshold determined for the same set of kernel sizes and, in this case, at a threshold of 99.90%. It is important to note that the selection of 99.90% as the threshold is arbitrary at this stage. Figure 8a displays the NDVI image of the Yellow Sea, on 31 May 2008, along with the selected ocean regions that were used to generate the algae coverage and exclusion threshold profiles presented in Figure 8b. Note, the black "ocean" pixels south of the algal bloom are sand flats along the coast of Jiangsu Province.

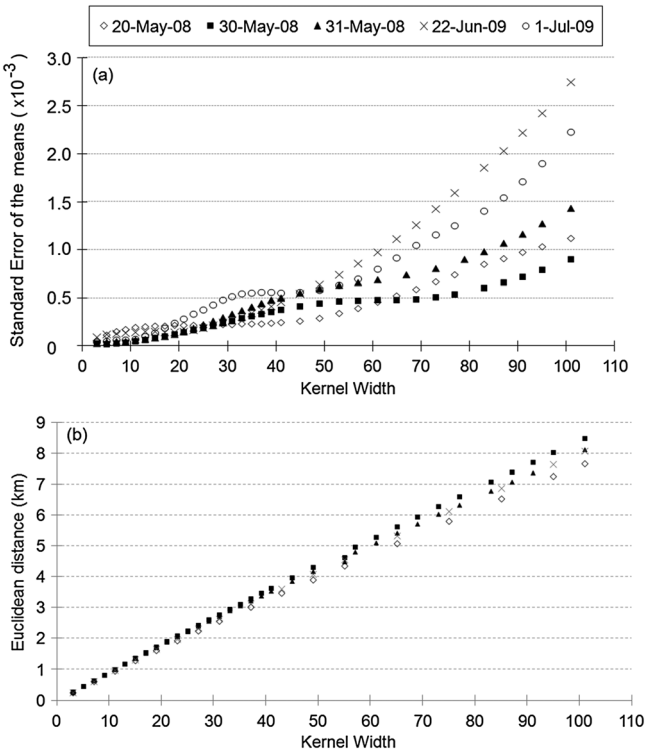


**Figure 6.** (a) The SEMs from 10, randomly located, ocean regions in SAI imagery for kernel sizes ranging between $3 \times 3$ and $101 \times 101$. (b) The average straight line, Euclidian, distance (km) between the median and central pixel for kernel sizes ranging between $3 \times 3$ and $101 \times 101$. Dates analyzed were 20 May 2008 (diamond), 30 May 2008 (solid square), 31 May 2008 (solid triangle), 22 June 2009 (cross), and 1 July 2009 (open circle). Note the Euclidean distance (Figure 6b) was averaged over all ocean and algae pixels.

[30] The algae coverage profile using the nonrepresentative, highly variable, ocean regions of Figure 8a, forms an inverse parabola (solid diamonds), where maximum algae coverage occurs for a kernel width of approximately 21. For larger kernel widths the spatial coverage of algae decreases substantially, coincident with the increase in the exclusion threshold. This is contrary to the asymptotic algae coverage profile (Figure 8b) obtained using regions 1, 2, and 3, where the effect of larger kernel sizes is lower such that the spatial coverage of algae experiences only a minor increase, despite the continual increase in the exclusion threshold. This analysis also shows that selecting the appropriate ocean regions in SAI imagery can be done by visually examining its respective NDVI image. Specifically, selecting those ocean regions (in close proximity to the bloom) that appear by "eye" to have similar NDVI values (i.e., gray-scale) as the ocean pixels among the algae bloom and surrounds.

[31] Table 1 shows the central latitude and longitude coordinates of each ocean region used to determine the exclusion threshold for SAI images of the Yellow Sea in 2008 and 2009. Table 1 also displays the number of pixels in each ocean region and the total number of ocean pixels selected for each SAI image. Note that each ocean region was square in shape, and the number of regions selected depended on cloud cover in the vicinity of the algae bloom.

### 3.3. Kernel Size and Threshold Sensitivity Analysis

[32] The sensitivity of the estimated spatial coverage of algae to changes in kernel size and exclusion threshold are shown in Figure 9. These algae coverage profiles, taken from $SAI_{NDVI}$ scenes of the Yellow Sea, were determined for exclusion thresholds ranging between 99.50 and 99.99%.

[33] The algae coverage profiles (Figure 9) highlight the vast influence the kernel size has on the estimated spatial

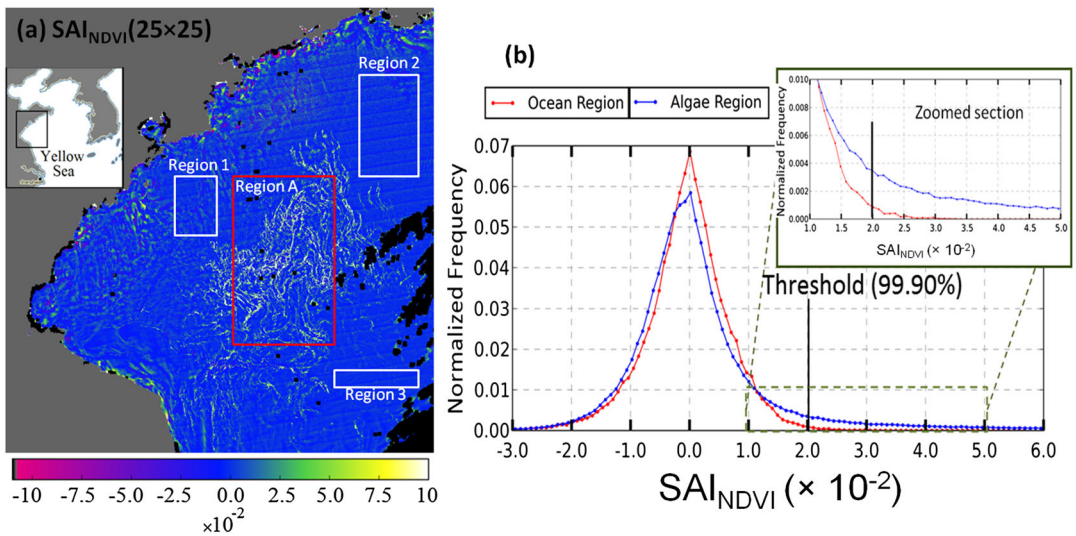


**Figure 7.** (a) $SAI_{NDVI}(25 \times 25)$ image of the Yellow Sea on 22 June 2009. Land and cloud pixels are assigned as grey and black pixels, respectively. (b) Histogram comparison of the ocean (red curve) and ocean-algae (blue curve) distributions with an exclusion threshold of 99.90% also displayed. The ocean-algae distribution was formed from pixels within region A shown in Figure 7a, while the ocean distribution was generated from the combined pixels of regions 1, 2, and 3. Note: pixels of regions 1, 2, and 3 were used to calculate the exclusion threshold (Figure 7a), and the normalized frequency is the frequency divided by the total number of pixels in the distribution (Figure 7b).

