CHAPTER 28

OPTICAL CLASSIFICATION

Monolisha S., Grinson George, Shubha Sathyendranath and Trevor Platt

Fishery Resources Assessment Division
ICAR-Central Marine Fisheries Research Institute

Marine Optics

Optical oceanography or Marine optics is the study of light propagation in the ocean surface through absorption or scattering processes. Marine bio-optics is the term used when the absorption and scattering by particles and dissolved substances are of biological origin. Ocean color is defined as the spectral variation of the water leaving radiance that can be related to the optical constituents present in the medium (Jerlov, 1976; Morel, 1974). Visible Spectral radiometry or Ocean colour remote sensing is the study on spectral signals of optically active materials using satellite observations. When sunlight reaches the upper water column or the photic zone of the ocean surface, the light propagation is determined by the optical properties of seawater. It depends on concentration of the optical constituents of seawater containing Coloured dissolved organic matter (CDOM), suspended sediments and phytoplankton (IOCCG, 2000). The contribution of particulate and dissolved constituents to the variability of optical properties and ocean color in coastal waters requires a better understanding of the linkages between the concentration of these constituents, the inherent optical properties (IOPs) of absorption and scattering coefficients, and the apparent optical properties (AOPs) such as the spectral attenuation for downward irradiance $K_d(\lambda)$ and remote sensing reflectance $R_{rs}(\lambda)$. Knowledge of these relationships is important for characterizing the marine optical environment and developing remote sensing ocean color algorithms for coastal waters.

The approach of Optical classification is the analysis on identification of a water type in terms of dissolved and suspended organic and inorganic materials, biological substances or the phytoplankton diversity. It has practical applications including the quantitative description of ocean color and the satellite remote sensing of chlorophyll (Wozniak and Pelevin, 1991). Based on the optical properties, different versions of optical classification of water types were proposed by many authors (Jerlov, 1976; Kirk, 1976; Morel and Prieur, 1977; Sathyendranath and Morel, 1989; Smith and Baker, 1977). Jerlov (1976) introduced a classification of water bodies based on their spectral optical attenuation depth. Jerlov classified his observations into a set of five typical oceanic spectra and nine typical coastal spectra. Morel and Prieur, (1977) carried out an independent analysis of spectral irradiance with the aim of describing ocean color in terms of dissolved and suspended material, in particular, phytoplankton pigment concentrations. Morel's analysis makes use of the inherent optical properties of water (Preisendorfer, 1976). Morel and Prieur (1977) presented the rationale for separating all water masses into two types: Case 1 and Case 2 waters. According to the classification system used in remote sensing studies, Case-1 represents the phytoplankton-dominated waters and the clear waters with algal or biological materials, and Case-2 represents all other possible water bodies rich in organic and inorganic substances. In Case-1 water, simple algorithms to retrieve pigment concentrations are used globally, but for Case-2 waters, the use of site-specific algorithms is necessary. Smith and Baker, (1977) used the apparent optical property K_T, the optical parameter that relates the spectral irradiance just beneath the ocean surface E_d (0, A) to the downwelling spectral irradiance at depth E_d(Z,A). Their classification also provides direct input to mathematical models of phytoplankton dynamics. Sathyendranath and Prieur, (1989) used a triangular plot, on which a point would represent the relative contributions from phytoplankton, non-algal particles, and CDOM to the total absorption coefficient (excluding that of pure seawater), at a specific wavelength. A point at the center of the triangle designates equal contributions by each component, while each point at the apex designates cases where all the contribution is from a single component. There are several studies on optical classification of the water types based on insitu optical measurements by Hoepffner, 2005; McKee and Cunningham, 2006; Reinart *et al.*, 2003; Roff *et al.*, 2003. Babin *et al.*, 2003 gave results based on absorption at various wavelengths for European coastal waters. The study explains how at different wavelengths different constituents (dissolved material, algae and non-algal particles) dominate absorption. Mélin and Vantrepotte, (2015) classified the global ocean into sixteen optical classes using the normalised remotely sensed reflectance data, an apparent optical property. He also classified the water types of each class into Case-1 and Case-2 using the mean spectral reflectance values. Recent studies on optical classification in the Indian coast using the empirical algorithms for the retrieval of chl-*a* from CZCS, MOS-B, IRS-P4-OCM and Sea-WiFS were carried out in the southeastern Arabian Sea (Nagamani *et al.*, 2008; Chauhan *et al.*, 2002; Sathe and Jadhav, 2001). Study on the phytoplankton community characteristics using absorption properties in the coastal waters of the southeastern Arabian Sea was done by Minu *et al.*, 2014; Minu *et al.*, 2016, measured the *in-situ* remote sensing reflectance (R_{rs}) and optically active substances (OAS) using hyper spectral radiometer and classified the coastal waters of the southeastern Arabian Sea. The authors proposed three distinct water types: Type-I, Type-II and Type-III based on variability in Optically Active Substance (OAS) such as chlorophyll-*a* (chl-*a*), chromophoric dissolved organic matter (CDOM) and volume scattering function at 650 nm (β_{650}).

