An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean

Nick J. Hardman-Mountford [a,b,*], Takafumi Hirata [a,b], Kirsten A. Richardson [b], Jim Aiken [a,b]

[a] Centre for observation of Air-Sea Interactions and Fluxes (CASIX), UK
[b] Plymouth Marine Laboratory, Plymouth, PL1 3DH, UK

A B S T R A C T

Based on the view of ecological pattern being steady state, rather than an equilibrium phenomenon, we assert that, if real, ecological biomes and provinces in the ocean should be detectable in surface fields obtained from satellite data as coherent, co-varying spatial regions with a high degree of permanence. Likewise, hierarchy is an important property of ecological systems that should be exhibited by such patterns if they represent real ecological structure. In this paper, we apply a combination of multivariate statistics and classification techniques to a time series of satellite-derived, surface-ocean chlorophyll data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), (a) to provide an objective characterisation and classification of ecological pattern in the ocean and (b) to investigate the characteristic system properties (persistence, hierarchy) of the broad-scale patterns observed to test whether they behave as autonomous ecological systems. Comparison of this classification with Longhurst [Longhurst, A. (1998), Ecological Geography of the Sea. San Diego: Academic Press, 398 pp.] suggests the need for equatorial forcing processes to be taken into account to explain observed ecological pattern at the biome level.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Ecological patterns (in time, space and organizational complexity) represent structures produced by interactions between organisms and with their environment (Hutchinson, 1953). As such, pattern is an emergent property of complex ecological systems and an important descriptor of these (eco)systems. In terrestrial landscapes, the association of ecological patterns with bounded units of land, at a variety of spatial scales, forms the basis of the ecosystem concept in ecology (Ducklow, 2003). At the macro-scale, the classification of these units into major biomes (e.g. boreal forest, savannah, desert, etc.), characterised by distinct climatic regimes, land cover and animal populations, is generally recognised (e.g. Bailey, 1998; Olson et al., 2001; Udvardy, 1975). In contrast, the question of whether such well defined biomes exist in the ocean is still largely unresolved.

2. System properties of ecological units

By definition, all ecological systems are thermodynamically-open, allowing exchange of energy and matter across the system bound-aries. Nonetheless, their internal dynamics are, to a greater or lesser extent, independent of the interactions between systems, conferring stability. Indeed, it is a truism that all systems are inherently stable over the scales of their existence (i.e. in time, space and complexity). Hence, one property of this stability is persistence through time (Grimm & Wissel, 1997), or the steady-state condition (Hutchinson, 1953). This condition of persistence is clearly met for terrestrial biomes (Platt & Sathyendranath, 1999), providing the steady-state spatial representation (pattern) of the underlying ecosystem. However, the fluid, highly dynamic nature of the pelagic marine realm makes the detection of spatially-coherent, persistent ecological units more challenging. Their spatial extent varies significantly through time, so a description needs to account for spatial and temporal variability together.

Another accepted property of complex systems is hierarchical structure, with each system composed of inter-related subsystems at nested scales (Simon, 1962). Ecological hierarchies provide a specific linkage between scale and heterogeneity in ecological patterns (Wu & Loucks, 1995). This is clearly exhibited in ecological systems and implicitly recognised in spatial classification schemes, from the largest to the smallest spatial scales, i.e. the planetary level biosphere system is composed of primary level ecological domains (biomes), which in turn comprise regional ecosystems (provinces), made up of communities, and so on, down to the scale of individual organisms.

3. Classification of pelagic ecosystems from ocean colour data

There have been many conceptual attempts at providing a comprehensive classification for the pelagic marine environment, based mainly on either the biogeography of distinct taxonomic groups (e.g. copepods) or the spatial variability of physical properties, such as temperature, salinity and turbulence (reviewed by Longhurst, 1998). Although severely restricted by historical oceanographic undersampling (Munk, 1997), these have nevertheless shown some limited success in providing classification schemes, both for ocean physical properties and ocean biology, due largely to the overwhelming importance of bottom-up control in ocean ecosystems (Cullen et al., 2002; Platt & Sathyendranath, 1999). The advent of ocean colour remote sensing, with the launch of the Coastal Zone Color Scanner (CZCS) in 1978, stimulated attempts to address this limitation and was used as the basis for a series of classifications based on empirically-derived chlorophyll and primary production estimates from this sensor (Longhurst, 1995; Longhurst et al., 1995; Longhurst, 1998; Sathyendranath et al., 1995).

The classification of terrestrial landscapes into biomes is generally based on indicative vegetation types, a manifestation of the characteristic physical environment that determines their distribution (broadly based on geology and climate) and the distinctive biota with which they are generally associated (Campbell, 1996; Platt & Sathyendranath, 1999). For the oceans, chlorophyll biomass is arguably the best single variable to characterise the spatial distribution and extent of pelagic ecosystems. It represents the base trophic level, accumulating the impacts of environmental forcing and propagating them up the food-web (bottom-up control), while also integrating the impacts of top-down control, e.g. through grazing. Different ranges of chlorophyll concentration have been identified as representative of different phytoplankton size classes (micro, nano, pico), so may be indicative of the dominant groups in the underlying phytoplankton community (Aiken et al., 2007; Fishwick et al., 2006; Uitz et al., 2006). Ocean colour satellite data provide fields of chlorophyll biomass throughout the oceans at high spatial resolution and high repeat frequency.

Longhurst's (1998) partition of the oceans into four major biomes, based on observed or inferred physical processes affecting stability and mixing in the upper ocean (polar, westerlies, trade winds, tropical), realised as more than 50 ocean provinces (regional expressions of different biomes), remains the most widely-accepted ocean classification to date and has been of much use as a framework in the estimation of oceanic primary production and biogeochemical budgets. Nonetheless, it has several limitations which have yet to be addressed:

1. Both the designation of provinces and their spatial extent are based on subjective interrogation of CZCS patterns and in situ data;
2. It is assumed that these units (provinces) constitute ecological entities, equivalent to ecosystems, but no rigorous test of the characteristic system properties of these regions has been undertaken that would confirm their identity as ecosystems.
3. Only nominal, fixed boundaries between provinces are provided. To be of operational use in oceanographic studies, instantaneous province boundaries should be delineated for any part of the ocean at any time (Platt & Sathyendranath, 1999).

4. Objectives of this study

In this paper, we apply a combination of multivariate statistics and classification techniques to satellite-derived, surface-ocean chlorophyll data, (a) to test an objective method for characterisation and
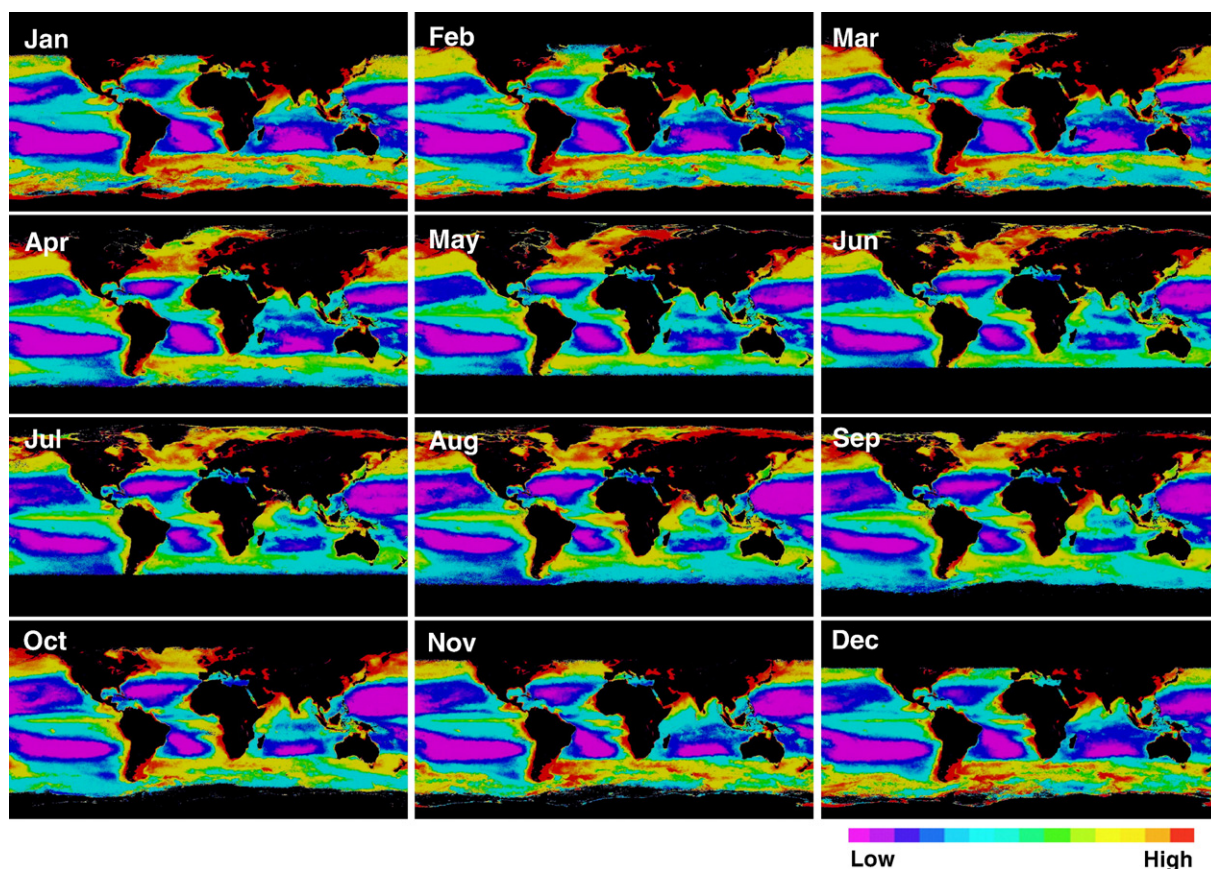