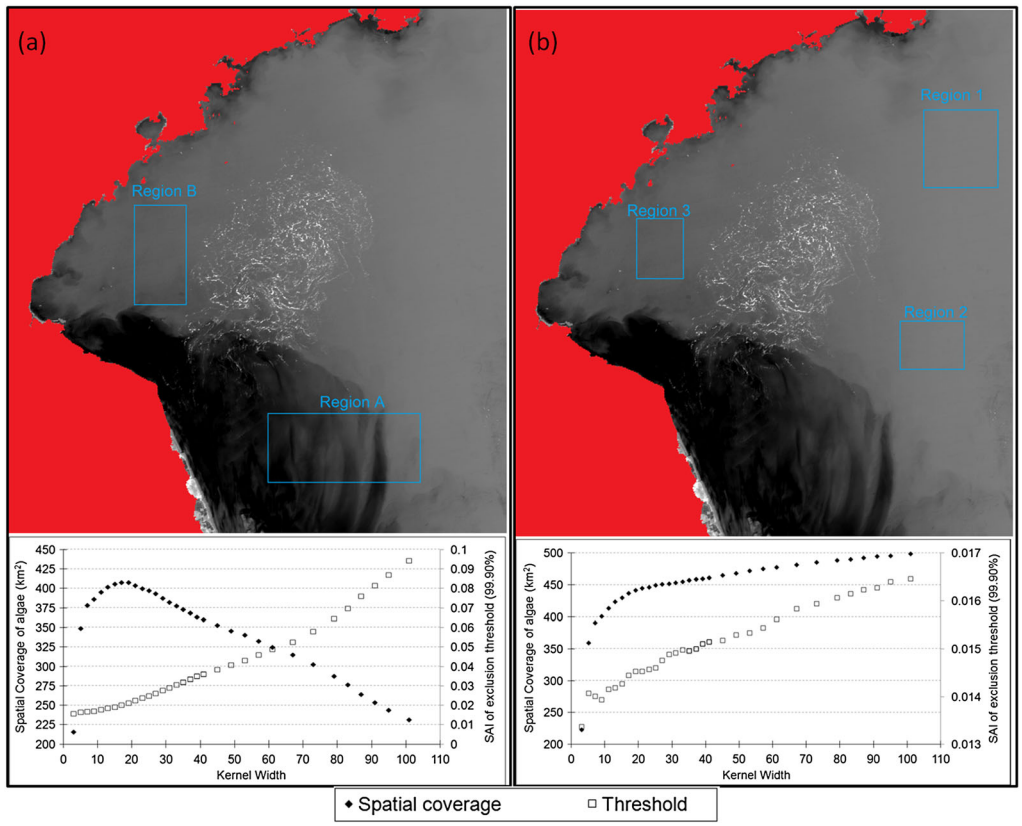


**Figure 8.** Effect of selecting highly variable ocean regions on the resultant algae coverage profile. (a, b) NDVI image of the Yellow Sea on 31 May 2008. The left panel shows the SAI of the 99.90% exclusion threshold (open square), obtained from the ocean pixels selected from regions A and B, and the resultant spatial coverage of algae (solid diamond) as the kernel size is increased. The right panel shows the same, except that ocean regions 1, 2, and 3 were used to obtain the exclusion threshold at each kernel size. Note (1) kernel width refers to the number of pixels forming the width of the kernel, e.g., the width of a $3 \times 3$ kernel is 3 etc., and (2) the spatial coverage of algae, for each kernel width, was obtained using an exclusion threshold of 99.90%.

coverage of algae at a given threshold. Extensive analysis has shown that these profiles tend to be asymptotic, in that, the rate of increase in the spatial coverage is high at low kernel sizes before decreasing dramatically beyond kernel sizes of approximately $21 \times 21$. This initial rate of increase can be quite high for some scenes—particularly for those dates where the algae blooms appear large. Our analysis has shown that an optimum range of kernel sizes falls within the bounds of $21 \times 21$ and $45 \times 45$. The choice of these lower and upper bounds are explained in sections 3.4 and 4.3.

[34] Between these lower and upper bounds the spatial coverage of algae actually experiences a minor change at a given threshold, as illustrated in Table 2. This table shows the average spatial coverage of algae between kernel sizes of $21 \times 21$ and $45 \times 45$ for exclusion thresholds ranging between 99.50% and 99.99%. Note that the uncertainties (i.e., variability) of the average values are equal to the differences between the maximum and minimum spatial coverage across the specified range of kernel sizes.

[35] Table 2 shows that the range in the spatial coverage, across the specified kernel sizes, can be as high as $\pm 18.8\%$ of the average value, e.g., $(155.4 \pm 29.3)\,\mathrm{km}^2$ for 24 June 2009 at an exclusion threshold of 99.50%, and as low as $\pm 0.3\%$, e.g., $(347.1 \pm 0.9)\,\mathrm{km}^2$ for 1 July 2009 for an exclusion threshold of 99.60%. Although for the majority of dates analyzed, the range is less than 10% of the average value, indicating that the relative change in the spatial coverage of algae as the kernel width increases from 21 to 45 is minor. It is also apparent from Table 2 that the largest change in the spatial coverage occurs with changing exclusion threshold.

[36] Sensitivity analysis of the spatial coverage of algae to changes in exclusion threshold shows an inverse relationship, where increasing the exclusion threshold generates lower spatial coverages (illustrated in Table 2). Additional analysis shows that the spatial coverage of algae undergoes a steady decline as the exclusion threshold increases from 99.50% to 99.80%. Between 99.80% to 99.90% this rate of decline increases twofold for some dates. The spatial coverage of algae, however, is most sensitive when the exclusion threshold is increased from 99.90% to 99.99% with increases in the rate of decline by factors of up to 5.

### 3.4. Over-scaling

[37] Scaled algae index imagery at low kernel sizes shows a high degree of uniformity as illustrated in Figure 6a. Unfortunately, at these kernel sizes a phenomena termed over-scaling occurs. We define over-scaling as the scaling

**Table 1.** The Central Latitude and Longitude Coordinate of Each, Square, Ocean Region Used to Determine the Exclusion Threshold for a Given $SAI_{NDVI}$ Image. Note That the Numbers in Brackets Represent the Number of Pixels Within the Given Ocean Region

| Date | Ocean Region | | | | | Total Number of Ocean Pixels |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | |
| 30 May 2008 | 35.177°N /120.244°E (34,544) | 35.265°N /122.341°E (12,771) | 36.356°N /121.674°E (20,160) | - | - | 67,475 |
| 31 May 2008 | 35.303°N /120.249°E (28,416) | 34.615°N /122.588°E (31,212) | 36.0139°N /122.83°E (56,810) | - | - | 116,438 |
| 5 Jun 2008 | 35.078°N /122.415°E (14,872) | 36.023°N /122.672°E (66,690) | 35.627°N /120.578°E (13,677) | 34.896°N /120.296°E (8908) | - | 104,147 |
| 28 Jun 2008 | 36.335°N /122.314°E (47,344) | 35.955°N /121.416°E (5733) | 34.991°N /119.579°E (5328) | 35.076°N /119.802°E (3382) | - | 61,787 |
| 17 Jul 2008 | 36.605°N /121.701°E (2925) | 36.451°N /121.533°E (2160) | 36.16°N /121.063°E (2296) | 31.892°N /122.466°E (22,820) | 33.032°N /122.933°E (7954) | 38,155 |
| 22 Jun 2009 | 34.464°N /121.97°E (8798) | 36.223°N /122.138°E (54,054) | 35.629°N /120.589°E (17,201) | - | - | 80,053 |
| 24 Jun 2009 | 33.623°N /122.01°E (30,334) | 35.692°N /122.832°E (81,450) | 35.719°N /120.797°E (32,712) | - | - | 144,496 |
| 1 Jul 2009 | 35.384°N /120.396°E (159,720) | 36.493°N /123.191°E (22,101) | 33.999°N /123.508°E (27,242) | - | - | 209,063 |
| 15 Jul 2009 | 34.869°N /122.208°E (18,900) | 35.625°N /123.448°E (51,917) | 34.986°N /120.141°E (26,364) | - | - | 97,181 |

of a given NDVI or FAI pixel by the index of an algae pixel rather than an ocean pixel. This causes algae pixels to be scaled to values near zero leading to underestimations in the spatial coverage of algae. Figure 10 illustrates over-scaling with a transect over a region densely populated with algae pixels in the NDVI image of the Yellow Sea (35.31°N/120.76°E to 35.31°N/121.98°E) on 30 May 2008. This figure also shows the NDVI of the median pixel chosen by the SAI algorithm for kernel sizes of $3 \times 3$, $9 \times 9$, $17 \times 17$, $23 \times 23$, $43 \times 43$, $75 \times 75$ and $101 \times 101$.

