

IDENTIFICATION AND CLASSIFICATION OF VERTICAL CHLOROPHYLL PATTERNS IN THE BENGUELA UPWELLING SYSTEM AND ANGOLA-BENGUELA FRONT USING AN ARTIFICIAL NEURAL NETWORK

N. F. SILULWANE*, A. J. RICHARDSON*, F. A. SHILLINGTON*
and B. A. MITCHELL-INNES†

Information on the vertical chlorophyll structure in the ocean is important for estimating integrated chlorophyll *a* and primary production from satellite. For this study, vertical chlorophyll profiles from the Benguela upwelling system and the Angola-Benguela front were collected in winter to identify characteristic profiles. A shifted Gaussian model was fitted to each profile to estimate four parameters that defined the shape of the curve: the background chlorophyll concentration (B_0), the height parameter of the peak (h), the width of the peak (σ) and the depth of the chlorophyll peak (z_m). A type of artificial neural network called a self-organizing map (SOM) was then used on these four parameters to identify characteristic profiles. The analysis identified a continuum of chlorophyll patterns, from those with large surface peaks ($>10 \text{ mg m}^{-3}$) to those with smaller near-surface peaks ($<2 \text{ mg m}^{-3}$). The frequency of occurrence of each chlorophyll pattern identified by the SOM showed that the most frequent pattern (~12%) had a near-surface peak and the least frequent pattern (~2%) had a large surface peak. These characteristic profile shapes were then related to pertinent environmental variables such as sea surface temperature, surface chlorophyll, mixed layer depth and euphotic depth. Partitioning the SOM output map into environmental categories showed large peaks of surface chlorophyll dominating in water with cool temperature, high surface chlorophyll concentration and shallow mixed layer and euphotic depth. By contrast, smaller peaks of subsurface chlorophyll were in water with warmer temperature, lower surface chlorophyll concentration, intermediate mixed layer and deep euphotic depth. These relationships can be used semi-quantitatively to predict profile shape under different environmental conditions. The SOM analysis highlighted the large variability in shape of vertical chlorophyll profiles in the Benguela. This suggests that an ideal typical chlorophyll profile, as used in the framework of biogeochemical provinces, may not be applicable to this dynamic upwelling system.

Key words: Benguela system, biogeochemical province, characteristic profiles, chlorophyll profiles, pattern recognition, self-organizing maps

Information on the vertical chlorophyll structure is important not only for estimating integrated chlorophyll (Platt *et al.* 1988, Sathyendranath *et al.* 1995), but also as input to analytical models of primary production that require depth-dependent values of chlorophyll concentration (Sathyendranath *et al.* 1995). This subsurface chlorophyll structure needs to be extrapolated spatially and temporally for the estimation of integrated pigment content and primary production (Sathyendranath *et al.* 1995). However, because satellite ocean colour sensors can only measure near-surface chlorophyll, satellites cannot detect deep chlorophyll maxima (André 1992, Kameda and Matsumura 1998). Primary production is then under-estimated when a homogeneous vertical chlorophyll structure is assumed (Platt *et al.* 1991).

To address this problem, Platt *et al.* (1988) partitioned the ocean into four primary domains, each with their own characteristic physical processes and

biological characteristics. These primary domains were further subdivided into biogeochemical provinces based on regional oceanographic processes (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). One such biogeochemical province within the coastal domain is the Benguela upwelling system, from Cape Peninsula (34°S) north to Cape Frio (18°S). Features of this province include coastal upwelling, anticyclonic eddy fields and a shelf-break front (Longhurst *et al.* 1995). It has been assumed that each biogeochemical province is characterized by a group of typical (usually seasonal) chlorophyll profiles.

Considering the dynamic physical oceanographic forcing within upwelling systems, it is likely that there is considerable variability in chlorophyll profile shapes. Previous studies in the Benguela upwelling system suggest that vertical patterns of chlorophyll distribution differ spatially and temporally. Often there is a cross-shore change in chlorophyll profiles,

* Remote Sensing Unit, Department of Oceanography, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa.
E-mail: nonkqu@mcm.wcape.gov.za

† Marine & Coastal Management, Private Bag X2, Rogge Bay 8012, Cape Town, South Africa

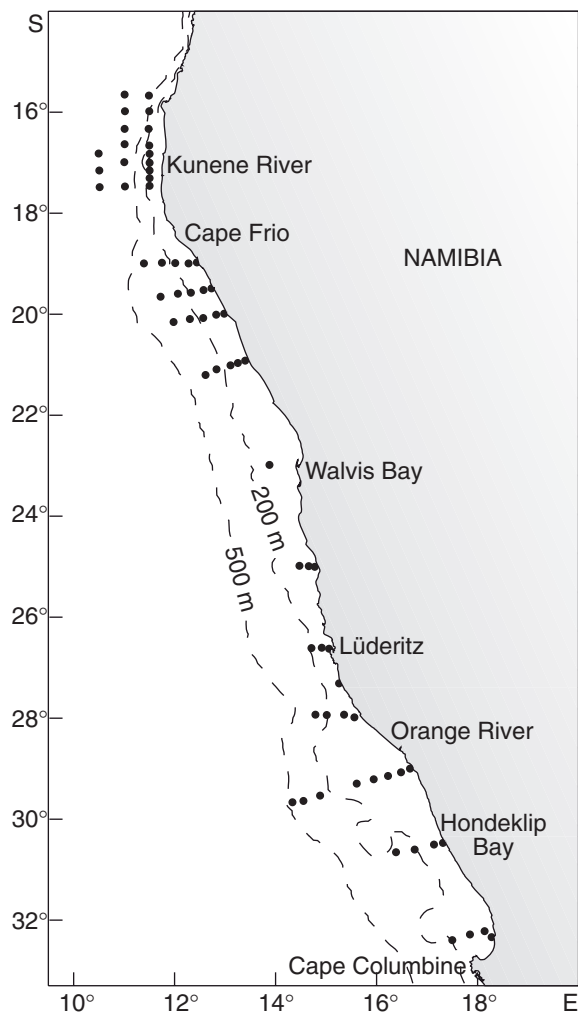


Fig. 1: Cruise map showing the station positions where vertical chlorophyll profiles were collected during the BENEFIT research cruise in June/July 1999

