Identifying characteristic chlorophyll a profiles in the coastal domain using an artificial neural network

A. J. Richardson1,2,*, M. C. Paff1,2, J. G. Field2, N. F. Silulwane3 and F. A. Shillington4

1OCEANOGRAPHY DEPARTMENT AND 2ZOOLOGY DEPARTMENT, UNIVERSITY OF CAPE TOWN, RONDEBOSCH 7701, CAPE TOWN, SOUTH AFRICA
3PRESENT ADDRESS: THE SIR ALISTER HARDY FOUNDATION FOR OCEAN SCIENCE, THE LABORATORY, CITADEL HILL, PLYMOUTH PL1 2PB, UK
*CORRESPONDING AUTHOR: anr@mail.pml.ac.uk

To estimate primary production in the marine environment, knowledge of the vertical distribution of phytoplankton biomass is needed. The measurement of ocean colour by satellite remote sensing makes it possible to map the near-surface phytoplankton distribution, although the subsurface vertical structure cannot be measured. In this study, we investigated the shape of vertical chlorophyll profiles from the Benguela upwelling system seasonally and with respect to environmental variables such as surface temperature and chlorophyll, and mixed layer and water column depth. We used an artificial neural network technique called a self-organizing map to identify characteristic classes of vertical chlorophyll profiles, and then classify existing profiles into these representative classes. The self-organizing map identified a continuum of patterns, ranging from those with small deep peaks (~1 mg m–3, 45 m) to those with large near-surface peaks (~9 mg m–3, 2 m). Although profile shape varied seasonally, profiles were very variable within each season, making chlorophyll profiles averaged seasonally meaningless. A canonical succession in profile shape following upwelling of cold water in spring and summer could be identified, with large surface peaks in cool water and small deep peaks in warm water. The approach presented here can be used in a semi-quantitative manner to predict the subsurface chlorophyll field from known (water column depth) or easily measured variables from satellites (surface temperature or surface chlorophyll), as the relative frequency of each characteristic profile under different environmental conditions is presented. This approach enables prediction of profile shapes in the dynamic coastal domain and thus superior regional estimates of primary production.

© Oxford University Press 2002
primary production, a typical seasonal Chl profile was then derived for each biogeochemical province. However, the assumption of a typical seasonal profile is problematic in the coastal domain because of the dynamic nature of coastal oceanography and its impact on profile shape. This is especially true in highly productive, wind-driven upwelling areas, where the extremely variable hydrology on all spatial and temporal scales leads to considerable variability in vertical phytoplankton distribution (Brown and Hutchings, 1982a; Bailey and Chapman, 1991; Pitcher et al., 1992; Mitchell-Innes et al., 1999). In such active coastal regions, a dynamic biogeochemical approach is preferable. Thus, within a province, sub-provinces with elastic boundaries can be estimated from satellite imagery (Sathyendranath et al., 1991; Watts et al., 1999), an approach similar to that used by Morel and Berthon (Morel and Berthon, 1989) for relating profile shape to Chl concentration in the surface layer. This improved approach, however, still uses a single profile within this sub-province, whereas a suite of characteristic profiles may be more realistic.

In this study, we have used an artificial neural network (ANN) technique to identify a suite of characteristic profiles for a set of dynamic sub-provinces. ANNs are computer algorithms that simulate the information-processing abilities of the brain by mimicking its basic structure (Dayhoff, 1990). They consist of a network of interconnected simple processing units or nodes that process information in parallel, rather than sequentially. A feature of biological neural networks is that they have the ability to learn, rather than being pre-programmed. This ability has made ANNs valuable applied mathematical tools (Hewitson and Craine, 1994). Moreover, ANNs have a number of advantages over traditional statistical methods (Wasserman, 1989). First, they can solve non-linear problems of almost infinite complexity (Dayhoff, 1990). Second, they are more robust in handling noisy and missing data than traditional methods. Finally, they do not require prior knowledge and assumptions about the data, such as normality or equality of variances (Chen and Ware, 1999).