Fig. 1. SeaWiFS chlorophyll-a mean monthly climatology. The colour scale used is based on histogram equalization for each month in order to highlight the relative patterns and gradients, and does not correspond to consistent chlorophyll values between months.
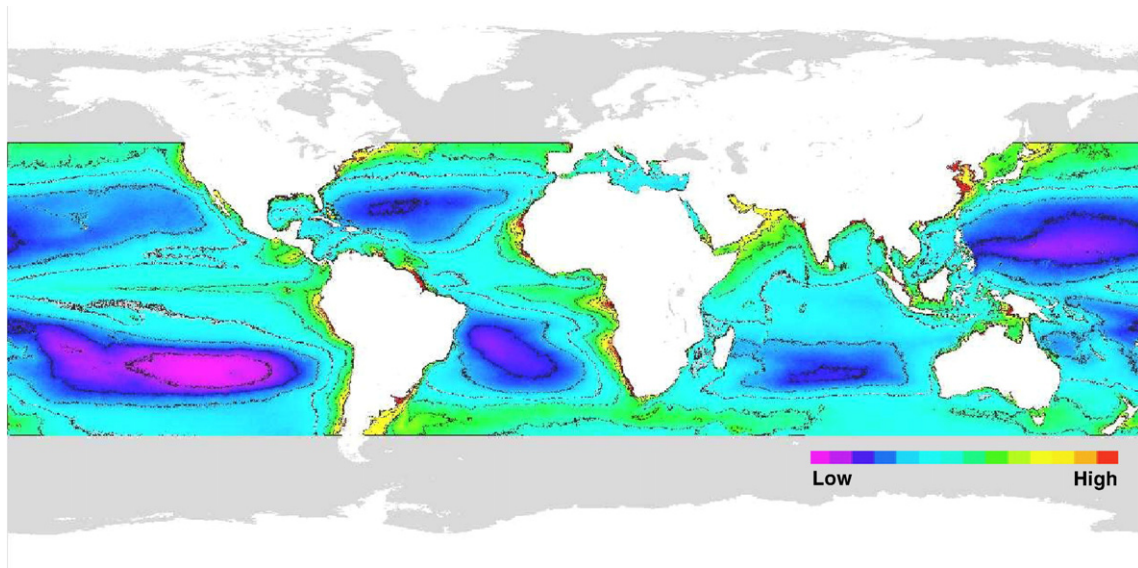
Fig. 2. PC1 from PCA of the global SeaWiFS chlorophyll-a mean monthly climatology. The colour scale used is a linear stretch (arbitrary units following normalisation and PCA transformation).

classification of ecological pattern in the ocean and (b) to investigate whether the broad-scale patterns observed exhibit the properties of persistence and hierarchy over a seasonal cycle, consistent with macro-scale ecological systems. The third limitation, identified for existing classifications, of providing instantaneous boundaries for marine provinces, is beyond the scope of this study but is considered in the discussion.

5. Data and methods

5.1. EO data

Monthly global composites of satellite chlorophyll-a level 3 products, derived from visible spectral radiometer (ocean colour) data from the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) on board the OrbView-2 satellite, were obtained from the US National Aeronautics and Space Administration (NASA). The composites have an 18 km spatial resolution. They were averaged at each pixel location over the time period Jan 1998–Dec 2004 to produce climatological means for each month (Fig. 1). The latitudinal extent of data retrieval by SeaWiFS is limited for the winter season in each hemisphere. For spatial Principal Components Analysis (PCA), the same spatial extent must be used for each variable (time step), so only data between 50°N and 50°S were used for each month in the creation of climatological means and for subsequent multivariate analyses.

5.2. Multivariate statistics

5.2.1. Spatial PCA

PCA is a commonly-used multivariate statistical technique related to factor analysis, which attempts to disentangle complicated inter-relationships by reducing the number of variables and transforming them to independence (Kendall & Stuart, 1966). It projects the variance structure of the data to linearly uncorrelated, or orthogonal, axes. Total variance of the data is preserved during the transformation and the mean square approximate errors are minimised (Fung & LeDrew, 1987).

Spatial (or T-mode) PCA (Preisendorfer, 1988) was applied to the prepared data, in which pixel locations represent the samples and time steps represent the variables in the input matrix. A normalised covariance matrix was used as the dispersion matrix. The output from the PCA consists of principal component (PC) fields, with associated eigenvalues and eigenvectors. Eigenvalues give the proportion of the total variance explained by each PC field, providing a measure of the relative importance of each PC. As such, they have been used to provide a significance measure to determine the number of principal components worth retaining for further analysis. Two common measures were used in this study to test for the number of principal components to retain: the log-eigenvalue (LEV) plot (Craddock & Flood, 1969) and the Guttman criterion (Guttman, 1954). Eigenvectors give a measure of the projection of each sample on each principal axis and in the study can be interpreted as the correlation of each PC field with each input time step. The PC image provides the spatial output from the analysis and is calculated as the sum of each of the sample values (pixels) through time, weighted by the eigenvector. The absolute values and the position of the zero line in these fields have no meaning; rather, it is the relative spatial gradients that describe the variance patterns of the underlying data set. As such, contour and legend values are not given for the PC fields.

5.2.2. Sub-sampling

To undertake the more computationally-expensive hierarchical cluster and multi-dimensional scaling (MDS) analyses, sub-samples were taken throughout the Atlantic Ocean for a regular grid of boxes with area of 2°×2° square, spaced approximately 8° apart, except for the highly heterogeneous region of the North Equatorial Counter Current systems where the distance between boxes was reduced to around 4° in the latitudinal. Mean monthly chlorophyll values were calculated for each of 106 sub-sample boxes.

5.2.3. Hierarchical cluster analysis

Cluster analysis is a multivariate method that uses a similarity matrix, in this case a correlation matrix of the sub-samples, to

Table 1
Eigenvalues and percentage of the total variance explained for each of the first 5 principal components from spatial PCA of SeaWiFS chlorophyll-a monthly climatology

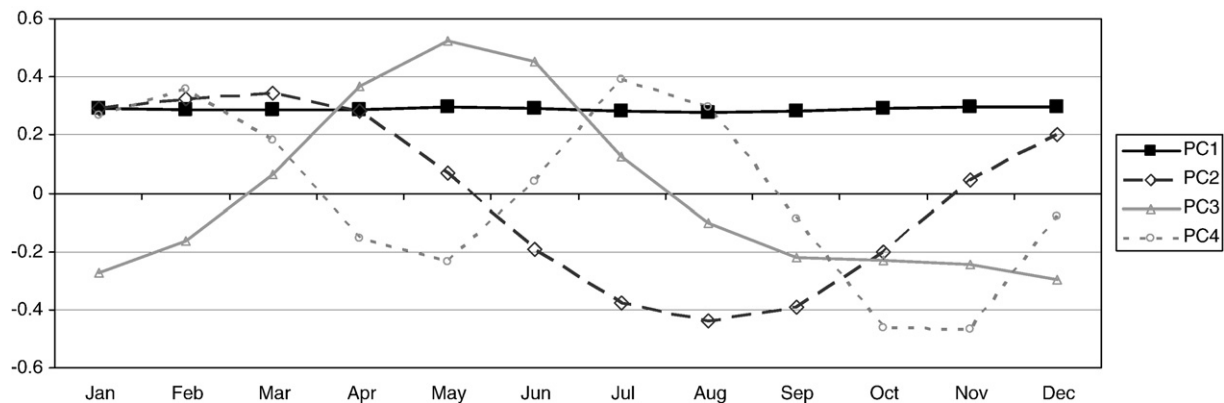| Principal component	Eigenvalue	% total variance	Cumulative % total variance
1	10.57	88.09	88.09
2	0.85	7.06	95.15
3	0.23	1.88	97.03
4	0.14	1.15	98.18
5	0.07	0.56	98.74

Fig. 3. Loadings (eigenvectors) from PCA of the global SeaWiFS chlorophyll-a mean monthly climatology.

determine natural groupings of samples, such that samples within a group are more similar to each other than to samples in different groups. For this study, a hierarchical agglomerative method was used which takes the similarity matrix as a starting point and successively fuses the samples into groups and the groups into larger clusters, starting with the highest mutual similarities, then gradually lowering the similarity level at which groups are formed. The results are displayed as a dendrogram. The analysis was implemented using routines provided by the PRIMER 6 software package for ecological statistics (Clarke & Warwick, 2001). Group average was used as the linkage rule because this is based on the average distance between samples so gives the best representation for all samples in a cluster.

