Remote Sensing of Environment 115 (2011) 2578–2593

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean

Elena Torrecilla [a], Dariusz Stramski [b], Rick A. Reynolds [b], Eduardo Millán-Núñez [b,c], Jaume Piera [a,*]

[a] Unitat de Tecnologia Marina, Centre Mediterrani d'Investigacions Marines i Ambientals (CSIC), Pg. Marítim Barceloneta 37, Barcelona 08003, Spain
[b] Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA
[c] CICESE, Ecología Marina, Carretera Ensenada-Tijuana No. 3918 Fraccionamiento Zona Playitas, Ensenada, B. C., Código Postal 22860, Mexico

ARTICLE INFO

Article history:
Received 5 October 2010
Received in revised form 24 March 2011
Accepted 23 May 2011
Available online 1 July 2011

Keywords:
Ocean optics
Ocean reflectance
Absorption coefficient
Phytoplankton pigments
Cluster analysis

ABSTRACT

Optical measurements including remote sensing provide a potential tool for the identification of dominant phytoplankton groups and for monitoring spatial and temporal changes in biodiversity in the upper ocean. We examine the application of an unsupervised hierarchical cluster analysis to phytoplankton pigment data and spectra of the absorption coefficient and remote-sensing reflectance with the aim of discriminating different phytoplankton assemblages in open ocean environments under non-bloom conditions. This technique is applied to an optical and phytoplankton pigment data set collected at several stations within the eastern Atlantic Ocean, where the surface total chlorophyll-*a* concentration (TChl*a*) ranged from 0.11 to 0.62 mg m^{-3}. Stations were selected on the basis of significant differences in the ratios of the two most dominant accessory pigments relative to TChl*a*, as derived from High Performance Liquid Chromatography (HPLC) analysis. The performance of cluster analysis applied to absorption and remote-sensing spectra is evaluated by comparisons with the cluster partitioning of the corresponding HPLC pigment data, in which the pigment-based clusters serve as a reference for identifying different phytoplankton assemblages. Two indices, cophenetic and Rand, are utilized in these comparisons to quantify the degree of similarity between pigment-based and optical-based clusters. The use of spectral derivative analysis for the optical data was also evaluated, and sensitivity tests were conducted to determine the influence of parameters used in these calculations (spectral range, smoothing filter size, and band separation). The results of our analyses indicate that the second derivative calculated from hyperspectral (1 nm resolution) data of the phytoplankton absorption coefficient, $a_{ph}(\lambda)$, and remote-sensing reflectance, $R_{rs}(\lambda)$, provide better discrimination of phytoplankton pigment assemblages than traditional multispectral band-ratios or ordinary (non-differentiated) hyperspectral data of absorption and remote-sensing reflectance. The most useful spectral region for this discrimination extends generally from wavelengths of about 425–435 nm to wavelengths within the 495–540 nm range, although in the case of phytoplankton absorption data a broader spectral region can also provide satisfactory results.

1. Introduction

In situ and remotely-sensed optical observations of ocean waters provide information regarding the concentrations of optically significant constituents in seawater, and offer the ability to observe important biological and biogeochemical variables (e.g., Chang et al., 2006). Numerous studies over the past three decades have focused on the development of bio-optical algorithms linking measurable optical properties to the primary pigment in phytoplankton, chlorophyll-*a*, a proxy for the phytoplankton biomass (e.g., Bricaud et al., 1998; Morel, 1988; O'Reilly et al., 2000; Reynolds et al., 2001). In recent years, efforts to expand the use of optical measurements for estimating other biogeochemically important ocean variables and phenomena have increased considerably. For example, optical measurements including satellite remote sensing have been used to detect harmful algal blooms (Craig et al., 2006; Cullen et al., 1997; Stumpf et al., 2003), surface concentrations of particulate inorganic and organic carbon (Balch et al., 2005; Stramski et al., 2008), particle size distribution (Kostadinov et al., 2009), phytoplankton community composition and size structure (Aiken et al., 2007; Alvain et al., 2005; Ciotti & Bricaud, 2006; Nair et al., 2008; Uitz et al., 2006), and phytoplankton class-specific primary production (Uitz et al., 2010).