Optical classification of Northern Indian Ocean

We made an initial attempt to optically classify the coastal waters of the Northern Indian Ocean using remotely-sensed ocean colour datasets. Monthly climatological dataset of remote sensing reflectance for the years 1998 – 2013 was obtained from the Ocean Colour

Climate Change Initiative (OC-CCI, www.oceancolour.org). Normalization and log-transformation of remote sensing reflectance values were done according to the method of Mélin and Vantrepotte (2015). Optical classification implemented uses the fuzzy logic classification method, based on Moore *et. al.,* (2009). Optimal cluster validity methods such as Xie-Beni Index and Partition Coefficient are computed to determine the optimal cluster (class) number to perform the classification.

Optical classification techniques used for the study:

Cluster validity measures

Cluster validity measures are chosen to validate the quality of clustering algorithms. Cluster validity methods are statistical functions that determine the performance of a clustering procedure. Criteria of merit for a clustering method includes the distance between clusters (separation) and the distribution of points around a cluster (compactness) (Deborah *et al.*, 2010). We can rely on multiple validity functions to aid selection of the optimal cluster number. The principal strategy used is to cluster the data over a range of cluster values (n_c) and evaluate each clustering result with each validity function (Moore *et al.*, 2009).

The Partition co-efficient and the Xie-Beni index are cluster validity methods designed specifically for use with fuzzy algorithms. These two methods are preferred to select the optimal number of clusters in fuzzy classification (Halkidi *et al.*, 2002).

Xie-Beni Index

Xie-Beni index is used to determine the best cluster number for the fuzzy classification method in a particular application. Xie-Beni index depends on the geometric properties of the dataset and the membership matrix. This index is defined as the ratio of the mean quadratic error to the minimum of squared distances between all points in the cluster. In cluster validation, the quadratic error is defined as the mean of the squared distances of all the points with respect to the centroid of the cluster they belong to (Xie and Beni, 1991).

The variation in cluster i, (i = {1,…,n_c}) is designated as $ó_{i,}$ it is the sum of the squares of the fuzzy deviation of the data points in dataset X. The average variation in cluster i, η_i= (σ_i/η_i), η_i is the number of points in the cluster belonging to the cluster i. Xie-Beni index is defined as XB = η/N*$D_{min,}$, N is the number of points in the dataset and D_{min} is the minimum distance between the centroids of the clusters.

The smallest value of index indicates the optimal cluster number. When Xie-Beni index is monotonically decreasing, the number of clusters n_c becomes very large and close to n_i. One way to eliminate the decreasing tendency of the index is to determine a starting point, the maximum of the cluster number (c_{max}) of the monotonic behavior and to search for the minimum value of Xie-Beni in the range [2, c_{max}]. Moreover, the values of the Xie-Beni index depend on the fuzzy membership values (F) and the maximum cluster number of the datasets (Halkidi *et al.*, 2002; Zhao *et al.*, 2009).

Partition Co-efficient

The Partition Coefficient is a validity function that uses the membership values (F_{ij}) to provide the best cluster number. It measures the amount of "overlap" between clusters. Partition co-efficient (PC) is defined as follows:

$$PC = \frac{1}{N} \sum_{i=j}^{N} \sum_{j=1}^{n_c} F_{ij}^{2}$$

The PC index values lie in the range [$1/n_c$, 1], where n_c is the number of clusters. The closer this value is to one the better the data are classified. The cluster number with a maximum partition coefficient is said to be the best cluster number to choose for classification. In case of a hard partition, we obtain the maximum value PC(n_c) = 1. The disadvantages of the partition coefficient are its monotonic decrease with cluster numbers n_c (Bezdek, 1974; Bezdek *et al.*, 1984).

Fuzzy C Mean Classification

Fuzzy classification evolved from classical set theory. The classical clustering approach determines whether the object is a member or non-member of a given set of that system. In contrast, fuzzy logic allows that an object or data may have partial memberships of more than one set. This method allows for overlap between boundaries of particular classes or sets, and recognizes that more than one class may be represented at a particular location at any given time. The membership F_{ij} of a class in the data from a particular point is given by ($1-F_{ij}(Z_{ij}^{2})$) where Z_{ij} is the Mahalanobis distance given by (X-M)/S where M is the mean and S is the standard deviation, and F is a cumulative χ^2 distribution (Zadeh, 1965).