There are many different types of ANN, each having its own architecture and being suited to a specific application. A common ANN used for pattern recognition and classification is the self-organizing map [SOM; Kohonen, 1984, 1997]. A SOM identifies patterns in the input data, based on the underlying structure of the data (Dayhoff, 1990). The SOM is a dimension-reducing procedure whereby a multidimensional input is mapped onto a lower (usually two) dimensional continuous output space. The output space consists of a number of patterns characteristic of the data, with similar patterns neighbouring and dissimilar patterns further apart. The SOM technique can be particularly useful when dealing with dynamic processes that produce a continuum of patterns, the situation we might expect when identifying characteristic vertical Chl patterns within an upwelling region. As the output patterns resemble the input format, they are often more easily interpreted than output from conventional multivariate techniques.

In this study, we identify characteristic Chl profiles from the western Agulhas Bank region of the southern Benguela upwelling system. This broad shelf region is an important spawning area for commercially exploited pelagic fish such as anchovy and sardine (Barange et al., 1999). Estimates of primary production are important for understanding recruitment fluctuations of these fish. We intend to show that Chl profiles averaged seasonally are insufficient to capture the variability in profile shape in a dynamic upwelling region. The SOM is then used to identify a suite of characteristic Chl profiles. The probability of occurrence of these characteristic Chl profiles under different environmental conditions (dynamic sub-provinces) is then described.

METHOD

Data collection

Oceanographic data were collected monthly (August–March) in 1993/1994 and 1994/1995 along two transects off Quoin Point and Walker Bay on the western Agulhas Bank as part of the South African Sardine and Anchovy Recruitment Programme (Figure 1). At each of the six stations along the two transects, continuous fluorescence profiles were collected to 100 m depth or to within 10 m of the seabed. To convert the fluorescence data into Chl equivalents, in situ fluorescence readings were related to extracted Chl concentrations collected in Nikkin bottles at the surface and the depth of maximum fluorescence. These water samples were filtered using Whatman GF/F filters and then extracted in 90% acetone at 20°C for 24 h. Samples were then measured fluorometrically before and after addition of hydrochloric acid to adjust for the phaeopigments using a Turner Designs Model 10-000R fluorometer.

In situ fluorescence readings were then related to extracted Chl concentrations by linear regression analysis of log-transformed Chl values against fluorescence. Low fluorescence values measured by the profiling fluorometer corresponding to high extracted Chl values were considered photoinhibited and were excluded from the calibration regression. The corresponding profiles were also discarded from all subsequent analyses, leaving a total of 184 profiles.

At each station, pertinent environmental variables were
collected, including sea surface temperature (11.4–22.4°C), surface Chl $a$ concentration (0.1–22.8 mg m$^{-3}$), depth of the water column (22–1058 m) and depth of the upper mixed layer (2–101 m). The upper mixed layer was defined as the depth where the difference in temperature from the surface is 0.5°C (Longhurst et al., 1995).

Parameterization of Chl $a$ profiles

Direct comparison of Chl profiles was difficult because inshore and offshore profiles varied greatly in depth, and the continuous fluorescence readings of a profile were not taken at standard depths because of differing cable speeds, ship roll and ocean currents during profile measurement. To classify profiles, however, requires that the input data have a consistent format so that they are comparable. Although the profiles could be standardized by interpolation of fluorescence values to standard depths and then adjusted for water column depth, a simpler approach was taken.

Chlorophyll profiles were parameterized by the shifted Gaussian model (Platt and Sathyendranath, 1995; Sathyendranath et al., 1995). This approach has the advantage of being relatively easy incorporated into primary production models. The four-parameter model can be expressed as:

$$B(z) = B_0 + \frac{h}{\sigma \sqrt{2\pi}} \left(\frac{z - z_m}{\sigma}\right)^2 e^{-\frac{(z - z_m)^2}{2\sigma^2}}$$

where $B(z)$ is Chl biomass as a function of depth (mg m$^{-3}$), $B_0$ is background Chl concentration (mg m$^{-3}$), $h$ is the total Chl within the peak (mg m$^{-2}$), $\sigma$ is the standard deviation around the peak (m) and $z_m$ is the depth of the Chl peak (m). The effect on the shape of the Chl profile of changing each of these parameters is shown in Figure 2.