[38] Over-scaling is clearly evident at a kernel size of $3 \times 3$, where the NDVI of the median pixel is similar to the NDVI of the central pixel for the majority of algae pixels in the transect. These algae pixels would experience scaling by the NDVI of an algae pixel rather than an ocean pixel leading them to have SAI values near zero, and hence their misclassification as "nonalgae" when segmenting the SAI image. Notice that as the kernel size is increased the NDVI of the median pixel gradually decreases until a kernel size of $43 \times 43$, beyond which there is negligible reduction in the median value. The presence of over-scaling and its reduction with increasing kernel size explains why the algae coverage profiles displayed in Figure 9 increase with the kernel size until it asymptotes.

[39] Figure 11 displays the standard deviation of the SAI for selected regions in SAI imagery, consisting of both algae and ocean pixels, as the kernel size is increased from $3 \times 3$ to $101 \times 101$. These standard deviation curves show a rapid increase with increasing kernel size until a plateau (roughly) is reached. At a kernel size of $3 \times 3$, the standard deviation of the ocean-algae region is small due to over-scaling, where a significant amount of algae pixels are scaled to indexes near zero. As the kernel size is increased more algae pixels are correctly scaled relative to the indexes of true ocean pixels, which has the effect of increasing the standard deviation of SAI. The plateau region indicates that only insignificant changes are occurring to their SAI values, and we have interpreted this as where the majority (if not all) of the algae pixels are being correctly scaled by the indexes of ocean pixels, i.e., the minimization of over-scaling. These standard deviation curves reach the plateau region at different kernel sizes; however, typically the larger the algal bloom (e.g., 30 May 2008, 28 June 2008) the larger the kernel size needed to minimize over-scaling.

### 3.5. SAI Limitations

[40] Quasi true color, NDVI, and binary images of the Yellow Sea, China, on 1 July 2009 and 10 March 2011 are displayed in Figures 12 and 13, respectively. These binary, "algae" - "nonalgae", images were generated from the application of an exclusion threshold of 99.90% to $SAI_{NDVI}(21 \times 21)$ and $SAI_{NDVI}(45 \times 45)$ imagery. During 1 July 2009, floating macroalgae appears as a large bloom (white slicks in Figure 12b) in the north central region of the Yellow Sea. Quasi true color and NDVI imagery on 10 March 2011 (Figure 13), however, did not indicate the presence of floating macroalgae in the Yellow Sea region. This figure is presented to illustrate the effectiveness of the exclusion threshold method in correctly classifying open ocean and slightly turbid waters from SAI imagery as "nonalgae".

[41] Figures 12c and 12d illustrates that the SAI image kernel with the application of the exclusion method not only

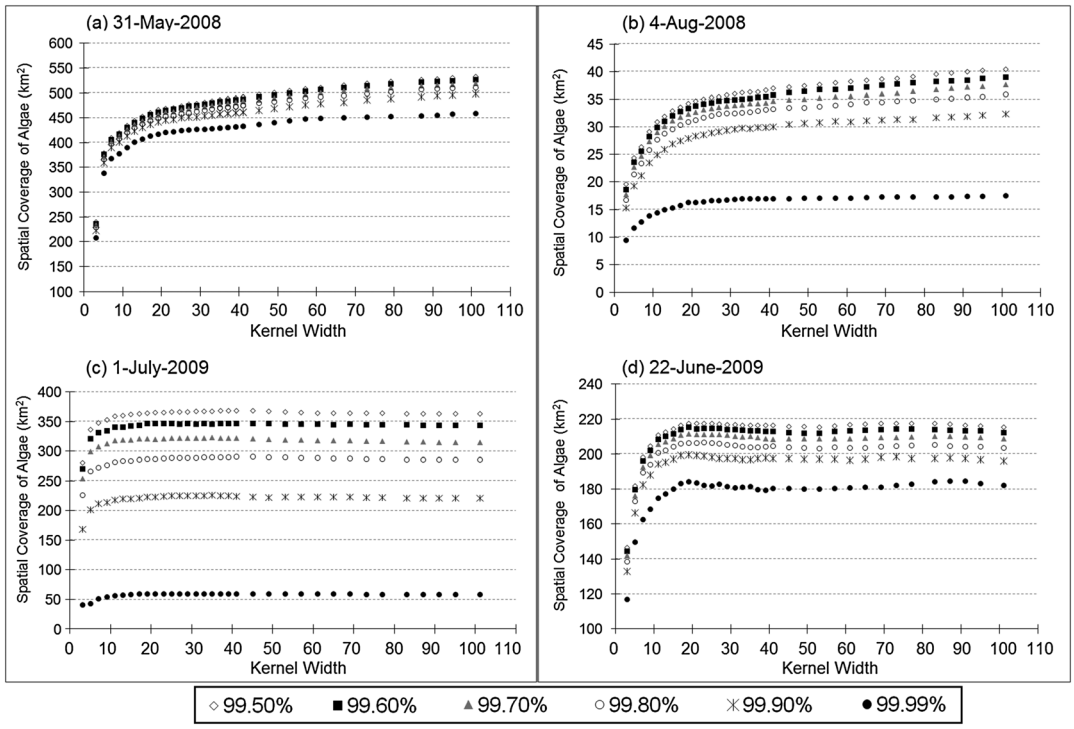


**Figure 9.** Algae coverage profiles for $SAI_{NDVI}$ scenes of the Yellow Sea, China, on (a) 31 May 2008; (b) 4 August 2008; (c) 1 July 2009; and (d) 22 June 2009. The spatial coverage of algae at each kernel size was obtained from a set of exclusion thresholds ranging from 99.50% to 99.99%. Note the algae coverage profiles on the left panel had associated with them highly variable NDVI scenes whose histograms contained multiple peaks, contrary to the algae coverage profiles on the right panel where the NDVI scenes were uniform displaying monomodal histograms (see Figure 5).

segments floating macroalgae pixels, but also bright turbid ocean pixels near the coast of Jiangsu Province (highlighted in Figure 12a). Misclassification of such turbid pixels as "algae" is also highlighted in the binary images of 10 March 2011 (Figures 13c and 13d)—a scene where no algae is present—and shows a potential limitation of the SAI image kernel. Notice though that these particular bright turbid ocean pixels are adjacent to or at the interface to a darker, less turbid ocean region. When the SAI image kernel is centered on these turbid pixels, the median index value would correspond to an ocean pixel from the darker region resulting in a high SAI, similar to that of an algae pixel. This effect may not necessarily occur in every SAI image—only in those whose NDVI shows rapid and significant changes (due to atmosphere/water turbidity) over very small spatial areas, i.e., very sharp gradients in turbidity such as those observed on 10 March 2011 (Figure 13b). In addition, these binary images show that increasing the size of the SAI image kernel also increases the amount of turbid ocean pixels classified as "algae", thus indicating that smaller kernel sizes are more effective in correctly classifying "algae" and "nonalgae" pixels. It should be noted that the majority of pixels in open and less turbid ocean regions, are correctly classified as "nonalgae" using the $21 \times 21$ and $45 \times 45$ kernel sizes.

