Fuzzy Classification of Ocean Color Satellite Data for Bio-optical Algorithm Constituent Retrievals

Timothy S. Moore, Janet W. Campbell, and Hui Feng

Ocean Process Analysis Lab
University of New Hampshire
ABSTRACT

The ocean has been traditionally viewed as a 2 class system. Morel and Prieur (1977) classified ocean water according to the dominant absorbent particle suspended in the water column. Case 1 is described as having a high concentration of phytoplankton (and detritus) relative to other particles. Conversely, case 2 is described as having inorganic particles such as suspended sediments in high concentrations. Little work has gone into the problem of mixing bio-optical models for these different water types. An approach is put forth here to blend bio-optical algorithms based on a fuzzy classification scheme. This scheme involves two procedures. First, a clustering procedure identifies classes and builds class statistics from in-situ optical measurements. Next, a classification procedure assigns satellite pixels partial memberships to these classes based on their ocean color reflectance signature. These membership assignments can be used as the basis for a weighting retrievals from class-specific bio-optical algorithms. This technique is demonstrated with in-situ optical measurements and an image from the SeaWiFS ocean color satellite.
**Introduction**

In remotely-sensed ocean color data, boundaries between different water types are uncertain (or fuzzy). This is particularly true in the nearshore environment where suspended sediments and dissolved organic matter influence the optical signature of the water. It is generally agreed that site-specific algorithms will have to be developed to account for differences in the optical properties of particles and dissolved substances found in a particular region. Once these algorithms are developed and parameterized, one must decide when or where to apply a particular algorithm.

Fuzzy classification techniques are well suited for ocean color remote sensing applications and can solve the problem of blending bio-optical algorithms. A fuzzy classification scheme can mix algorithms by extending partial membership of a pixel to one or more water classes. This is done by computing memberships to predetermined classes by a membership function which uses the remotely-sensed reflectance spectra for a given pixel and generated class statistics from in-situ measurements. These memberships can then be used to weight the retrievals of class-specific algorithms. This allows a graded transition between different water types in an image scene, and multiple class-dependent algorithms can be effectively mixed or blended. The work presented here will demonstrate a fuzzy classification scheme using a 2-step procedure. (Step 1) The clustering of in-situ reflectance data is first segregated into distinct classes and (Step 2) the membership determination of multispectral satellite data is computed for these classes using a membership function.

**Fuzzy Logic**

Fuzzy logic was first developed by Zadeh (1965) as a mathematical way to represent vagueness contained in imprecise information. It is a superset of classical set theory which contains objects that are required to satisfy precise boundaries for set membership. Fuzzy theory extends beyond this type of hard precision and allows for partial set membership and boundaries which are not sharply defined (which is more typical of environmental data). The crisp event (one pixel-one class) is
replaced by an event or (class membership) that can have a real value between 0 and 1. In the context of remote sensing, a set (or class) is defined by the partitioning of spectral space (feature space), and classification becomes a matter of an object’s position in feature space. These membership grades describe the extent to which a pixel may belong to a class, which involves quantifying the nearness or proximity of a pixel’s position in spectral space to a pre-defined class vector. The membership values assigned may belong to one of three categories (label types) described in the literature (Bensaid et al, 1996): crisp (or “hard”; non-fuzzy), fuzzy and probabilistic. Let $c$ denote the number of classes, $1 \leq c \leq n$, and define three sets of label vectors as follows:

$$N_{fcu} = \{ y \in \mathbb{R}^c \mid y_i \in [0,1] \forall i \}$$ (unconstrained) fuzzy;

$$N_{fc} = \{ y \in N_{fcu} \mid \sum_{i=1}^{c} n_i = 1 \}$$ (constrained) fuzzy/probabilistic;

$$N_c = \{ y \in N_{fc} \mid y_i \in \{0,1\} \forall i \}$$ crisp (or hard).

