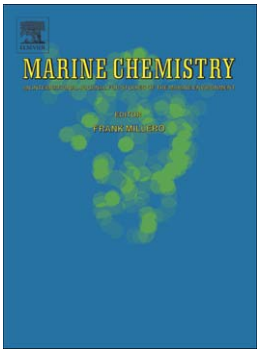

Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy

Kathleen R. Murphy, Colin A. Stedmon, T. David Waite, Gregory M. Ruiz

PII: S0304-4203(07)00235-6
DOI: doi: 10.1016/j.marchem.2007.10.003
Reference: MARCHE 2549

To appear in: *Marine Chemistry*

Received date: 21 February 2007
Revised date: 5 October 2007
Accepted date: 5 October 2007

Please cite this article as: Murphy, Kathleen R., Stedmon, Colin A., Waite, T. David, Ruiz, Gregory M., Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy, *Marine Chemistry* (2007), doi: 10.1016/j.marchem.2007.10.003

Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy

Kathleen R. Murphy[*1,3] , *Colin A. Stedmon*[2], *T. David Waite* [3], *Gregory M. Ruiz* [1],

[1]*Smithsonian Environmental Research Center, P.O. Box 28, 647 Contees Wharf Rd, Edgewater Maryland 21037, USA*
[2]*National Environmental Research Institute, Department of Marine Ecology, University of Aarhus, P.O. Box 358, Frederiksborgvej 399, DK-4000, Roskilde Denmark*
[3]*The University of New South Wales, School of Civil and Environmental Engineering, Sydney, New South Wales 2052, Australia*

Keywords: Dissolved organic matter; Tracer techniques; Fluorescence spectroscopy, Parallel Factor Analysis (PARAFAC).

* Corresponding author: Phone: +61-2-9385-5778, FAX: +61-2-9385-6139, Email: murphyka@si.edu

Abstract

The optical properties of chromophoric dissolved organic matter (CDOM) are frequently used as tracers of water masses in bays and estuaries but present unique challenges in the ocean due to the small quantities of organic matter present and the similarities between spectra from coastal and offshore environments. Samples collected on trans-oceanic cruises in the Pacific and Atlantic oceans were used to investigate the optical characteristics of dissolved organic matter in waters with limited freshwater influence (salinity > 30). Parallel Factor Analysis (PARAFAC) of fluorescence spectra revealed that coastal and oceanic dissolved organic matter (DOM) fluorescence could be separated into at least eight separate components: 4-5 humic-like and 3-5 protein-like signals. Two of the humic components were identified as representing terrestrial organic matter and their signals could be traced in the open ocean (Pacific and Atlantic) at levels of approximately 1.5% of riverine concentrations. An additional humic component, traditionally identified as the “marine” or “M” peak, was found to be both sourced from land and produced in the ocean. These results demonstrate that the supply, mixing and removal of terrestrial organic matter in oceanic waters can be observed with relatively simple measurement techniques, suggesting that fluorescence spectroscopy could play a useful role in future studies of the origin and fate of DOM in oceanic environments.

1. Introduction

The optical properties of chromophoric dissolved organic matter (CDOM) in natural waters have been studied for many decades by researchers of ocean color remote-sensing and aquatic optics (Kalle, 1949; Liu et al., 2006). As a major absorber of photochemically-active radiation in the oceans, CDOM has a profound effect on light penetration to the underwater environment, influencing biological productivity, the formation of chemical reactants and the speciation, transport and availability of trace elements. The ability to differentiate and quantify sources of CDOM in the oceans and the factors underlying its variability is hence fundamentally important to understanding biogeochemical cycles in the oceans.

The fluorescent constituents of dissolved organic matter are known to include humic substances and amino-acids, both free and bound in proteins (Mopper and Schultz, 1993; Coble, 1996). Although humic substances are among the most widely distributed organic materials on the planet, attempts to identify their chemical composition have had limited success due to their complex and heterogeneous nature, and even basic assumptions regarding their structure remain controversial (Sutton and Sposito, 2005). Since optical measurements are relatively simple to obtain, alterations in CDOM are sometimes used as a proxy for changes in composition of the wider dissolved organic matter (DOM) pool (McKnight et al. 2001; Zepp et al. 2004).

CDOM in natural waters has traditionally been quantified using absorption spectroscopy and, more recently, fluorescence spectroscopy, which has greater sensitivity (Duursma, 1974; Coble et al., 1990; Blough and Del Vecchio, 2002). Since the early 1990s, Excitation-Emission Matrix Spectroscopy (EEMS) has enabled fluorescence measurements to be performed across a range of excitation and emission wavelengths and presented as a 3-dimensional contouring fluorescence-intensity landscape (Coble et al., 1990; Coble, 1996). This presented a significant increase in

optical resolution over absorption spectra by enabling differentiation of substances that absorb at the same wavelength but emit in separate spectral regions.

Despite the advantages of this approach, a major drawback in earlier studies was that EEMs are data-intensive and difficult to parameterize. Hence, researchers have tended to focus upon a handful of visually distinguishable peaks which were characterized in terms of their positions, heights and ratios. Coble (1996) developed a nomenclature for CDOM fluorescence peaks which has gained widespread use, however, since the locations of individual peaks vary between environments, and may be obscured by overlapping spectra, a degree of subjectivity in data interpretation has been unavoidable. Recently, the characterization of CDOM fluorescence has advanced with the application of a multivariate tool called parallel factor analysis (PARAFAC) (Stedmon et al., 2003). PARAFAC enables decomposition of an EEM dataset into the least squares sum of several mathematically independent components, parameterized by concentrations (loadings) and excitation and emission spectra and corresponding, ideally, to a chemical analyte or group of strongly covarying analytes (Bro, 1997). The main advantage of PARAFAC over standard bilinear decomposition methods such as Principal Components Analysis (PCA) is simplicity and ease of interpretation; PARAFAC solutions are unique and have many fewer degrees of freedom, making them more readily interpretable. Since PARAFAC can isolate spectra from overlapping components, this may expose patterns within a dataset that are visually obscure and overlooked by traditional interpretive methods.

CDOM in coastal regions is often dominated by terrestrially-derived organic matter, enabling relationships between fluorescence, absorption and salinity to be used to trace coastal water mass mixing (Blough et al., 1993; Coble et al., 1998; Stedmon et al., 2003). In marine and coastal waters receiving little freshwater input, tracing the origin and fate of CDOM poses unique

challenges due to the relatively small quantities of organic material present and the strong similarity between spectra from coastal and offshore locations (Green and Blough, 1994; Coble, 1996; Blough and Del Vecchio, 2002). Few studies have focused explicitly upon discriminating CDOM sources in marine environments receiving little freshwater influence (c.f., Nelson et al., 1998; Stedmon and Markager, 2001). Recently, however, biosecurity concerns surrounding the illegal importation of ships' ballast waters have led to efforts to identify global tracers of seawater originating in high-salinity, oligotrophic ports (Murphy et al., 2004; Murphy et al., 2006). Quarantine rules by the US Coast Guard (2004) mandate that coastal ballast water be replaced with "oceanic" seawater (defined as originating from >200 nautical miles from shore where water depth exceeds 200 m) prior to discharge in US waters, to reduce the risk of harmful biological invasions (I.M.O., 2004; Carlton and Ruiz, 2005). Several other nations, including Australia, New Zealand and Canada, have similar requirements (Firestone and Corbett, 2005). While low salinity is diagnostic of coastal ballast water, a reliable method for differentiating between high-salinity coastal versus oceanic ballast water is also necessary to verify compliance on many vessels.

In this study, CDOM optical properties were measured in surface waters during transoceanic cruises through predominantly high salinity marine environments. We deduce herein changes in the dominant source (terrestrial vs. planktonic) of colored organic material, and the factors influencing its production and removal, while transitioning between coastal and open ocean sampling sites. We further identify the optical features of CDOM that are indicators of proximity to land. It is intended that this research will increase our ability to discriminate between CDOM derived from terrestrial and oceanic sources using routine optical measurements.

2. Methods

2.1. Sampling

CDOM samples were collected in summer/fall (June – October) in 2001-2004 during cruises on commercial ships operating between high salinity ports in the north Pacific and Atlantic oceans (cruises K, BN, and Fos, Figure 1). Seawater samples were collected via each ship's engine cooling systems, which circulate ambient seawater from a depth of approximately 4 -7 meters through steel pipes. The pipes were constantly flushed with seawater at flow rates of around 200 m^3 per hour, and were accessed as near as possible to the seawater intake (sea chest) and upstream of the engine machinery. Water samples from each ship's ballast tanks were also collected on these cruises and included in the PARAFAC models described below. In contrast to bilge-water, which contains significant quantities of run-off and oils from the ship, ballast water represents clean ambient seawater used only to aid stability and reduce stresses on the ship structure. It has been shown to be chemically similar to the water from which the ballast water was drawn (Murphy et al., submitted). Its inclusion in models increases resolution (through larger *N*) without introducing significant bias (Section 2.3). Methodologies and results specific to the ballast water samples are published elsewhere (Murphy et al. 2006); the discussion in this paper is thus limited to the seawater samples.

To sample, a short length of plastic tubing was attached to a sampling pipe at the sea chest and directed through an inline capsule filter (Osmonics Inc., Memtrex™ CMMP, 0.45 μm) to pre-baked amber glass collection bottles. Before sampling, the pipe and tubing were flushed with 4 L of water. More extensive flushing, although desirable, was usually prevented by the logistics of disposing of large quantities of waste water. We have used these methods extensively to collect seawater samples since 2001 (Murphy et al. 2004; 2006) and have found no evidence of

systematic CDOM contamination. CDOM samples were frozen on the vessel, shipped to the laboratory at the completion of the cruise and spectral analysis completed within 2 months. Salinity measurements were obtained using a YSI-85 conductivity probe (Hydrolab Incorp.), calibrated with distilled water and NIST-traceable seawater conductivity standard (YSI 3169, 50 000 μScm^{-1}).

Since sampling was performed while the ship was underway, GPS positions for each sample were calculated by interpolating waypoints recorded within 2-10 minutes of sampling, with a maximum expected error of approximately 3 nautical miles (nmi) at full ship speed. Coastal boundaries were obtained from the coastline extractor website of the National Oceanographic and Atmospheric Administration (NOAA) and represent mean location at mean tide (http://www.ngdc.noaa.gov/mgg/shorelines/shorelines.html). The minimum distance between these boundaries and each sampling position was determined in Matlab (Mathworks Inc., Version 6.5) using the seawater toolbox (CSIRO, Version 1.4). The nautical mile (1 nmi = 1852 m) was chosen as the reference unit for distance to facilitate the interpretation of these data in relation to territorial boundaries established under the 1982 United Nations Convention on the Law of the Sea. Ocean bathymetry was estimated from the depth (in meters) at the nearest position in the online database of Smith and Sandwell (1997).