**Table 2.** The Average and Range in the Estimated Spatial Coverage Between Kernel Sizes of $21 \times 21$ and $45 \times 45$ for Various $SAI_{NDVI}$ Images of the Yellow Sea, China, for Exclusion Thresholds Ranging Between 99.50% and 99.99%

| | Average ± Range ($km^2$) | | | | | |
|---|---|---|---|---|---|---|
| Date | 99.50% | 99.60% | 99.70% | 99.80% | 99.90% | 99.99% |
| 20 May 2008 | 26.5 ± 3.8 | 26.0 ± 3.6 | 25.5 ± 3.4 | 24.7 ± 3.1 | 23.5 ± 2.8 | 19.3 ± 1.7 |
| 30 May 2008 | 416.7 ± 25.6 | 413.8 ± 27.1 | 410.9 ± 26.7 | 406.2 ± 25.1 | 399.6 ± 24.6 | 377.8 ± 26.6 |
| 31 May 2008 | 481.3 ± 23.8 | 477.3 ± 22.1 | 472.8 ± 22.2 | 464.7 ± 21.5 | 454.0 ± 16.4 | 427.8 ± 13.0 |
| 5 Jun 2008 | 131.3 ± 1.1 | 129.2 ± 1.0 | 126.3 ± 1.5 | 122.0 ± 3.2 | 115.9 ± 4.8 | 101.6 ± 4.2 |
| 28 Jun 2008 | 237.7 ± 2.0 | 236.6 ± 1.9 | 235.4 ± 2.3 | 233.9 ± 2.3 | 230.3 ± 2.7 | 212.3 ± 2.3 |
| 17 Jul 2008 | 53.7 ± 3.5 | 51.2 ± 3.2 | 49.1 ± 1.7 | 45.4 ± 3.2 | 41.5 ± 2.8 | 36.4 ± 2.5 |
| 4 Aug 2008 | 36.0 ± 2.3 | 35.0 ± 2.0 | 33.9 ± 1.9 | 32.5 ± 2.0 | 29.5 ± 1.7 | 16.8 ± 0.7 |
| 22 Jun 2009 | 216.9 ± 1.2 | 213.9 ± 1.9 | 210.3 ± 2.8 | 205.1 ± 2.8 | 197.9 ± 2.3 | 181.5 ± 4.2 |
| 24 Jun 2009 | 155.4 ± 29.3 | 149.1 ± 26.4 | 141.3 ± 24.4 | 131.8 ± 23.4 | 119.0 ± 19.8 | 94.3 ± 11.0 |
| 1 Jul 2009 | 367.7 ± 3.5 | 347.1 ± 0.9 | 322.6 ± 2.2 | 290.1 ± 3.7 | 225.6 ± 3.0 | 60.3 ± 0.2 |
| 15 Jul 2009 | 197.5 ± 18.6 | 173.8 ± 17.5 | 146.1 ± 13.4 | 113.4 ± 12.1 | 70.6 ± 8.0 | 14.7 ± 1.9 |

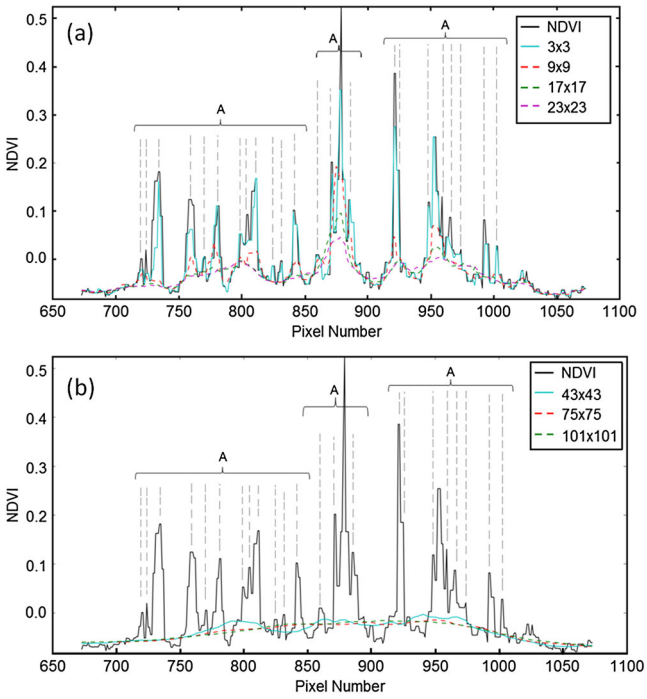


**Figure 10.** Horizontal transect (constant latitude) over a highly dense algae region in the MODIS (250 m resolution) NDVI image of the Yellow Sea, China, 30 May 2008. The transect shown is 401 pixels in length and starts at 35.31°N/120.76°E on the left and finishes at 35.31°N/121.98°E. (a) and (b) The NDVI of each pixel in the transect and the NDVI of the median pixel chosen from the SAI algorithm at kernel sizes of $3 \times 3$, $9 \times 9$, $17 \times 17$ and $23 \times 23$ for Figure 10a, and $43 \times 43$, $75 \times 75$ and $101 \times 101$ for Figure 10b. Note that peaks labeled with "A" are subjectively classified as algae pixels based on visual examination of the NDVI image.

## 4. Discussion

### 4.1. Characteristics of SAI Imagery

[42] In most images examined of the Yellow Sea, the index (in NDVI and FAI imagery) of ocean pixels in locations with high turbidity (atmospheric or water related) are greater than those of algae pixels in locations of lower turbidity. This is illustrated in the histogram comparison of regions A and B in the FAI and NDVI images of the Yellow Sea on 28 June 2008 (Figures 3b and 3d respectively). These figures show the presence of two separate histograms in Figures 3b and 3d and serves to demonstrate the effect of spatially varying radiance contributions from turbid water and/or aerosols. Moreover, the majority of algae pixels from region A had lower indexes than the ocean pixels of region B leading to segmentation issues. This variable NDVI and FAI scene would require the application of two separate thresholds to adequately determine the number of algae pixels for coverage estimates. In the case of Figures 3a and 3c, applying the "low-threshold" (Figures 3b and 3d) will isolate all algae pixels in the dark regions, but will include ocean pixels having higher indexes than this threshold. Conversely, the "high-threshold" will exclude all ocean pixels including algae pixels that have lower indexes. For a quantitative estimate, 3127 algae pixels (equivalent to 195.4 $km^2$) were isolated in region A of the NDVI image (Figure 3c) using the "low-threshold" compared to 530 (33.1 $km^2$) with the "high-threshold". This is a reduction by a factor of 6 and highlights the potential for significant underestimations when applying a global threshold to a scene suffering from a highly variable background.