Significance of groupings in the cluster analysis was tested using SIMPROF, a permutation method within the PRIMER 6 software. This tests for structure in the data after making a similarity profile by ranking the similarity matrix. A mean profile is then calculated by randomising the order of each variable value and re-calculating the profile. The pi statistic is calculated as the deviation of the actual data profile from the mean one. This is compared with the deviations of further randomly generated profiles to test for significance (Clarke & Warwick, 2001).
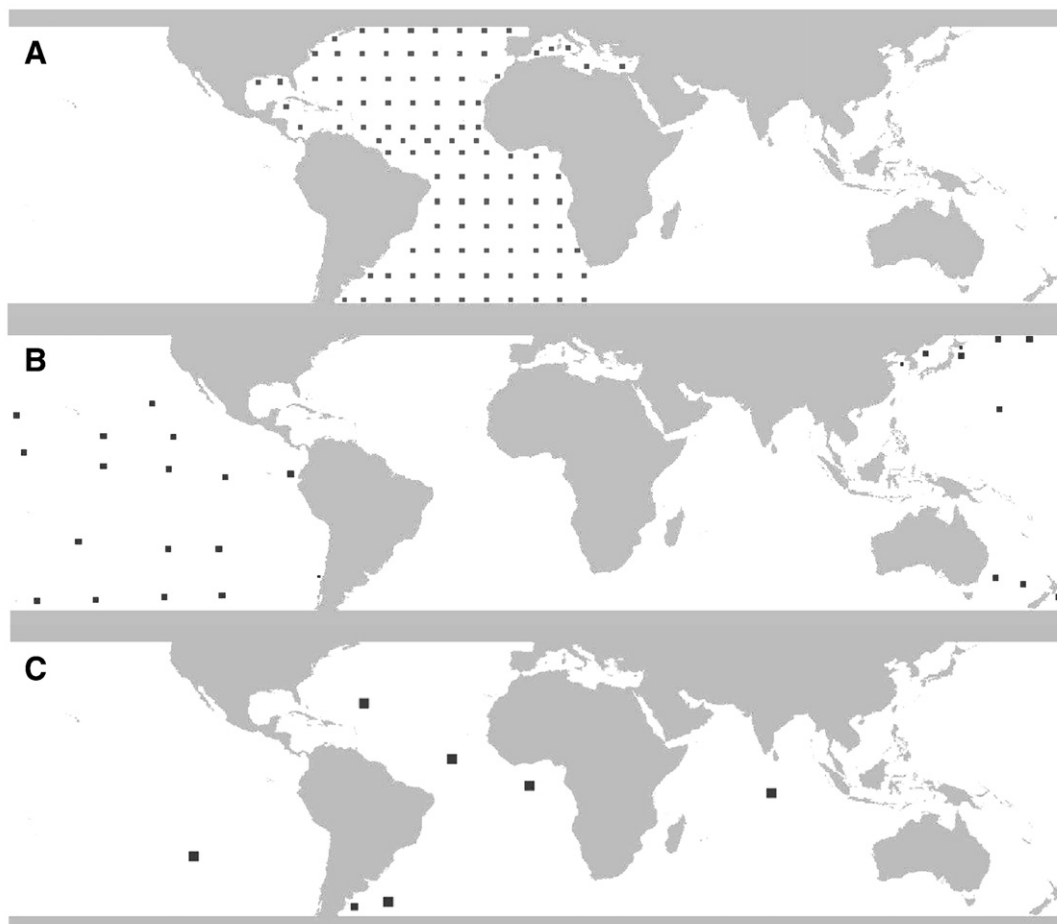

Fig. 4. Location of sub-sample boxes used (black squares). A) Atlantic: grid samples for use in cluster analysis and MDS, B) Pacific: replicates for each putative biome for use in MDS and ANOSIM test, C) training sets for supervised classification.
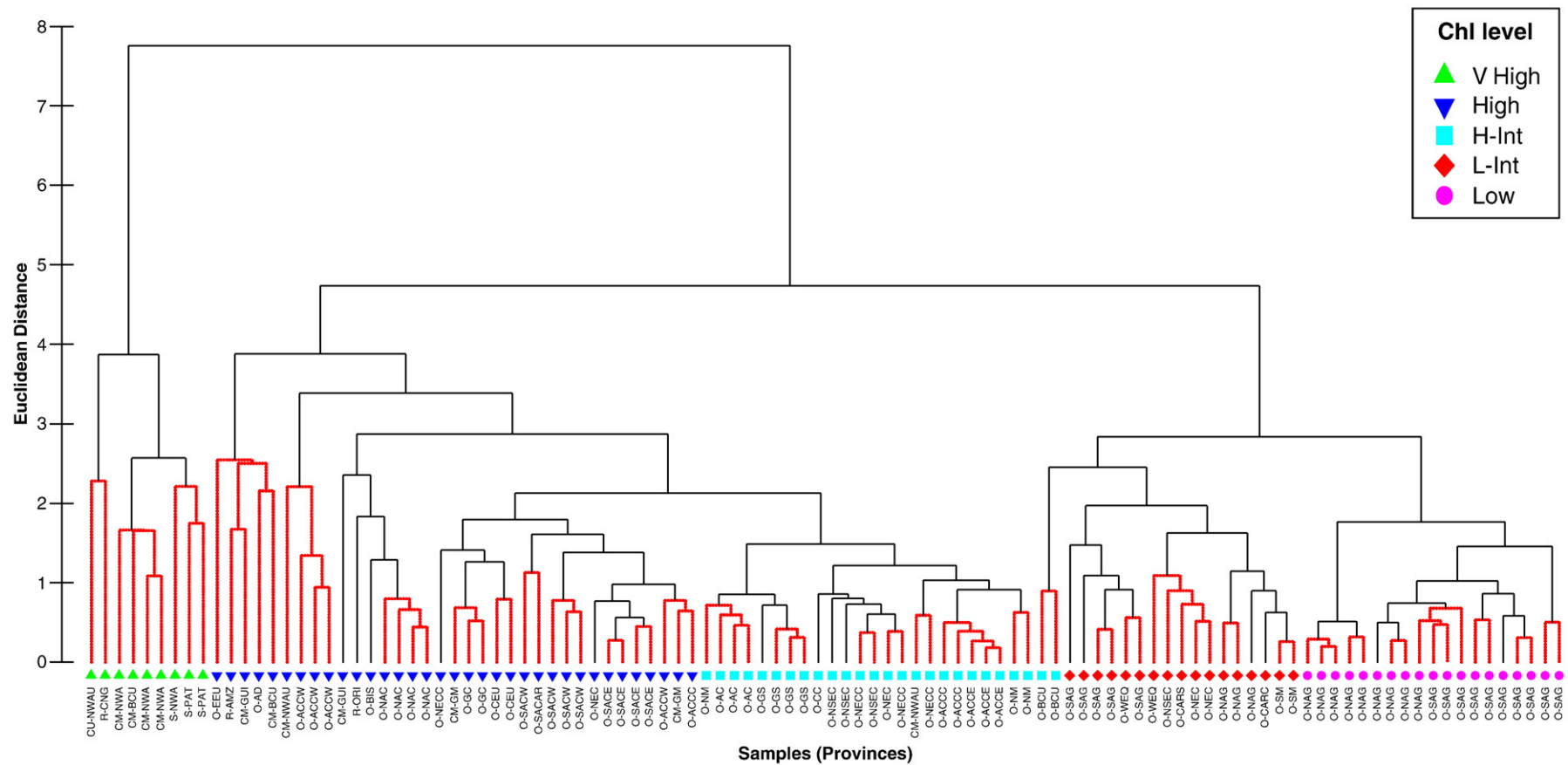
Fig. 5. Cluster dendrogram of Atlantic grid sub-sample data. Symbols along the *x*-axis represent chlorophyll level and labels correspond to nominal province codes (see Table 2). Solid black lines in the dendrogram represent clusters significant at the 95% level using SIMPROF test; dotted red lines represent clusters not significant at this level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2.4. Multi-dimensional scaling (MDS)

Multi Multi-dimensional scaling is an ordination technique whose goal is to present a configuration of the samples from a multivariate data set in a specified number of reduced dimensions, usually 2 or 3. Like cluster analysis, MDS acts on a similarity matrix but, whereas cluster analysis attempts to group samples into discrete clusters, MDS displays their inter-relations on a continuous scale. The method of non-metric MDS, used in this study, was introduced by Shepard (1962) and Kruskal (1964). The analysis was implemented for the correlation matrix of sub-samples using routines provided by the PRIMER 6 software package for ecological statistics. In this study results are presented as a 2-D ordination plot, with closeness of points representing their similarity. The results of cluster analysis are superimposed to assess the level of agreement between the two techniques.

