

APPLICATION OF CLUSTERING TECHNIQUES TO MULTISPECTRAL OPTICAL DATA OVER THE OCEAN

Margaret Joseph¹, Graham D. Quartly², Colette Robertson², Valborg Byfield² and Ray Harris¹

¹ Department of Geography, University College London, 26, Bedford Way, London, WC1H 0AP

² National Oceanography Centre, Empress Dock, Southampton, Hants, SO14 3ZH, UK

ABSTRACT

MERIS, on *Envisat*, provides high-resolution radiometric data at nine discrete channels in the visible band. This paper looks at the potential of an unsupervised classification technique for utilizing these multi-spectral data to provide better discrimination between water masses according to their optical properties, and in particular whether phytoplankton groups can be distinguished. Although the majority of data do show a spectral peak associated with chlorophyll's red fluorescence line, clustering using only the red bands was found to separate out coastal waters according to their sediment content. Red-end classification also appeared to identify sub-pixel cloud, and demonstrate that the smile correction had not removed all the striping from the data. Classification using bands from the blue-green end showed a response to changes in chlorophyll concentration, but also indicated other variations. However, without *in situ* data no firm conclusions can be drawn on which phytoplankton groupings are present.

1. INTRODUCTION AND METHODOLOGY

The surface waters of the ocean are optically complex, with reflectances that are the result of absorption, scattering and fluorescence by plankton cells belonging to different phytoplankton taxa, which contain various pigments in addition to chlorophyll and vary widely in size and reflectivity. Pigment concentrations relative to biomass, and the relationship between chlorophyll fluorescence and chlorophyll content may vary from group to group, as well as with physiological state. In Case I waters, particle scattering and absorption by yellow substance (coloured dissolved organic material or CDOM) is directly related to plankton concentrations, and therefore co-vary with chlorophyll concentrations. In Case II waters (e.g. waters influenced by river-input) suspended sediment and yellow substance of terrestrial origin will also be present in varying concentrations and contribute to reflectances in a way that does not co-vary with chlorophyll concentration; see IOCCG [1] for further details on Case II waters. It is therefore to be expected that the reflectance spectra obtained from different types of water, containing different plankton communities and different contributions from CDOM or suspended sediments, should vary in a way that may be used to discriminate between different water types, and potentially between different plankton communities. Thus, optical algorithms based on clustering techniques may be expected to be capable of discriminating between Case I and Case II waters, and could also provide a method for distinguishing between different phytoplankton communities. Potentially this could lead to new techniques for the monitoring of harmful algal blooms, quantification of species associated with halocarbon production in the marine environment, and studies of how the dominant groups in a region change on seasonal and interannual time scales.

Other satellite-borne colour sensors (CZCS, SeaWiFS, MODIS) have recorded radiances at a number of wavelengths in the visible band (400 to 700 nm); however, it is hoped that new terrestrial and marine applications will be developed with the advent of MERIS, with its extra bands in the orange-red part of the spectrum, the 620nm band designed for increased sensitivity to particle scattering, and the three bands (7,8 and 9) designed to measure the height of the chlorophyll fluorescence peak at ~681 nm.

In this study, instead of using knowledge of the optical spectra associated with different seawater constituents, such as chlorophyll and yellow substance, we analyse the data using unsupervised classification. This is a statistical technique that identifies the key structures within the data in a purely objective manner. The principal idea is to divide the pixels into clusters based on their measured reflectance spectra, with minimum variability within each cluster but maximum variability between clusters. The specific technique trialled here was ISODATA unsupervised classification, implemented within the *ENVI* software package. Unsupervised classification has been extensively used in terrestrial remote sensing, but rarely applied in the marine environment, with a few notable exceptions [2,3].

For the purpose of this short paper we concentrate on a MERIS pass of 14th Feb. 2005 running down the western edge of the Indian Ocean from the Arabian Sea, past the island of Socotra at the mouth of the Gulf of Aden, along the east coast of Madagascar and on to about 40°S in the Southern Ocean. This dataset thus spans a diverse range of conditions from the oligotrophic waters of the central Indian Ocean to the various coastal regions. In this paper we first look at classification using the red end of the spectrum (section 2), then the blue end (section 3), and finally with all the visible channels (section 4). In section 5 we summarise the findings and discuss further extensions of this work.