[43] For accurate and reproducible spatial coverages, it is necessary to apply a scene-wide threshold that isolates all the algae pixels. This in turn requires an image that has a normal frequency distribution of ocean and algae pixels. SAI imagery meets this requirement, as indicated in the histogram comparisons of Figures 4b and 4d, where ocean and algae pixels of two regions experiencing different atmospheric/water turbidity have been normalized to the same spread of indexes. Histogram analysis of SAI imagery, displayed in Figure 5, showed normal frequency distributions of both ocean and algae pixels. These types of distributions—absent in the majority of NDVI imagery—further highlights the normalizing effect of the SAI algorithm, particularly when precursor NDVI images are highly variable, such as 1 July 2009 (Figure 5c).

[44] The variability within an SAI image has been found to be dependent on the size of the image kernel used to generate the SAI. At low kernel sizes (less than $21 \times 21$), the variability in the SAI of the ocean pixels is very low as indicated in Figure 6a. Increasing the kernel size results in a nonlinear increase in the intravariability of SAI imagery—due to the increase in the Euclidean distance between the median and central pixel (Figure 6b)—and results in greater differences in the NDVI (or FAI) between these two pixels. This analysis demonstrates that the SAI algorithm at larger kernel sizes is less effective at reducing the variability caused from atmospheric/water turbidity.

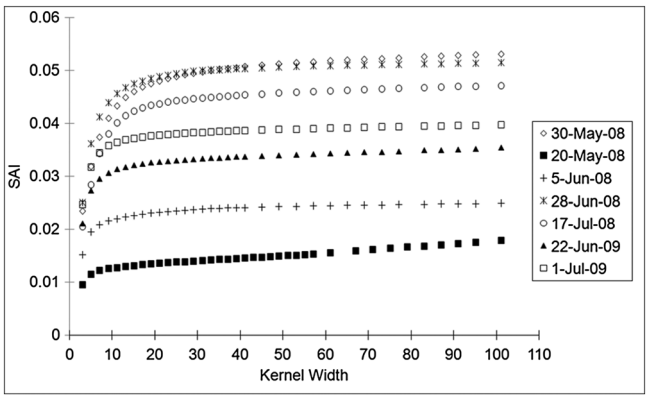


**Figure 11.** The standard deviation of the ocean-and-algae region vs. kernel width for 20 May 2008, 30 May 2008, 5 June 2008, 23 June 2008, 17 July 2008, 22 June 2009, and 1 July 2009.

### 4.2. Image Segmentation of SAI Imagery

[45] The generation of normal frequency distributions of the target and background pixels is the first step in computing the coverage of floating macroalgae. The second and equally important step is the application of a suitable threshold to isolate these algae pixels. Selecting a threshold using traditional image segmentation techniques assumes that the pixels of interest collectively form a distinctive and easily identifiable feature in a scene-wide

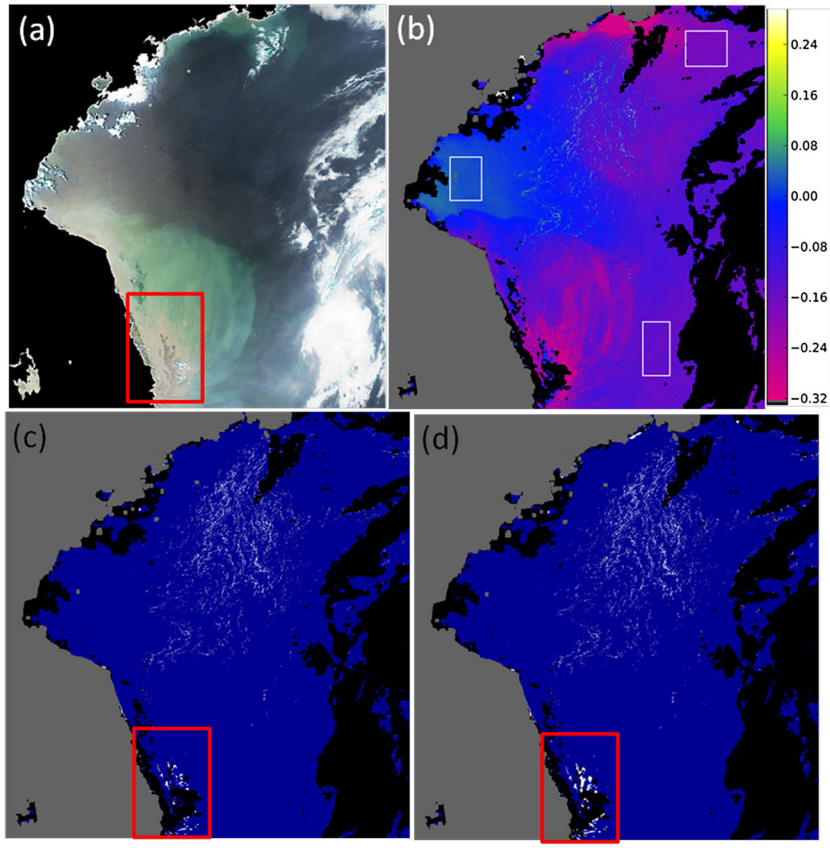


**Figure 12.** MODIS (a) true color composite and (b) NDVI images of the Yellow Sea on 1 July 2009. (c, d) Binary images produced from the application of 99.90% exclusion threshold to $SAI_{NDVI}(21 \times 21)$ and $SAI_{NDVI}(45 \times 45)$, respectively. The three ocean regions used to generate the exclusion threshold are shown as white squares in the NDVI image (Figure 12b). In Figures 12c and 12d, blue pixels represent "nonalgae", white pixels represent "algae", while land and cloud pixels are grey and black pixels, respectively.

histogram. For instance a "peak-trough-peak" (or bimodal) histogram where the background and target pixels each form one mode and where the threshold is located in the trough. However, scene-wide histograms of SAI imagery are not of this nature (see Figure 5). The algae pixels in SAI imagery generally constitute less than 1% of the pixels in an image of the Yellow Sea, and collectively do not form a distinctive feature in the histogram. Given that the algae pixels form the leading tail, selecting a threshold based on visual examination would be precarious. As such, the exclusion method was developed to maximize the accuracy and reproducibility in computing the spatial coverage of algae.

[46] The exclusion method uses the distribution of ocean pixels from SAI imagery to determine the threshold value. Obtaining this distribution requires an operator to manually select ocean regions in SAI imagery based on visual examination of its respective NDVI image. By visual examination, we mean whether or not a particular ocean region (in the NDVI image) looks to have similar grey-scale values (i.e., NDVI) to the ocean pixels among the algae bloom and its surrounds. Ideally ocean pixels among the algae bloom should be used to determine the exclusion threshold. However, developing an automated procedure that isolates these ocean pixels has proved difficult. Thus, to keep the SAI algorithm simple we have opted for the manual selection of ocean regions that are in close proximity to the algae bloom.