5.2.5. Significance testing with ANOSIM test

The Analysis of Similarities (ANOSIM, R statistic) test in PRIMER 6 is a combined permutation/Monte Carlo method which provides a non-parametric alternative to MANOVA-type analyses (Clarke & Warwick, 2001). In this study it was used to test the significance of the groups identified by both MDS and clustering. The test was applied to an independent set of monthly chlorophyll sub-samples taken from the Pacific Ocean. For the open ocean, boxes covering an area of $2° \times 2°$ were sampled, whereas for the less extensive coastal regions, boxes with an area of $1° \times 1°$ were used. The locations of the sub-samples were selected a priori, according to their expected biome based on visual inspection of annual mean chlorophyll levels. For putative biomes with a very large, discontinuous geographic coverage, two different regions representative of the putative biome were sampled (e.g. for the high-intermediate biome, samples were taken in the south subtropical convergence zone, SSTZ, and the equatorial upwelling). Three replicates were taken for each region sampled.

5.2.6. Supervised classification

Supervised classification of the PC1 field was undertaken using the routine from ERDAS IMAGINE 8.6. Training fields were derived from $3° \times 3°$ sample boxes situated within relatively homogeneous areas characteristic of each class in the PC1 field, for a priori classes identified by MDS. First a signature file was generated for each training field providing the mean vector and covariance matrix. This was then used to assign pixels in the image to classes using a maximum likelihood algorithm. The separability of the classes was assessed using the Jeffries–Matusita distance (Swain & Davis, 1978).

6. Results

6.1. Initial characterisation

Fig. 1 shows the monthly mean chlorophyll concentration from SeaWiFS. This climatology shows that a consistent structure is present throughout the year. Fig. 2 shows PC1 of chlorophyll from the SeaWiFS climatology. The spatial pattern represented by this PC dominates the variance of the data set, explaining 88.1% of the total variance (Table 1). Within this pattern, distinct large-scale ocean regions of similar transformed-chlorophyll values are differentiated by distinct gradients. The strongest gradients have been superimposed as non-smoothed contours. These highlight four distinct regions: a) coastal and shelf regions (including coastal upwellings and river mouths) with very high chlorophyll concentration and rapid gradients in chlorophyll (red to yellow); b) oceanic upwellings, frontal regions and extended river plumes with moderate to high chlorophyll concentrations (yellow to green); c) low chlorophyll central ocean gyres (blue to magenta) and d) intermediate chlorophyll regions (cyan), which can be separated into high-intermediate and low-intermediate areas by the central contour line.

6.2. System properties

6.2.1. Persistence

The eigenvectors from PCA (Fig. 3) show that the temporal amplitude of the pattern in PC1 exhibits very little variability, with the pattern being strongly present throughout the year. Higher order PCs show temporal amplitudes related to different modes of intra-annual variability, but these modes each represent only a small part of the total variance (Table 1), with a combined contribution of less than 12% of the total variance. Consideration of the Guttman criterion and LEV plot for the eigenvalues suggests only the first two to four PCs carry sufficient information to be useful. Eigenvectors for PC2 and PC3 show different annular modes of variability whereas PC4 shows a biannual mode. Together, these modes represent 10.1% of the total variance. Although this is not an insignificant contribution, it is minor compared with the variance explained by PC1. Thus, although seasonal variability of chlorophyll patterns is clearly important, the dominant spatial pattern is persistent over an annual time scale. Furthermore, the highly constant eigenvector for PC1 shows that variability described by higher order PCs acts upon this underlying pattern.

Table 2
Nominal Atlantic provinces, together with their abbreviated codes, used for labeling of samples for cluster analysis

Nominal province	Code	Chl level	Longhurst province
Shelf-Patagonian	S-PAT	V. high	BRAZ/FKLD
Shelf-Northwest Atlantic	S-NWA	V. high	NWCS
Coastal Upwelling-Northwest African Upwelling	CU-NWAU	V. high	CNRY
River-Congo	R-CNG	V. high	GUIN
River-Orinoco	R-ORI	High	GUIA
River-Amazon	R-AMZ	High	GUIA
Coastal Margin-Benguela Current Upwelling	CM-BCU	V. high/high	BENG
Coastal Margin-Guiana	CM-GUI	High	GUIA
Coastal Margin-Gulf of Mexico	CM-GM	High	CARB
Coastal Margin-Northwest African Upwelling	CM-NWAU	High/H-Int	CNRY
Coastal Margin-Northwest Atlantic	CM-NWA	V. high	NWCS
Ocean-Angola Dome	O-AD	High	GUIN
Ocean-Antarctic Circumpolar Current Central	O-ACCC	High/H-Int	SSTC/SANT
Ocean-Antarctic Circumpolar Current East	O-ACCE	H-Int	SSTC/SANT
Ocean-Antarctic Circumpolar Current West	O-ACCW	High	SSTC/SANT
Ocean-Azores Current	O-AC	H-Int	NASW/E
Ocean-Bay of Biscay	O-BIS	High	NADR
Ocean-Benguela Current Upwelling	O-BCU	H-Int	SATL
Ocean-Canary Current	O-CC	H-Int	NATR
Ocean-Caribbean Central	O-CARC	L-Int	CARB
Ocean-Caribbean South	O-CARS	L-Int	CARB
Ocean-Central Equatorial Upwelling	O-CEU	High	ETRA
Ocean-Eastern Equatorial Upwelling	O-EEU	High	ETRA
Ocean-Guinea Current	O-GC	High	GUIN
Ocean-Gulf Stream	O-GS	H-Int	GFST/NASW
Ocean-North Atlantic Current	O-NAC	High	NADR
Ocean-North Atlantic Gyre	O-NAG	L-Int/low	NATR/NASW/E
Ocean-North Equatorial Counter Current	O-NECC	High/H-Int	E/WTRA
Ocean-North Equatorial Current	O-NEC	High/H-Int/L-Int	NATR/ETRA
Ocean-North Mediterranean	O-NM	H-Int	MEDI
Ocean-Northern South Equatorial Current	O-NSEC	H-Int/L-Int	E/WTRA
Ocean-South Atlantic Convergence Agulhas Retroflection	O-SACAR	High	EAFR
Ocean-South Atlantic Convergence East	O-SACE	High	SATL
Ocean-South Atlantic Convergence West	O-SACW	High	SATL
Ocean-South Atlantic Gyre	O-SAG	L-Int/low	SATL
Ocean-South Mediterranean	O-SM	L-Int	MEDI
Ocean-West Equatorial	O-WEQ	L-Int	SATL

Comparative Longhurst (1998) provinces are given for reference.

6.2.2. Hierarchy

The locations of 106 spatially-discrete samples for the Atlantic Ocean, are shown in Fig. 4A. The mean satellite chlorophyll value for each month was extracted for each of these sample locations. Cluster analysis (Fig. 5) of the normalised matrix (locations = samples, time = variables) grouped the samples into three main clusters. These three main clusters broadly correspond to *very high*, *high* and *low* chlorophyll levels. The two largest of these clusters can each be further decomposed into at least two more major groups, separating *high-intermediate* from *high* and *low-intermediate* from *low* chlorophyll levels. Only two samples did not cluster according to their identified chlorophyll level, both were identified as *high-intermediate* but clustered as *low-intermediate* and both were located on the oceanic edge of the Benguela Current and the South Atlantic subtropical gyre. All of these clusters were significant at the 99% level when analysed using the SIMPROF test in PRIMER 6.

Each of the samples was further labelled using a combination of geographic region (e.g. South Atlantic) and known physical character-istics (e.g. coastal upwelling). The names and abbreviated codes of these nominal 'provinces' are given in Table 2. The labels were superimposed on the results of hierarchical clustering. The clustering generally showed provinces with similar characteristics to group together within the higher order clusters related to chlorophyll level. For example, within the *low-intermediate* cluster, two sub-clusters generally separated samples located in the southern hemisphere from those located in the northern hemisphere/equatorial region. Thus, while the hierarchical structure shown by this analysis corresponds firstly to patterns of chlorophyll concentration (from now on termed biomes), it also reflects regional patterns of oceanographic structure (from now on termed provinces).