2. CLASSIFICATION USING THE RED END OF THE SPECTRUM

Initial development was done using bands 7, 8 and 9 (665, 681 and 709 nm) from the red end of the spectrum. This is because chlorophyll not only acts as a scatterer of green light, but can also fluoresce at ~ 681 nm. Thus it is to be expected that the height of the fluorescence peak at band 8 relative to the background given by the neighbouring bands 7 and 9, provides a measure of the presence of active phytoplankton.

2.1 Subjective selection

A simple scatter plot of the reflectances in bands 8 and 9 (Fig. 1a) shows there to be two distinct groupings. The set of data encircled in red not only have higher values in both channels, but also belong to a completely separate envelope of points. Within geographical space these high reflectance points are found to belong to two cohesive groupings on the edges of islands (Fig. 1b). Examination of further MERIS images showed the persistence of these high reflectance pixels around the islands.

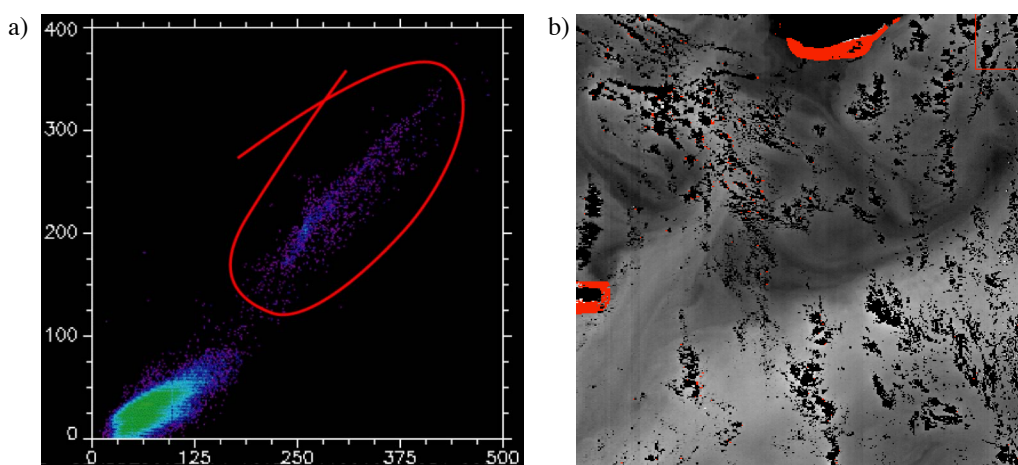


Fig. 1. a) Scatterplot of reflectances in channel 8 (x-axis) and 9 (y-axis) using all the data from the MERIS pass. Circled points are unlikely to be associated with chlorophyll fluorescence, as they have similar reflectance values in both channels. b) Zoom of channel 8 reflectances south of Socotra (black island at middle of top), with red indicating the points highlighted in the scatterplot.

2.2 Unsupervised classification

The ISODATA unsupervised classification finds an optimum set of groupings under chosen constraints. The user has to supply a range for the number of possible classes, plus the maximum number of iterations to be performed. For most cases, we required the process to produce between 5 and 10 classes, using at most 50 iterations. Under such conditions it always found 10 classes (with only 5 iterations, the process determined 7 classes). When automated clustering using bands 7, 8 & 9 was applied to the whole pass, two of the ten classes were found to represent 97.4% of the pixels. These two, indicated in red and green in Figs. 2-4 covered all the open ocean, with the green cluster predominating in the north (Fig. 2a) and the red in the south (Fig. 2b).

The other groupings only existed in coastal areas, where there is the greatest variability in the reflectances in the three channels, with the order of progression from open ocean to the coastline being the same in the two very separate regions focussed on in Figs 3b,c. The mean spectra of the red and green clusters have among the lowest reflectances of all the groupings (Fig. 3a), with most of the others showing a peak at $0.56\mu\text{m}$, indicative of suspended sediment.