[47] This threshold method must therefore assume that the manually selected ocean regions have the same SAI distribution as the ocean pixels within the algae bloom. At low kernel sizes, this assumption is generally met, because the variability of the ocean pixels in SAI imagery is very low (Figure 6a). However, increasing the kernel size causes an increase in the variability of the ocean pixels and the possibility that this assumption no longer holds. In this respect the locations of the ocean regions needs to be carefully considered. This is demonstrated with the inverse parabolic algae coverage profile (Figure 8a) obtained when variable, nonrepresentative, ocean regions are used—such as the sand flats along the coast of the Jiangsu Province (e.g., region A in Figure 8a). In this region, the ocean pixels do not have the same SAI distribution as those within the algae bloom, and this therefore causes the estimated spatial coverage of algae to decrease significantly with increasing kernel size. Avoiding these sand flats and choosing other ocean regions around the algae bloom (e.g., regions 1, 2, and 3 in Figure 8b), generates an asymptotic algae coverage profile, where the spatial coverage of algae experiences a minor increase with

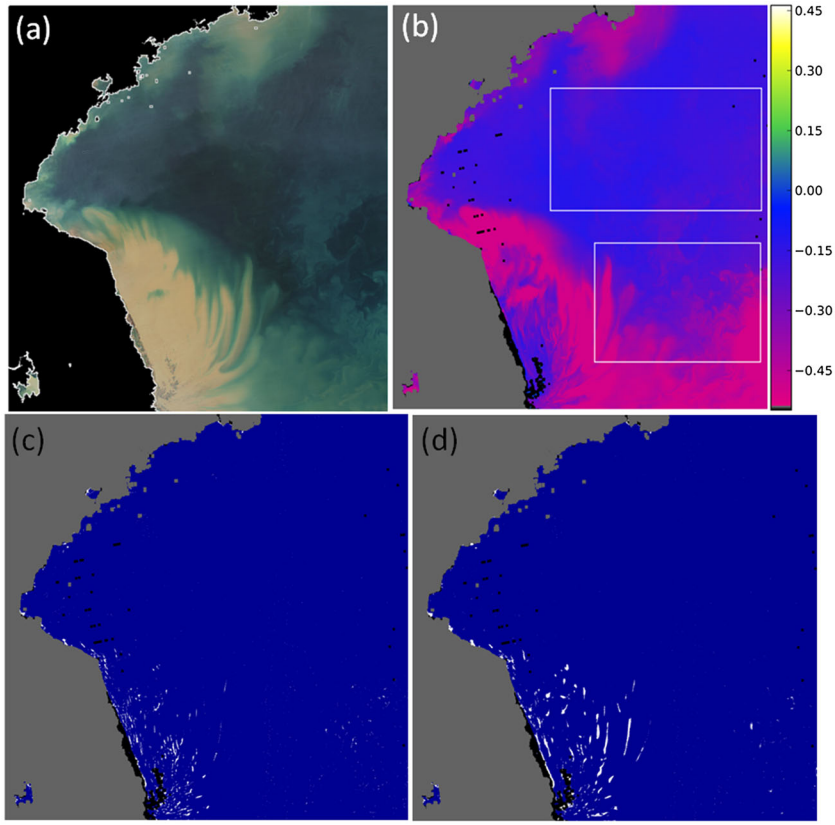


**Figure 13.** MODIS (a) true color composite and (b) NDVI images of the Yellow Sea on 10 March 2011. (c, d) Binary images produced from the application of 99.90% exclusion threshold to $SAI_{NDVI}(21 \times 21)$ and $SAI_{NDVI}(45 \times 45)$, respectively. The two ocean regions used to generate the exclusion threshold are shown as white squares in the NDVI image (Figure 13b). In Figures 13c and 13d, blue pixels represent "nonalgae", white pixels represent "algae", while land pixels are displayed as grey pixels.

increasing kernel size. Consequently, to minimize inaccuracies in estimates of the spatial coverage of algae, it is important to select multiple ocean regions located adjacent to the algae bloom, as well as avoiding variable regions that are not representative of ocean pixels within the algae bloom—this can effectively be done by visually examining the NDVI image.

### 4.3. Sensitivity Analysis and Optimal SAI Parameters

[48] The absence of ground truth data makes a definite selection and validation of an exclusion threshold and kernel size for accurate spatial coverages difficult. This is particularly so since the spatial coverage of algae can be sensitive to changes in kernel size and threshold (e.g., see Figure 9 and Table 2). Ground truth data, in this case, could possibly comprise of airborne hyperspectral imagery of several different blooms that the MODIS sensor can "see". The advantage of hyperspectral imagery is the ability to classify "algae" based on a pixels visible-to-NIR spectrum rather than relying on thresholds or other image manipulations. Given that no accurate in situ data exist to compare with, it is currently not known which of these two parameters (kernel size or exclusion threshold) would consistently generate the most accurate result.

[49] The most influential parameter of the SAI algorithm in regards to estimating the spatial coverage of algae is the kernel size. It directly controls the scaling of a given pixel in an NDVI or FAI image, which in turn affects the resultant distribution of algae and ocean pixels in the SAI image. The latter distribution also influences the threshold value chosen from the exclusion method and hence affects the estimated spatial coverage of algae. The net sensitivity of the spatial coverage to changes in kernel size is seen in Figure 9, where it typically increases substantially as the kernel size is increased from $3 \times 3$ to $21 \times 21$. The spatial coverage then asymptotes where large increases in kernel size cause only small changes. Analysis of the SAI image kernel (discussed below) revealed that accurate estimates of algae occurs in this region—specifically between kernel widths of 21 and 45. Table 2 shows that in this region the spatial coverage of algae can increase by as little as 0.9 $km^2$ to as high as 29.3 $km^2$. However, compared to the average spatial coverage, these increases are typically below 10%, and thus within the lower $(21 \times 21)$ and upper $(45 \times 45)$ bounds the kernel size would not influence the spatial coverage substantially.

[50] The choice of these lower and upper kernel sizes (i.e., optimal range of kernel sizes) is made based on the presence of over-scaling and high image variability at low and high kernel sizes, respectively. Over-scaling occurs when algae pixels constitute more than 50% of the pixels within an SAI image kernel. The index of the median pixel for such a kernel would correspond to an algae rather than an ocean pixel. This would cause the scaling of potential

algae pixels to SAIs near zero, producing misclassification and hence underestimation in estimates of spatial coverage. Over-scaling is clearly evident in Figure 10, where the NDVI of the median pixel for a $3 \times 3$ kernel equals to or is very similar to the NDVI for the majority of algae pixels in the transect. Increasing the kernel size decreases the value of the median pixel until it follows the general trend of the background ocean signal, such as for kernel sizes of $43 \times 43$, $75 \times 75$, and $101 \times 101$ (Figure 10b).

[51] Figure 11 indirectly shows the reduction of over-scaling with increasing kernel size. The lower bound of the optimal range in kernel size is based on where these plots asymptote, i.e., where we have assumed over-scaling is minimized. Note that the asymptote is reached at a different kernel size for different dates. As a general rule, the larger the bloom, the larger the kernel size needed to minimize over-scaling. To accommodate large algae blooms, a kernel size of $21 \times 21$ is suggested as a lower bound for the optimal kernel range.

[52] For kernel sizes greater than $21 \times 21$, the spatial coverage of algae becomes relatively insensitive to increases in kernel size (illustrated in Figure 9 and Table 2). As such there is no need to process SAI imagery using larger sized kernels, particularly since the intravariability of SAI imagery increases with kernel size. We therefore state $45 \times 45$ as the upper bound. This particular size was chosen because above this, the intravariability of SAI imagery begins to increase nonlinearly (Figure 6a).

[53] Therefore, to avoid over-scaling and yet to have an SAI image where the atmospheric and/or water turbidity contributions have been reduced, we propose an optimal kernel range of $21 \times 21$ to $45 \times 45$. To standardize this method, we propose the use of a kernel size in the middle of this range, i.e., $33 \times 33$ and an exclusion threshold of 99.90%, to determine the spatial coverage of algae. The choice of this exclusion threshold is based on the significant reduction observed in the estimated spatial coverage of algae as the exclusion threshold is increased from 99.90% to 99.99%. Our analysis has also shown that binary ("algae" and "nonalgae") images generated using a threshold of 99.99% excludes patches of pixels that appear to be "algae" in NDVI imagery, which would otherwise be included using lower exclusion thresholds.