from larger surface peaks inshore to smaller subsurface peaks offshore (Pitcher *et al.* 1992). Drogue studies, following the path of newly upwelled water, have highlighted the temporal change in profile shape during the maturation of upwelled water. Newly upwelled water generally is low in chlorophyll throughout the water column (Brown and Hutchings 1987). Following warming and stratification, surface phytoplankton blooms occur. After 6–8 days, the bloom declines and the chlorophyll maximum be-

comes subsurface. Therefore, the main control of phytoplankton bloom formation is the stability of the water column (Pitcher *et al.* 1996, Brown and Hutchings 1997). Consequently, if shapes of chlorophyll profiles can be related to pertinent environmental variables, improved regional estimates of primary production in the Benguela would be possible.

A number of techniques can be used to identify and characterize chlorophyll profile shapes. One such technique is artificial neural networks. Artificial neural networks are computer algorithms that are good at identifying and classifying patterns. Their architecture and functioning are inspired by the massive interconnectivity of the animal brain and its ability to solve non-linear problems without prior assumptions about the data (Chen and Ware 1999). Artificial neural networks, like their biological counterparts, do not require an explicitly programmed solution to a problem, but rather learn patterns within the data. Artificial neural networks have been fruitfully applied in a variety of disciplines, including speech recognition, target identification for military purposes, stock market prediction and weather forecasting, although there have been few applications in marine ecology and fisheries (but see Jarre-Teichmann *et al.* 1995, Aoki and Komatsu 1997, Chen and Ware 1999, Storbeck and Daan 2001).

In this study, a type of artificial neural network called a self-organizing map (SOM), which is particularly adept at pattern recognition (Kohonen *et al.* 1995), is used. The SOM has the ability to capture non-linear patterns and to span breakpoints in the data. For identifying groups in a dataset, the primary advantage of SOMs over traditional multivariate techniques such as cluster analysis is that the relationship among the identified patterns can be visualized (Hewitson and Crane *in press*). Therefore, output patterns are arranged in a two-dimensional array (the "self-organizing map"), in which similar patterns are found closer to one another and dissimilar patterns farther apart. The number of output patterns identified is set by the user, depending on the detail required. SOMs, as in other artificial neural networks, also have the advantages of being able to handle large datasets and missing values.

In this study, a preliminary assessment of whether the Benguela Current biogeochemical province is characterized by a single characteristic profile in winter is made from *in situ* data. A self-organizing map has been used to identify characteristic vertical chlorophyll profiles objectively. These patterns were then related to easily measured environmental variables. The study forms part of an ongoing project to identify characteristic chlorophyll profiles throughout the Benguela upwelling system in an attempt to improve regional estimates of primary production from satellite.

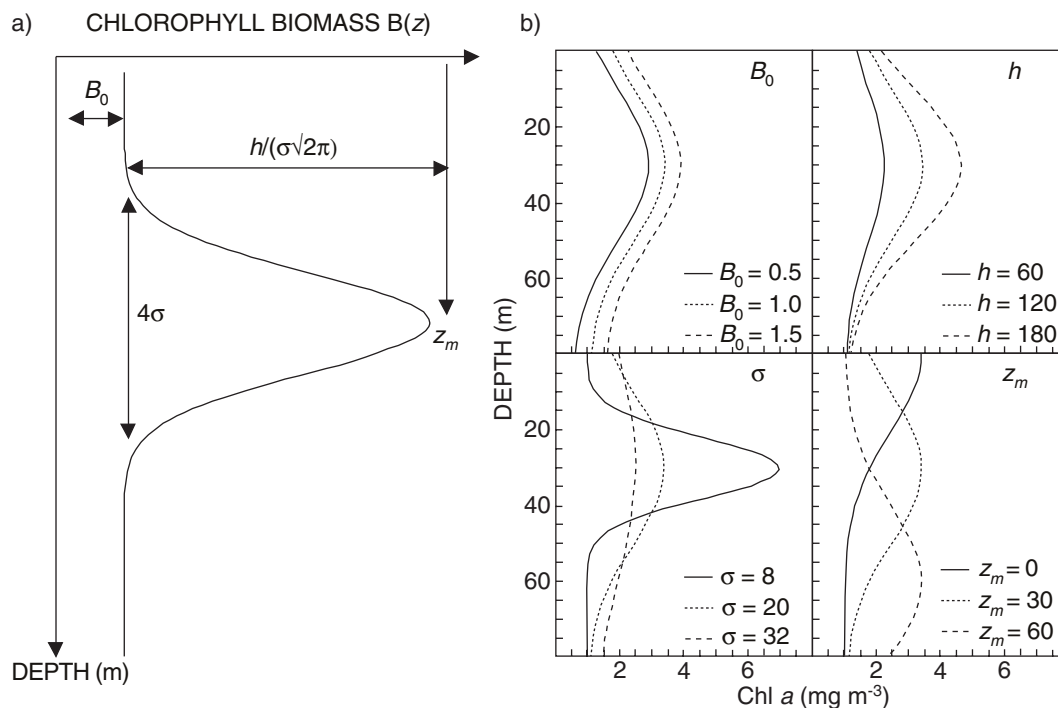


Fig. 2: (a) A shifted Gaussian curve showing the four parameters (B_0 , h , σ and z_m) used to describe *in situ* vertical chlorophyll profiles, (b) the effect of changing B_0 , h , σ and z_m on the shape of the Gaussian curve

