

# Finding Green River in SeaWiFS Satellite Images

W. Yao\*, L.O. Hall, D.B. Goldgof, and F. Muller-Karger\*

Department of Computer Science and Eng., ENB 118

\*Department of Marine Science

University of South Florida

Tampa, Fl. 33620

{hall,goldgof}@csee.usf.edu, {carib,wyao}@carbon.marine.usf.edu

## Abstract

*Understanding oceanic primary production on a global scale can be enhanced by methods that are able to automatically track phytoplankton blooms from color satellite images. In this paper, unsupervised clustering and rule learning are combined to track green river, a plume of discolored water that forms every March-May offshore along the edge of the west Florida Shelf, from the Sea Viewing Wide Field of View Sensor which began flying in late 1997. Spatial information and sea surface temperature can be integrated into the approach to improve performance. Using cross-validation experiments over a series of 59 multi-spectral images, it is shown that the developed system is able to reliably discriminate between images with green river from those with no phytoplankton blooms or other kinds of blooms. It is also effective in identifying the region which the green river covers.*

## 1. Introduction

Over the past two decades, research has focused on refining algorithms for estimating chlorophyll in the ocean using oceanic color imagery. Most previous studies estimated the phytoplankton concentration using a simple ratio of two or three wavelength bands, namely ratios of a blue or blue-green sensor channel (e.g. 440 or 520 nm) to a green channel (e.g. 550 nm) [5]. This simple band ratio may lead to serious errors estimating biomass when the algorithm is applied indiscriminately to all marine environments. For example, the algorithm fails in environments where the color of water is not a simple function of the amount of green pigment (chlorophyll) available within phytoplankton.

The results reported here are the first (that we know of) describing a system which can automatically segment phytoplankton blooms from previously unseen images

from the SeaWiFS (the Sea-Viewing-Wide-Field-of-View-Sensor) [6] satellite which began flying in late 1997. Related work can be found in [9, 2, 11]. The focus here is segmenting phytoplankton blooms off the West Florida shelf. The so-called *green river*, a plume of discolored water that forms every March-May offshore along the edge of the west Florida Shelf, is the main focus of these segmentation efforts. It extends from Cape San Blas to the Florida Keys, lasts for several weeks to two months, and then dissipates.

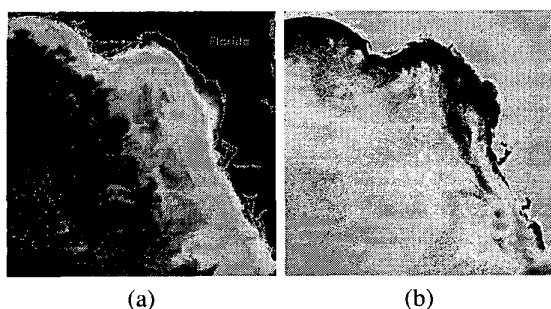
Automatic identification and segmentation of green river would facilitate the analysis of the ocean dynamics and biological effects of this phenomenon. We show an approach which utilizes some domain knowledge with clustering and supervised learning to build a classifier for *green river* blooms as seen in 59 SeaWiFS satellite images.

## 2. Processing SeaWiFS Data

The SeaWiFS satellite provides 8 spectral bands centered at 412, 443, 490, 510, 555, 670, 765, and 865 nm each with bandwidth of 20nm and a spatial resolution of 1.1km. Each pixel records a 10 bit intensity. The SeaWiFS level-2 products also includes a calculated Chlorophyll band [7]. Figure 1 shows the level-2 Chlorophyll and 412 nm band from a SeaWiFS image. Level-2 images were configured to be 512x512 pixels, with each pixel containing the water-leaving radiance after atmospheric correction, sensor calibration and any bio-optical algorithms have been applied.

We use the 5 shortest atmospheric bands and a calculated value for Chlorophyll-a concentration. Hence, there are 10-bit intensities for the five bands and a 6<sup>th</sup> feature which is an approximate calculation of the amount of Chlorophyll-a on the water covered by a pixel. These features are clustered with a fast fuzzy clustering algorithm, mrFCM [1].

Decision trees were used to classify the cluster centroids produced by mrFCM in the case that the clusters are homogeneous enough to be labeled by expert marine scientists.