2.2. Spectral measurements

Spectroscopic analysis of CDOM samples was performed at one of three regional laboratories: 1. the Marine Spectrochemistry Laboratory of the University of South Florida (USF) for samples from cruise Fos; 2. the Darling Marine Center at the University of Maine for samples from cruise K; and 3. the Department of Marine Ecology, National Environmental Research Institute

(NERI), Denmark, for samples from cruise BN. At USF, fluorescence was measured using a SPEX FluoroLog-2 and sample absorption (cruise Fos) using a 1-cm quartz cell and a Hitachi U-3300 absorption spectrophotometer. At the University of Maine, fluorescence was measured using a SPEX FluoroMax-2 and absorption using a Cary-50 UV-Vis spectrophotometer (Varian, Inc.) connected via a fiber optic dip probe coupled to a capillary waveguide (LWCC-2, World Precision Instruments), producing a path-length of 48.37 cm. At NERI, fluorescence was measured using a Varian Cary Eclipse spectrofluorometer and absorption in a 10-cm cuvette with a Shimadzu UV-2401PC UV-Vis spectrophotometer.

Excitation-emission matrices (EEMs) were generated for each sample over excitation wavelengths between 240-600 nm in 5-nm intervals and emission wavelengths between 300-600 nm in 2-nm intervals, with 5 nm bandwidths on excitation and emission modes. EEMs were corrected for instrumental differences with user-generated excitation, emission and lamp intensity correction factors and blank subtraction and normalized to quinine sulfate equivalent units (QSE) using standard methods (Coble, 1996; Mobed et al., 1996). An intercalibration experiment was conducted during a two week period in October 2004. Using Quinine Sulfate Standard Reference Material (SRM 936) (50 ppb) and their standard EEMs analysis and correction protocols, each laboratory produced fully corrected EEMs at 20°C. When normalized to peak height, the emission scans at λ_{ex} = 350 nm deviated from the NBS relative technical emission spectrum for quinine sulfate dihydrate (Velapoldi and Mielenz, 1980) by less than ± 10 % across the range of wavelengths unaffected by Raman scatter (Murphy et al., 2006). Measurements of replicate environmental samples during the same intercalibration experiment (Boehme & Stedmon, unpublished data) also showed congruent measurements from the three labs at fixed wavelength pairs.

Differences between the three instruments that are known to have affected spectral measurements are as follows. The Cary Eclipse was operated at lower sensitivity than the other instruments, greatly decreasing analysis time but resulting in noticeably noisy data. This would not affect PARAFAC modeling which is insensitive to random scatter (Bro 1997) but would decrease the accuracy of point measurements or their ratios (e.g.. the fluorescence index, Section 3.2). EEMs from the single-monochrometer FluoroMax-2 were affected by significant secondary Raman and Rayleigh scatter at emission wavelengths > 518 nm, necessitating separate modeling of these data (Section 2.3). The Cary Eclipse and the FluoroLog-2 use double monochrometers on the emission side which minimize secondary scatter.

Absorption measurements were corrected against Milli-Q water blanks and the slope of the absorption curve, *S*, calculated by non-linear regression (Stedmon et al., 2000) over the 300 - 650 nm wavelength range, according to the equation:

$$a_l = a_{l_0} e^{S(l_0 - l)} + K$$

where a_l is the absorption coefficient at wavelength l, a_{lo} is the absorption coefficient at a reference wavelength l_o, *S* is the slope coefficient (spectral slope) and *K* is a background parameter accounting for baseline shifts or attenuation due to factors other than CDOM. In this study l and l_o were chosen to be 375 and 400 nm, respectively. Samples that were fitted poorly by the regression equation or for which spectral slopes were unrealistically high (S >> 0.03 nm^{-1} (Blough and Del Vecchio, 2002; Siegel et al., 2002)) were excluded from subsequent analyses.

In concentrated samples, reabsorption of emitted radiation can affect the linearity of the fluorescence response (inner filter effect), potentially causing the underestimation of CDOM concentrations (Green and Blough, 1994). Since marine environments typically have low organic matter concentrations, we did not dilute our samples. Absorbance measurements confirmed all

samples were optically dilute (n = 176) with the exception of a handful of coastal samples on cruise Fos (3 samples near Norfolk) and cruise K (12 samples from San Francisco Bay, Seattle and coastal California). Fluorescence may be underestimated for these samples. Otherwise, absorbance at 300 nm was below 0.02 (cruise BN, 10 cm cell), below 0.002 (cruise Fos, 1 cm cell) and below 0.096 (cruise K, 48.37 cm cell).

2.3. PARAFAC modeling

Fluorescence EEMs were modeled using parallel factor analysis (PARAFAC) which uses an alternating least squares algorithm to minimize the sum of squared residuals across the dataset and, in so doing, enables estimation of the underlying structure of the EEMs (Harshman and Lundy, 1994; Bro, 1997). The data signal is decomposed into a set of trilinear terms and a residual array:

$$x_{ijk} = \sum_{f=1}^{F} a_{if}\, b_{jf}\, c_{kf} + e_{ijk} \qquad i = 1, \ldots, I \; ; j = 1, \ldots, J \; ; k = 1, \ldots, K$$

where x_{ijk} is the intensity of the i^{th} sample at the j^{th} variable (emission mode) and at the k^{th} variable (excitation mode) in the F-component model; a_{if} is directly proportional to the concentration of the f^{th} analyte at emission wavelength j, b_{jf} is a scaled estimate of the emission spectrum of the f^{th} analyte, c_{kf} is linearly proportional to the specific absorption coefficient at excitation wavelength k and e_{ijk} is the residual noise, representing the variability not accounted for by the model.

PARAFAC has previously been shown to be capable of recovering accurate spectra and concentrations of known fluorescent materials in mixtures including uncalibrated spectral interferents (Bro, 1997; Murphy et al., 2006). In general, however, components identified by PARAFAC in complex mixtures like seawater probably represent combined spectra of fluorophore groups that are chemically similar or chemically dependent (Del Vecchio and

Blough, 2004) as well as strongly covarying. Also, because the PARAFAC algorithm solves loadings for each component simultaneously (unlike, for example, PCA), successful modeling is sensitive to specifying the correct number of components. While diagnostic tools are available to assess PARAFAC models, these can give ambiguous or conflicting results (Smilde et al., 2004); consequently, it is not always easy to ensure that the correct number of components is selected in the absence of knowledge regarding the chemistry of a mixture. In this study, numerous tools were used to refine PARAFAC models and ensure the selection of the appropriate number of components, including jack-knifing (Riu and Bro, 2003), split-half analysis (Bro, 1997) and examination of residual error plots. Additional diagnostics included with the NWAY- (Andersson and Bro, 2002) and PLS- Matlab toolboxes, particularly core-consistency (Bro and Kiers, 2003) and influence plots were also examined. Ultimately the congruence of components identified in completely independent studies provides the most convincing evidence that PARAFAC models can capture underlying chemical phenomena.

We were able to draw upon an existing PARAFAC model that was developed from a dataset that included all samples collected during this study except the Kauai (cruise K) dataset (Murphy et al., 2006), along with ballast water samples from these and additional cruises (N = 687). This existing model is referred to hereafter as the BWE7 model. Signatures of the nine independent fluorescence components comprising the BWE7 model are presented in Figure 2. The Kauai dataset was not included in the BWE7 model since, as described below, EEMs from this cruise were severely affected by second order scatter that disrupted the original modeling efforts. To allow the Kauai EEMs to be included in this study, a separate PARAFAC model was developed for these EEMs using a restricted wavelength range.

The Kauai dataset consisted of EEMs from 130 samples, including 63 ballast water samples, all of which were used to develop the Kauai PARAFAC model. Modeling was performed on excitation wavelengths ranging from 260-450 nm at 5 nm intervals and from 302-518 nm at 4 nm intervals. Excitation wavelengths below 260 nm were deleted due to high residuals; low excitation wavelengths are strongly influenced by the condition of the xenon lamp and the physical environment, resulting in low signal to noise ratios (Smilde et al., 2004). Data from emission wavelengths > 518 nm were removed since they were dominated by secondary Rayleigh and Raman scatter as well as secondary protein fluorescence peaks propagated from intense fluorescence near $\lambda_{ex}/\lambda_{em}$ = 275/300 nm. Even after removing these wavelengths, secondary Rayleigh and Raman scatter still dominated a region of the EEM within the modeled range (Figure 3). Scatter bands are non-trilinear in structure and must be minimized or removed prior to PARAFAC modeling (Bro, 1997). We chose to smooth the affected data (Zepp et al., 2004), although valid alternatives would be to insert missing numbers or down-weight their importance in the model (Andersen and Bro, 2003).

An eight-component PARAFAC model of the Kauai dataset (n = 130) was generated using the PLS toolbox (Eigenvector, Inc.) for Matlab (Mathworks, Inc.) with non-negativity constraints applied on all modes (Andersen and Bro, 2003). To reduce the dominance of high intensity signals from coastal water the EEMs were first scaled to unit variance within the sample mode. The model explained > 99.6 % of the variation within the dataset, with variation explained by each component decreasing sequentially from P1-P8. Relatively high residual errors remained at low excitation wavelengths (< 300 nm) and in a narrow band of emission wavelengths (315-330 nm) and core consistencies were low. While these characteristics do not necessarily invalidate a model (Smilde et al., 2004), they can indicate that it is not well specified. To independently

confirm the robustness of the Kauai model, three additional models were therefore developed using subsets of the data consisting of (a) only seawater samples collected more than 100 miles from any coast, (b) all seawater samples regardless of distance from shore; and (c) only ballast water samples. All three models agreed closely, indicating that the Kauai model was both robust and representative of samples from disparate environments.

3. Results

3.1. Agreement between PARAFAC models

As in previous PARAFAC models of CDOM fluorescence (Table 1), the spectra of Figures 2-4 suggest the presence of humic and fulvic acids (components C2, C3, C8, C9, P1, P3 and P8), free and protein-bound amino acids (C1, C6, C7, P2, P5, P6 and P7) and an unknown material (C4) that was previously tentatively identified as a polycyclic aromatic hydrocarbon (PAH) contaminant (Murphy et al., 2006). Seven of eight components from the Kauai PARAFAC model (P1-3 and P5-8) (Figure 4) corresponded directly with five of nine components from the BWE7 model (components C1, C2, C3, C6, C8) (Figure 2), either singly or else when combined with the spectra from a second component.