MATERIAL AND METHODS

Data

Vertical fluorescence profiles were collected during the BENguela Environment Fisheries Interaction and Training (BENEFIT) research cruise in June/July 1999. The cruise covered the Benguela Current and the Angola-Benguela front (Fig. 1). Fluorescence profiles were measured (down to 100 m) using a profiling fluorometer (Chelsea Instruments AquaTracka MKIII). To convert the fluorescence readings into chlorophyll equivalents, *in situ* fluorescence readings were related to extracted chlorophyll concentrations. Samples for extracted chlorophyll *a* were collected in Niskin bottles at the surface and the depth of maximum fluorescence. Water samples were filtered using Whatmann GF/F filters and extracted in 90% acetone at 20°C for 24 h. Samples were then measured fluorometrically using a Turner designs Model 10-000R fluorometer before and after addition of hydrochloric

acid to adjust for phaeopigments.

In situ fluorescence readings were then related to the extracted chlorophyll concentrations by linear regression analysis of log-transformed values of chlorophyll *a* against fluorescence. High values of extracted chlorophyll at the surface corresponding to low fluorescence values measured by the profiling fluorometer were considered photo-inhibited and were excluded from the calibration regression. Chlorophyll profiles that corresponded to these photo-inhibited values were discarded from all subsequent analyses. After exclusion of the profiles that appeared photo-inhibited, a total of 60 chlorophyll profiles remained.

Underwater irradiance was measured concurrently with the fluorescence profiles using a LI-COR Underwater Quantum sensor attached to the sampling rosette. The euphotic depth was the depth where the irradiance was 1% of the surface irradiance. The mixed layer depth was calculated as the depth where the difference in temperature from the surface was 0.5°C (Longhurst 1995).

Parameterizing vertical chlorophyll profiles

To describe the shape of each vertical chlorophyll profile, a shifted Gaussian curve was fitted (Fig. 2a). This equation has four parameters: the background chlorophyll concentration (B_0), the height parameter (h), the width of the peak (σ), and the depth of the chlorophyll peak (z_m ; Lewis *et al.* 1983, Platt *et al.* 1988, 1991, Sathyendranath *et al.* 1995). The equation for the shifted Gaussian curve is given by

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

The effect that each of these parameters has on the shape of the shifted Gaussian curve is shown in Fig. 2b.

The shifted Gaussian curve was fitted to each profile using the quasi-Newton algorithm. The background chlorophyll concentration, the height parameter and the width of the peak were constrained to be positive, because they are only meaningful for positive values. The depth of the chlorophyll peak was also constrained to be positive, even though negative values are meaningful (Platt *et al.* 1991), because the h parameter was sometimes unrealistically large ($h > 10\,000$ mg m⁻²). Constraining the z_m parameter to be positive in situations when it was negative effectively meant that it was ~ 0 (surface peak). This constrained model had no effect on models with positive z_m values and very little effect on the “goodness of fit” of models for profiles with negative z_m values (r^2 values changed by <3%).

Using a self-organizing map to identify patterns

To characterize *in situ* chlorophyll profiles, a self-organizing map (SOM) was used on the four model-derived profile parameters. Input data consisted of the four parameter values (columns) and a number of profiles (rows). Column normalization was performed prior to the SOM analysis to give each parameter equal weighting.

The SOM consists of two layers, each with a number of processing units or nodes. The first is the input layer, which consists of the input data (here the parameters for each profile), and the second is the output layer or map, which are the parameter values for the vertical patterns identified (these can be visualized directly). These layers are interconnected by connection weights, in which each node of the input layer is connected to every node in the output layer (Dayhoff 1990). The basic SOM algorithm is briefly outlined below, but more details can be found in Kohonen (1997) and Hewitson and Crane (in press).

The shape or topology of the output map can be chosen to be either rectangular or hexagonal. A rectangular topology places adjacent nodes on the corners of a square, whereas a hexagonal topology places adjacent nodes at the apices of a hexagon. Hexagonal topology produces more continuous output patterns, whereas rectangular topology is easier to plot. The dimensions of the output map determine the number of output patterns. Too many nodes do not adequately reduce the data to characteristic patterns, but too few nodes do not adequately differentiate underlying patterns. In the present analysis, a map of four columns by three rows with rectangular topology was used. The 4×3 SOM provided sufficient resolution of the underlying patterns, although 3×2 and 6×4 SOMs were also investigated.

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 the present analysis). These can either be chosen to be random, or they can be initialized with the first two orthogonal components of a principal components analysis of the input data. In this analysis, the four-dimensional input data (normalized parameter values for each profile) were initialized with a principal components analysis on the input data, ensuring that the major diagonals of the SOM were axes of most variation.

Input data are presented to the SOM sequentially in a training cycle. The first input vector (parameters of the first profile in the present 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) learn from the same input and their weights are adjusted by a spatial decay function to be similar (although not identical) to that input vector. The spatial decay function can either be a 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, resulting in a continuum of patterns across the node space. The training procedure is repeated for all input data.