6.3. Objective classification of patterns observed

6.3.1. Deduction and testing of classes

The MDS 2-D ordination plot for these samples (Fig. 6A) shows a gradient along the horizontal axis from very high to low chlorophyll
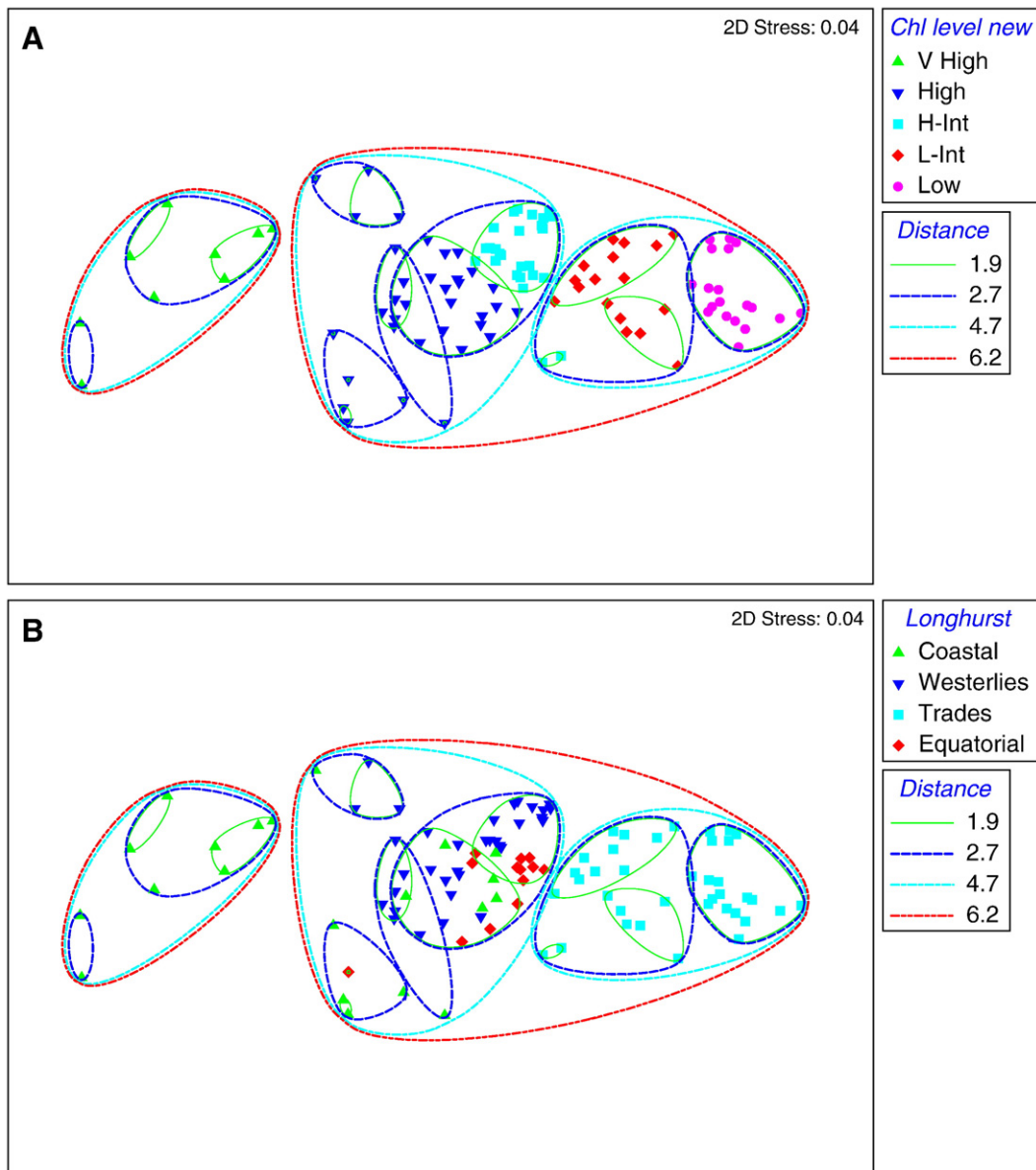

Fig. 6. MDS 2-D ordination plot of Atlantic grid sub-sample data with cluster analysis groupings superimposed. Symbols correspond to (A) chlorophyll-a levels, (B) Longhurst biomes (including new 'equatorial' biome from 'trades'); coloured lines correspond to separation distances, superimposed from cluster analysis.

Table 3
Sampling locations for putative Pacific biomes taken from a range of geographic locations; SSTCZ is the South Subtropical Convergence Zone

Biome	Location
Very high	Coast 1. Yellow Sea (124E, 37N)
	Coast 2. Hokkaido (Pacific Coast; 144E, 42N)
	Coast 3. Chile (Humboldt upwelling; 74W, 35S)
High (shelf)	Shelf 1. Sea of Japan (132E, 40N)
	Shelf 2. Pacific Ocean (Japanese shelf; 144E, 39N)
	Shelf 3. New Zealand shelf (177E, 42S)
High (ocean)	Temperate North Pacific 1 (157E, 45N)
	Temperate North Pacific 2 (167E, 45N)
	Equatorial upwelling 4 (84W, 1S)
High-intermediate (higher latitude)	SSTCZ 1. West Tasman Sea (156E, 36S)
	SSTCZ 2. East Tasman Sea (165E, 38S)
	SSTCZ 3. South Pacific (170W, 44S)
High-intermediate (low latitude)	Equatorial upwelling 1 (147W)
	Equatorial upwelling 2 (125W)
	Equatorial upwelling 3 (106W)
Low-intermediate (N. hemisphere)	Tropical North Pacific 1 (174W, 7N)
	Tropical North Pacific 2 (147W, 12N)
	Tropical North Pacific 3 (123W, 12N)
Low-intermediate (S. hemisphere)	Temperate South Pacific 1 (150W, 43S)
	Temperate South Pacific 2 (126W, 42S)
	Temperate South Pacific 3 (107W, 42S)
Low (N. hemisphere)	North Pacific Gyre 1 (157E, 21N)
	North Pacific Gyre 2 (177W, 19N)
	North Pacific Gyre 3 (130W, 23N)
Low (S. hemisphere)	South Pacific Gyre 1 (156W, 24S)
	South Pacific Gyre 2 (125W, 26S)
	South Pacific Gyre 3 (108W, 26S)

(left to right), with the five biomes identified from cluster analysis (*very high, high, high-intermediate, low-intermediate, low*) clearly separable. Repeating the analysis with random subsets of the samples showed this biome pattern to be consistent and independent of sample size. Superimposition of boundaries from cluster analysis confirms the similar grouping by both techniques: a) two top-level clusters, of which the largest is divided into two, giving the three major sub-divisions (*very high, high, low*), b) further sub-division of the high and low clusters to provide intermediate groups (*high-intermediate, low-intermediate*). It is notable that the sub-clustering within the *high* cluster into *high* and *high-intermediate* is at a very low level of separation compared to other distinct clusters within the *high* group.

To provide an independent test of these biomes, sub-samples from the Pacific Ocean were grouped based on the five regimes identified in Fig. 6A. The locations of the samples and their putative biomes are given in Fig. 4B and Table 3. Application of the ANOSIM test to these samples showed the regimes to be significantly distinct (Global $R = 0.916$, $P < 0.001$), confirming the biome classification. The lowest correlation in pairwise tests between each of the biomes was between *low-intermediate* and *low* ($R = 0.794$, $P < 0.002$, all other pairwise tests falling in the range $0.941 \leq R \leq 1$). Examination of MDS ordination and clustering for the Pacific sub-samples (Fig. 7A) explained this lower pairwise result, showing a high degree of separability within the *low* biome between the North Pacific and South Pacific gyres. This is consistent with the chlorophyll patterns shown in Fig. 2, in which the South Pacific gyre has exceptionally low chlorophyll. Therefore, a sixth *very low* biome was recognised.

6.3.2. Comparison with Longhurst's biomes

Samples in the MDS plots for Figs. 6A and 7A were also labelled according to their corresponding Longhurst's biomes for comparison (shown in Figs. 6B and 7B). Boundaries between Longhurst's biomes were identified from strong features within the chlorophyll images, e.g. frontal boundaries between the 'westerly' and 'trades' biomes. Additionally, the Mediterranean was split into northern and southern

regions, belonging to the westerly and trades biomes respectively, and an area of 'coastal' biome was identified in the Gulf of Mexico.

Some strong similarities were seen between the MDS/cluster patterns and Longhurst's biomes. For Atlantic samples, all the *very high* cluster belonged to the 'coastal' biome and all the *low* and *low-intermediate* cluster were identified with the 'trades' biome. However the *high* and *high-intermediate* clusters were dominated by the 'westerly' biome but with a mixture of samples from 'coastal' and 'trades' biomes throughout. The 'trades' samples in this cluster could be identified with the region of equatorial current systems belonging mainly to the Longhurst ETRA and WTRA provinces, so were re-labelled as 'equatorial'. This is also the case for the Pacific samples. Additionally, in the Pacific, three of the samples in the *low-intermediate* cluster corresponded to the 'westerly' biome. These were all from a reduced chlorophyll region (putative high-nutrient–low chlorophyll, HNLC) of the subtropical convergence zone in the eastern South Pacific.