Corresponding spectra from the two models are compared visually in Figure 5 and quantitatively in Supplementary Information S1. Quantitative similarity was indicated by (a) values of Tucker's congruence component exceeding 0.92 (Tucker, 1951; Lorenzo-Seva and Berge, 2006) and by (b) linear correlations between the scores of corresponding components in the two models. High similarity according to both criteria was considered justification for treating the models as interchangeable with respect to five components (C1-3, C6 and C8). This allowed us to redefine much of the fluorescence in the Kauai dataset to fluorescence by components C1-3, C6 and C8, and so include the Kauai dataset in comparisons throughout the

remainder of this paper. Note that component C5 in the BWE7 model was due to rhodamine dye present in some ballast water samples included in the model (Murphy et al., 2006); since the dye does not occur in the ocean and was not added to ballast tanks during cruise K, it is absent from the Kauai model.

3.2. CDOM source discrimination

Fluorescence properties in seawater along the ocean tracks of cruises K, Fos and BN are depicted in Figure 6. Locations on the horizontal axis in this figure are described in Table 2. Despite salinities at most sampling locations falling within a small range (34 - 37), changes in mean CDOM concentrations and compositions were evident along each cruise track. Humic-like CDOM fluorescence generally followed a terrestrial-type distribution, with higher intensities in near-shore waters decreasing to baseline levels in the open ocean. This general distribution was shown by humic-like components C2, C3, C8 and C9. The distribution of unknown component C4 was similarly consistent with an anthropogenic source. Even on cruise Fos, between the Mediterranean Sea where salinities were very high (> 36) and fluorescence correspondingly low (< 1 ppb QSE), and the northeast Atlantic, small decreases in intensity were measured for C3 and C4. In samples collected from the tropical Atlantic on Cruise BN, significant increases were observed for components C2 and C9 that were not mirrored by the other humic-like components.

Protein-like components followed variable distributions. Fluorescence by the tyrosine-like component C1 was highest in Californian coastal waters on cruise K, in the Mediterranean and the NW Atlantic on cruise Fos, and in the English Channel and the Tropical Atlantic on cruise BN. Tryptophan-like component C6 followed a terrestrial – type distribution on cruises K and Fos, whereas on cruise BN, it mirrored the undulations by component C1. Another tryptophan-like component, C7, was not detected on cruise K and followed very different distributions on

the two Atlantic cruises. On cruise Fos, it followed an inverse-terrestrial distribution with minima in the Mediterranean and coastal waters and maxima in the mid-Atlantic. On cruise BN, the highest C7 fluorescence was in near-shore waters of the North Sea and English Channel, then decreased steadily along the cruise track to essentially undetectable levels in the mid-Atlantic and coastal Brazil.

The humic-like components C3 and C8, and the protein-like components C1 and C6 were linearly correlated on all cruises, but are presented as log-log plots in Figure 7 due to the measured output spanning over 3 orders of magnitude. In Figure 7A, the log(C8):log(C3) ratios (i.e. the slopes of regression lines) for the Atlantic, Mediterranean and the Pacific samples were not significantly different from each other ($F_{2,178} = 0.97$, $p = 0.38$) or from 1.0 (slope = 0.98 ± 0.04). This indicates that C8 and C3 concentrations were directly proportional, suggesting that concentrations were controlled by a conservative process, presumably mixing, possibly in concert with further processes (e.g. degradation) that acted upon both components at identical rates. The log(C3) intercept for the Pacific samples was significantly lower than for the Atlantic and Mediterranean samples ($F_{2,178} = 33.8$, $p < 0.0001$), indicating compositional differences in the original CDOM. Similarly, the log(C1):log(C6) ratios for samples from the Atlantic, Mediterranean and the Pacific (Figure 7B) were not significantly different from each other ($F_{2,177} = 0.47$, $p = 0.63$) or from 1.0 (slope = 0.96 ± 0.08), indicating that C6 and C1 concentrations were controlled by processes occurring at near-identical rates. The log(C6) intercept for the Atlantic and Mediterranean samples was higher than for the Pacific samples ($F_{2,179} = 23.0$, $p < 0.0001$) indicating different relative compositions of C1 and C6 in the two regions.

Fluorescence as a function of distance from land, averaged for each component across the same three cruises, reveals some overall trends in the composition and abundance of CDOM in

the oceans (Figure 8). To limit the influence of freshwater in this analysis, we restricted the dataset to samples with salinity exceeding 30; mean salinities for included samples are presented in Figure 9. Decreasing fluorescence with distance from land was most pronounced for the terrestrial and anthropogenic components C3, C4 and C8. A similar distribution was seen for components C2 and C9, although the relative decrease from coasts to the ocean was less (Figure 8A). Amino-acid fluorescence (components C1, C6 and C7) was relatively variable, following no single trend with increasing distance from shore. The average distribution of components in EEMs from varying distances from shore offer snapshots or fingerprints of the compositions of different water masses (Figure 8B). The overall impression obtained from this picture is the removal or conservative mixing of humic-like components, and increasing importance and variability of amino-acid like fluorescence, with increasing distance from land. The fluorescence results are thus consistent with absorption measurements (Figure 9) showing (a) decreasing absorption (a_{375}) and increasing spectral slopes; and (b) decreasing variance of a_{375} but generally increasing variance of S in offshore waters.

In Figures 10 - 12, the ratio of fluorescence by pairs of components is used to examine how fluorescence relationships vary between coastal and offshore environments. Changes in ratios reflect alterations in the chemical composition of CDOM, and may indicate transitions between waters dominated by allochthonous sources versus autochthonous production. The ratio of emission intensities (450 nm/500 nm) at 370 nm-excitation has been used to distinguish between different organic matter precursor materials in inland waters (McKnight et al. 2001). The utility of this index in identifying proximity to land in marine waters was examined for cruises Fos and K, but not Cruise BN due to low measurement sensitivity (Section 2.2). Overall, the mean value of the ratio was lower on cruise Fos (1.20 ± 0.13 QSE) than cruise K (1.29 ± 0.1 QSE). Although

the variability in ratios increased with distance from terrestrial sources (Figure 11), mean values were unaffected ($p > 0.85$).

In PARAFAC models, ratios between the intensities of different components may be more broadly meaningful than absolute intensities, since the latter could vary between models depending upon how many components a fluorescence signal is distributed across and their precise spectral loadings. Since component C3 on all voyages appeared to be an exclusively terrestrial signal, increases in the ratios of other components to C3 may highlight non-terrestrial sources of fluorescence. The average fluorescence of each component relative to C3 is depicted in Figure 11 as a function of distance from land with the analysis limited to samples for which salinities exceeded 30. Overall, protein-like components significantly increased relative to C3 at greater distances from shore. Humic components C8 and C9 were variable but generally increased relative to C3, and component C4 was highly variable, reflecting differences in its distribution among cruises. The ratio of C2/C3 increased steadily from a mean of ~1 near the coast to ~1.5 in the open ocean.

The steady increase in humic component C2 relative to humic component C3 with increasing distance from shore indicates differences in their production or removal rates. Overall, concentrations of C2 and C3 were strongly correlated (Figure 12A): among four broad sample sources (harbor, coast, shelf, ocean) the slope of linear regressions varied from 0.83 (harbor) to 1.0 (coastal) but were not statistically different ($F_3 = 2.25$, $p = 0.08$), with an overall slope of 0.92 (R^2=0.84). At C3 concentrations below 1 ppb, C2 fluorescence intensities generally lay above the conservative mixing curve (Figure 12A) indicating that, at low concentrations, the two components became decoupled. This is also depicted in Figure 12B, where variation in the ratio of the two components is largely driven by changes in C2

4. Discussion

In the few years since PARAFAC was first used to characterize CDOM fluorescence, similar characteristic spectra have emerged in models of geographically and ecologically disparate datasets. This provides compelling evidence that PARAFAC can reveal underlying chemical phenomena and illuminate differences between organic matter sources. Our analyses of CDOM in the surface Pacific and Atlantic oceans suggests that fluorescence in marine environments can be described by the distribution and behavior of a limited number of components common to coastal and offshore waters. We chose to present our data in terms of proximity to terrestrial sources (distance from shore) and focused upon regions where freshwater influences were small. In many other parts of the world, river plumes extend far offshore; consequently, discretion is needed in applying the relationships we observed to new locations.

Researchers of CDOM fluorescence have traditionally characterized CDOM sources using a handful of visually-distinguishable peaks in EEMs, for which a standard terminology is now widely adopted (Coble et al., 1993; Coble, 1996; Blough and Del Vecchio, 2002). Two typical peaks associated with terrestrial CDOM are named "A" ($\lambda_{ex}/\lambda_{em}$ ~ 260/380-420) and "C" ($\lambda_{ex}/\lambda_{em}$ ~ 350/420-480). A third humic-like peak, apparently marine in origin is named peak "M" ($\lambda_{ex}/\lambda_{em}$ ~ 312/380-420 nm). Additional peaks due to fluorescence of aromatic amino acids are named T (tyrosine-like: $\lambda_{ex}/\lambda_{em}$ ~ 220/303 nm) and B (tryptophan-like: $\lambda_{ex}/\lambda_{em}$ ~ 275/345 nm).

Coble (1996) proposed that the intensity, positions and ratios of fluorescence at peaks "A" and "C" may be used as tracers of terrestrial sources, noting that (1) the ratio of peak A:C fluorescence in seawater varied between coastal and oceanic samples, with higher A:C ratios in near-shore samples, and (2) the position of peak "C" moved towards shorter excitation and

emission wavelengths at increasing distances from freshwater sources. In this study, however, components C3 and C8 were responsible for nearly all of the fluorescence in both the "A" and "C" peak regions (Figure 2). As a result, the ratio of intensities of peaks A:C and the position of peak "C" varied little between coasts and the ocean, making A:C ratios redundant as a tracer of distance to shore. We believe that this contradiction arises due to the lack of uniquely "A" peak fluorophores in our study, which was limited to high salinity (>30) marine waters, compared to Coble's (1996) study, which also included samples from freshwater lakes and rivers. For example, Stedmon et al (2003) identified two fluorescent "A" components that dominated "C" components in samples from forest streams and wetlands but were less prevalent at higher salinities.