**Figure 1. The calculated Chlorophyll (a) and 412nm (b)**

C4.5 release 8 [8] was used as the decision tree learning algorithm with C4.5rules used to create classification rules.

### 2.1. Integrating knowledge into the classification

In this study, relevant domain knowledge consists of the various types of waters expected in an image which should be identified as different types of clusters in a specific region.

River plumes are a phenomenon of extensive pigment concentration on the West Florida Shelf. The larger plumes occur mainly during spring. High pigment concentrations persist 1-6 weeks in a pattern which extends more than 250 km southward along the shelf from Cape San Blas. Gilbes [4] suggests the plume formation may be associated with one or more of several processes. During dinoflagellate blooms, biomass in surface waters varies from  $2 - 30 \text{ mg m}^{-3}$  Chl-a. Otherwise, concentrations range  $0.1$  to  $2 \text{ mg m}^{-3}$  Chl-a in the Gulf of Mexico, with higher concentrations over the shelf areas (near shore). The river plume over the West Florida Shelf has been referred to as the “green river”. The terms river plume and “green river” are equivalent here.

The current SeaWiFS algorithm [3, 6] which pre-processes the raw data is often erroneous in turbid coastal waters. Therefore, case II water at depth less than 30m is not processed.

There are 3 classes of interest in this study, namely case I water, green river, and unknown. The “unknown class” could be a possible mixture of case I water, case II water, green river and other types of blooms.

### 2.2. Image description

Fifty nine SeaWiFS images from September 1997 to April 1999 were used in this study. The images which cover the West Florida Shelf are within the latitude from  $24.5 \text{ N}$  to  $30.5 \text{ N}$  and the longitude from  $80.5 \text{ W}$  to  $86.5 \text{ W}$ . Images are

SeaWiFS Level-2 files generated by NASA standard processing software (SeaDAS 3.2). These images have been mapped onto latitude/longitude space with an image size of  $512 \times 512$  pixels. The main data contents of a Level-2 file are the water-leaving radiances for each pixel derived by applying sensor calibration, atmospheric corrections, and bio-optical algorithms to the raw data.

The Level-2 files consist of water leaving radiance at 412, 443, 490, 510, and 555 nm and the chlorophyll-a concentration. We used two conversion techniques, linear and nonlinear to convert the (originally) 10-bit data into 8-bits before clustering the SeaWiFS images with the mrFCM algorithm [10]. The conversion preserves the precision of  $0.01 \text{ mg m}^{-3}$  for low chlorophyll-a concentrations, which is close to the limit of actual field measurements.

### 2.3. Ground Truth for Cluster labeling

Fifty nine SeaWiFS images were given to oceanographers for identification of different water types (or ground truthing). These oceanographers have first-hand knowledge of the various water types in the region. Ground-truthed images were used to label the clusters generated by mrFCM. The generated clusters form a multi-dimensional feature space. Centroid values of labeled clusters of training images were used to train the decision tree and extract heuristic rules, which are then used to identify case I water and green river (after applying clustering) in the testing images. Our knowledge-based clustering approach is based on over-clustering the satellite image, and labeling regions of interest produced by the over-segmentation. Our over-clustering approach is designed to provide homogeneous clusters. There are 3 objects of interest in our SeaWiFS images (case I water, green water, and unknown water type), but we partitioned the data set into 11 classes at the cost of splitting some true classes into multiple clusters.

### 2.4. SeaWiFS image classification

Our knowledge-based system operates as follows. The six bands/features are provided to the system and segmented into 11 clusters by the mrFCM clustering algorithm. After clustering labels are assigned by comparing clusters with ground truth. A cluster of the training images is labeled as an object (one of the three classes in this study) if the majority of pixels in the cluster belong to that specific object in the ground truth image. The labeled fuzzy cluster centroids of the training images form a feature space, and cluster labeling rules can be extracted based on the distribution of the labeled cluster centroids. These rules are used to classify new clusters. Other expert rules can also be incorporated to identify clusters, as shown in the proceeding.