Results of the study

The cluster validity measures showed that the cluster number 8 gave the best compromise, with low numbers of both under-classified and over-classified pixels. Therefore, eight classes were selected as the optimal cluster number for further analyses. Preliminary analyses of optical classification showed month-to-month variations (See Figure 1.Optical classification for the summer monsoon seasons - June to September). The mean spectra of the eight selected optical classes were calculated and shown in Figure 2. The optical classes also relates to Case-1 and Case-2 waters as defined by Morel and Prieur (1977); Prieur and Sathyendranath, (1989) based on spectral shapes. From the shapes of the spectra, it appears that classes 1-6 are representative of Case-1 waters and classes 6-8 of turbid Case-2 waters. On further, the resultant optical classes are analyzed to explore their biological significance using the available distribution datasets of different taxonomic groups of phytoplankton and zooplankton from literature.

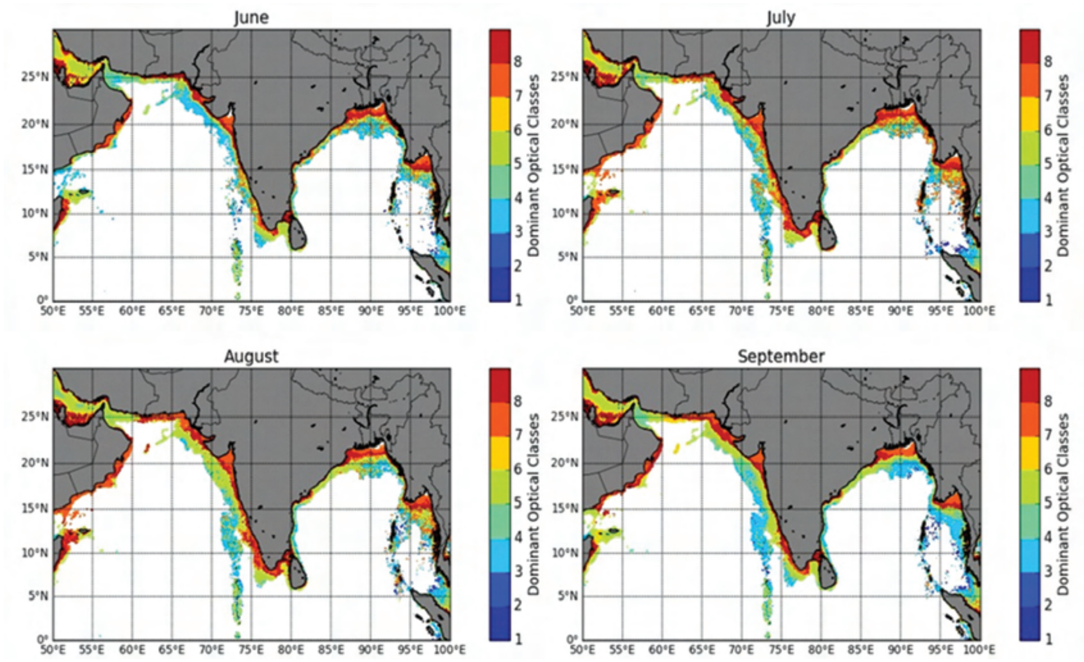

Fig. 1. Optical classification of the monthly climatology datasets of summer monsoon season

Although there have been many biogeographic studies of the ocean with various approaches and many applications of remote-sensing to partition the oceans into ecological zones or provinces, studies that integrates conventional biogeography with the results from remote sensing have been relatively few. Application of remotely sensed data to classify any optically distinct regions can result in providing understanding on the various bio-geochemical cycles and dynamics of the ocean regions.

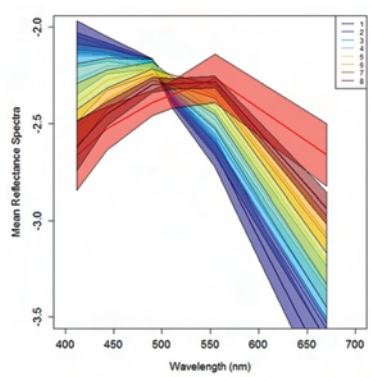

Fig. 2. Mean Reflectance Spectra of the Log10-normalised Rrs values

References

Babin, M., Stramski, D., Ferrari, G. M., Claustre, H., Bricaud, A., Obolensky, G., *et al.,* (2003). Variations in the light absorption coefficients of phytoplankton , nonalgal particles and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res. Ocean.* 108, 1–20. doi:10.1029/2001JC000882.

Bezdek, J. C. (1974). Cluster validity with fuzzy sets. *J. Cybern.* 3, 58–73. doi:10.1080/01969727308546047.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* 10, 191–203. doi:10.1016/0098-3004(84)90020-7.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002). Clustering validity checking methods. *ACM SIGMOD Rec.* 31, 19. doi:10.1145/601858.601862.