McKnight et al. (2001) showed that sources of organic matter precursors in lakes and rivers could be distinguished using the ratio of fluorescence emitted at two wavelength positions (the "fluorescence index"). Microbially-derived fulvic acids exhibited a greater ratio (~1.9) of emission intensity (450 nm/500 nm at 370 nm-excitation) than terrestrially-derived fulvic acids (~1.3). In our samples, the value of this index predominantly overlapped with ranges classified as terrestrially derived end-members (1.3-1.4) and large river samples (1.4-1.5) (McKnight et al., 2001), even where samples were obtained far off-shore. During measurements of open ocean samples, all spectrofluorometers were operated near their limits of detection and some of the observed variability in the index would have been related to noise. In a variety of PARAFAC models (Table 1), however, the secondary peaks of fluorescent components C8 ($\lambda_{ex}/\lambda_{em}$ = 355-380/415-440 nm) and C3 ($\lambda_{ex}/\lambda_{em}$ = 360-388/478-504 nm) each contribute to emission at 450 and 500 nm when excited by light at 370 nm (see overlapping spectra in Figure 5), resulting in the 'buffering' of C8/C3 ratios that is consistent with the fluorescence index results (Figure 11).

These factors suggest that this index will be less sensitive for identifying precursor material in marine environments than in inland lakes and streams.

4.1. PARAFAC components as tracers in the oceans

Murphy et al. (2006) attributed component C4 to PAH sources; several PAHs (e.g. pyrene, phenanthracene) have notable spectral overlap (Beltran et al., 1998). In coastal samples, C4 fluorescence was clearly dominated by sources external to the ship and its distribution was consistent with an anthropogenic source: i.e. high concentrations in harbors, the English Channel and on the Atlantic shelf approaching Delaware and Chesapeake Bay outflows. On cruise Fos, intensities increased steadily beginning ~200 nmi offshore in the direction of the Chesapeake Bay, which collects runoff from highly PAH-impacted urban catchments (Gustafson and Dickhut, 1997). Persistent organic pollutants, including PAHs, have been detected in surface waters of the Atlantic ocean (Schulz-Bull et al., 1998), however, comparative data collected with traditional clean-sampling methods are needed to confirm that component C4 was exclusively a tracer of terrestrial pollutants in this study.

Tryptophan and tyrosine-like fluorescence in this study was represented by three components (C1, C6, C7), with C1 and C6 further resolved in the Kauai model into P2 + P5 and P6 + P7, respectively. Components C1 and C6 were strongly linearly correlated, suggesting they have a common source and similar removal processes (Figure 7). Tryptophan and tyrosine-like fluorescence are frequently attributed to autochthonous production associated with the biological degradation of DOM (Mopper and Schultz, 1993; Coble et al., 1998; Stedmon and Markager, 2005b). Their distributions in this study at times correlated with terrestrial humic fluorescence and at others varied independently of it, suggesting they are produced from both terrestrial and autochthonous substrates. Tryptophan-like component C7 was not resolved in the Kauai dataset.

On cruise BN, C7 concentrations decreased monotonically with decreasing latitude and correspondingly increasing solar radiation. In a laboratory study of samples from the Baltic Sea (Stedmon et al. 2007), a component spectrally identical to C7 (Tucker congruence ~0.99 on excitation and emission) was found to be removed under exposure to UVA light; furthermore, in previous studies from the same region this component was absent from PARAFAC models (Stedmon et al., 2003; Stedmon and Markager, 2005a,b). Component C7 may thus represent fluorescence from a photo-sensitive CDOM constituent sourced from the European subcontinent.

Component C9 in this study increased in concentration relative to C3 in offshore waters (Figure 11). A component spectrally similar to C9 (Tucker congruence >0.92 on excitation and >0.96 on emission) has been identified in a recent laboratory study where it was produced following the UVA irradiation of Baltic seawater (Stedmon et al., 2007). The irradiation of refractory DOM can result in its direct photochemical mineralization as well as the production of relatively reactive (labile) constituents (Moran and Zepp, 1997; Kieber, 2000). In the current study, the distribution of C9 suggests it could be a long-lived product of the photodegradation of organic matter. The fact that component C9 was not resolved in the Kauai dataset and accounted for the least variation of all components in the BWE7 model suggests that it is present in surface seawater at low ambient concentrations.

Overall in this study, tryptophan- and tyrosine-like fluorescence followed a terrestrial type distribution overlaid with significant local production events, particularly in high-productivity regions of the northeast-Atlantic. Dissolved free amino acids are exuded directly from algae (Nagata, 2000), and tryptophan-like fluorescence has been observed to increase in association with algal blooms (Wolfbeis, 1985; Stedmon and Markager, 2005a).The elevated CDOM levels in samples from the eastern tropical Atlantic and coastal Brazil suggests that these signals had

planktonic sources that may include direct exudation by algae and production by bacteria utilizing a range of substrates, including terrestrial and marine organic matter associated with active or expired plankton blooms.

Humic-like fluorescence in this study was dominated by components C2 and C3, both of which decreased in magnitude at increasing distances from shore. This, coupled with a relative increase in protein-like components, was correlated to increasing spectral slopes in absorption curves. Averaged across three cruises and two oceans, S increased from ~ 0.018 nm^{-1} near the coast to ~0.023 nm^{-1} in the open ocean (Figure 9). A similar trend of increasing spectral slopes has been observed for offshore waters in the Mid-Atlantic Bight (Blough and Del Vecchio, 2002) and in North Sea/Skagerrak waters (Stedmon and Markager, 2001).

Ocean-color satellite observations indicate that significant regional variability exists in the global surface distribution of colored materials (predominantly CDOM), suggesting that the dynamics of CDOM in the ocean are controlled by the local balance of oceanographic, photochemical and biological forces (Nelson et al., 1998; Siegel et al., 2002). SeaWiFS satellite ocean imagery and pigment distribution studies show increased phytoplankton productivity in biogeographic provinces associated with the Canary Current, North Equatorial Current and Guinea Dome, west of North Africa (Gibb et al., 2000; Kitidis et al., 2006). Global trends observed via satellite (Siegel et al., 2002) are also evident in this study (Figure 6), including (i) significantly higher baseline humic fluorescence (components C2, C3 and C8) in the north Atlantic vs Pacific open oceans, possibly due to the ~3x greater annual river discharge to the Atlantic; (ii) strong latitudinal variation in humic concentrations in the Atlantic ocean (compare the east-west transect of the Atlantic at ~38 °N on cruise Fos with the north-south transect on cruise BN); (iii) measurable effects of river discharge, particularly near outflow from the

Chesapeake Bay on cruise Fos and the Amazon River on cruise BN; and (iv) locally elevated concentrations of oceanic CDOM, e.g. in productive regions of the eastern tropical Atlantic.

Studies of the humic material in marine surface waters suggest that only a small fraction is of terrestrial origin, indicating that sources of CDOM must exist in the ocean (Harvey et al., 1983; Hedges et al., 1992). The commonly termed "marine" humic peak "M" is believed to originate from autochthonous sources and is usually taken to represent the marine humic end-member, despite usually strong positive correlations between peak "M" intensity and terrestrial humic fluorescence (Coble, 1996; Nieto-Cid et al., 2005). Coble (1996) recorded a mean fluorescence maximum for marine shallow transitional waters of $\lambda_{ex}/\lambda_{em}$ = 310/423 nm. In this study, component C2 ($\lambda_{ex2^{o}}/\lambda_{em}$ ~ 310/422) shared spectral characteristics both with peak "M" and a component ($\lambda_{ex2^{o}}/\lambda_{em}$ ~ 325/416, "component 4") identified by Stedmon et al. (2003) from a Danish catchment, where it was attributed to terrestrial DOM. No other candidates for a humic 'marine' component were identified despite the predominance in this study of fully marine sampling sites.

Interestingly, in this and the Danish study, peak "M" followed a predominantly "terrestrial-type" distribution; i.e. was most abundant in coastal waters near the Americas and Europe, least abundant in the open ocean and displayed intermediate concentrations in transitional waters (Figure 8). Coble (1996) previously questioned the status of peak-M as a true marine end-member, citing conflicting trends in its distribution between low- and high-salinity coastal environments. In our dataset the ratio of components C2/C3 increased from ~1.0 in near-shore samples to around 1.5 in open ocean samples (Figure 11). Possible explanations for this trend include (a) C2 and C3 are both terrestrially derived but C3 is removed more rapidly than C2 in the oceans, or (b) C2 is produced autochthonously from terrestrial and marine organic substrates,

implying that C2 is new organic material regardless of origin. Either explanation could account for the strongly correlated C2 and C3 concentrations in near-shore waters, with increasing relative concentration of C2 at increasing distance from terrestrial sources. An increasing dominance of component C2 over C3 would account for the apparent shift at higher salinities in the position of the emission maximum peak "C" toward shorter wavelengths, as has been described in previous studies (Coble, 1996; Del Castillo et al., 1999; Del Castillo et al., 2000).

4.2. Distinguishing between production and degradation processes

To investigate possible factors driving the apparent decoupling between C2 and C3, simple models were developed (Table 1), reflecting four possible scenarios: (i) conservative mixing only; (ii) mixing + constant differential removal; (iii) mixing + constant background production of C2; (iv) oceanic production of C2 against background C3. In scenarios (i)-(iii), C2 and C3 were assumed to behave conservatively subject to conservative mixing at rate *D*. In scenario (ii), we wished to investigate the effect of differing susceptibilities to photochemical and/or biological degradation as has been observed in experimental studies (Stedmon and Markager, 2005b), with the components additionally subject to removal (*r*) at different rates ($r_3 > r_2$). In scenario (iii), the concentration of terrestrial C2 was augmented by autochthonously-produced C2 at a low constant 'background' rate, p_2. In scenario (iv), C3 was set at a realistic background oceanic level, *m*, with autochthonous production of C2 allowed at each model iteration. The equations governing the four model scenarios are provided in Table 3, with initial concentrations of C2 and C3 set to 14 QSE to emulate the data in this study.

If conservative mixing were the only important process operating on C2 and C3 concentrations then log(C2) and log(C3) will be perfectly correlated with a slope of 1 and intercept equal to C2/C3. Conversely, if component C3 were always removed at a faster rate than component C2,

due, for example, to greater susceptibility to photochemical degradation, then a linear log(C2)/log(C3) relationship is predicted with slope < 1 and varying according to the relative rates of r_2, r_3 and D. Conversely, if it is assumed that the amount of autochthonously-produced C2 is independent of terrestrial C2 and C3 loading, then a departure from log(C2)/log(C3) linearity is predicted at low concentrations of C2 (Figure 12B, dashed line). If C3 in the ocean is set to a background level m whereas C2 is allowed to vary freely, then the ratio of C2/C3 plotted against C2 will lie upon a straight line with slope 1/m (scenario iv, Figure 12D). Sections S2 and S3 of the Supplementary Information illustrate the effect of varying removal and production rates upon the modeled data. It should be noted that this technique is useful for illuminating only *differential* rates of production or removal, and can not be used to estimate absolute rates. If in addition to conservative mixing the two components were removed or produced at identical rates, the resulting correlative relationship would be indistinguishable from conservative mixing alone.