The crisp or hard label allows for full membership in a single class, and no membership in any other class. The constrained fuzzy label allows for partial membership to any or all classes with the restriction that the sum of all partial memberships must equal 1.0. The unconstrained label allows for partial membership to any of the classes, but there is no restriction on the sum of the class memberships. Both the constrained and unconstrained labels can allow for full membership in a given class, or can be “hardened” to produce a crisp label set by changing the maximum class membership value to “1” and setting all others to “0”.

**The Complete Fuzzy Model**

The advantage of fuzzy partitioning is that is allows for intermediate situations and class mixtures to be described without loss of information. A flowchart is depicted in Figure 1 which illustrates the complete scheme of using fuzzy classification with bio-optical algorithm parameterization to retrieve in-water constituent
concentrations. According to this scheme, in-situ optical data are clustered and accompanying in-water concentration measurements are separated according to class. Algorithms can be parameterized for each class (Feng et al, unpublished). Class means and other statistics are calculated and pixel memberships can be computed from a satellite image. The membership function then can be used to weight the retrieval of each class algorithm. The weighted sum of all algorithm retrievals becomes the blended retrieval for that pixel.

Methods

Cluster Analysis - Step One

An in-situ optical data set (Kishino et al., 1985) was used in this work to demonstrate the clustering step. The measurements made for this data are irradiance reflectance (every 5 nm from 400 to 750 nm), Chl $a$, total suspended matter (TSM), colored dissolved organic matter absorption ($a_t$), and Secchi depth.

Unsupervised clustering is the process of identifying and characterizing natural subgroups (creating classes) that exist within the data. This procedure was performed with the MultiSpec software (Landgrebe and Biehl, 1997), which uses the ISODATA method, a self-iterative algorithm based on a minimum Euclidean distance measure to cluster data, and other in-water measurements were grouped according to these clusters.

Fuzzy Classification - Step Two

Class membership grades are determined for each pixel in a satellite image. The membership grade is a relative measure of how close the pixel is to the class mean radiance vector. The type of class membership function used determines not only the membership grade, but shapes the decision-boundaries produced by the classifier. For example, the Minimum Euclidean Distance classifier produces a hyper-plane boundary as opposed to Guassian Maximum Likelihood classifiers which produce quadratic boundaries (Shahshahani and Landgrebe, 1994). The decision boundary shape should reflect the complexity of the class shape which varies and is
dependent on the class statistics, such as the variance and covariance.

A Chi-square probability function was used for the class membership function. This function is given as follows:

\[ P_{I}(\chi^2 > Z^2) = \int_{0}^{Z^2} f(\chi^2) d\chi^2 \]

where \( \chi^2 \) is the Chi-square distribution function given by:

\[ f(\chi^2) = \frac{\chi^{(n/2)-1}e^{-\chi/2}}{2^{n/2}\Gamma(n/2)} \]

and \( Z^2 \) is the Mahalonobis distance from the pixel radiance vector \( x \) to the \( i \)th class mean radiance vector. The result is the probability \( P \) of observing the radiance or reflectance vector \( x \) in water of class \( I \). The probability was based on a 4 band vector (at 412, 443, 490 and 555 nm) to simplify the integral.

**Image Pixel Classification**

A satellite image from SeaWiFS has been used to demonstrate the fuzzy classification scheme. The calculations were performed based on remote sensing reflectance. The in-situ optical measurements were converted from irradiance reflectance to remote sensing reflectance with the following equation:

\[ R_n = \frac{R}{Q} \]

where \( R \) was the measured irradiance reflectance. A constant "Q" factor of 4.5 was used, although it is known to vary from 3 to 5 (Morel *et al.*, 1995).

Satellite water-leaving radiance values were also converted to remote sensing reflectance with the following equation (Gordon *et al.*, 1988):
where \( L_{wN} \) is the normalized water-leaving radiance and \( M \), \( F_0 \) and \( rQ \) are wavelength-dependent parameters.