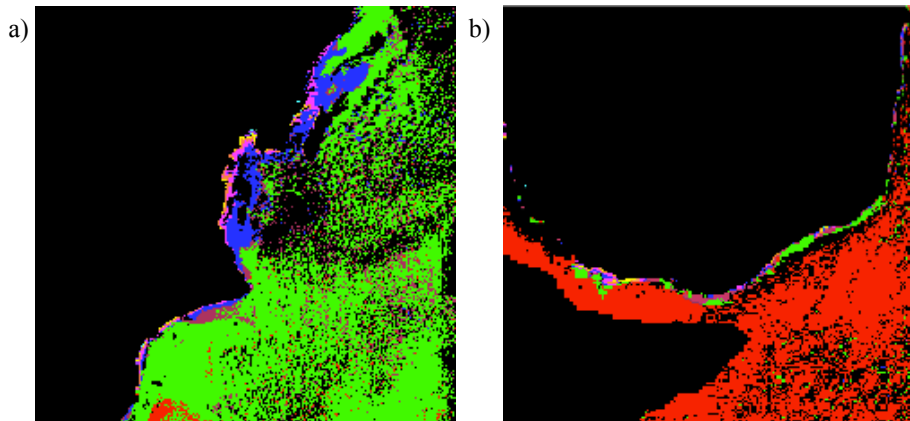


Fig. 2 Unsupervised classification of the whole pass using the red bands gave ten clusters indicated by different colours. Zooms are shown of a) Oman, and b) Southern Madagascar.

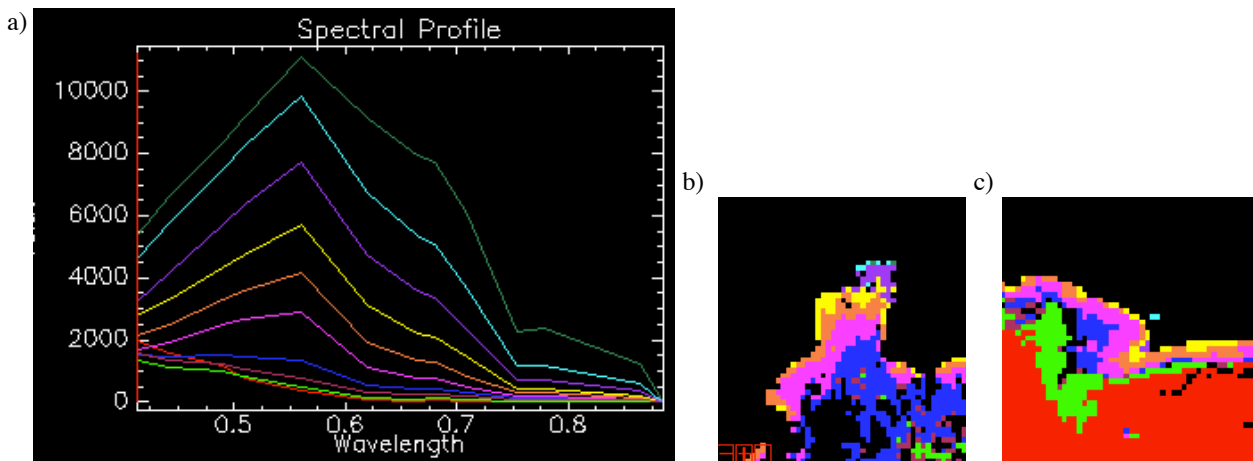


Fig. 3.a) Mean spectra associated with each of the 10 clusters. b), c) High-resolution segments from Fig. 2 showing coastal waters.

2.3 Non-oceanographic features

The spectra of the red and green clusters are very similar compared with other groupings, but, in some locations, the variations in spatial distribution between these two clusters are also suggestive of two factors of non-oceanographic origin. Most of the water in Fig. 4a falls within the red cluster; however, many of the pixels on the edge of clouds (black) are within the green cluster. It seems probable that these locations correspond to sub-pixel cloud or thin cirrus that have not been adequately flagged within the MERIS product. Figure 4b shows a section east of Socotra, where both red and green clusters are prominent. However, the pronounced vertical boundary between classes probably relates to a small step change in reflectances caused by inaccuracies in the 'smile' correction used to compensate for MERIS's five parallel camera swaths. Thus the clustering algorithm responds not only to ocean optical properties, but also to atmospheric conditions and instrumental artefacts. Consequently, red-end classification could be used to pinpoint problems within the MERIS product.