The similarities and differences between the modeled and measured data in Figure 12 offer some insights into processes that may be driving the observed behavior. Conservative mixing (scenario (i)) appears to be the dominant force driving the concentrations of both C2 and C3, as evidenced by a large proportion of samples in both coastal and oceanic environments lying on a straight line with slope =1 (Figure 12A). In scenario (ii), it is assumed that C2 and C3 were removed at 1% and 1.5% of the mixing rate respectively, however, these and other relative rates that can approximate the observed data can not reproduce the steep inflection at low C2 concentrations characteristic of the curve in Figure 12B (see also the Supplementary Information). In the constant C2 production scenario (iii), production is assumed to occur at 2% of the mixing rate. The steeper inflection in this production model, which more closely emulates

the data than the differential degradation model, is due to the assumption that degradation must act upon existing terrestrial material but production is independent of it. In the oceanic C2 production scenario (iv), C3 is set to the mean concentration that we measured in fully oceanic samples (0.28 ppb), whereas C2 is allowed to vary randomly between 0 and 1.8 ppb, encompassing the range of C2 concentrations we measured in the open ocean. This model, which most closely reproduces the oceanic portion of the dataset in Figure 12B, provides support for the interpretation that the only sources of C3 in our surface samples were terrestrial, whereas C2 was sourced from terrestrial and oceanic environments.

Simple models such as this have obvious limitations; chiefly that it is not possible to conclusively distinguish between differential production and removal of components on the basis of ratios alone. Furthermore, production and removal processes can act simultaneously and may be spatially and temporally variable. The cruises of this study took place in the warmer months between June and late September, when higher solar irradiance would have enhanced both biological activity and the photochemical destruction of fluorophores. Also during summer in the mid-latitudes, stratification inhibits the exchange of surface DOM with deeper pools, trapping terrestrial material in the shallow surface layer and reducing the opportunity for subsurface renewal.

Our data suggest that low-level marine production of component C2 is driving the observed decoupling of C2 and C3, although there may also be small differences in their susceptibilities to degradation. Independent lines of evidence support an autochthonous source for C2. The stable isotope composition of oceanic humic substances indicates a predominantly marine origin (Hedges et al., 1992) and while this does not implicate any single component, C2 explained more variation in our PARAFAC model than any other humic-like component. More direct evidence

comes from mesocosm experiments designed to investigate the production and degradation of autochthonous organic matter fractions (Stedmon and Markager, 2005b), in which a component largely overlapping with C2 was produced by bacteria at rates that were enhanced by the UV exposure of precursor material. In this study, several samples collected from high productivity regions of the tropical North Atlantic on Cruise BN featured relatively high concentrations of components C9 and C2, but not the terrestrial component C3 (Figure 6). Sources of autochthonous DOM in highly productive regions could include algae and macrophytes (Nagata, 2000); alternatively, upwelling regions provide opportunity for "deep" CDOM to be transported vertically to the surface. These results support the interpretation that C2 is new organic material produced autochthonously in inland and marine aquatic environments; possibly as a degradation product of non-colored DOM. However, given the evidence that terrestrial and marine DOM are chemically distinct (Williams and Druffel, 1987; Hedges et al., 1992), it is as yet unclear how a spectrally identical component would be produced via the two pathways.

4.3. Remnant terrestrial CDOM in the ocean

Studies of CDOM absorption in the open ocean have reported a nearly constant background signal, interpreted as a possible remnant of terrestrial organic materials (Blough and Del Vecchio, 2002). Our data confirm that terrestrial signals are detectable far from known coastal influences. For example, we were able to resolve the characteristic spectra of humic-like components C2 and C3 in both the Pacific and Atlantic oceans via PARAFAC even when restricting datasets to full-strength salinity samples obtained at least 100 nautical miles from any coast or island. In such models, the spectra of component C3 appeared ragged. While possibly influenced by low signal to noise ratios, this may reflect its status as a photo-degraded refractory remnant of the original terrestrial signal.

Coble (1996) reported mean concentrations and wavelengths of humic peak "C" in rivers ($\lambda_{340}/\lambda_{448}$, F_{max} = 18.5 ppb QSE, N = 21) and marine shallow oligotrophic waters ($\lambda_{347}/\lambda_{451}$, F_{max} = 0.32 ppb QSE, N = 22). Using the mean fluorescence intensities she reported in Table 4, and assuming that this fluorescence was due to components C2, C3 and C8 in the same average proportions that we observed in coastal and fully marine sampling sites (1.1:1:0.7 and 1.65:1:0.8, respectively), we estimate that the average quantity of C3 present in these samples was 18.7 and 0.26 ppb QSE, respectively. The latter value compares favorably with our measurements of fully oceanic samples, in which the mean concentration of component C3 was 0.28 ppb QSE (s = 0.16, N = 63). Assuming that conservative mixing of terrestrial CDOM is the major factor driving the concentrations of C3 in the surface oceans, as is suggested by the results of this study, we deduce that we were able to detect, on average, fluorescence from ~ 1.5% residual terrestrial organic matter in the surface Pacific and Atlantic oceans. This is within the range (0.7 - 2.4%) calculated from the distribution of lignin biomarkers (Opsahl and Benner, 1997), the measurement of which necessitates considerably more complicated analytical techniques. While further data are required to ascertain whether the approximately 3-fold differences between terrestrial organic matter concentrations in the Atlantic and Pacific Oceans implied by lignin distributions are also reflected in optical measurements, our dataset also indicated higher background CDOM in the Atlantic.

The above calculation assumes that the deep ocean was not a significant source of component C3 during our study, as is suggested by the data in Figure 12. We could not test this assumption explicitly since we did not collect samples from below the euphotic zone and there are as yet no published PARAFAC decompositions of deep ocean EEMs. However, previous researchers have noted similarities between deep ocean and terrestrial EEMs (Mopper and Schultz, 1993;Coble,

1996; Coble et al., 1998). If C3 is found to be produced in the deep ocean as well as on land, this would favor the interpretation that like C2, C3 is a ubiquitous product of degrading organic matter rather than material of specifically terrestrial or marine origin.

5. Conclusions

The ability to discriminate between terrestrial and microbial seawater organic matter constituents has applications for identifying dominant ecological processes and tracing water sources in a range of aquatic ecosystems. Detailed spectral analyses of CDOM in the north Pacific and Atlantic oceans suggests that existing rules of thumb for distinguishing terrestrial and autochthonous organic matter sources via fluorescence are not widely applicable in the surface ocean. PARAFAC analysis allowed the sensitive spectral decomposition of independent chemical components that exhibited variable distributions in the oceans. Several of these components have now been resolved in multiple unrelated studies. In this study, the relative concentration of the two most widely-distributed humic-like components was a robust indicator of distance from shore. It also seems clear from this study that the commonly-termed marine humic peak “M” is produced in terrestrial as well as marine environments, suggesting that prevailing interpretations of the origin of major CDOM fluorescence peaks may need revisiting. Using broad relationships between characteristic organic matter components inferred across ocean basins and inter-annual timescales, we estimate that fluorescence of terrestrial humic material was detectable in the surface open ocean at levels around 1.5% of freshwater concentrations. In characterizing CDOM fluorescence spectra in the surface oceans, we present an additional tool for analyzing the supply, mixing and removal of terrestrial organic matter to the worlds’ oceans.

6. Acknowledgements

We thank P. Coble, J. Boehme, R. Conmy, and W. Martinsen for EEMs analyses and R. Bro for training on PARAFAC analyses. This paper has benefited greatly from insightful criticism by anonymous reviewers. We are indebted to NYK Bulkship (USA) LTD., Gateway Maritime Corp. / Sincere Industrial Corp., Matson Navigation Company, Bergesen DY ASA., Sea River Maritime, the Alaska Tanker Company, BP Amoco PLC and Krupp Seeschiffahrt GmbH for hosting this research on their ships. Funding was by the US Coast Guard's Research and Development Center and the Columbia River Aquatic Nuisance Species Initiative (CRANSI) and Danish Research Council grant to CS (274-05-0064).

References

2004. U.S. Coast Guard. Mandatory Ballast Water Management Program for U.S. Waters: Final Rule, 33 CFR 151, Subpart D, pp. 44952-44961.

Andersen, C.M. and Bro, R., 2003. Practical aspects of PARAFAC modeling of fluorescence excitation - emission data. Journal of Chemometrics, 17: 200-215.

Andersson, C.A. and Bro, R., 2002. The N-way toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems, 52: 1-4.

Beltran, J.L., Ferrer, R. and Guiteras, J., 1998. Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation-emission fluorescence spectra. Analytica Chimica Acta, 373(2-3): 311-319.

Blough, N.V. and Del Vecchio, R., 2002. Chromophoric DOM in the coastal environment. In: D.A. Hansell and C.A. Carlson (Editors), Biogeochemistry of marine dissolved organic matter. Academic Press, pp. 509-546.

Blough, N.V., Zafiriou, O.C. and Bonilla, J., 1993. Optical-Absorption Spectra of Waters from the Orinoco River Outflow - Terrestrial Input of Colored Organic-Matter to the Caribbean. Journal of Geophysical Research-Oceans, 98(C2): 2271-2278.

Bro, R., 1997. PARAFAC: Tutorial and applications. Chemometrics and Intelligent Laboratory Systems, 38: 149-171.

Bro, R. and Kiers, H.A.L., 2003. A new efficient method for determining the number of components in PARAFAC models. Journal of Chemometrics, 17(5): 274-286.

Carlton, J.T. and Ruiz, G.M., 2005. The magnitude and consequences of bioinvasions in marine ecosystems: implications for conservation biology. In: E.A. Norse and L.B. Crowder (Editors), Marine Conservation Biology: The Science of Maintaining the Sea's Biodiversity. Island Press, Washington, pp. 123-148.

Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine Chemistry, 51: 325-346.

Coble, P.G., Del Castillo, C.E. and Avril, B., 1998. Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. Deep-Sea Research Part Ii-Topical Studies in Oceanography, 45(10-11): 2195-2223.

Coble, P.G., Green, S.A., Blough, N.V. and Gagosian, R.B., 1990. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. Nature, 348(6300): 432-435.

Coble, P.G., Mopper, K. and Schultz, C.S., 1993. Fluorescence contouring analysis of DOC intercalibration experiment samples: A comparison of techniques. Marine Chemistry, 41: 173-178.

Del Castillo, C.E., Coble, P.G., Morell, J.M., Lopez, J.M. and Corredor, J.E., 1999. Analysis of the optical properties of the Orinoco River plume by absorption and fluorescence spectroscopy. Marine Chemistry, 66(1-2): 35-51.