6.3.3. Classification

These six chlorophyll biomes were taken as the *a priori* number of classes for an objective supervised classification of the steady-state pattern shown in PC1. The locations of sample boxes used to extract training data for this classification are shown in Fig. 4C. Two training samples were taken for the *high* biome to reflect the complex clustering and wide spread in ordination patterns for this biome, as seen in Fig. 6A. The resulting classification is shown in Fig. 8.

The overall mean chlorophyll field will, by definition, approximate the steady-state pattern represented by PC1. Therefore, to extend the classification of pelagic ecosystems to higher latitudes where missing data preclude the use of PCA, supervised classification of the mean chlorophyll field was undertaken, using training data from the same regions as the PC1 classification. This classification is shown in Fig. 9.

These classifications show the six biomes to be generally associated with distinct regions:

1. The *very high chlorophyll* biome is mainly associated with coastal and shelf waters, upwelling regions and areas of large riverine inputs.
2. The *high chlorophyll* biome is generally seen in oceanic areas with elevated chlorophyll, especially mid-latitude westerly and polar regions, the offshore edge of eastern boundary currents, equatorial upwellings and monsoon areas of the Arabian Sea and around the Indian subcontinent. It also occurs in the shelf seas of South-East Asia, in some higher latitude, seasonally-stratified shelf seas and toward the offshore edge of some coastal margins.
3. The *high-intermediate chlorophyll* biome is associated with equatorial regions outside of the upwelling zones, and areas more generally around the edge of the *high chlorophyll* biome. It is also present towards the eastern side of the Pacific, Atlantic and Indian Ocean sectors of the Southern Ocean.
4. The *low-intermediate chlorophyll* biome mainly occurs around the edges of the central tropical gyres of the Atlantic, Indian Ocean and Pacific Oceans, but also in the Caribbean, southern Mediterranean and in putative high-nutrient, low chlorophyll regions of the Southern Ocean (e.g. Drake's Passage).
5. The *low chlorophyll* biome occurs in the central tropical gyres of the Atlantic Ocean, Indian Ocean and North Pacific, around the edge of the South Pacific gyre and in a small patch of the Coral Sea, off the NE coast of Australia.
6. The *very low chlorophyll* biome occurs only in the central tropical Pacific gyre and a small area of the North Pacific tropical gyre.

Fig. 10 shows mean satellite chlorophyll levels throughout the year for each biome. In each month, the mean chlorophyll levels for each biome are distinct but their standard deviations sometimes overlap. This reflects changes in the position of boundaries between the
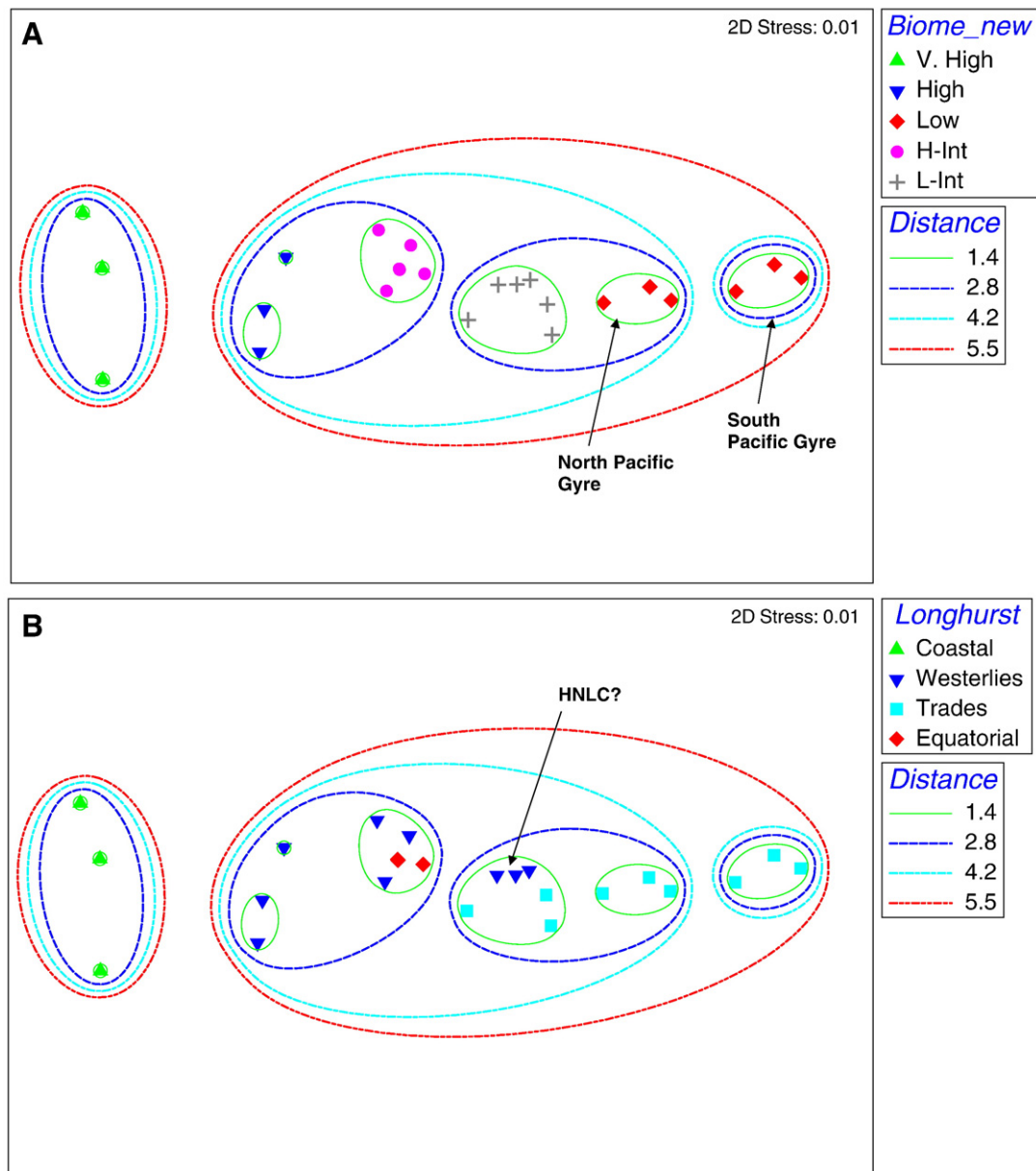
Fig. 7. MDS 2-D ordination plot of Pacific biome sub-sample data with cluster analysis groupings superimposed. Symbols correspond to (A) chlorophyll-a levels, (B) Longhurst biomes (including new 'equatorial' biome from 'trades'); coloured lines correspond to separation distances, superimposed from cluster analysis. In (A), the North and South Pacific Gyres are labelled to highlight the strong separation between the South Pacific Gyre and other *low* (North Pacific Gyre) and *low-intermediate* samples. In (B), putative HNLC samples are labelled.

biomes in different months that are not taken account of by the steady-state classification method.

7. Discussion

In this study we investigated the ecological partition of the ocean; specifically, the hypothesis that the global picture of ecological pattern in the ocean, provided by visible spectrum (ocean colour) satellite sensors, enables us to describe the spatial extent of macro-scale ecological systems. To test this hypothesis we have examined two properties of thermodynamically-open systems: persistence and hierarchy. These properties were investigated using the SeaWiFS global chlorophyll product over a mean annual cycle, the dominant mode of temporal variability over the time series.

Spatial PCA showed that the majority of the spatio-temporal variance was explained by an overwhelmingly dominant spatial