The training cycle is repeated until convergence is reached. This typically involves a large number of cycles (10^5 – 10^6). During each training cycle, the learning

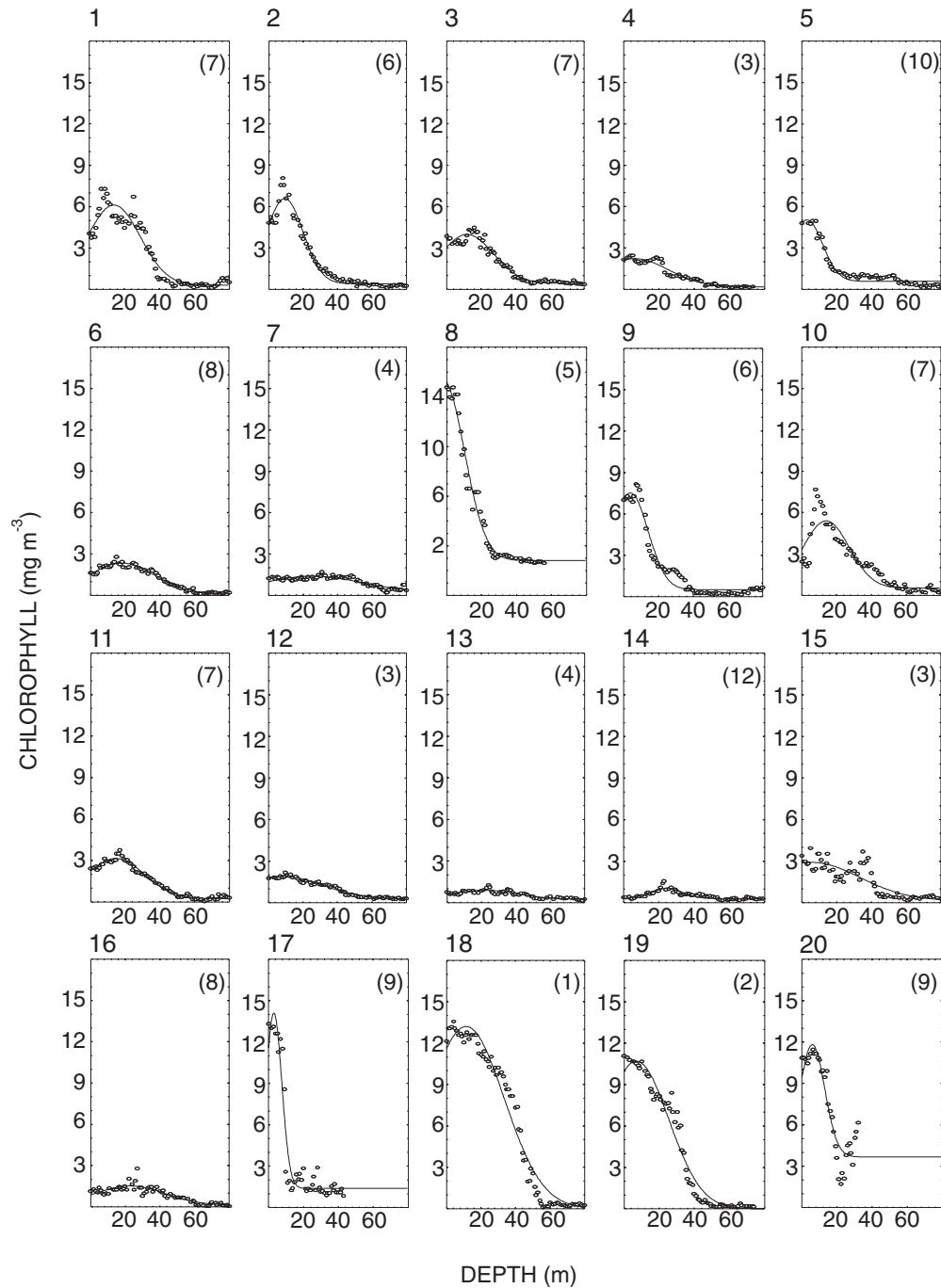


Fig. 3: Each vertical chlorophyll profile collected (points) fitted with a shifted Gaussian curve (line). At the top left of each figure is the profile number (see Table 1), and in the top right is a number (in parenthesis) indicating the pattern that the profile mapped to in the SOM