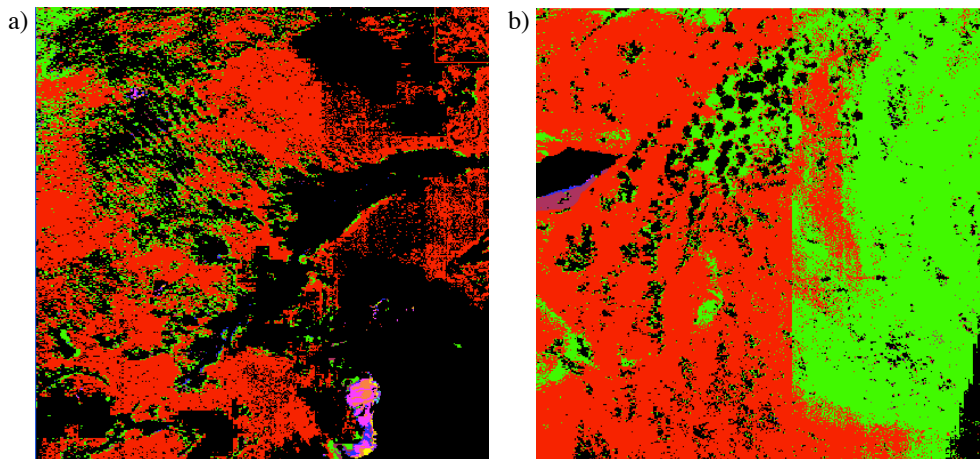


Fig. 4. a) Zoom of classification in area to north of Madagascar, showing patchy cloud (black). b) Zoom of region east of Socotra, showing near-linear boundary between red and green classes.

3. CLASSIFICATION USING THE BLUE END OF THE SPECTRUM

As chlorophyll is a strong absorber of blue light, the ratio of reflectances in blue and green is commonly used as the basis for algorithms estimating chlorophyll concentration [4]. However, a scatterplot of the reflectances in bands 2 and 5 (443 and 560 nm) shows a clumping of points (Fig. 5a) rather than a smooth continuum. There are also small clusters of points (not highlighted) lying above the main population.

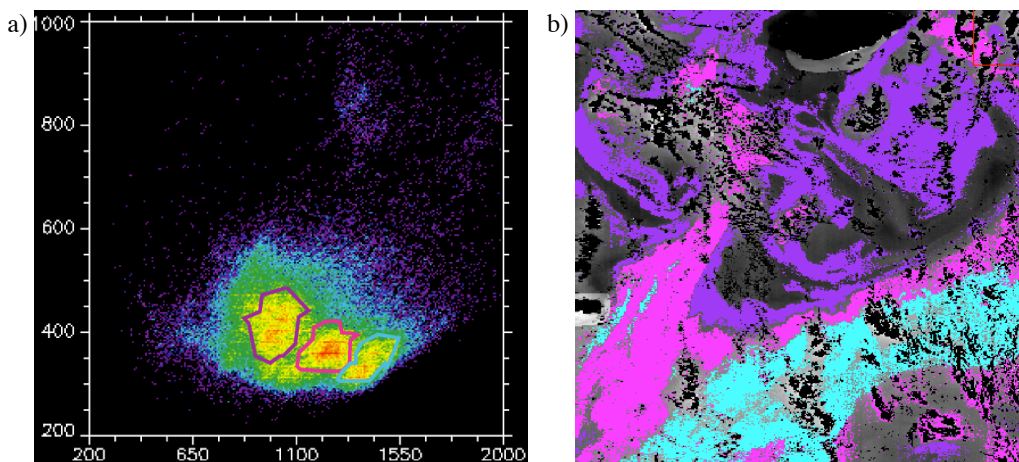


Fig. 5 a) Scatterplot of reflectances in channel 2 (x-axis) and 5 (y-axis), with 3 groupings highlighted in purple, pink and light blue. b) Location of these clusters within the zoom region south of Socotra.