structure with a near-constant presence throughout the year. Therefore, the patterns observed in oceanic chlorophyll clearly support the view that a strong spatial structure is persistent throughout the annual cycle. Cluster analysis of 106 spatially-discrete samples from the Atlantic Ocean showed the samples to cluster hierarchically, first according to chlorophyll concentration and then, in most cases, according to known geographic (e.g. northern or southern hemisphere) or oceanographic characteristics (e.g. upwelling region). Thus, there is a clear hierarchical-link from regional to global scales consistent with known biological and physical characteristics. As both of these system properties have been clearly demonstrated to hold for the oceanic chlorophyll fields investigated, we assert that the macro-scale structures identified in this data set represent the spatial extent of pelagic ecological systems and are suitable for providing the basis of an objective classification scheme for ecological ocean provinces. However, care needs to be taken when interpreting some
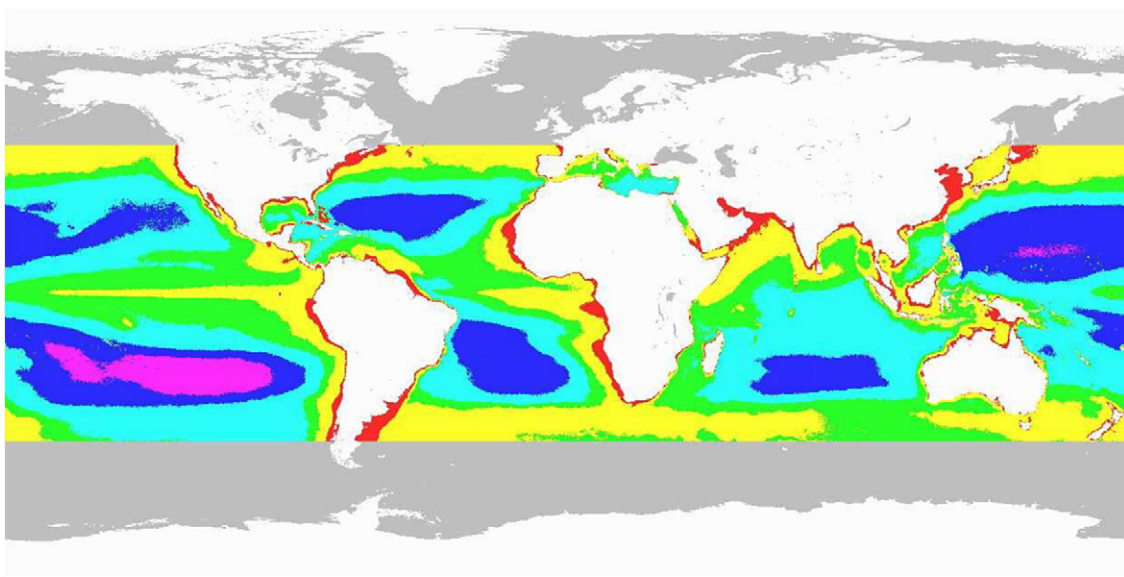
Fig. 8. Classification of biomes from PC1 of chlorophyll-a: red=*very high*, yellow=*high*, green=*high-intermediate*, cyan=*low-intermediate*, blue=*low*, magenta=*very low*.

of the high-latitude patterns; they are only representative of a few months observation per year. Furthermore, patterns in the Arctic may be influenced by strong sediment loads from glacial run-off.

Classification of ocean ecosystems according to their chlorophyll concentration is not new. The classical division into oligotrophic, mesotrophic and eutrophic waters follows this concept. However, the divisions between these categories are subjectively defined along a continuum (nominally <0.1 mg m^{-3}, 0.1–1 mg m^{-3}, >1 mg m^{-3}, respectively; Antoine et al., 1996). The classification scheme developed here has identified three top-level classes (*very high*, *high* and *low*) that broadly fit this schema, with sub-divisions into six biomes. However, the divisions between biomes have arisen as emergent properties from the data themselves. The classification is defined by the spatial distribution and variability of chlorophyll concentrations rather than by predefined bands of fixed concentration. The categories are still distributed along a continuum of chlorophyll values but the boundaries are determined by natural horizontal gradients. Indeed, the consistent ecological patterns that have been used to define these

biomes are clearly related to underlying physical processes, which vary between regions and through time (e.g. upwelling, river run-off, solar insolation, wind mixing). Therefore, chlorophyll concentrations within a region may vary significantly over seasonal and interannual time scales, given a change in the intensity of the underlying physical dynamics, but the stable biome pattern could remain if the spatial pattern of physical forcing was unchanged.

Comparison of this chlorophyll-based classification with Longhurst's (1998) physically-forced biomes show general consistency only if a 5th 'equatorial' biome is introduced, but there is not a direct match, particularly for the coastal biome and putative HNLC regions. This is to be expected because coastal regions are highly heterogeneous and show variability in physical forcing and ecological response at finer spatial scales than described by Longhurst's bulk classification of these regions. HNLC regions, by definition, will have lower chlorophyll than expected from known physical forcing and associated macro-nutrient distributions. The consistency between the two approaches is encouraging, suggesting that a general understanding of forcing processes can broadly
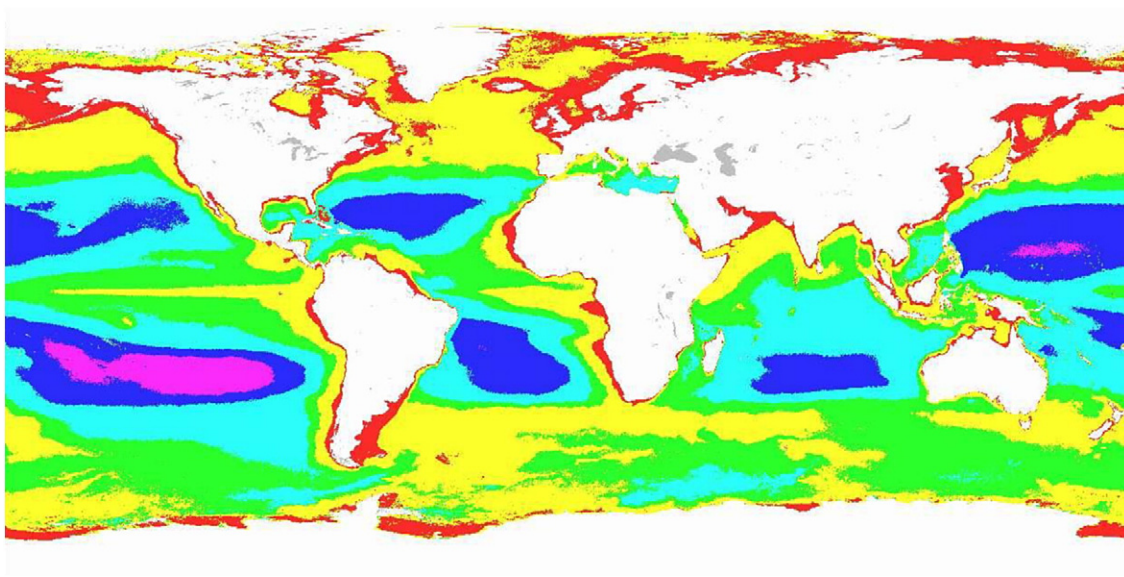

Fig. 9. Classification of biomes from mean chlorophyll-a: red=*very high*, yellow=*high*, green=*high-intermediate*, cyan=*low-intermediate*, blue=*low*, magenta=*very low*.

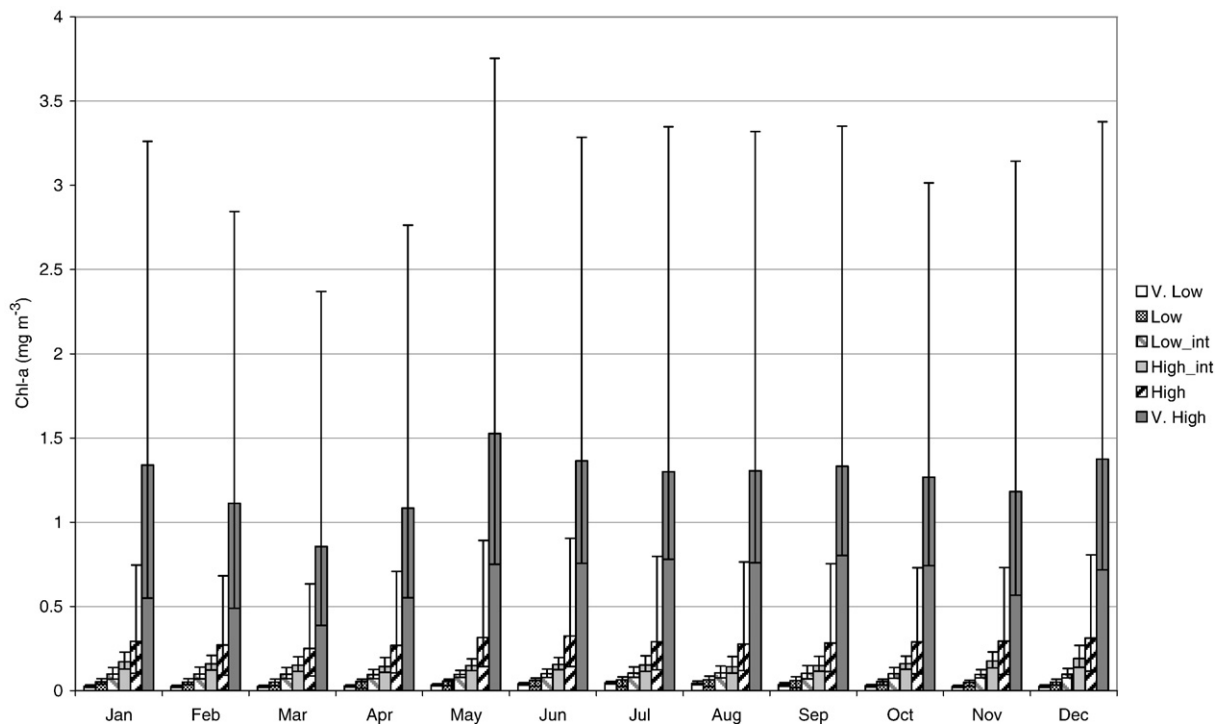**Fig. 10.** Plot of mean chlorophyll-a±1 standard deviation, for each biome for each month.

describe empirically-derived patterns. We suggest future modifications of the Longhurst biome classification take account of the different oceanic response to trade-wind forcing in the equatorial regions compared with the gyre regions. Further empirical investigation, comparing patterns of physical ocean processes from satellite (SST or altimeter) data with the chlorophyll-derived patterns from this study may also help elucidate the spatial relationship between physical forcing and ecological response for some regions.