**Results**

**Irradiance Reflectance Clustering**

The clustering process resulted in 4 clusters being identified. Their reflectance characteristics can be seen in Figure 2. The spectral shapes for clusters 1-3 show that these waters are predominantly green in color (cluster 4 exhibits a blue-green water color). This is typical of coastal waters and turbid estuarine environments. Spectral shape and magnitude varied amongst the four classes. The differences and similarities are more discernible when looking at the mean class spectral values plotted together (Figure 4). Class 1 had the highest mean values from 400 to 550 nm, as well as a pronounced plateau between 490 to 570 nm. Class 2 and 3 were similar in shape throughout the spectrum, but with different magnitudes. These 2 classes exhibited a rising reflectance curve peaking around 500 nm, then declining. Class 4 exhibited the only declining reflectance shape from 400 to 500 nm, which is typical of a more oceanic-type reflectance shape (Roessler and Perry, 1995). All classes exhibited similar shape after 500 nm. The spectral shapes of these classes can be explained by examining the in-water constituents grouped by cluster.

**Tokyo Bay In-situ Data**

The clustered in-situ data are shown in Figure 3. Chl \( a \) concentrations ranged from 0.25 \( \mu g/l \) to 34.48 \( \mu g/l \), while TSM concentrations ranged from 0.10 to 6.4 mg/l. Generally, Chl \( a \) and TSM concentrations tended to covary. Thus, phytoplankton dominate the particulate fraction for most stations. However, cluster 1 tended to have a lower Chl \( a \) to TSM ratio, perhaps indicating that suspended sediments contribute significantly to the radiances distribution. This effect can be seen in the mean reflectance curve
for cluster 1 (Figures 2 and 4). Secchi depths ranged from 2.5 m to 27.0 m, and also correlated with high and low levels of TSM and Chl a.

The \( a_g \) absorption coefficient for \( \lambda=375\text{nm} \) ranged from a low of 0.55 to a high of 1.022. The background oceanic value for \( a_g \) has been previously assigned a value of 0.06 (Gordon et al., 1988). The \( a_g \) values measured in Tokyo Bay are much higher in comparison the open ocean, and is not an insignificant absorber in the water column. The \( a_g \) concentration also showed a covarying relationship with Chl a (Feng et al., unpublished).

**Satellite Image Pixel Classification**

The class probability maps are shown in Figure 5. These maps show the probability (in percent) that a given pixel belongs to the classes given in Figure 4. There is partial membership to class 1 along the coast near the Bay of Fundy, and along the western edge of Nova Scotia. There is also some membership over George’s Bank. When viewing the true color image of the scene (Figure 6), the pixels which show membership in class 1 are from the areas colored yellow. These are highly turbid areas in which the reflection from suspended sediments influences the optical signal.

A cluster of pixels showing membership to class 2 is located over George’s Bank. There is also a group of pixels near Nantucket Shoals which shows partial membership to this class. Weak membership or probability is given to pixels associated with class 1 (see above). This class is associated with orange-brown colored pixels from Figure 6.

Membership in class 3 is strong along the New England coast, particularly high in what appears to be an offshore plume extending from a Maine estuary (Penobscot). There is partial membership from waters circling George’s Bank, and a parcel of water near the Bay of Fundy. The pixels showing membership in class 3 associate with the brown-colored pixels in Figure 6.

Class 4 membership is high throughout the interior of the Gulf of Maine and south of Cape Cod. Notably absent are pixels from George’s Bank and the offshore plume associated with class 3, as well as near-shore waters. Membership in this class is associated with the dark green-blue areas of Figure 6. This membership is consistent with the type of water that class 4 represents which can be gathered
from Figure 3. This is the lowest chlorophyll a-laden water type from Tokyo Bay and the class mean spectral shape resembles typical "oceanic" water.