**Table 1. Number of SeaWiFS images with areas of interest (max. 59).**

	Number of Images
Green River	35
Case I Water	59
Unknown	47
Only Case I Water	7

**Table 2. Cluster-level confusion matrix for 10-fold Cross validation of 59 images, where GR = green river, Case I = case I water and UNK = unknown.**

Class	GR	Case I	UNK
GR	58	0	5
Case I	1	373	5
UNK	3	3	98

### 3. Experiments

A total of 59 training images generate 649 clusters after applying the mrFCM clustering algorithm to create 11 clusters per image. We manually excluded 103 cloud edge clusters from all experiments. C4.5 followed by C4.5rules [8] (release 8), which generates a classifier in the form of a rule set, was applied to the 546 labeled cluster centroids with all C4.5 parameters at their default values. Eight rules were generated.

A robust estimate of accuracy on unseen cases can be obtained by cross-validation. Hence, 10-fold cross validation is used on both the image and cluster level in the experiments on the 59 acquired images.

The number of images with green river, case I water and the unknown water type are summarized in Table 1. The unknown type of water here refers to what is generally believed not to be green river and possibly a mixture of case I water and case II water or green river.

The cluster classification performance of a 10-fold cross validation experiment is shown as a confusion matrix in Table 2. The average standard error was 3.2 %. There were 5 green river clusters misclassified and 4 clusters misclassified as green river.

Of special interest in this study is the automatic identification of green river in images along the West Florida Shelf. Out of a total fifty-nine images, thirty-five images contain green river. We partitioned the total 59 images (instead of the total 546 labeled clusters as above) into 10 folds and performed a 10-fold cross-validation to evaluate the accuracy of our knowledge-based system in identifying the images with green river. Table 3 shows the confusion matrix on the

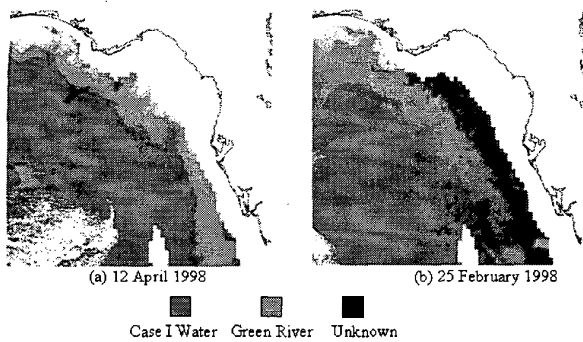
**Table 3. Cluster-level confusion matrix on the image level for 10-fold Cross validation of 59 images, where GR = green river.**

Class	GR	no GR
GR	33	2
no GR	3	22

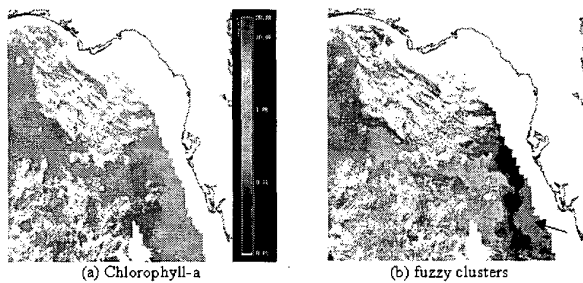
image level.

In Table 3, we adopt a threshold of 1 cluster to determine whether to label an image as containing green river: if an image has 1 or more than 1 cluster labeled as green river, it is classified as containing green river. The threshold of 1 cluster is used since out of 35 images with green river, 13 images have only one green river cluster. Another possible threshold is on the number of pixels in a cluster. The size of green river clusters ranges from 3,567 to 17,549 pixels. It is difficult to give a threshold on the pixel number. As shown in Table 3, the false negative image error rate is 5.7% (2/35), the false positive image error rate is 8.3% (2/24). The results show that our system can successfully recognize most images with green river (33 out of 35). The higher false positive image error rate is attributable to two factors. First, it is difficult for a marine biologist to recognize the phytoplankton blooms from the coastal case II water based on satellite images. In this study, we adopted a conservative approach and only labeled pixels that were absolutely green river. Therefore, the mislabeled clusters may, in fact, have quite a bit of green river in them as the unknown water type could be a mixture of case I water, case II water, green river and other types of phytoplankton blooms. Second, other types of phytoplankton blooms, which can be attributed to the upwelling and local river discharge along the west coast of Florida, might also be confused with green river, due to many similarities which make it difficult for humans to separate them.