The three highlighted groups in feature space map on to coherent regions in geographical space (Fig. 5b). Implementation of the ISODATA unsupervised classification on data from the whole MERIS pass picks out similar groupings and features within this region (Fig. 6a). Mean spectra for the five clusters present in this region all show a similar gradual decrease in reflectance from 500 to 650 nm. At shorter wavelengths these clusters are easily separated, with the region highlighted in yellow having significant reflectance below 500 nm, with this being less for the dark blue region, then green, purple and finally red. This is consistent with the idea of absorption of more blue light by increasing chlorophyll concentrations. The values at band 8 (681 nm) are in roughly the reverse order, as would be expected for an increased contribution via fluorescence. A plot of the MERIS algal_1 product (Fig. 6b) shows some similarity in geographical structures, which is not surprising given that it is based on the blue-green part of the spectrum used to calculate chlorophyll concentrations in Case I waters. [4]. However, there are also clearly differences, which augur that the classification routine is responding to more than just chlorophyll concentration.

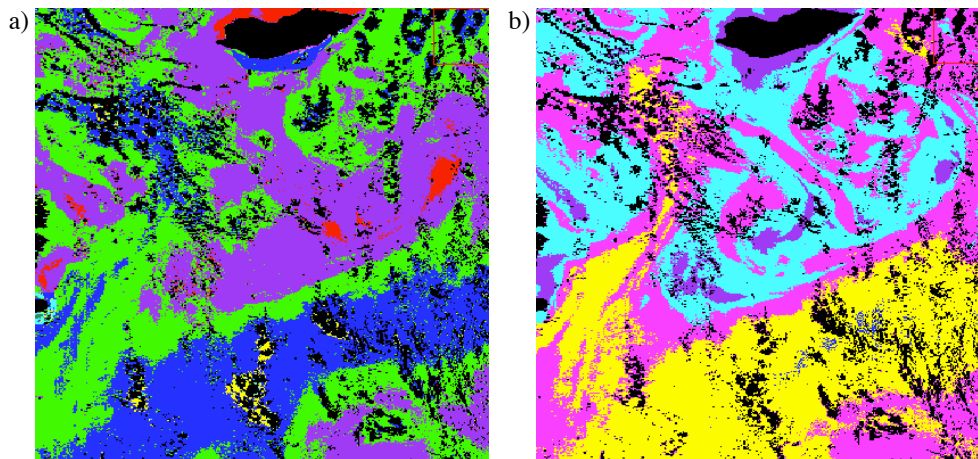


Fig. 6. a) Geographical pattern of clusters identified by unsupervised classification of channels 2,3,4 & 5. Zoom of region south of Socotra. b) Algal_1 product (chlorophyll estimate for Case I waters), with broad gradation from purple (high values) to light blue, pink, yellow, and then grey (low values).

4. CLASSIFICATION USING ALL VISIBLE BANDS

Having demonstrated that the ISODATA routine usually selects features in the limited band case that were similar to the subjective views of the operator, we then applied the technique to all nine visible bands to ascertain the key groupings present in a multi-dimensional space not easily studied by separate scatter plots. The classifier identified ten groupings with 4 of them covering 92% of observations, with great similarity in the geographical pattern (Fig. 7a) to the results of the blue-green classifier. (Note, the region highlighted in dark green in Fig. 7a is akin to the purple patch in Fig. 6a.)

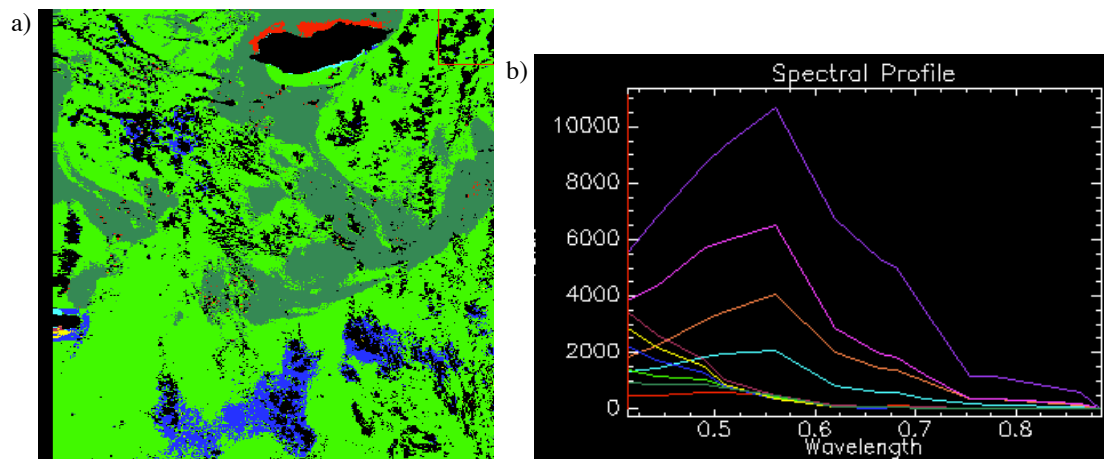


Fig. 7. a) Geographical display of clusters derived using all visible channels, for region south of Socotra. b) Mean spectra of the 10 clusters.