Del Castillo, C.E., Gilbes, F., Coble, P.G. and Muller-Karger, F.E., 2000. On the dispersal of riverine colored dissolved organic matter over the West Florida Shelf. Limnology and Oceanography, 45(6): 1425-1432.

Del Vecchio, R. and Blough, N.V., 2004. On the origin of the optical properties of humic substances. Environmental Science & Technology, 38(14): 3885-3891.

Duursma, E.K., 1974. The fluorescence of dissolved organic matter in the sea. In: N.G. Jerlov and E. Steeman Nielsen (Editors), Optical Aspects of Oceanography. Academic Press, New York, pp. 237–256.

Firestone, J. and Corbett, J.J., 2005. Coastal and Port Environments: International Legal and Policy Responses to Reduce Ballast Water Introductions of Potentially Invasive Species. Ocean Development & International Law, 36(3): 291 - 316.

Gibb, S.W. et al., 2000. Surface phytoplankton pigment distributions in the Atlantic Ocean: an assessment of basin scale variability between 50[deg]N and 50[deg]S. Progress In Oceanography, 45(3-4): 339-368.

Green, S.A. and Blough, N.V., 1994. Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. Limnology and Oceanography, 39: 1903-1916.

Gustafson, K.E. and Dickhut, R.M., 1997. Distribution of polycyclic aromatic hydrocarbons in southern chesapeake bay surface water: evaluation of three methods for determining freely dissolved water concentrations. Environmental Toxicology and Chemistry, 16(3): 452-461.

Harshman, R.A. and Lundy, M.E., 1994. Parafac - Parallel Factor-Analysis. Computational Statistics & Data Analysis, 18(1): 39-72.

Harvey, G.R., Boran, D.A., Chesal, L.A. and Tokar, J.M., 1983. The structure of marine fulvic and humic acids. Marine Chemistry, 12(2-3): 119-132.

Hedges, J.I., Hatcher, P.G., Ertel, J.R. and Meyers-Schulte, K.J., 1992. A comparison of dissolved humic substances from seawater with Amazon River counterparts by 13C-NMR spectrometry. Geochimica et Cosmochimica Acta, 56(4): 1753-1757.

I.M.O., 2004. International Convention for the Control and Management of Ships Ballast Water and Sediments. BWM/CONF/36, 16 February 2004. International Maritime Organisation, London.

Kalle, K., 1949. Fluoreszenz und Gelbstoff im Bottnischen und Finnischen Meerbusen. Deutsche Hydrographische Zeitschrift, 2: 117-124.

Kieber, D.J., 2000. Photochemical production of biological substrates. Chapter 5. In: S.J. Mora, S.J.S. Demers and M. Vernet (Editors), The effects of UV radiation in the marine environment. Cambridge University Press, New York, pp. 130-148.

Kitidis, V. et al., 2006. Variability of chromophoric organic matter in surface waters of the Atlantic Ocean. Deep-Sea Research Part Ii-Topical Studies in Oceanography, 53(14-16): 1666-1684.

Liu, C.C. et al., 2006. Estimating the underwater light field from remote sensing of ocean color. Journal of Oceanography, 62(3): 235-248.

Lorenzo-Seva, U. and Berge, J.M.F.t., 2006. Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. Methodology, 2(2): 57-64.

Mobed, J.J., Hemmigsen, S.L., Autry, J.L. and McGown, L.B., 1996. Fluorescence characterization of IHSS humic substances: Total luminescence spectra with absorbance correction. Environmental Science & Technology, 30(10): 3061-3065.

Mopper, K. and Schultz, C.A., 1993. Fluorescence as a possible tool for studying the nature and water column distribution of DOC components. Marine Chemistry, 41(1-3): 229-238.

Moran, M.A. and Zepp, R.G., 1997. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. Limnology and Oceanography, 42(6): 1307-1316.

Murphy, K., Ruiz, G. and Sytsma, M., 2004. Standardized sampling protocol for verifying mid-ocean ballast water exchange: Revision II March 4 2004., Smithsonian Environmental Research Center, Edgewater.

Murphy, K.R., Ruiz, G.M., Dunsmuir, W.T.M. and Waite, T.D., 2006. Optimized parameters for fluorescence-based verification of ballast water exchange by ships. Environmental Science and Technology, 40(7): 2357-2362.

Nagata, T., 2000. Production mechanisms of dissolved organic matter. In: D.G. Kirchman (Editor), Microbial Ecology of the Oceans. Wiley-Liss, New York, pp. 121-152.

Nelson, N.B., Siegel, D.A. and Michaels, A.F., 1998. Seasonal dynamics of colored dissolved material in the Sargasso Sea. Deep-Sea Research Part I-Oceanographic Research Papers, 45(6): 931-957.

Nieto-Cid, M., Alvarez-Salgado, X.A., Gago, J. and Perez, F.F., 2005. DOM fluorescence, a tracer for biogeochemical processes in a coastal upwelling system (NW Iberian Peninsula). Marine Ecology-Progress Series, 297: 33-50.

Opsahl, S. and Benner, R., 1997. Distribution and cycling of terrigenous dissolved organic matter in the ocean. Nature, 386(6624): 480-482.

Riu, J. and Bro, R., 2003. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. Chemometrics and Intelligent Laboratory Systems, 65: 35-49.

Schulz-Bull, D.E., Petrick, G., Bruhn, R. and Duinker, J.C., 1998. Chlorobiphenyls (PCB) and PAHs in water masses of the northern North Atlantic. Marine Chemistry, 61(1-2): 101-114.

Siegel, D.A., Maritorena, S., Nelson, N.B., Hansell, D.A. and Lorenzi-Kayser, M., 2002. Global distribution and dynamics of colored dissolved and detrital organic materials. Journal of Geophysical Research-Oceans, 107(C12).

Smilde, A., Bro, R. and Geladi, P., 2004. Multiway analysis with applications in the chemical sciences. John Wiley & Sons Ltd, Chichester, 384 pp.

Smith, W.H.F. and Sandwell, D.T., 1997. Global sea-floor topography from satellite altimetry and ship depth soundings. Science, 277: 1956–1961.

Stedmon, C.A. and Markager, S., 2001. The optics of chromophoric dissolved organic matter (CDOM) in the Greenland sea: An algorithm for differentiation between marine and terrestrially-derived organic matter. Limnology and Oceanography, 46(8): 2087-2093.

Stedmon, C.A. and Markager, S., 2005a. Resolving the variability of dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnology and Oceanography, 50: 686-697.

Stedmon, C.A. and Markager, S., 2005b. Tracing the production and degradation of autochthonous fractions of dissolved organic matter using fluorescence analysis. Limnology and Oceanography, 50(5): 1415-1426.

Stedmon, C.A., Markager, S. and Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Marine Chemistry, 82: 239-254.

Stedmon, C.A., Markager, S. and Kaas, H., 2000. Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. Estuarine, Coastal and Shelf Science, 51: 267-278.

Stedmon, C.A. et al., 2007. Photochemical production of ammonium and transformation of dissolved organic matter in the Baltic Sea. Marine Chemistry, 104(3-4): 227-240.

Sutton, R. and Sposito, G., 2005. Molecular structure in soil humic substances: The new view. Environmental Science & Technology, 39(23): 9009-9015.

Tucker, L.R., 1951. A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984), Department of the Army, Washington D.C., Washington D.C.

Velapoldi, R.A. and Mielenz, K.D., 1980. Standard reference materials: A fluorescence standard reference material- quinine sulfate dihydrate. NBS Special Publication 260-64, National Bureau of Standards, Washington,DC.

Williams, P.M. and Druffel, E.R.M., 1987. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. Nature, 330(6145): 246-248.

Wolfbeis, O., 1985. The fluorescence of organic natural products. In: S. Schulman (Editor), Molecular Luminescence Spectroscopy. Part I: Methods and Applications. Wiley, New York, NY, pp. 167-370.

Zepp, R.G., Sheldon, W.M. and Moran, M.A., 2004. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices. Marine Chemistry, 89(1-4): 15-36.

Table 1. Fluorescence characteristics of Seawater EEMs: Primary (1°) and secondary (2°) excitation (λ_{ex}) and emission (λ_{em}) maxima, compared with previously identified spectra. References (*Ref.*): 1. Coble (1996); 2. Murphy et al. (2006); 3. Stedmon et al. (2003); 4. Stedmon et al. (2005a); 5. Stedmon et al. (2005b); 6. Stedmon et al. (2007); 7. Wolfbeis (1985)

BWE7 Model *(Ref 2)*	**Kauai Model** l_{ex} / l_{em} **1° (2°)**		**Description and probable source (*Ref*)**
			Amino acids, free or bound in proteins
	P2	<260 /<300	Tyrosine: 275/303 (*Ref. 7*)
C1: 275 / <300	P5	270 / 310	"B" peak: 275/305 (*Ref. 1*)
			Component 8: 275/304 (*Ref. 4*)
			Component 4: 275/306 (*Ref. 5*)
			(1) Marine and terrestrial humic materials
			Marine humic "M" peak (*Ref. 1*)
C2: 315 / 418	P1	(<260) 310 / 414	Terrestrial component 4: (<250) 325/416 (*Ref. 3*)
			Microbial component 3: 295/398 (*Ref. 5*)
			(2) Anthropogenic humic materials, agricultural:
			Component 5: 325/428 (*Ref. 4*)
			Terrestrial humic substances, widespread
C3: 260 (370) / 490	P3	<260 (380) / 498	"A" and "C" peaks (*Ref. 1*)
			Component 3: 270 (360)/478 (*Ref. 3*)
			Component 2: 250 (385)/504 (*Ref. 4*)
C4: 250 (320) / 370			Origin uncertain, possibe PAH (*Ref 2*)
C5: 255 (280) / 580			Rhodamine WT dye (*Ref. 2*)
			Amino acids, free or bound in proteins
	P6	275 / 318	Tryptophan: 278/354 (*Ref. 7*)
C6: 280 / 328	P7	280 / 342	"T" peak: 275/340 (*Ref. 1*)
			Component 6: 280/338 (*Ref. 5*)
			Component 7: 280/344 (*Ref. 4*)
C7: 240 (300) / 338			Amino acids, free or bound in proteins (*Ref. 6*)
			Terrestrial humic substances
C8: 250 (380) / 416	P8	<260 (355) / 434	"A" and "C" peaks (*Ref. 1*)
			Component 4:250 (360)/440 (*Ref. 4*)
C9: <240 / 422			Photochemical product of terrestrial organic matter (*Ref. 6*)
absent	P4	<260 / -	Origin uncertain, possible fluorometer artifact

Table 2. Summary of locations and intensity of seawater sampling on the Pacific and Atlantic cruises. Sample frequency (*N*) refers to samples for which salinity exceeded 30. Numbers of samples with salinity below 30 (N_{30}) are indicated in brackets.