Figure 2 shows images correctly classified as having green river and includes in (a) an image with green river far from the coast. Figure 3 (b) shows an image in which it seems probable that another type of phytoplankton bloom near the Fl. coast was misclassified. The other typical error that occurred is not shown due to space limitations, but occurred away from the shore and was a result of a phytoplankton upwelling from the Yucatan peninsula. The cluster centers from this bloom are nearly identical to those from green river which indicates it cannot be distinguished with the SeaWiFS features [10]. However, if sea surface temperature is used from another satellite (AVHRR/NOAA14), at times other than summer and fall when it is uniform, the Yucatan upwelling can be reliably identified [10] as being warmer.



**Figure 2. Two test images correctly classified by our system.**



**Figure 3. An incorrectly classified image (b) with a bloom (see arrow). The center Bar is Chlorophyll, showing a concentration similar to green river in this region.**

#### 4. Summary

This paper shows an approach utilizing clustering and learning that was successfully applied to segment *green river* in SeaWiFS satellite images off the West Florida shelf. Almost all images containing green river were identified from a set of 59 images and almost all images without green river were correctly identified. The false positive errors involved other types of phytoplankton blooms and we found that sea surface temperature could be used to differentiate some of these. Likewise, the 2 (of 33) false negatives were images with phytoplankton in them, but were misrecognized as being without green river blooms.

The approach is promising. In several experiments clusters were found to be homogeneous [10]. The learned rules made significant use of the Chlorophyll feature, as expected, in identifying phytoplankton blooms. With additional information, this approach could be utilized as a

screening tool in searching for SeaWiFS images with green river and as an tool to point out the initial choice of likely regions which contain coastal or riverine waters such as those of the green river off Florida.

**Acknowledgements:** We'd like to thank Dr. Mingrui Zhang and Dr. Chunming Hu for their help in data analysis.

#### References

- [1] T. W. Cheng, D. B. Goldgof, and L. Hall. Fast fuzzy clustering. *Fuzzy Sets and Systems*, 93:49–56, 1998.
- [2] M. Clark, L. Hall, D. Goldgof, R. Velthuizen, F. Murtagh, and M. Silbiger. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17(2):187–201, 1998.
- [3] M.-K. F. E., C. R. McClain, R. N. Sambrotto, and C. G. Ray. Measurements of phytoplankton distribution in the southeastern bering sea using the czcs: A note of caution. *J. Geophys. Research*, C7(11):483–11,499, 1990.
- [4] F. Gilbes. *Analysis of Episodic Phytoplankton Blooming and Associated Oceanographic Parameters on the West Florida Shelf Using Remote Sensing and Field Data*. PhD thesis, University of South Florida, Dept. of Marine Science, St. Petersburg, FL, 1996.
- [5] H. Gordon and D. Clark. Atmospheric effects in the remote sensing of phytoplankton pigments. *Boundary-Layer Meteorol.*, 18:299–313, 1980.
- [6] S. Hooker, W. Hooker, W. Esaias, G. Feldman, W. Gregg, and C. McClain. Seawifs technical report series: Volume 1, an overview of seawifs and ocean color. Technical Report NASA Tech. Memo. 104566, NASA Greenbelt Space Flight Center, Greenbelt, Md., July 1992.
- [7] C. McClain, M. Cleave, G. Feldman, W. Gregg, and S. Hooker. Science quality seawifs data for global biosphere research. *Sea Tech.*, 39:10–16, 1998.
- [8] J. Quinlan. Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4:77–90, 1996.
- [9] F. Wang. Fuzzy supervised classification of remote sensing images. *IEEE Trans. Geosci. and remote sensing*, 28:194–201, 1990.
- [10] W. Yao. Knowledge-based classification of seawifs satellite images for monitoring phytoplankton blooms off west florida. Master's thesis, University of South Florida, 1999.
- [11] M. Zhang, L. O. Hall, D. B. Goldgof, and F. E. Muller-Karger. Fuzzy analysis of satellite images to find phytoplankton blooms. In *IEEE International Conference on Systems Man and Cybernetics*, 1997.