Recent advances in measuring ocean optical properties and light fields within and leaving the ocean have included a progressive shift from using multispectral to high spectral resolution (hyperspectral) acquisition systems (Chang et al., 2004). New technologies and the miniaturization of electro-optical components have permitted the development of accurate, low-cost, and energy-efficient hyperspectral

* Corresponding author. Unitat de Tecnologia Marina (CSIC), Spain.
E-mail address: jpiera@utm.csic.es (J. Piera).

0034-4257/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2011.05.014

sensors suitable for deployments from a variety of platforms such as in-water vertical profiling systems, moorings, drifters, autonomous vehicles, air-borne and space-borne platforms (Dickey et al., 2006; Perry & Rudnick, 2003). The capability to obtain measurements at hundreds of narrow and closely spaced wavelength bands from the ultraviolet to near-infrared, with a resolution better than 10 nm, has become one of the most powerful and fastest growing areas of technology in the field of ocean optics.

Hyperspectral optical data provide the opportunity for improvements in spectral shape analysis and subsequent extraction of environmental information compared with low spectral resolution optical data. Derivative spectroscopy is one powerful technique of spectral shape analysis which enhances subtle features in hyperspectral data, and has been used successfully to obtain information about optically significant water constituents. For example, Craig et al. (2006) assessed the feasibility of detection of a toxic bloom of the dinoflagellate *Karenia brevis* from the analysis of the fourth derivative of phytoplankton absorption spectra, estimated from in situ hyperspectral measurements of remote-sensing reflectance $R_{rs}(\lambda)$ (λ is light wavelength in vacuo). The advantages offered by hyperspectral measurements of $R_{rs}(\lambda)$ in combination with derivative spectroscopy for identifying algal blooms were also demonstrated by Lubac et al. (2008), who based their analysis on the position of the maxima and minima of the second derivative of the spectral $R_{rs}(\lambda)$. Louchard et al. (2002) assessed major sediment pigments of benthic substrates from derivative spectra of hyperspectral $R_{rs}(\lambda)$ measured in shallow marine environments. In general, the optical detection of specific algal blooms appears feasible because certain accessory pigments with specific absorption features are unique to individual phytoplankton taxa (e.g., Millie et al., 1995) and can be better differentiated in hyperspectral absorption data than in multispectral data with a limited number of wavelengths.

The advantages and increasing availability of high spectral resolution measurements suggest that the effectiveness of hyperspectral optical information for assessing phytoplankton diversity should be further explored. In particular, there is a need to test whether the hyperspectral approach, which has proven useful in inland and coastal waters (e.g., Hunter et al., 2008; Lee & Carder, 2004; Lubac et al., 2008), can be also effective for the identification of different phytoplankton assemblages at large spatial scales in open ocean waters. These tests are also especially important for the common situation in which various phytoplankton groups co-exist at significant concentrations, and no single species dominates the assemblage (i.e., a non-bloom condition).