Cruise	Location	Abbreviation (see Fig. 6)	*N* (N_{30})
Pacific Ocean			
K	San Francisco Bay	SF Bay	4
	Californian coast	Cal Cst	9
	Eastern Pacific	E Pac	21
	Hawaii coastal	Hawaii	14
	Northeastern Pacific	NE Pac	29 (2)
	Seattle, WA	Seattle	2
Atlantic Ocean			
Fos	Mediterranean Sea	Mediterr	16
	Straight of Gibraltar	Gibr Str	2
	Northeastern Atlantic	NE Atl	8
	Central North Atlantic	Mid Atl	6
	Northwestern Atlantic	NW Atl	6
	Atlantic Transitional	Atl Trans	8
	Mid-Atlantic Bight	MAB	7 (4)
	Norfolk, VA	Norfolk	2 (2)
BN	Rotterdam	Rottd	2
	N. Sea/English Channel	NS/EC	13
	European Shelf	Eur Shlf	3
	Bay of Biscay	B Biscay	4
	Northeastern Atlantic	NE Atl	9
	Tropical North Atlantic	Trop Atl	16
	Brazilian Shelf	Braz Shlf	7
	Sao Luis, Brazil	Sao Luis	2

Table 3: Four possible scenarios in a C2 and C3 decoupling model. The function rand(1) generates a random number between 0 and 1.

Scenario	Governing equations	Proposed mechanism
i	$C3_{(i+1)} = D*C3_i$ $C2_{(i+1)} = D*C2_i$	Conservative mixing
ii	$C3_{(i+1)} = (1- r_3)*D*C3_i$ $C2_{(i+1)} = (1- r_2)*D*C2_i$ $r_3 > r_2$	Conservative mixing + removal (faster for C3 than C2)
iii	$C3_{(i+1)} = D*C3_i$ $C2_{(i+1)} = D*(C2_i + p_2)$	Conservative mixing + C2 production
iv	$C3_{(i+1)} = 0.28$ $C2_{(i+1)} = 1.8*rand(1)$	Background C3 + variable C2 production in the ocean

Figure Legends

Figure 1: Sampling stations during cruises in A. the Pacific Ocean (cruise K) and B. the Atlantic Ocean (cruises BN and Fos).

Figure 2: Fluorescence spectra of nine previously identified PARAFAC components (C1-9) from seawater and ships' ballast water according to the BWE7 model (Murphy et al., 2006).

Figure 3: Example EEMs (corrected) from the two datasets: (A) Kauai (B) Fos. Kauai EEMs were affected by dominant bands of secondary Rayleigh and Raman scattering and a secondary protein peak near the 600 nm emission wavelength.

Figure 4: Fluorescence signatures of eight PARAFAC components identified in the Kauai EEM dataset. Contour plots of components P1-P8 are ordered by decreasing percent explained variation. Line plots at right show split-half validations in which each component's excitation (left) and emission (right) spectra are estimated from two independent halves of the dataset (thin lines) and the complete dataset (thick lines).

Figure 5: Comparison of PARAFAC components from the Kauai model and the BWE7 model of Murphy et al. 2006. Comparisons are limited to combinations of components with high (> 0.92) Tucker Congruence on both excitation and emission spectra.

Figure 6: Mean CDOM fluorescence intensities (ppb QSE) along cruise tracks in the Pacific (K) and Atlantic (Fos and BN) oceans. Error bars (standard errors) are bidirectional with lower limits omitted for clarity. Colored bars are in the same order as they appear in the key for each plot. See Table 2 for a description of sampling locations on the horizontal axes. 1st row: Salinity and fluorescence intensities of the humic-like components; 2nd row: fluorescence intensities of the protein-like components; 3rd row: ocean bathymetry and distance from shore.

Figure 7: Varying relationships between fluorescence intensities (ppb QSE) of pairs of components in Pacific, Atlantic and Mediterranean samples (3 cruises: K, Fos, BN): (A) Correlation of components C8 and C3; (B) Correlation of components C1 and C6.

Figure 8: Mean fluorescence intensity (ppb QSE) in seawater as a function of distance from the nearest land mass (3 cruises: K, Fos, BN; salinity > 30). Error bars (standard errors) are bidirectional with lower limits omitted for clarity. Colored bars appear in the same order as in the key. (A) Component-wise comparisons showing terrestrial-type (C2, C3, C4, C8, C9) and protein-type (C2, C6, C7) behaviors; (B) comparison of average fluorescence signatures by water mass.

Figure 9: Absorption, spectral slope and salinity in seawater as a function of distance from the nearest land mass. Spectral slope was calculated by non-linear regression over the 300 - 650 nm wavelength range. Data represent mean ± SE for samples from 3 cruises (K, Fos, BN) for which salinity > 30.

Figure 10: Relationship between the fluorescence index (ratio of emission at 500/450 nm, when excited at 370 nm) and distance from shore on cruises K and Fos.

Figure 11: Mean fluorescence ratios in seawater as a function of distance to the nearest land mass (3 cruises: K, Fos, BN; salinity > 30). Bar plots show mean and standard errors across all samples using each component's fluorescence divided by the fluorescence of component C3. Colored bars in each plot appear in the same order as in the title. Bottom plot shows corresponding mean salinity and ocean bathymetry.

Figure 12: Decoupling between C2 and C3 fluorescence (ppb QSE) in seawater. (A) At low C3 concentration, C2 concentrations frequently lie above the conservative mixing curve; (B) the ratio of C2/C3 in seawater is independent of C2 at high concentrations, but at low concentrations, it is driven by the concentration of C2; (C & D) modeled relationships assuming conservative mixing only (i), mixing and faster removal of C3 (ii), mixing and constant production of C2 (iii), constant C3 and production of C2 in the open ocean (iv).