Consideration of the associated spectra (Fig. 4b) shows 6 of the classes to have similar spectral tails beyond $0.5\mu\text{m}$, with differences only in their absorption of blue. These spectra are consistent with waters containing chlorophyll. The remaining 4 classes, encapsulating 0.75% of the data, have mean spectra typical of sediment-laden waters.

5. SUMMARY AND DISCUSSION

Scatterplots of the reflectance data from individual channels have shown that the data do not simply occupy an amorphous continuum in reflectance space, but that they have structure, with a number of clusters being easily identified. Unsupervised classification is able to highlight the structures in an objective manner. Often such clusters will map to coherent geographical features; ancillary data, such as locations of coastlines and river plumes, may enable us to infer the presence of suspended sediment. Red-end classification appears to distinguish between case I and case II waters, although we have no *in situ* measurements to confirm this.

Classification using the red end of the spectrum also highlights various processing problems. Individual pixels of a different classification to their neighbours, are typical of unflagged cloud. Long linear demarcations between classes coinciding with the boundaries between the 5 MERIS cameras show that the smile correction has left some residual striping. Using our standard set of processing options (50 iterations to produce between 5 and 10 classes) we noted that the red-end classification gave only 2 classes in the open ocean (97.4% of data), reserving the remaining 8 for different sediment loading in the coastal waters. Application to the blue-green end of the spectrum produced 5 classes encompassing 89.6% of the data (principally open ocean waters containing phytoplankton), with a further two having spectra consistent with chlorophyll absorption. Unsupervised classification using all the visible channels produced similar results to blue-green classification, except only 4 clusters were used to explain 93.7% of the data, and an extra cluster was used to discriminate different coastal waters. The precise values determined do vary with the particular pass and extent of data used, but the above proportions are indicative of how the classification algorithm responds to different parts of the spectrum.

The strength of unsupervised classification is that it permits a statistical interpretation of MERIS data that is not dependent upon preconceived expectations about the presence of different seawater constituents, such as chlorophyll, suspended sediment or yellow substance. The weakness is absence of clear linkage between statistical classes and optically active constituents. Thus there is still a requirement for subjective interpretation.

This technique has also been used successfully for waters surrounding the UK, with similar conclusions as to sensitivity to clouds, sediment and chlorophyll [5]. However, as each classification is based upon its own data, the frequencies of occurrence of classes are not readily comparable between different regions or for different times, making such a technique difficult to use for the purpose of detecting changes. A possible way forward is to use a few MERIS passes as a training set to develop a supervised classification that can be applied to all subsequent datasets.

6. ACKNOWLEDGMENTS

The MERIS data were requested from ESA via a Category-1 proposal in support of the MadEx cruise, and were provided via RSDAS. Help from Dave Poulter and Tristan Quaife is also gratefully acknowledged.

7. REFERENCES

1. International Ocean-Colour Co-ordinating Group, Remote-sensing of ocean colour in coastal, and other optically-complex waters, IOCCG Report No. 3, IOCCG, Dartmouth, Canada, 2000.
2. Esaias, W., Iverson, R.L. and Turpie, K., Ocean Province classification using ocean colour data: observing biological signatures of variations in physical dynamics. *Global Change Biology*, Vol. 6, 39-55, 2000.
3. Hilger, K.B., Nielsen, A.A., Anderson, O.B., Knudsen, P. and Knudsen, T, Classification of multispectral and multitemporal ocean colour images. Available: http://manicoral.kms.dk/PK/havf2002_clas.pdf, 2002.
4. MERIS Algorithm Theoretical Basis Document, section 2.9 http://envisat.esa.int/instruments/meris/atbd_2_09.pdf
5. Joseph, M., Investigation of MERIS data for phytoplankton discrimination, MSc dissertation [available from University College London], 2005.