The shifted Gaussian equation was fitted using the quasi-Newton algorithm. The parameters $B_0$, $h$ and $\sigma$ were constrained to be positive, as only positive values are valid (Platt and Sathyendranath, 1995). Note that $z_m$ was also constrained to be positive, although there is theoretically no such restriction, as negative values of $z_m$ led to unrealistically large values of $h$ (sometimes >10 000). By constraining $z_m$ to be positive, the $r^2$ values of the curves were virtually unaffected (<1%), and for the profiles with very large $h$, the new $z_m$ values were very close to zero (<0.1 m).

Profiles that were obviously polymodal or had an $r^2$ < 0.75 were excluded from the analysis. Profiles were also excluded that had Chl spikes caused by gelatinous masses of the colonial diatom Thalassiosira sp. (Mitchell-Innes et al., 1999). These profiles were common in winter and were not adequately represented by the shifted Gaussian model.
Of a total of 184 profiles, 29 (15.7%) did not fit the model and were discarded.

Application of the SOM

The four parameters from the fitting of the shifted Gaussian curve were used as input into the SOM. Input data consisted of a matrix of four parameters (variables) by 155 profiles (cases). Each parameter was standardized by subtracting the mean and dividing by the standard deviation to give equal weighting in the analysis to each model parameter. The SOM was performed using the SOM_Pak software Version 3.1 (Kohonen et al., 1992).
1995), which is produced by, and freely available from, the Neural Network Research Centre at the Helsinki University of Technology (http://www.cis.hut.fi/research/som_lvq_pak.shtml).

The general SOM procedure is briefly outlined below; for more details, consult Kohonen (Kohonen, 1997) or Hewitson and Crane (Hewitson and Crane, 2002).

1. Define shape and dimensions of output map

The shape or topology of the output map can be chosen to be either rectangular or hexagonal. The rectangular topology places adjacent nodes on the corners of a square, whereas the hexagonal topology places adjacent nodes at the apices of a hexagon. The dimensions of the output map determine the number of output patterns. Too many nodes do not adequately reduce the data to characteristic patterns, whereas too few nodes do not allow differentiation of underlying patterns.

2. Initialization

Prior to the training process, connection weights for each node need to be initialized (these are the initial values of the four Gaussian parameters in our analysis). These can either be chosen to be random values, or they can be initialized with the first two orthogonal components of a principal components analysis of the input data.

3. Training

Input data are presented to the SOM sequentially in a training cycle. The first input vector (parameters of the first profile in our analysis) is compared (using Euclidean distance) with the vector on each node on the output map. The node that is most similar to the profile (smallest Euclidean distance) is declared the 'winning' node and the centre of the ‘update neighbourhood’. Weights for all nodes that are topologically close (within the update neighbourhood) will learn from the same input and their weights will be adjusted by a spatial decay function to be similar (although not identical) to that input vector. The spatial decay function can be either bubble (hat shaped) or Gaussian (bell shaped). With the bubble function, the winner and surrounding nodes within the update neighbourhood are adjusted to the same extent. The Gaussian function updates the winning node and the surrounding nodes according to a Gaussian function, with the degree of update decreasing with distance from the winning node. The radius of the update neighbourhood determines the spatial extent of the update function. The update neighbourhood creates a relationship between neighbouring nodes, which results in a continuum of patterns across the node space. The training procedure is repeated for all input data.

4. Convergence

Repeat the training cycle in step 3 until convergence. This typically involves a large number of cycles (10⁵–10⁶). During each training cycle, the learning rate parameter (\(\eta\)) determines the extent that weights are adjusted. Both the learning rate and the update radius decrease throughout the entire training process. This aids convergence by establishing rough patterns early in the training process, which improve in detail later in training. Convergence is achieved when the overall error (the sum of the squared Euclidean distances between all input data and the pattern they are mapped to) is a minimum. The error is dimensionless after standardization of input data.