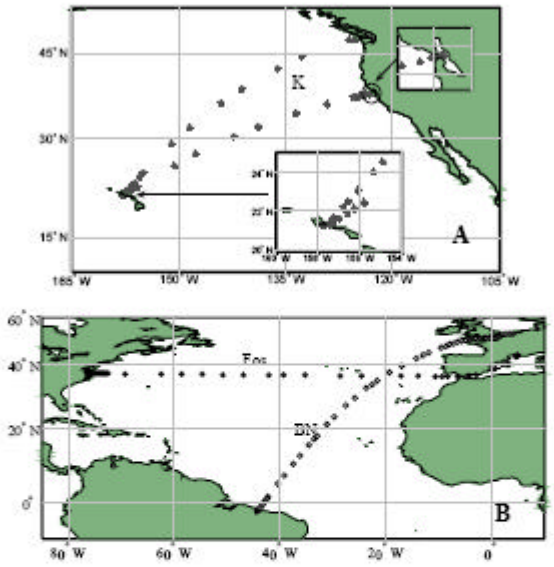

Figure 1. *K. R. Murphy et al.*

Fig 2

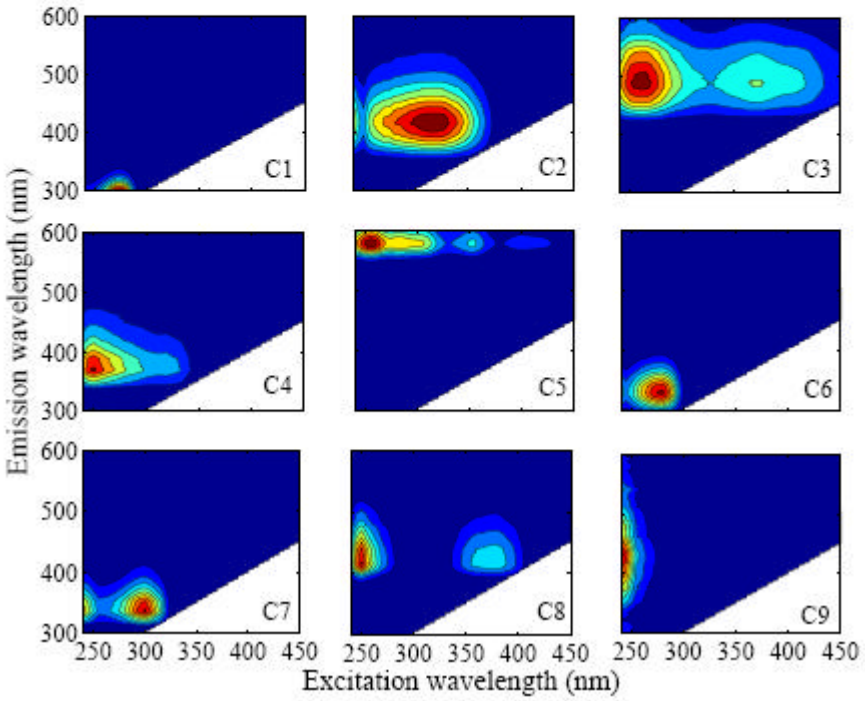

Fig 3

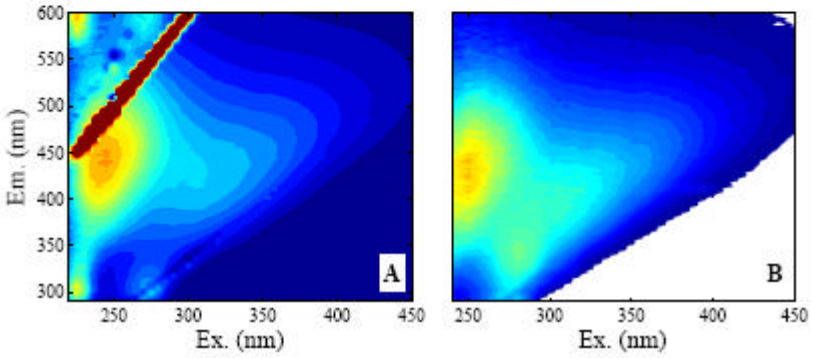

Fig 4

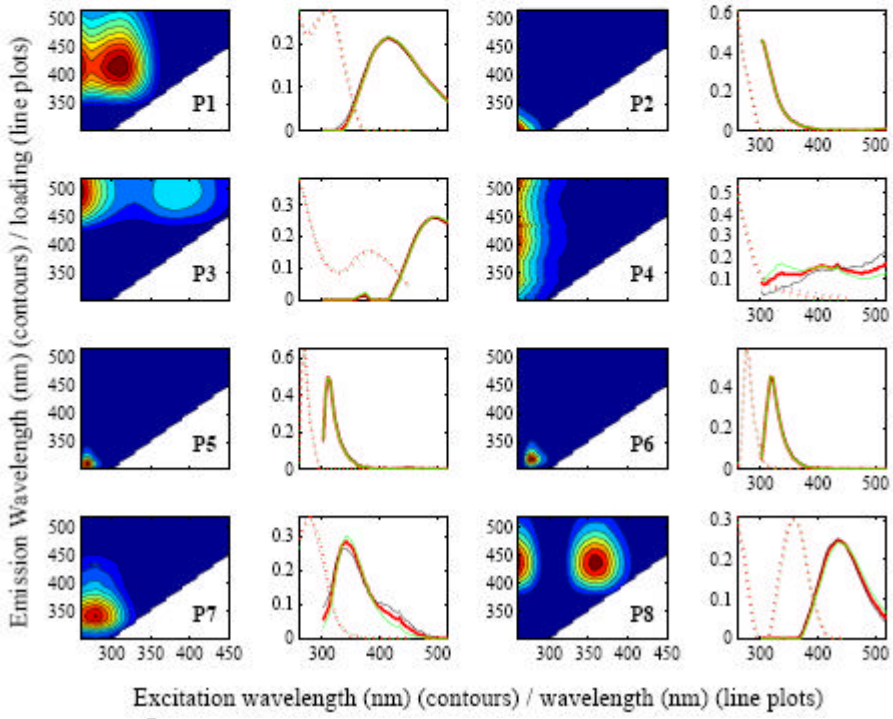

Fig 5

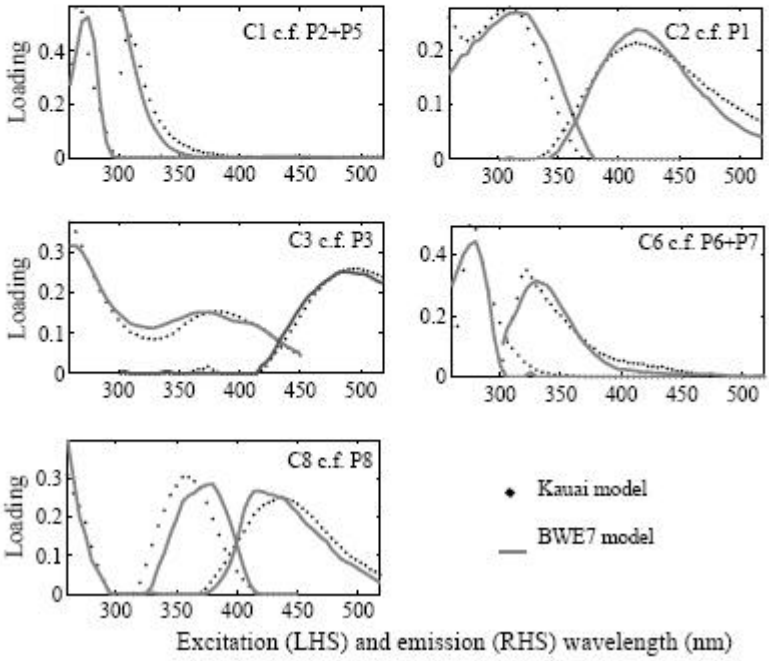

Fig 6

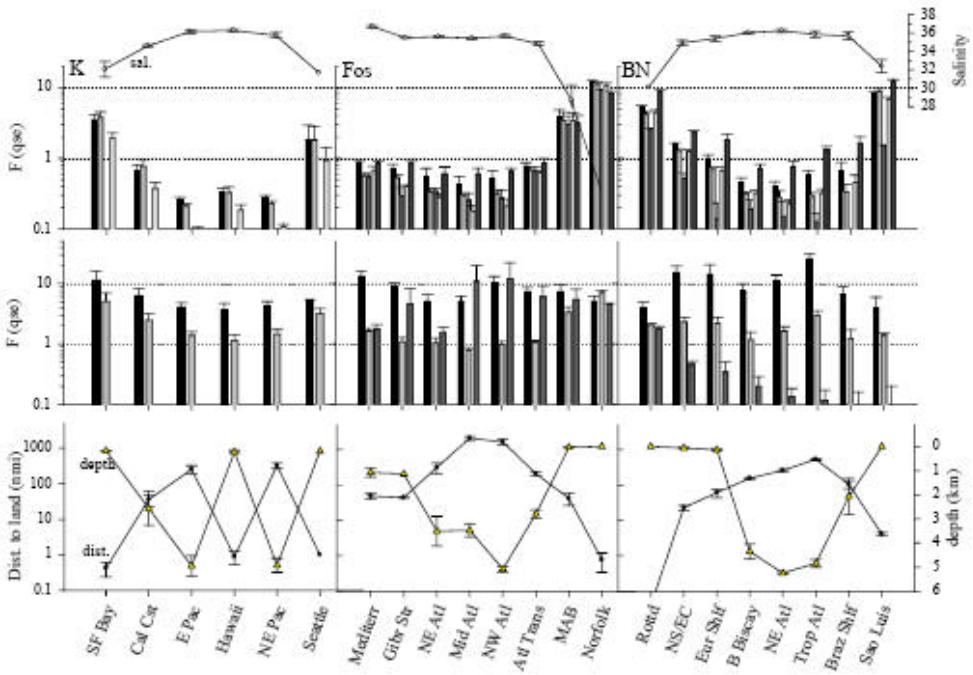

Fig 7

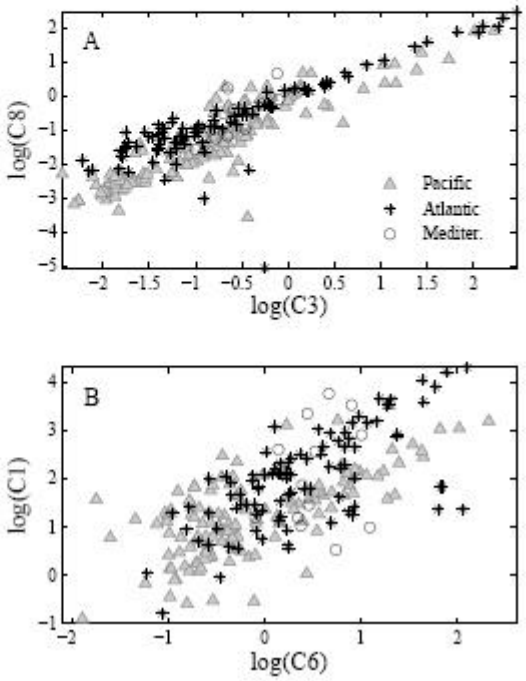

Fig 8

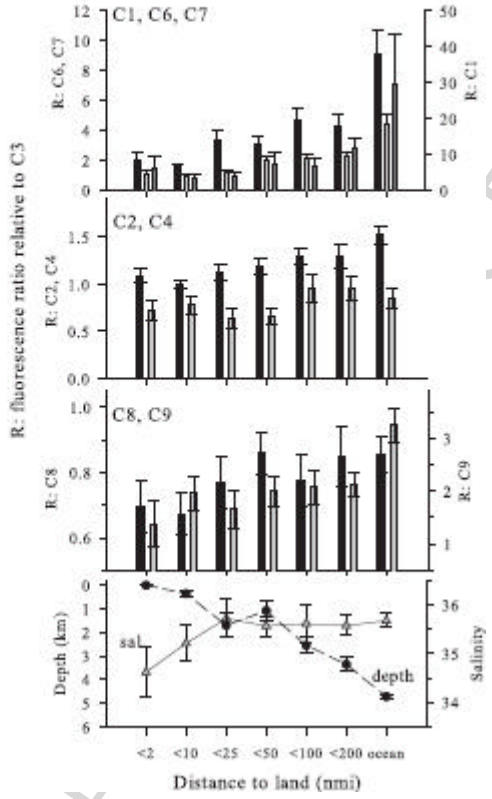

Fig 9

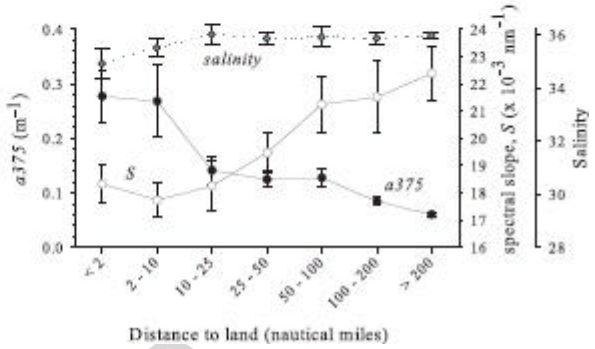

Fig 10

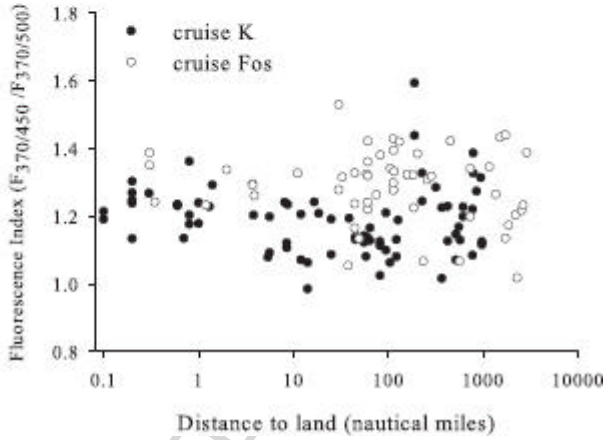

Fig 11

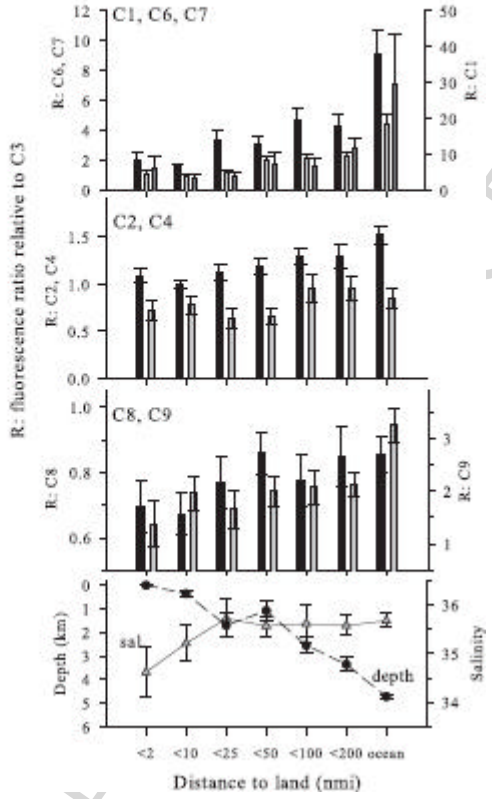

Fig 12