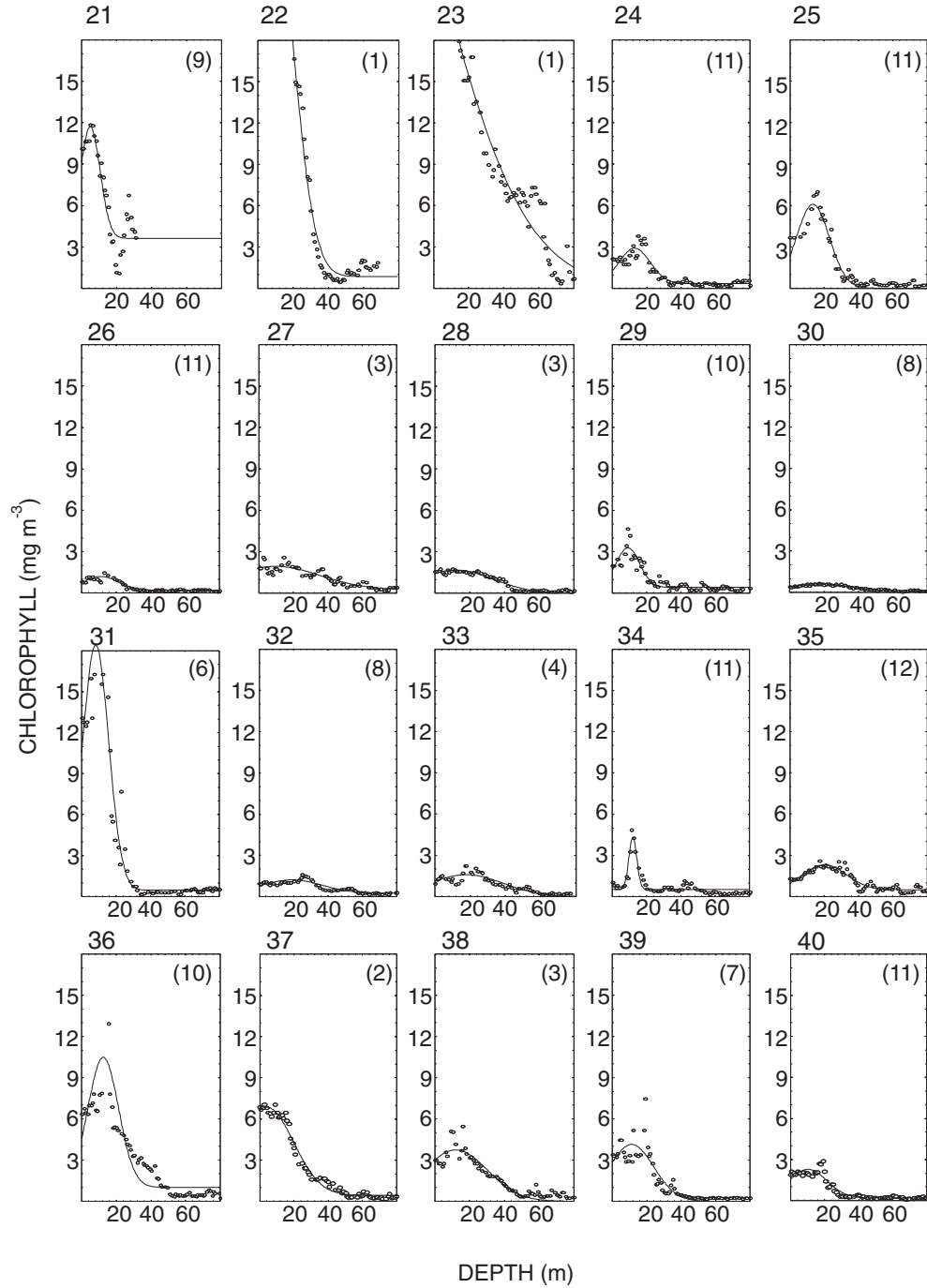


Fig. 3 (continued)

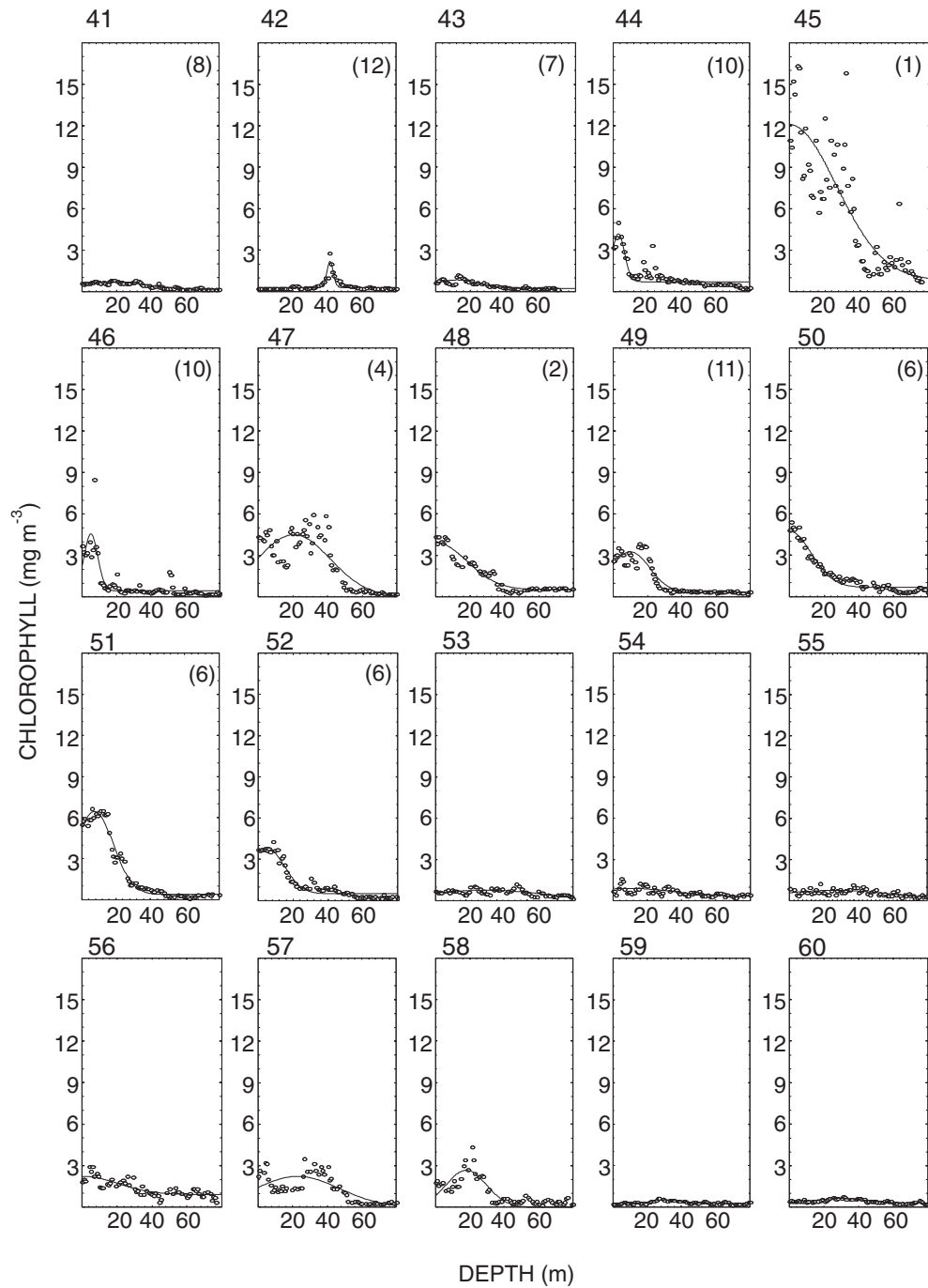


Fig. 3 (continued)

rate parameter (α) determines the extent to which 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 increase in detail later in training. Convergence is achieved when the overall error (the mean of the Euclidean distances for all input data and the pattern to which they are mapped) is a minimum (the error is dimensionless after normalization of input data). In the present analysis, the training consisted of 100 000 iterations with an initial learning rate (α) of 0.2, a bubble neighbourhood function and an initial update neighbourhood of 6.