In this study, we analyze phytoplankton pigment data in conjunction with optical data of absorption coefficients and remote-sensing reflectance, which were determined along a north-to-south transect in the eastern Atlantic Ocean. Our primary goal is to examine the feasibility of classifying different open ocean environments under non-bloom conditions in terms of phytoplankton pigment assemblages from analysis of hyperspectral absorption and remote-sensing reflectance measurements. In order to address this question, an unsupervised hierarchical cluster analysis is applied to the pigment data set obtained from High Performance Liquid Chromatography (HPLC) analysis of seawater samples and to the optical data sets including the spectra of absorption coefficients and remote-sensing reflectance and their second derivative spectra. For this analysis, the pigment data and the corresponding optical data were selected to represent distinct differences in major accessory pigments present in the samples. We view our analysis basically as a proof-of-concept study in which our approach is to use a relatively small but carefully selected set of data which exhibits significant contrasts in the composition of pigments, rather than to indiscriminately use large data sets. The pigment-based clusters provide a reference for partitioning the selected data set into distinct subsets, each characterized by different phytoplankton pigment composition. Two indices, cophenetic and Rand, are examined to quantify the degree of similarity between the pigment-based clusters and optical-based clusters, and are ultimately used to illustrate the effectiveness of optical classification. The degree of similarity between clusters was evaluated for calculations involving different spectral ranges of optical data. Because the quality of derivative analysis also depends on parameters involved in data processing and computations, especially smoothing filter size and derivative band separation (Lee & Carder, 2002; Tsai & Philpot, 1998; Vaiphasa, 2006), a sensitivity of cluster analysis to the choice of these parameters was performed.

2. Measurements and data analysis

The approach in this study consists of three main components: (i) collection of field data of phytoplankton pigments and ocean optical properties and selection of a subset of data characterized by distinct differences in major accessory pigments for the cluster analysis, (ii) radiative transfer modeling to compute hyperspectral remote sensing-reflectance, and (iii) cluster analysis of pigment and optical data. The methodology of each component is described below.

2.1. Field measurements

Measurements of phytoplankton pigment composition and seawater optical properties were obtained during the ANT-XXIII/1 expedition of the R/V Polarstern along a north-to-south transect in the eastern Atlantic Ocean during October and November, 2005 (Fig. 1). The investigated area spanned a wide range of different oceanic environments between the English Channel and the waters off the African coast of Namibia. Typically, one full station was conducted daily near local

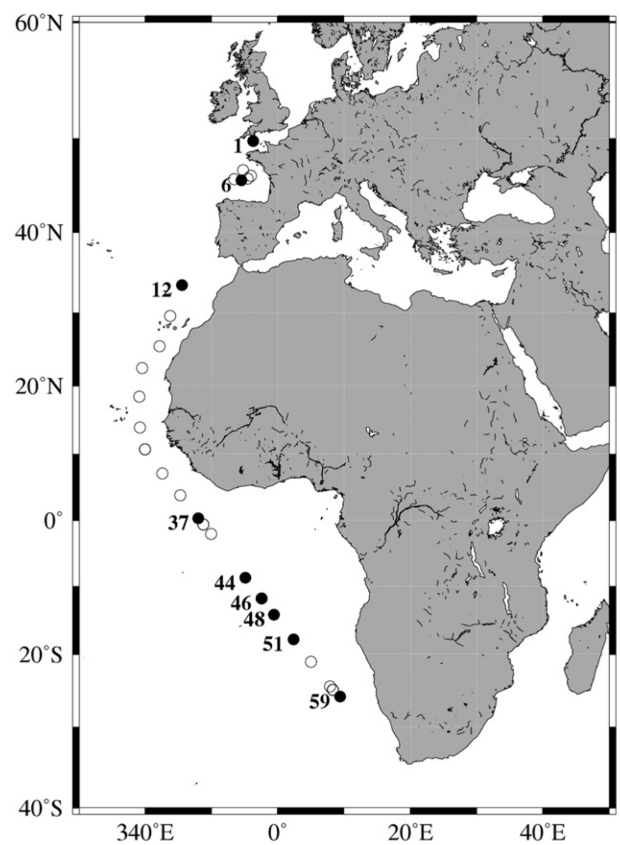

Fig. 1. Map depicting the location of full stations sampled along the north-to-south ANT-XXIII/1 cruise track in the eastern Atlantic during October and November, 2005. Each full station consisted of in situ optical measurements accompanied by discrete water sample analyses. Stations chosen for use in the cluster analysis are identified by filled circles and labeled with the station ID.