5. Classification

Once the underlying patterns (classes) have been identified, the SOM can be used to classify the input data (profiles in our analysis) into these classes. For each input vector there is also an error value, which is the Euclidean distance between the input vector and the pattern in the SOM output map it is mapped to. Thus, large errors identify input data that are not well represented by the output map.

In the current analysis, we chose a rectangular topology with five columns by three rows (5 \(\times\) 3), giving 15 patterns. Both smaller (3 \(\times\) 2) and larger (5 \(\times\) 4) dimension maps were explored. The 5 \(\times\) 3 map was chosen because it was necessary to subdivide the 155 profiles into those collected at different levels of environmental variables (described in the next section). The question of choosing the number of nodes is examined more fully in the Discussion.

The SOM was initialized by performing a principal components analysis on the input data, which places axes of most variation along the diagonals on the SOM output map. The SOM was trained with different values of the learning rate, initial update radius and the number of cycles to determine the best overall map (i.e. the minimum overall error and thus the best fit to the input data). Parameter values for the best map were a learning rate of 0.2, an initial update radius of 2 and 100 000 cycles. Frequencies of profiles mapped to each node (class) were visualized by using maps of relative frequencies (all frequencies on a map sum to 100%). These frequency maps have the dimensions of the output map of the SOM, and depict frequencies as grey levels for each output profile pattern.

We refer readers interested in assessing either the adequacy of the Gaussian curves for describing Chl profiles or the capability of the SOM to identify characteristic vertical Chl patterns to the paper by Silulwane et al. (Silulwane et al., 2001). In their pilot study, a small dataset is presented with all the raw profiles and the
corresponding SOM output map: this was not possible in this study because of the large number of profiles.

Relating SOM patterns to season and environmental variables

To assess the seasonal variability of the vertical distribution of Chl, input data were separated into three seasons, i.e. winter (August/September), spring (October–December), and summer (January–March), based on hydrographic conditions and upwelling activity (Richardson et al., 1998; Mitchell-Innes et al., 1999). In winter, strong westerly winds dominate and there is deep mixing of the water column, which has a uniform temperature of ~14°C over the shelf. In spring, the mixed layer depth shallows and the surface water warms to 16–19°C. Despite frequent upwelling-favourable south-easterly winds, surface upwelling is relatively infrequent, as the mixed layer is still too deep to permit upwelling to outcrop at the surface. By summer, surface shelf water has warmed to >20°C and the mixed layer has shallowed further. Frequent south-easterly winds cause intense surface upwelling of cold water inshore. A more detailed description of the hydrography of the region can be found in Boyd et al. (Boyd et al., 1985), Largier et al. (Largier et al., 1992), Probyn et al. (Probyn et al., 1994) and Boyd and Shillington (Boyd and Shillington, 1994).

To investigate the influence of environmental variables on the shape of Chl a profiles, maps of relative frequencies for different ranges of environmental conditions were analysed. Environmental variables considered were sea surface temperature, surface Chl a concentration, depth of the upper mixed layer and the depth of the water column. These variables were chosen because they can be estimated from satellites (surface and surface Chl), are known (water column depth), or may be important to profile shape (upper mixed layer depth). Three categories of values (low, medium and high) were chosen for each environmental variable to have roughly equal numbers of profiles in each.

RESULTS

Seasonally averaged Chl profiles

Profiles averaged seasonally, together with their parameter estimates, are shown in Figure 3. All profiles had peaks between 15 and 25 m depth. The average winter profile has a broad, moderately high near-surface peak. The mean summer profile is very different from that of winter or spring, having a large and relatively narrow peak.

Inspecting frequency histograms of the model parameters within each season (Figure 4) assessed the adequacy of these seasonally averaged profiles for describing the seasonal profile shape. The considerable spread of parameter values within each season suggests that profiles averaged seasonally are inadequate. For example, the depth of the peak ($z_m$) varied from 0 to >55 m depth in spring and summer. Moreover, the spread of the peak, $H$, also varied substantially, from 10 to 25 m in spring (categories with a relative frequency >10%). This changes the profile from one having a very sharp peak to one that is almost uniform throughout the water column (Figure 2c).