There are areas of the image that do not show membership in any of the classes represented. A map of the sum of class probabilities (Figure 7) shows which areas are poorly represented. The most conspicuous areas are the offshore open-ocean waters along the bottom and near the bottom right of the Gulf of Maine image scene. The Tokyo Bay measurements did not contain any true "open-ocean" stations and it is not surprising that these pixels are not represented in the image. The other areas of low probability sums are pixels from George's Bank and near the coast around Nova Scotia. These are waters from very turbid environments and contain suspended sediments. This type of water also was not well represented in the Tokyo Bay data set, although class 1 does show a high total suspended matter to chlorophyll a ratio which may indicate a stronger effect on the optical signal from suspended sediments.

In contrast, there are pixels which have more than 100% probability (red pixels) which are found in the central portion of the Gulf of Maine and along certain areas of the coast. This effect is caused by a lack of separation between classes or over-representation. The classification maps that overlap in these areas are class 3 and 4 (Figure 5). The mean spectral curves for these classes in Figure 3 show that they are both lower than the other 2 class curves, but have opposite slopes in the SeaWiFS band range and appear separate. However, the Q factor is also influencing the class membership values as it is not constant but varies as seen from in situ measurements (Zibordi et al., 1997).

Discussion

Classification of ocean water types was introduced by Jerlov (1951) and was based on the transmittance of downwelling irradiance in the surface layer. Jerlov (1976) discusses 12 optically different classes of ocean water ranging from oceanic to coastal. Morel and Prieur (1977) classified ocean water into 2 cases based on the type of absorbent particle suspended in the water column. Other classification schemes based on other criteria have also been
put forth (Kirk, 1980; Pevelin and Rutkovskaya, 1977; Smith and Baker, 1978). The classification scheme introduced here is built upon clustering optical data into classes based on irradiance spectral characteristics. These classes are used to extend partial or full memberships to ocean color satellite data using a membership function. This serves to allow for mixtures of water types that naturally occur in the ocean.

The methodology presented here is one that is commonly used in the classification of land remote sensing images (Wang, 1990; Jia and Richards, 1994; Jenson, 1996). The fuzzy probability results presented here are intermediary in the sense that the pixel probability values will be used as weighting-factors for the class-specific algorithms. The final output of the processing stream will be constituent concentration maps such as chlorophyll $a$ and $a_t$ absorption. However, the classification maps themselves contain some intriguing information. The notion of separable ocean classes and class mixtures raises the question of exactly what do we mean by ocean classes and class membership? Some insight into this can be gathered by comparing land classification with ocean classification [as used in the context of this paper].

The ocean is not a static or rigid environment unlike the terrestrial environment, but a fluid constantly moving carrying particles that are changing in terms of their concentration and composition. These changes cause the spectral nature of ocean pixels to rapidly change over space and time. The spectral response of land pixels change as well, but the time scales for changes much greater. Thus, class structure in the ocean varies dynamically with space and time. There is no easy way of verifying class structure within the ocean. Land scenes can be compared with airplane photographs and topographic maps, but these are not easily obtained for the ocean. Since the ocean is constantly changing, sea-truth information would only be valid for a concurrent remote-sensing overpass. Despite these limitations, the spectral classes for the ocean may be far fewer than the land. A typical land scene may have over 60 known classes, compared to the 4 classes found in this analysis. Although only 4 classes were identified, it would be expected for these to change from region to region. We performed the same unsupervised clustering on an optical data set from the North Sea and 4 clusters were coincidentally identified, but were vastly different in terms of spectral shape. Not only are there fundamental differences in the
way that classes are thought of and measured between the land and ocean, the way these classes are interpreted on a pixel and sub-pixel level are also very different.