The comparison of these chlorophyll-based biomes and provinces with biogeographic patterns in the phytoplankton is beyond the scope of this initial study but is worthy of further investigation. Current knowledge of phytoplankton species distributions is insufficient globally for a province-by-province analysis. However, the use of diagnostic pigments (Vidussi et al., 2001; Uitz et al., 2006) to define phytoplankton groups, aggregated at a higher taxonomic level (phytoplankton type, e.g. diatom, dinoflagellate, flagellate, pico-eukaryote, prokaryote) or by size class (pico, nano, micro), allows for some biogeographic patterns to be identified. Aiken, Pradhan, et al. (in press) analysed 10 years of high-performance liquid chromatography (HPLC) derived pigment data from the Atlantic Meridional Transect (AMT) cruises (1995–2005). These showed picoplankton to dominate in the oligotrophic, low chlorophyll, tropical gyres, whereas micro-plankton dominated in the eutrophic, highest chlorophyll regions (e.g. upwellings) and nanoplankton in intermediate regions. Several other studies have shown a similar relationship between chlorophyll-a concentration, environmental conditions and phytoplankton size class or type (reviewed by Aiken, Hardman-Mountford, et al., 2008).

Although this investigation has focused on identifying steady-state ecological structures, we do not deny the importance of temporal variability in the ecosystems identified. The fluid nature of the ocean environment results in large temporal changes in the spatial distribution of chlorophyll and this is greatest over seasonal time scales (at least within a secular time frame). The PCA results show that this seasonal variability explains only a minor part of the total variance and acts upon the dominant, persistent, underlying spatial structure. Nonetheless, examination of the seasonal maps of chlorophyll distribution shows that the spatial extent of these biomes changes

throughout the year, marginally in some cases and more significantly in others. Chlorophyll levels for each biome show some overlap in their ranges throughout the year, again reflecting these changes in spatial extent. Thus, there is also a clear need for instantaneous realisations of such classifications that take account of spatio-temporal variability in the extent of regional ecosystems. Devred et al. (2007) have developed such a method for the North West Atlantic coastal, shelf, slope and ocean regions. However, the formulation of an instantaneous global ecosystem classification scheme is beyond the scope of our investigation.

Similarly, while this study has focused on pattern at the macro-scale, we do not underestimate the importance of variability at the micro-scale. Rather, the demonstration of hierarchical patterns scaling regional chlorophyll variability into global categories reinforces the view that micro-scale variability significantly affects system function-ing at the macro-level. However, macro-scale ecosystem descriptions are of key importance for understanding the Earth system. Important changes at the micro-scale will be reflected at the macro-scale as changes in stability or other inherent properties of the system. Investigation at the finer scale may then be required to elucidate this change. This has important implications for the monitoring of ecosystems facing climate stresses, possibly allowing early identifica-tion of those most vulnerable.

In conclusion, this study has shown that macro-scale ecological patterns in ocean chlorophyll exhibit properties consistent with their being characterised as ecological systems. From these results, a classification scheme has been developed that uses these patterns to objectively define top-level biomes and their regional expression as provinces, synonymous with ecosystems, for the pelagic marine environment.

Acknowledgements

References

Aiken, J., Fishwick, J. R., Lavender, S. J., Barlow, R., Moore, G. F., Sessions, H., et al. (2007). Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October, 2002: Preliminary validation and new products for phytoplankton functional types and photosynthetic parameters. *International Journal of Remote Sensing*, *28*, 497–516.

Aiken, J., Hardman-Mountford, N. J., Barlow, R., Fishwick, J., Hirata, T., & Smyth, T. (2008). Functional links between bioenergetics and bio-optical traits of phytoplankton taxonomic groups; an overarching hypothesis with applications for ocean colour remote sensing. *Journal of Plankton Research*, *30*, 165–181.

Aiken, J., Pradhan, Y., Barlow, R. G., Lavender, S., Poulton, A., Holligan, P. M., et al. (in press). Phytoplankton pigments and functional types in the Atlantic Ocean: a decadal assessment, 1995–2005. *Deep Sea Research II*.

Antoine, D., André, J. -M., & Morel, A. (1996). Oceanic primary production 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Global Biogeochemical Cycles*, *10*, 57–69.

Bailey, R. G. (1998). *Ecoregions: the ecosystem geography on the oceans and continents.* New York: Springer Verlag 176 pp.

Campbell, N. A. (1996). *Biology*, 4th Edition Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc.

Clarke, K. R., & Warwick, R. M. (2001). *Changes in marine communities: an approach to statistical analysis and interpretation*, 2nd edition Plymouth, UK: PRIMER-E.

Craddock, J. M., & Flood, C. R. (1969). Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere. *Quarterly Journal of the Royal Meteorological Society*, *95*, 576–593.

Cullen, J. J., Franks, P. J. S., Karl, D. M., & Longhurst, A. (2002). Physical influences on marine ecosystem dynamics. In A. R. Robins on, J. J. McCarthy, & B. J. Rothschild (Eds.), *The Sea, Vol. 12* (pp. 297–336). New York: John Wiley and Sons.

Devred, E., Sathyendranath, S., & Platt, T. (2007). Delineation of ecological provinces using ocean colour radiometry. *Marine Ecology Progress Series*, *346*, 1–13.

Ducklow, H. W. (2003). Biogeochemical provinces: Towards a JGOFS synthesis. In M. J. R. Fasham (Ed.), *Ocean biogeochemistry: A synthesis of the Joint Global Ocean Flux Study (JGOFS)* (pp. 3–17). Berlin Heidelberg: Springer-Verlag.

Fishwick, J. R., Aiken, J. A., Barlow, R., Sessions, H., Bernard, S., & Ras, J. (2006). Relationships between phytoplankton pigments, photosynthetic parameters and optical properties of the southern Benguela ecosystem. *Journal of the Marine Biological Association of the UK*, *86*, 1267–1280.

Fung, T., & LeDrew, E. (1987). Application of principal components analysis to change detection. *Photogrammetric Engineering and Remote Sensing*, *53*, 1649–1658.

Grimm, V., & Wissel, C. (1997). Babel, or the ecological stability discussions: An inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia*, *109*, 323–334.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*, 149–161.

Hutchinson, G. E. (1953). The concept of pattern in ecology. *Proceedings of the Academy of Natural Sciences, Philadelphia*, *105*, 1–12.

Kendall, M. G., & Stuart, A. (1966). *The advanced theory of statistics (vol. 3).* London: Charles Griffin and Company Ltd. 552 pp.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.

Longhurst, A. (1995). Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, *36*, 77–167.

Longhurst, A. (1998). *Ecological Geography of the Sea.* San Diego: Academic Press 398 pp.

Longhurst, A. R., Sathyendranath, S., Platt, T., & Caverhill, C. M. (1995). An estimate of global primary production in the oceans from satellite radiometer data. *Journal of Plankton Research*, *17*, 1245–1271.

Munk, W. (1997). Sampling the oceans from above and beneath. *IAPSO President's invited lecture. IAMAS/IAPSO Joint Assembly, 1997. Melbourne, June 1997* http://www.olympus.net/IAPSO/JointAssembly97/munk/munk.html (accessed 29/01/2008).

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, *51*, 933–938.

Platt, T., & Sathyendranath, S. (1999). Spatial structure of pelagic ecosystems in the global ocean. *Ecosystems*, *2*, 384–394.

Preisendorfer, R. W. (1988). *Principal component analysis in Meteorology and Oceanography.* Amsterdam: Elsevier 425 pp.

Sathyendranath, S., Longhurst, A. R., Caverhill, C. M., & Platt, T. (1995). Regionally and seasonally differentiated primary production in the North Atlantic. *Deep Sea Research*, *42*, 1773–1802.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, *27*, 125–140.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*, 467–482.

Swain, P. H., & Davis, S. M. (1978). *Remote Sensing: The Quantitative Approach.* New York: McGraw Hill Book Company.

Udvardy, M. D. F. (1975). A classification of the biogeographical provinces of the world. *Prepared as a Contribution to UNESCO's Man and the Biosphere Programme Project No.8. IUCN Occasional Paper No. 18* Morges, Switzerland: International Union for Conservation of Nature and Natural Resources.

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research — Oceans*, *111*, C08005. doi:10.1029/2005JC003207

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & Marty, J. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research — Oceans*, *106* (C9), 19939–19956.

Wu, J., & Loucks, O. L. (1995). From balance of nature to hierarchical patch dynamics: A paradigm shift in ecology. *The Quarterly Review of Biology*, *70*, 439–466.