Another reason for the inadequacy of seasonally averaged profiles is that the frequency distributions are not all unimodal. For example, there are two types of profiles in summer: those with shallow peaks <15 m and those with deeper peaks >30 m. Thus, despite very few profiles (3%) having peaks between 15 and 30 m depth, the mean summer profile has its peak at 21 m (Figure 3).

Characteristic Chl profiles from the SOM

The SOM technique produces a more intuitively realistic description of profile shapes than profiles averaged seasonally. The SOM analysis produces a continuum of Chl patterns (profiles), representing the range of profile types in the input data and displays them in a two-dimensional output map (Figure 5). The notation describing the position of the output patterns is ($x$, $y$), with the origin at the top left corner of the output map. The
By contrast, the background Chl concentration, \( \rho \), generally increased from the bottom to the top of the output map. The depth of the peak, \( z \), increases along the diagonal, from ~7 m at pattern (4,2) to ~27 m at pattern (0,0). By contrast, the background Chl concentration, \( \rho \), generally increased from the bottom to the top of the output map.

**Classification of Chl profiles**

To obtain the frequency of occurrence of the different patterns (classes) identified, a relative frequency map of the SOM output map was produced (Figure 6). This map can be visualized as being superimposed on the SOM output map, so that the co-ordinates of the relative frequency map correspond to those of the output map. The most common pattern in the data was pattern (0,0), representing 11.6% of the profiles and having the smallest and deepest peak. Two other patterns had relative frequencies >10%: pattern (0,5), representing 11.0% of the profiles, has a low Chl concentration evenly distributed throughout the water column; pattern (3,1), representing 10.3% of the profiles, has a moderate (4 mg m\(^{-3}\)) near-surface peak. Remaining patterns were less common; pattern (4,1), representing only 2.6% of the profiles, has very high near-surface Chl distribution. Importantly, all patterns identified in the output map represent several profiles (non-zero relative frequency), so that there seems to be a continuum of profile shape rather than only a few discrete classes.

The error associated with a profile mapping to a particular pattern (i.e., belonging to a particular class) indicates how well this pattern represents that profile. Error distributions for each pattern (Figure 7) show that the majority have relatively low errors (<1), suggesting that the patterns fit most of the profiles adequately.

---

**Fig. 4.** Relative frequency histograms of (a) \( R_0 \), (b) \( h \), (c) \( \sigma \) and (d) \( z_m \) for winter (a = 10), spring (a = 10) and summer (a = 57) profiles collected on the western Agulhas Bank. Note that the y-axis scale for \( h \) in spring is different than for all other histograms.
Exceptionally large errors associated with patterns (4,0) and (4,2), which have high surface Chl peaks, suggest that some of the profiles that mapped to these patterns were not well represented by these patterns. The two outliers mapping to pattern (4,0) were the profiles with the highest background Chl concentrations in the data ($B_0 = 1.93$ and 1.99 mg m$^{-3}$, compared with the $B_0 = 0.71$ for this pattern, see Figure 5). Similarly, the outlier that mapped to pattern (4,2) has an extremely large area underneath the curve ($h = 948.14$ mg m$^{-2}$, compared with the $h = 256.3$ mg m$^{-2}$ for this pattern, see Figure 5). A powerful aspect of the SOM is that by analysing the errors for each profile, outliers can be easily identified.

Seasonal variability in profiles

The seasonal variability in profile shape was identified by investigating the frequency of occurrence within winter, spring and summer of the different patterns from the SOM output map. Marked seasonal differences in the frequency of occurrence of these Chl patterns were observed.

Winter

Only 18 out of the 154 profiles were from the winter period. The most common patterns were in the bottom left corner of the output map (Figure 8a), with 38.9% of the profiles mapping to pattern (0,2) and 22.2% mapping to pattern (1,2). These patterns have low ($\lesssim 1$ mg m$^{-3}$) Chl concentrations distributed fairly evenly throughout the water column. Profile peaks were broader than in other seasons. There were three exceptions, however, which are on the right side of the output map and have high near-surface peaks.