### 4.4. SAI Algorithm Improvements and Limitations

[54] Since the introduction of the SAI algorithm shown by *Keesing et al*. [2011], several improvements and modifications have been made, the most important of which were the development of the exclusion threshold, and the use of a scaling procedure to determine the relative fractional algae cover. In the SAI algorithm presented by *Keesing et al*. [2011], the threshold was calculated by using the distribution of manually selected clear ocean regions in SAI imagery. Specifically, the threshold was chosen to equal 3.5 standard deviations above the mean ($3.5\sigma + \mu$) of this ocean distribution. If the distribution is normal then this threshold would exclude approximately 99.86% of the ocean pixels. However, analysis has shown that the ocean distributions follow a Laplace distribution; moreover, the statistics of these distributions change with kernel size and therefore the proportion of ocean pixels that are actually excluded. For this reason we chose a technique that determines an SAI value that excludes a certain proportion of ocean pixels.

[55] The SAI algorithm presented by *Keesing et al*. [2011] also calculates the spatial coverage of algae by multiplying the number of algae pixels in a given SAI image by the spatial area of an image pixel (250 m × 250 m). We term this method as the total affected area procedure. In the SAI algorithm presented here, a scaling is performed such that the SAI of the threshold has 0% cover, and the algae pixel having the maximum SAI has 100% cover of an image pixel. We term this method as the fractional coverage procedure.

[56] The total affected area procedure in *Keesing et al*. [2011] effectively states that all pixels (in the SAI image) above a particular threshold covers 100% of a 250 $m^2$ area. This is rather impractical as it is unlikely that an "algae" pixel having an SAI just above the threshold would cover the same area as an "algae" pixel whose SAI is, say 80% greater. We base this on the knowledge that the NDVI is sensitive to changes in vegetation cover and biomass [*Carlson and Ripley*, 1997; *Huete et al*., 2002]. Previous investigations have shown that the NDVI increases linearly with increasing (terrestrial) leaf area index (LAI) before reaching a plateau region, where a large increase in LAI causes a small increase in NDVI [*Carlson et al*., 1990; *Huete et al*., 2002]. With this same relationship observed between NDVI and biomass [*Hobbs*, 1995; *Gilbert et al*., 1996], further investigations will be required to quantify the relationship between $SAI_{NDVI}$ and $SAI_{FAI}$ to floating macroalgae aerial coverage and/or biomass.

[57] Estimates of the spatial coverage of algae, for the floating macroalgae blooms that occurred in the Yellow Sea between 2008 and 2009, obtained from the fractional coverage and total affected area procedures are presented in Table 3. Note that the standardized parameters (i.e., 99.90% exclusion threshold and an SAI kernel size of $33 \times 33$) were used.

[58] Table 3 shows that estimates of the spatial coverage of algae calculated with the fractional coverage procedure are substantially lower than those determined using the total affected area procedure. Analysis has shown that this difference relates to the substantial amount of algae pixels whose SAIs are near the value of the threshold. Such algae pixels have a low coverage of a 250 m × 250 m area, and thus they collectively sum to miniscule amounts of algae cover.

[59] Although the SAI image kernel is able to normalize and efficiently separate the indexes of ocean and algae pixels from an NDVI or FAI image, it does suffer from two limitations: (1) the scaling routine for algae coverage estimation (equation (4)) only works when algae pixels are present in a scene, and (2) the misclassification of those highly turbid ocean pixels as "algae" (see Figures 12 and 13).

[60] Limitation (1) arises from the fact that the exclusion threshold and scaling routine of equation (4) are based on the statistics of an image's background and target pixels. For example, an exclusion threshold of 99.90% will classify 99.90% of the ocean pixels in a given scene as "nonalgae", and 0.10% as "algae" (e.g., in an image with $1.0 \times 10^6$ ocean pixels, 1000 of them will be classed as "algae"). These pixels, being close to the threshold value, would, using the scaling of equation (4), collectively cover a minute amount of a 250 $m^2$ area. Consequently, ocean pixels misclassified as "algae" would only marginally affect estimates of algae

**Table 3.** The Spatial Coverage of Algae ($km^2$) for the Floating Macroalgae Blooms That Occurred in the Yellow Sea, China, Between 2008 and 2009. These Spatial Coverages Were Calculated Using the Fractional Coverage and the Total Affected Area Procedure, Each Using an Exclusion Threshold of 99.90% and an SAI Kernel Size of 33 × 33. Note That the Number of Pixels Classified as Algae are Also Presented

| Date | Number of Algae Pixels Extracted | Spatial Coverage of Algae ($km^2$) | |
|---|---|---|---|
| | | Fractional Coverage Procedure | Total Affected Area Procedure [*Keesing et al*., 2011] |
| 20 May 2008 | 9364 | 23.7 | 585.25 |
| 30 May 2008 | 66,010 | 402.2 | 4125.6 |
| 31 May 2008 | 93,366 | 455.8 | 5835.4 |
| 5 Jun 2008 | 32,827 | 114.3 | 2051.7 |
| 28 Jun 2008 | 36,827 | 230.0 | 2301.7 |
| 17 Jul 2008 | 3563 | 41.20 | 222.7 |
| 4 Aug 2008 | 8032 | 29.9 | 502.0 |
| 22 Jun 2009 | 37,711 | 197.0 | 2356.9 |
| 24 Jun 2009 | 26,962 | 117.3 | 1685.1 |
| 1 Jul 2009 | 29,896 | 226.1 | 1868.5 |
| 15 Jul 2009 | 5630 | 72.3 | 351.9 |
| 14 Aug 2009 | 12,702 | 47.9 | 793.9 |

coverage. However, applying this scaling procedure to scenes where floating algae is absent will cause scaling bounded by the SAI threshold, as 0% cover, and the ocean pixel with the maximum SAI, as 100% cover of 250 $m^2$ area. This would generate potentially large, false algae coverage estimates, and as such the current scaling routine should only be used when algae is visually observed in NDVI and true color composite imagery.

[61] Limitation 2 is caused by the sensitivity on the NDVI to changing sediment loads. This is clearly illustrated in quasi true color (Figures 12a and 13a) and NDVI imagery (Figures 12b and 13b) of the coastal area near Jiangsu Province. This particular coastal region appears—from these sets of images—to be shallow and influenced by very high suspended sediment loads and thus equates to an extreme example of case-2 waters. Note that standard MODIS bio-optical algorithms for chlorophyll-*a* concentration and diffuse attenuation coefficient are typically inaccurate in these turbid case-2 waters [*Darecki and Stramski*, 2004] and are flagged (i.e., masked) during operational Level 2 MODIS ocean color processing [*Patt et al*., 2003]. It is therefore possible to circumvent this limitation by identifying and masking these turbid pixels prior to SAI processing using the L2 flag product from SeaDAS processing [*Patt et al*., 2003]. Alternatively known "turbid" regions—based on quasi true color imagery or knowledge of the region—can be manually excluded. In the present work, the algae coverage estimates in Tables 2 and 3 had such pixels manually flagged.