The concept of fuzzy classification in land remote sensing is typically used as a way of subdividing the pixel into more than one class coverage (a class coverage is a labeled class such as “woods”, “grassland”, or “pasture”). The membership values assigned for a pixel to given classes represent the percent composition of that coverage within the pixel. In contrast to a mixed land pixel, a mixed-class ocean pixel can truly be a blend of water types which share characteristics down to the Nisken bottle sample (it also can be two or more water types occupying the same pixel). These class mixtures then represent real blends of water in between ocean classes. The resulting classification maps in fact show convincing patterns of distinct ocean water masses and the gradual (in some cases abrupt) transition with adjacent water masses. The use of a classification scheme may also then provide additional information then as a tracer of water mass movement from a time-series of satellite images that one may not readily see in a constituent concentration map. While the results presented here do show that, it is important to consider that this is really a demonstration of the application of a fuzzy classification scheme. There are limitations in this present example.

The in-situ optical data from Tokyo Bay seems to be reasonably representative with water types in this particular image scene from Gulf of Maine. However, as evident from Figure 7, not all water types are accounted for in the in-situ data. This can be explained by the lack of any open ocean stations in the Tokyo Bay data set, and the presence of open-ocean water in the sample image scene. It is also important to consider that the in-situ data was irradiance reflectance which was converted to remote sensing reflectance. The choice of a constant Q-factor is not optimal in this conversion. The Q-factor varies with wavelength and with radiance distribution. Thus, the mean class spectral values for remote sensing reflectance were not entirely accurate, and were a distortion of true class mean spectral shape.

There were 4 classes determined from the Tokyo Bay data set. Landgrebe (1994) lists three basic requirements for the training of a classifier. These are 1) the number of classes must be exhaustive, or that all pixels in a scene can be assigned to a class; 2) the classes
must be separable by their spectral characteristics; and 3) the classes must of informational value, or be classes of interest to the user. It is the purpose of a training set to provide a heterogeneous population that adequately represents the true classes that naturally exist for that region. The samples from Tokyo Bay were restricted to a fairly small spatial domain of a few 10's of kilometers, compared to the image scene used in this analysis which is 100's of kilometers. While the data set may adequately describe the range in water types in Tokyo Bay, it may not a complete representation of other regions. Obviously, the water types from Tokyo Bay are not totally representative of every water type from the Gulf of Maine scene. Every region may have its own unique water types. Models developed from regional data sets become tuned to those areas. An example of this can be seen in primary productivity models (Campbell et al., 1998). Similarities of water types between regions are evidenced in the Gulf of Maine class memberships maps created from the Tokyo Bay classes. However, more comparison is needed between reflectance measurements and bio-optical models from various regions before cross-regional models can be deemed reliable.

It is desirable to have the membership sum for a given pixel to have a value of 1.00 to fully account for all probabilities. Using the methods we employed, each class probability was allowed to be unrestricted and take on whatever value was returned by the probability function. It is common in land coverage classification (using Landsat TM imagery) to normalize individual class probabilities to the sum of all class probabilities and force memberships to total to 1.00 (Jenson, 1996). This can distort the true probability of a given vector belonging to a labeled class, especially when the pixel vector is far away from class mean vectors and returned probabilities are of very small magnitude. This can assign class membership to a pixel when in reality none exists. However, it can be seen that membership sums can exceed 1.00. It may be advantageous to normalize in these circumstances, or distortions may also arise when applying these weights to the class algorithms.

With all these considerations, fuzzy logic can discriminate between different water types and allow for class mixture situations. This can be particularly beneficial to bio-optical algorithm retrievals, which up to now have not adequately responded to the challenge of blending different algorithms suited to different water types.
Although the results presented here are prone to errors previously discussed, the promise of real progress on this front is not far away in the future. By creating and parameterizing bio-optical algorithms based on local in-situ data sets which contain remote sensing reflectance measurements and concurrent IOP measurements, the implementation of a fuzzy classification scheme can improve constituent retrieval accuracy.
Conclusions

Fuzzy classification of ocean color satellite images has been demonstrated. Based on the clustering of in-situ optical measurements, pixels can be assigned partial memberships to these classes. The class memberships can be used as to derive weighting factors for class-specific bio-optical algorithms. This method allows for algorithm blending in a way that avoids the "patch-work quilt" effect associated with non-weighted or hard-partitioned classification schemes. The initial choice of the number of clusters to be extracted from the in-situ optical data is crucial. This will vary from region to region. In our analysis, 4 classes were determined. Further work on the blending of these domain classes remains as one of our tasks. In any case, fuzzy classification seems well suited to the needs of ocean color remote sensing.