Spring

Spring (57 profiles) had a greater variety of patterns than summer and winter, with all patterns occurring, except those with the largest surface blooms (Figure 8b). The majority of patterns [38.9%, patterns (0,0), (0,1), (0,2), (1,0), (1,1), (1,2)] were characterized by small relatively deep peaks (<1.5 mg m$^{-3}$, >25 m). Patterns with moderate (1.5–5 mg m$^{-3}$) Chl peaks at intermediate (10–22 m) depths were also relatively common [37.6%, patterns (2,0), (2,1), (2,2), (3,0), (3,1)]. Only 3.8% of the profiles mapped to patterns with high (>5 mg m$^{-3}$) surface (<10 m) Chl [patterns (3,2), (4,0)].
small (< 1 mg m\(^{-3}\)) deep subsurface peaks \([\text{patterns (0,0)}, (1,0)]\) and 38.6\% of the profiles having more pronounced (>4 mg m\(^{-3}\)) near-surface peaks \([\text{patterns (3,1), (3,2), (4,2)}]\). No profiles had broad deep Chl maxima. In summer, there is a more defined (small \(\sigma\)) subsurface peak than in spring or winter, which both have some profiles that are almost uniform throughout the water column.

Relating seasonal patterns to environmental variability

There were not only marked changes in profile shape with each level of the environmental variables in question, but the evolution of profile shapes was different in spring and summer. Because of the very small sample size in winter, it was not possible to investigate the possible influence of environmental variables in this season.

Spring

There is a general trend of having smaller and deeper Chl peaks as water warms (Figure 9a). There is, however, considerable variation in the profile shapes, especially at intermediate temperatures. The same general trend occurs as mixed layer and water column depth increases, with a migration of the most frequent patterns in the output map from the top right to the bottom left. This shows a change from the reasonably large but narrow \(\sigma = 6.6\) near-surface peaks to small broad \(\sigma = 27.1\), reasonably deep subsurface peaks. Surface Chl shows the opposite trend, with larger surface peaks as surface Chl increases. The most frequent patterns in this season \([\text{patterns (0,1), (0,2), (1,1), (2,1)}]\) were found in warm surface water, with low surface Chl concentrations, deep mixed layers and water columns.

Summer

In summer, changes in profile shape for different levels of the environmental variables were generally similar to those identified in spring. Generally, as sea surface temperature, upper mixed layer depth and water column depth increase, there is a shift from the right to the left of the output map (Figure 10), reflecting a change in shape from large near-surface peaks to the smallest and deepest peaks. By contrast, increasing surface Chl shows the opposite trend, with a general movement across the output map from the top left to the bottom right. This reflects a change in profile shape from those with small

---

**Fig. 6.** Overall relative frequency map of the SOM. Note that the relative frequency for Node (0,0) in Figure 5 is shown on Node (0,0) in the current figure to be 11.6\%.

---

[Graph and table showing relative frequency map of the SOM]
and very deep peaks to those with the shallowest and largest peaks. Thus, the two groups of frequently occurring patterns in summer (see Figure 8) occur under vastly different environmental conditions. The small deep peaks of patterns (0,0) and (1,0) commonly occur at warm temperatures, low surface Chl concentrations, and intermediate mixed layer depths and deep water. Conversely, the large near-surface peaks of patterns (3,1), (3,2) and (4,2) commonly occur at cool to intermediate temperatures, high surface Chl concentrations, shallow mixed layers and shallow to intermediate water column depths. Although changes in profile shape for different levels of the environmental variables are similar in summer and spring, the exact patterns represented were quite different.

Fig. 7. Error frequencies for input profiles that were mapped to each output pattern. The error is the Euclidean distance between a profile and its associated pattern, and is a measure of how well the profile is characterized by this pattern.

Fig. 8. Relative frequency map of profile patterns in (a) winter, (b) spring and (c) summer. Note that the relative frequencies are calculated separately for each period.
For example, at cool and intermediate temperatures, Chl peaks are generally larger and closer to the surface in summer than spring. Moreover, large near-surface Chl peaks occur in deep water in summer but not in spring.