Hoepffner, N. and Dowell, M. (2005). Assessing the Dynamics of Ecological Provinces in the European Seas. European Commission – Joint Research Centre Institute for Environment and Sustainability Inland & Marine Waters Unit, 1-43.

Deborah, L.J., Baskaran, R., and Kannan, A. (2010). A Survey on Internal Validity Measure for Cluster Validation. *Int. J. Comput. Sci. Eng. Surv.* 1, 85–102. doi:10.5121/ijcses.2010.1207.

Jerlov, N. G. (1976). *Marine Optics*. Elsevier Oceanography Series, 276. doi:10.1016/S0422-9894(08)70794-3.

Kirk J.T.O (1976). A theoretical analysis of the contribution of algal cells to the attenuation of light within natural waters. *New Phytol.* 77, 341–358. doi:10.1111/j.1469-8137.1976.tb01524.x.

McKee, D., and Cunningham, A. (2006). Identification and characterisation of two optical water types in the Irish Sea from in situ inherent optical properties and seawater constituents. *Estuar. Coast. Shelf Sci.* 68, 305–316. doi:10.1016/j.ecss.2006.02.010.

Mélin, F., and Vantrepotte, V. (2015). How optically diverse is the coastal ocean? *Remote Sens. Environ.* 160, 235–251. doi:10.1016/j.rse.2015.01.023.

Minu P., Shaju S. S, Muhammed Ashraf P., Meenakumari, B. (2014). Phytoplankton community characteristics in the coastal waters of the southeastern Arabian Sea. *Acta Oceanoglogia* 33, 170–179. doi:10.1007/s13131-014-0571-x.

Minu, P., Lotliker, A. A., Shaju, S. S., Ashraf, P. M., Kumar, T. S., and Meenakumari, B. (2016). ScienceDirect Performance of operational satellite bio-optical algorithms in different water types in the southeastern Arabian Sea. *Oceanologia* 58, 317–326. doi:10.1016/j.oceano.2016.05.005.

Moore, T. S., Campbell, J. W., and Dowell, M. D. (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sens. Environ.* 113, 2424–2430. doi:10.1016/j.rse.2009.07.016.

Morel, A. (1974). Optical Properites of Pure Water and Pure Sea Water, 1–24.

Morel, A., and Prieur, L. (1977). Analysis of variations in ocean color. *Limnol. Oceanogr.* 22, 709–722. doi:10.4319/lo.1977.22.4.0709.

Nagamani P.V., Chauhan, P. and Dwived, R.M. (2008). Development of chlorophyll a algorithm for Ocean Colour Monitor onboard OCEANSAT-2 satellite. *IEEE Geosci. Remote Sens. Lett.* 5, 527–531. doi:10.1109/LGRS.2008.923213.

Chauhan, P., Mohan, M., Sarngi, R.K., MeenaKumari, B., Nayak, S. and Matondkar, S.G.P. (2002). Surface chlorophyll-a estimation in the Arabian Sea using IRS-P4 Ocean Colour Monitor (OCM) satellite data. *Int. J. Remote Sens.* 23, 1663–1676. doi:http://dx.doi.org/10.1080/01431160110075866.

Preisendorfer, R. W. (1976). *Hydrologic Optics*. Honolulu/ : U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Pacific Marine Environmental Laboratory Available at: http://udspace.udel.edu/handle/19716/1577.

Reinart, A., Herlevi, A., Arst, H., and Sipelgas, L. (2003). Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. *J. Sea Res.* 49, 357–366. doi:10.1016/S1385-1101(03)00019-4.

IOCCG (2000). Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. Sathyendranath, S. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 3, IOCCG, Dartmouth, Canada.

Roff, J. C., Taylor, M. E., and Laughren, J. (2003). Geophysical approaches to the classiûcation, delineation and monitoring of marine habitats and their communities. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 13, 77–90.

Sathe, P.V and Jadhan, N. (2001). Retrieval of chlorophyll from the sea-leaving radiance in the Arabian Sea. *J. Indian Soc. Remote Sens.* 29, 97–106.

Sathyendranath, S. and Morel, A. (1989). A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *Int. J. Mar. Sci.* 10, 1373–1394.

Smith, R. C., and Baker, K. S. (1977). Optical Classification Of Natural Waters University Of California , San Diego Scripps Institution Of Oceanography Visibility Laboratory Lajolla, California.

Wozniak, B., and Pelevin, V. N. (1991). Optical classifications of the seasin relation to phytoplankton characteristics. *Oceanologia*, 25–55.

Xie, X. L., and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 13, 841–847. doi:10.1109/34.85677.

Zadeh (1965). Fuzzy Sets, 25.

Zhao, Q., Xu, M., and Fränti, P. (2009). Sum-of-squares based cluster validity index and significance analysis. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 5495 LNCS, 313–322. doi:10.1007/978-3-642-04921-7_32.