Acknowledgement

This work was supported under a MODIS Instrument Team Investigation (NASA Contract NAS5-96063).
References


Figure 1. Flowchart of a Fuzzy Classification and Parameterization scheme. In-situ reflectance measurements are grouped by a clustering algorithm to form unique classes. Other in-situ measurements are grouped according to spectral reflectance grouping. Bio-optical algorithms are parameterized with the class-specific measurements. Satellite image pixels are then assigned graded memberships from 0 to 1 to each class using a membership function. These class memberships are used to weight each class algorithm constituent retrieval. The weighted sums form the final retrieved constituent concentration for that pixel.
Figure 2. Tokyo Bay irradiance reflectance cluster results based on 45 measurements taken at 67 wavelengths (400 - 750 nm every 5 nm). Top left: Cluster 1, N=6; top right: Cluster 2, N=12; bottom left: Cluster 3, N=16; bottom right: Cluster 4, N=11. Mean curves are shown in red, individual stations in blue. These spectral curves were clustered using the ISODATA method using minimum Euclidean distance.
Figure 3. Tokyo Bay in-water measurements grouped according to spectral clustering (Figure 2). The open red circle indicates the maximum recorded value; the red star '*' indicates the minimum; the mean and 1 standard deviation are represented by the blue rectangle.
Figure 4. Tokyo Bay remote sensing reflectance (Rrs) class means. Tokyo Bay irradiance reflectance was converted to Rrs using a constant Q factor of 4.5. Reflectance values at the SeaWiFS bands (412, 443, 490, 510 and 555 nm) are shown with *. The values at 412, 443, 490 and 555 nm were used in the membership function.
Figure 5. Pixel membership to the 4 Tokyo Bay classes for a SeaWiFS image taken over the Gulf of Maine on Oct. 8, 1997. The image shows the probability (in percent) of each water pixel belonging to each class. Pixel membership was determined from a Chi-square distribution function. Top left: Class 1 membership; top right: Class 2 membership; bottom left: Class 3 membership; bottom right: Class 4 membership.
Figure 6. True color image for Oct. 8, 1997. The color of the water is closely associated with class memberships from Figure 5.
Figure 7. Probability Sum map of the 4 class memberships shown in Figure 5. Pixels range from low probability (black and dark blue pixels) to 100% probability (white). Red pixels are where the sum was greater than 100%. This map shows the coverage of class distribution in the image. Low probability pixels (bottom right) indicate a lack of representation in the in-situ measurement set.
This is a quarterly progress report for the period January through March 1998 for the Definition Phase of my MODIS Instrument Team investigator project. In this report, I will describe progress made by a research team consisting of myself, a full-time assistant research scientist (Timothy Moore), and a graduate research assistant (Hui Feng).

The ocean has been traditionally viewed as a 2 class system. Morel and Prieur (1977) classified ocean water according to the dominant absorbent particle suspended in the water column. Case 1 is described as having a high concentration of phytoplankton (and detritus) relative to other particles. Conversely, case 2 is described as having inorganic particles such as suspended sediments in high concentrations. Little work has gone into the problem of mixing bio-optical models for these different water types. An approach is put forth here to blend bio-optical algorithms based on a fuzzy classification scheme. This scheme involves two procedures. First, a clustering procedure identifies classes and builds class statistics from in-situ optical measurements. Next, a classification procedure assigns satellite pixels partial memberships to these classes based on their ocean color reflectance signature. These membership assignments can be used as the basis for weighting retrievals from class-specific bio-optical algorithms. This technique is demonstrated with in-situ optical measurements and an image from the SeaWiFS ocean color satellite.