## 5. Conclusion

[62] Difficulties in segmenting algae pixels from NDVI imagery, and in some cases FAI imagery, arise from the presence of irregular frequency distributions of ocean and floating macroalgae pixels. The aim of this research was to develop an image processing algorithm that normalizes the distributions of ocean and algae pixels but to still have enough contrast between these two pixels to enable segmentation using a global threshold.

[63] This goal was achieved with the development of the scaled algae index. This semiautomated algorithm is a scaling routine where every pixel in the NDVI or FAI image are scaled by their local ocean index. The monomodal histograms of SAI imagery are indicative of a highly uniform image, and analysis of the intravariability shows that the SAI algorithm has substantially reduced any variability introduced by atmospheric and/or water turbidity. The SAI is a necessary intermediate product used as a platform to quantify the spatial coverage of floating macroalgae observed in satellite imagery. The presence of normal frequency distributions of ocean and algae pixels in SAI imagery allows the segmentation of algae pixels through the use a global threshold. This threshold is determined by an operator assisted scene-by-scene procedure we have termed the exclusion method. Additional image programming improvements would need to be made to automate this threshold approach for its use in global data processing. The most important parameters in the SAI algorithm is the size of the image kernel and the proportion of ocean pixels that are excluded using the exclusion threshold. Our analysis has shown that the spatial coverage of algae is underestimated when kernel sizes of less than 21 × 21 are used, due to a phenomena we have termed "over-scaling". Given that the spatial coverage of algae experiences only a minor change between kernel sizes of 21 × 21 and 101 × 101, and since the variability of SAI imagery increases with kernel size, we propose an upper kernel size of 45 × 45. To standardize this algorithm, we propose that a kernel size of 33 × 33 be used.

[64] A major issue faced, which merits further work, is the lack of field data with which to compare floating macroalgae coverage estimates. These data, if available, will enable the selection of the appropriate exclusion threshold. However, based on sensitivity analysis, we propose an exclusion threshold of 99.90% as a suitable, standardized, threshold.

## References


Behrenfeld, M. J., T. K. Westberry, E. S. Boss, R. T. O'Malley, D. A. Siegel, J. D. Wiggert, B. A. Franz, C. R. McClain, G. C. Feldman, S. C. Doney, J. K. Moore, G. Dall'Olmo, A. J. Milligan, I. Lima, and N. Mahowald (2009), Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences 6*: 779-794.

Carlson, T. N., E. M. Perry, and T. J. Schmugge (1990), Remote estimation of soil moisture availability and fractional vegetation cover for agricultural fields, *Agric. For. Meteorol.*, *52*, 45-69.

Carlson, T. N., and D. A. Ripley (1997), On the relation between NDVI, fractional vegetation cover, and leaf area index, *Remote Sens. Environ.*, *63*, 241-252.

Darecki, M., and D. Stramski (2004), An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea, *Remote Sens. Environ.*, *89*, 326-350, doi:10.1016/j.rse.2003.10.012

Gilbert, M. A., S. Gandia, and J. Melia (1996), Analysis of spectral-biophysical relationships for a corn canopy, *Remote Sens. Environ.*, *55*, 11-20.

Gordon, H. R., and M. Wang (1994), Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary approach, *Appl. Opt.*, *33*, 443-452.

Gower, J. F. R., R. Doerffer, and G. A. Borstad (1999), Interpretation of the 685 nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS, *Int. J. Remote Sens.*, *20*, 1771-1786.

Gower, J., S. King, G. Borstad, and L. Brown (2005), Detection of intense phytoplankton blooms using the 709 nm band of the MERIS imaging spectrometer, *Int. J. Remote Sens.*, *26*, 2005-2012.

Gower, J., C. Hu, G. Borstad, S. King (2006), Ocean color satellites show extensive lines of floating sargassum in the Gulf of Mexico. *IEEE Trans. Geosci. Remote Sens. 44*(12), 3619-3625, doi: 10.1109/TGRS.2006.882258

Hobbs, T. J. (1995), The use of NOAA-AVHRR NDVI data to assess herbage productions in the arid rangelands of central Australia, *Int. J. Remote Sens.*, *16*, 1289-1302.

Hu, C., (2009), A novel ocean color index to detect floating algae in the global oceans, *Remote Sens. Environ.*, *113*, 2118-2129.

Hu, C., and He, M.-X. (2008), Origin and offshore extent of algae in Olympic sailing area, *Eos Trans. AGU*, *89(*33*)*, 301-302.

Hu, C., D. Li, C. Chen, J. Ge, F. E. Muller-Karger, J. Liu, F. Yu, and M.-X. He (2010), On the recurrent Ulva Prolifera blooms in the Yellow Sea and East China Sea, *J. Geophys. Res. 115*, C05017, doi:10.1029/2009JC005561.

Hu, C., Y. Wang, Q. Yang, S. He, L. Hu, and M. He (2007), Comparison of ocean color data products from MERIS, MODIS, and SEAWIFS: Preliminary results for the East China Seas. *Proc. 'Envisat Symposium 2007', Montreux, Switzerland*, 23-27.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002), Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sens. Environ.*, *83*, 195-213.

Kaufman, Y. J., and D. Tanre (1992), Atmospherically resistant vegetation index (ARVI) for EOS-MODIS, *IEEE Trans. Geosci. Remote Sens.*, *30*, 261-270.

Keesing, J. K., D. Liu, P. Fearns, and R. Garcia (2011), Inter- and intra-annual patterns of *Ulva prolifera* green tides in the Yellow Sea during 2007-2009, their origin and relationship to the expansion of coastal seaweed aquaculture in China, *Mar. Pollut. Bull.*, *62*, 1169-1182

Liu, D., J. K. Keesing, Q. Xing, and P. Shi (2009), World's largest macroalgal bloom caused by expansion of seaweed aquaculture in China, *Mar. Pollut. Bull.*, *58*, 888-895.

Liu, D., J. K. Keesing, Z. Dong, D. Di, Y. Shi, P. Fearns, and P. Shi (2010), Recurrence of the world's largest green-tide in 2009 in Yellow Sea, China confirms *Porphyra yezoensis* aquaculture rafts provide nursery for macroalgal blooms, *Mar. Pollut. Bull.*, *60*, 1423-1432.

Patt, F. S., R. A. Barnes, R. E. Eplee, Jr., B. A. Franz, W. D. Robinson, G. C. Feldman, S. W. Bailey, J. Gales, P. J. Werdell, M. Wang, R. Frouin, R. P. Stumpf, R. A. Arnone, J. R. W. Gould, P. M. Martinolich, V. Ransibrahmanakul, J. E. O'Reilly, and J. A. Yoder (2003), Algorithm updates for the Fourth SeaWiFS data reprocessing, *NASA Tech. Memo*. 206892, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD.

Shi, W., and M. Wang (2009), Green macroalgae blooms in the Yellow Sea during the spring and summer of 2008, *J. Geophys. Res.*, *114*, C12010.

Sneller, K. G., G. J. Doucette, and G. J. Kirkpatrick (2003), Harmful algal blooms: causes, impacts and detection. *J. Ind. Microbiol. Biotechnol.*, *30*, 383-406.

Tucker, C. J. (1979) Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sens. Environ.*, *8*, 127-150.

Wang, M. (2007), Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations, *Appl. Opt.*, *46*, 1535-1547.

Wang, X. H., F. Qiao, J. Lu, and F. Gong (2011), The turbidity maxima of the northern Jiangsu shoal-water in the Yellow Sea, China, *Estuar. Coast. Shelf. Sci.*, *93*, 202-211, doi:10.1016/j.ecss.2010.10.020.