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

A frequency map of the chlorophyll patterns from the SOM was then constructed to determine the relative frequency of each of the patterns in the dataset. Frequency maps have the dimensions of the SOM output and show the relative frequency (%) of each pattern. All relative frequencies on a self-organizing map sum to 100%.

Relating vertical chlorophyll patterns to environmental variables

Vertical chlorophyll patterns identified in the SOM were then related to environmental variables that may influence the profile shape. This was achieved by partitioning the overall frequency map of the SOM into categories based on different values of the environmental variable in question. Environmental variables used were sea surface temperature (SST), surface chlorophyll, mixed layer depth (the depth where the difference in temperature from the surface is 0.5°C) and euphotic depth (1% surface light depth). SST was chosen because it is related to the age of upwelled water in an upwelling system (Brown and Hutchings 1987, Platt *et al.* 1995). Surface chlorophyll was included because it has been related previously to the shape of chlorophyll profiles (Morel and Berthon 1989, Sathyendranath *et al.* 1995). Mixed layer depth was used because it is related to stability of the water column and to physical forcing mechanisms such as wind (Mann and Lazier 1991). Euphotic depth was included because the depth to which light penetrates is related to the vertical chlorophyll distribution. Chlorophyll and SST can be estimated from satellite imagery so they can be mea-

sured synoptically in real time. Whenever possible, categories were chosen to have approximately the same number of profiles within each category of an environmental variable.

RESULTS

Fitted vertical chlorophyll profiles

Gaussian curves generally fitted the raw chlorophyll profiles well (Fig. 3). Of the total of 60 profiles, 52 had an r^2 value for the Gaussian model $>70\%$ and were used for further analysis. The four model-derived parameter values for each profile are listed in Table I.

Characteristic profile shapes

The SOM analysis produced a continuum of chlorophyll patterns in a two-dimensional (4×3) array (Fig. 4a). To assess how well the patterns represent the raw profiles, the curves in Figure 3 can be compared with the SOM output map in Figure 4. Generally, the characteristic patterns identified by the SOM represent the chlorophyll profiles well. Chlorophyll patterns on the left side of the SOM had larger ($\sim 13 \text{ mg m}^{-3}$) surface chlorophyll concentration (e.g. #1 in the SOM map) than those with smaller subsurface peaks ($\sim 2 \text{ mg m}^{-3}$) on the right side (e.g. #12). Moreover, chlorophyll patterns on the left side of the SOM had surface peaks ($<5 \text{ m}$) and these graded to patterns with subsurface maxima ($\sim 20 \text{ m}$) on the right side (z_m increasing). Also, the total chlorophyll beneath the curve decreased from top to bottom and from left to right (h decreasing). The width of the peak also changed across the SOM output, peaks at the bottom left corner being narrower than those towards the top right corner (σ increasing). The background chlorophyll concentration increased from a minimum in the top right corner to a maximum in the bottom left corner (B_0 increasing). Therefore, the continuous change of shape of chlorophyll patterns across the SOM output map reflects the continuous change in the profile parameters.

The frequency of occurrence of each of the chlorophyll patterns in the SOM output map is shown in Figures 4b and c for the Benguela Current system and the Angola-Benguela front. Note that each map in Figures 4b and c has 12 grid squares: each corresponds to the pattern in the same position in the SOM in Figure 4a. In Figure 4b, profiles with moderate near-surface peaks ($2\text{--}4 \text{ mg m}^{-3}$) dominated. Pattern #3 was the most frequent chlorophyll pattern ($\sim 11\%$) identified by the SOM. Pattern #5 was the least frequent