**DISCUSSION**

This study shows that there is considerable variability in the shape of vertical Chl profiles in the western Agulhas Bank region of the southern Benguela upwelling system. The SOM identified a continuum of patterns, ranging from those with small (~1 mg m⁻³) subsurface peaks in warm water usually found offshore, to large (~9 mg m⁻³) near-surface peaks in cool water usually found inshore.

Chlorophyll profiles averaged seasonally give a misleading representation of the characteristic shape of profiles in this coastal domain. For example, the averaged Chl profile for summer has a moderately large peak at ~20 m, a profile shape that is not apparent in the patterns from the SOM analysis. By contrast, the SOM identified two groups of profiles in summer, one with a small subsurface peak and the other with a large near-surface peak, with very few intermediate forms. By averaging over a season, unrealistic profiles are created.

The SOM technique

SOMs have only recently been applied to vertical Chl profiles (Silulwane et al., 2001) and they have considerable potential for classifying vertical Chl profiles into characteristic patterns. Problems addressed by the SOM are similar to those for which traditional statistical methods, such as cluster analysis, principal components analysis or multidimensional scaling, are used (Dayhoff, 1990; Hewitson and Crane, 1994). The SOM, however, does not require the same assumptions as these techniques. Moreover, the SOM enables easy visualization of the underlying patterns in the same form as the input data (e.g. shapes of vertical Chl profiles), something that is not done with standard output from cluster analysis or multidimensional scaling. Another advantage of the SOM is that large datasets can be visualized easily. It was straightforward to use over a hundred profiles in this study; in fact, results are more robust with large datasets because the SOM can
learn from more data. The corresponding frequency maps would also be more robust. The SOM could be used on datasets with thousands of profiles; these would be more difficult to analyse with principal components analysis, as there are no significance levels for the patterns and does not give the number of patterns chosen is dependent upon the level of detail desired in the analysis. For instance, a very large SOM output map identifies a large number of patterns so that detailed structure can be identified, whereas a small SOM output map identifies only a few generalized patterns. However, if the data have a finite number of patterns, the SOM would still identify these: those patterns would have high frequency and other patterns that were produced in the SOM output map would have low or zero frequency. The decision on the number of groups is similar to cluster analysis where the researcher must make an arbitrary decision on the level of similarity required to identify clusters.

Performing the SOM analysis on the four parameters of the shifted Gaussian model rather than on the raw Chl profiles leads to some biases in the patterns observed. As ~16% of the profiles were discarded prior to the SOM analysis because they did not fit the Gaussian model, there is actually greater variability in Chl profiles than presented in this study. This is a common problem with studies that have parameterized vertical Chl profiles by the shifted Gaussian model (Longhurst et al., 1995; Sathyendranath et al., 1995). To avoid this shortcoming, it
is possible to perform a SOM analysis on raw profiles that have been interpolated to standard depths. This approach would capture more of the true shape and may aid interpretation, although the characteristic profiles would then not be derived from model parameters that can be used to calculate integrated primary production easily.

**Seasonal changes in profile shape on the western Agulhas Bank**

In winter, the most common profile type has a near-uniform distribution of Chl throughout the mixed layer. This is because of light limitation and turbulent mixing by strong westerly winds and tidal stirring, resulting in a well-mixed water column (Pitcher et al., 1992).

In summer, the profiles commonly have either large surface peaks or small deep peaks. This is because upwelling is intense owing to the strong south-easterly winds. Upwelling-favourable winds generally persist for 2–4 days (Largier et al., 1992), followed by a 2- to 4-day period of quiescence or wind reversal (Pitcher et al., 1992). The hydrographic structure changes rapidly in response to these pulsed periods of wind forcing and nutrient input, and the subsequent quiescence and stratification (Mitchell-Innes et al., 1999).

In spring, there is the greatest diversity of profiles, with profiles having small relatively deep peaks or moderate Chl peaks at intermediate depths. This is because spring is a season of transition, having characteristics of both winter and summer. For example, there were more deep and very broad Chl peaks than in summer, probably because the water column still has a deeper mixed layer than in summer (Richardson et al., 1998). However, there were relatively few profiles with large surface peaks, presumably because upwelling-favourable south-easterly winds are relatively infrequent (Jury 1988), the mixed layer is still too deep to permit surface upwelling, and involution is too weak to permit rapid development of surface blooms.