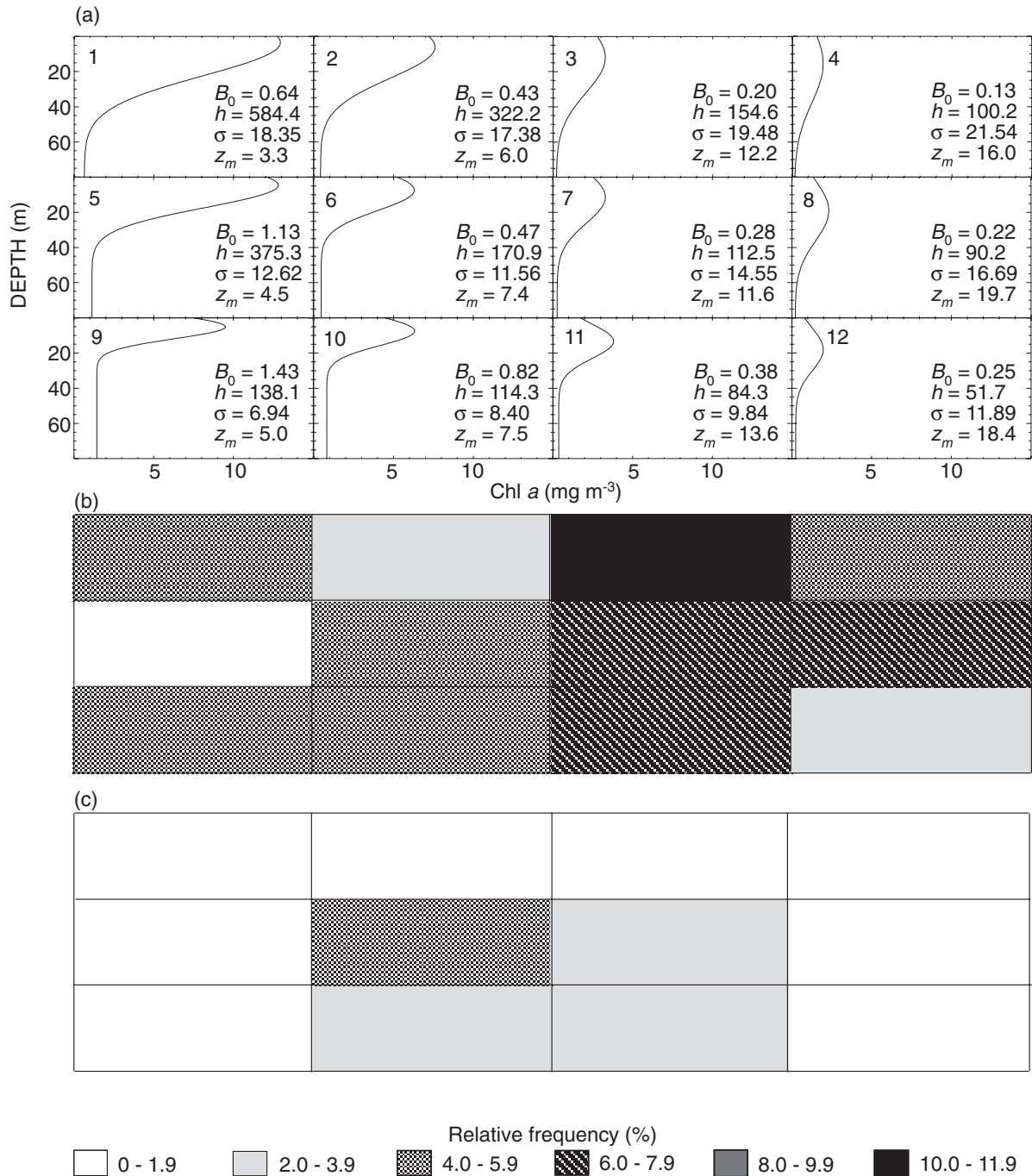


Fig. 4: (a) A 4 × 3 SOM showing the characteristic vertical chlorophyll patterns representing the 52 profiles used as input data; (b) relative frequency map of the SOM of the Benguela system. Note that each map has 12 grid squares, each corresponding to the pattern in the same position in the SOM in (a); (c) relative frequency map of the SOM of the Angola-Benguela front

Table I: Parameter values (B_0 , h , σ and z_m) for each profile obtained by fitting a shifted Gaussian model. The variance explained (r^2 value) is included to show how well the model fitted each profile. Only profiles with $r^2 \geq 70\%$ were used in further analysis

Profile number	Station grid	B_0	h	σ	z_m	r^2 value (%)	Pattern to which mapped
1	202	0.3	220.0	15.1	14.1	93	7
2	203	0.4	166.6	10.7	9.3	97	6
3	205	0.4	136.1	15.3	11.9	96	7
4	301	0.2	99.1	19.3	5.9	94	3
5	302	0.6	100.3	9.0	2.9	94	10
6	303	0.1	105.0	18.9	18.0	98	8
7	304	0.2	88.7	29.4	25.5	84	4
8	401	0.9	378.8	11.0	0.0	98	5
9	402	0.5	180.8	10.3	3.7	96	6
10	403	0.5	163.1	13.3	14.0	87	7
11	404	0.2	121.3	16.7	14.4	98	7
12	405	0.3	95.9	23.7	8.2	97	3
13	503	0.3	29.6	18.6	21.8	80	4
14	502	0.3	17.4	9.4	24.1	79	12
15	602	0.2	189.7	27.5	5.8	77	3
16	604	0.1	70.7	20.2	22.7	83	8
17	801	1.4	145.2	4.6	3.0	94	9
18	803	0.0	550.8	23.9	7.3	99	1
19	804	0.0	495.6	18.5	7.0	97	2
20	901	3.7	144.6	7.1	5.9	88	9
21	901A	3.6	113.8	5.6	4.9	87	9
22	902	0.8	744.3	11.7	9.3	99	1
23	903	1.6	1 329.3	25.8	0.0	95	1
24	1004	0.3	45.0	9.0	13.0	86	11
25	202	0.0	146.8	9.6	13.4	94	11
26	203	0.2	25.6	10.2	11.1	91	11
27	204	0.3	110.3	26.1	9.4	83	3
28	205	0.0	90.7	21.6	11.8	95	3
29	302	0.4	49.5	6.9	9.1	88	10
30	305	0.1	20.0	16.1	18.2	96	8
31	402	0.5	340.7	7.6	8.1	94	6
32	404	0.2	42.7	17.4	19.1	84	8
33	405	0.0	83.7	21.0	17.8	82	4
34	502	0.5	21.7	2.3	11.9	88	11
35	503	0.5	57.2	12.6	20.3	88	12
36	504	1.0	210.6	8.9	12.4	74	10
37	505	0.4	250.1	15.6	4.1	98	2
38	601	0.0	170.9	18.2	12.1	88	3
39	602	0.0	133.2	12.8	11.2	82	7
40	603	0.2	58.0	11.0	10.2	90	11
41	604	0.2	26.4	18.7	17.6	87	8
42	701	0.3	8.9	1.9	41.6	86	12
43	703	0.2	20.3	13.2	12.2	81	7
44	704	0.7	30.7	3.5	3.0	76	10
45	705	0.8	752.5	26.7	0.0	70	1
46	706	0.4	40.3	3.9	5.1	73	10
47	707	0.0	239.7	21.2	20.5	75	4
48	709	0.5	239.5	23.0	0.0	94	2
49	600	0.3	82.9	11.3	9.9	88	11
50	801	0.7	129.2	12.7	0.0	95	6
51	802	0.4	167.4	11.0	7.4	98	6
52	803	0.5	85.0	10.0	4.7	94	6
53	408	0.0	78.9	40.4	32.8	47	—
54	603	0.3	36.8	25.6	9.5	51	—
55	701	0.0	50.1	27.9	21.0	43	—
56	303	0.9	65.3	20.0	0.7	66	—
57	304	0.0	136.1	24.3	22.9	56	—
58	403	0.0	78.8	11.7	18.0	68	—
59	605	0.3	5.0	7.6	31.0	62	—
60	702	0.3	12.1	14.9	29.4	68	—