**Effect of upwelling on Chl profiles**

In this study, we identified strong relationships between profile shape and environmental variables using the SOM technique. Surface phytoplankton blooms were common in cooler surface water inshore with shallow mixed layers. These peaks became progressively smaller and deeper as the water warmed and the mixed layer deepened. These relationships are a consequence of the influence of physical processes such as upwelling, stratification and turbulence (Pitcher et al., 1992; Mitchell-Innes et al., 1999).

The primary control on phytoplankton bloom development in upwelling areas and hence the profile shapes identified by the SOM is stability following upwelling of nutrient-rich water (Brown and Hutchings, 1987a). Drogue studies following the maturation of newly upwelled water masses in the southern Benguela system have elucidated the temporal sequence of bloom development (Brown and Hutchings, 1987a,b). Phytoplankton biomass is low and growth is initially slow in cold, newly upwelled water. Stability of the water column increases as newly upwelled water warms. This increase in stability and the ample light field at the surface lead to an increase in phytoplankton growth and a surface phytoplankton bloom. This causes shading of phytoplankton deeper in the water column, resulting in low subsurface phytoplankton biomass. As the upwelled water warms further, the water column stratifies and the mixed layer deepens. When nutrients in the mixed layer are consumed, maximum primary productivity shifts deeper in the water column, resulting in a subsurface Chl maximum. Although nutrients are present below the mixed layer, light limitation at these depths causes low subsurface Chl concentrations.

Despite a canonical succession of phytoplankton development, there is considerable variability in the timing and magnitude of these events because of varying physical and biological factors. Even within a particular range of an environmental variable, the SOM did not identify a single typical profile, but rather a variety of shapes. Blooms do not develop in discrete water parcels moving offshore: there is considerable mixing of new and old water because of the effect of turbulence, horizontal advection and strong shearing motions between water parcels. Biotic factors such as the phytoplankton seed community (Pitcher, 1990) and the abundance of herbivorous zooplankton influence the rate of bloom development. In one study in the southern Benguela upwelling system, copepods cleared 42% of the primary production (Walker and Peterson, 1991).

**Implications for predicting integrated production**

The use of a SOM to identify characteristic profile shapes represents a dynamic approach compatible with that suggested by Platt and Sathyendranath (Platt and Sathyendranath, 1999). In the southern Benguela upwelling system, and probably other biogeochemical provinces in the coastal domain as well, there are marked seasonal differences in profile shape, although there are a wide variety of profiles within each season. It is suggested that in the coastal domain, it is insufficient to use profiles averaged seasonally: it is necessary to relate the variability in profile shape to environmental variables that can be measured from satellites. We found general relationships between the vertical phytoplankton distribution and environmental variables that capture the dynamic nature of upwelling regions. The approach presented here can
be used in a semi-quantitative manner to predict the subsurface Chl field from known (water column depth) or easily measured variables from satellites (sea surface temperature or surface Chl concentration), as the relative frequency of each characteristic profile under different environmental conditions is presented. Using this approach would enable improved estimates of profile shape in the dynamic coastal domain and thus superior regional estimates of primary production.

ACKNOWLEDGEMENTS

We would like to thank the Chief Directorate Marine and Coastal Management of the Department of Environmental Affairs and Tourism, South Africa, and especially Betty Mitchell-Innes, for providing the Chl profiles and environmental variables for the analysis. Bruce Hewitson, Jeremy Main and Geoff Brundrit are thanked for their useful discussions on the use of SOMs. We would also like to thank Professor Shubha Sathyendranath for her helpful and constructive comments on the manuscript. The EU ENVIFISH program, contract number IC18-CT98-0129, provided funding for N.F.S. and A.J.R.

REFERENCES


Received on January 2, 2001; accepted on August 9, 2002


Received on January 2, 2001; accepted on August 9, 2002