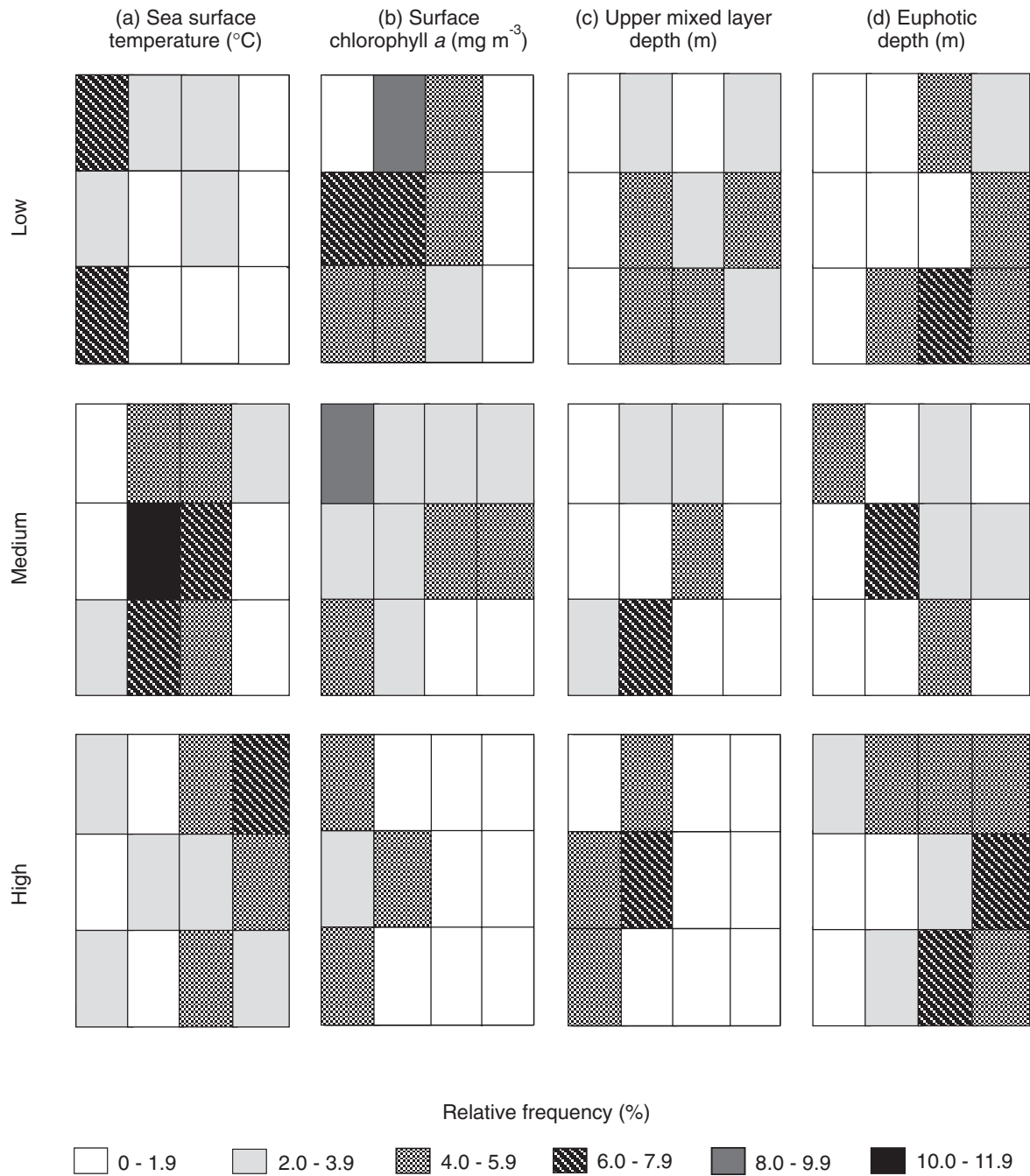


Fig. 5: Frequency maps of the vertical chlorophyll patterns for different values of (a) SST, (b) surface chlorophyll, (c) upper mixed layer depth and (d) euphotic depth. Relative frequencies for each of the variables sum to 100%. Note that each map has 12 grid squares, corresponding to the same position in the SOM in Figure 4(a)

(<2%) within the Benguela Current system (Fig. 4b). For the Angola-Benguela front, only patterns with moderate near-surface peaks (3–6 mg m⁻³) were identified, with pattern #6 being dominant (~5%; Fig. 4c). The frequency map showed patterns of the Angola-Benguela front to be similar to those of the Benguela Current system. Therefore, data from the two systems were combined because of the relatively small dataset, and will be henceforth referred to as the Benguela system for simplicity. Considering the overall frequency of occurrence of patterns for the Benguela system, pattern #7 was the most frequent vertical chlorophyll pattern identified by the SOM.

Relationship of profile shapes to environmental variables

There were marked changes in the shape of chlorophyll profiles with different levels of each environmental variable, as can be seen from differences in the relative frequency maps (Fig. 5). In terms of SST, there was a general deepening and decline in the size of the peak as water warmed. At cool temperatures (SST <14°C), chlorophyll profiles had large (>10 mg m⁻³) surface chlorophyll peaks. At intermediate temperatures (14–17°C) the peaks were still near the surface, but they had decreased in magnitude. At warm temperatures (SST >17°C), there was a variety of small (<2 mg m⁻³) and moderate peaks (~6 mg m⁻³) that were relatively deep in the water column.

To aid interpretation of the relationship between the SOM output map and environmental variables, relationships were identified between each profile parameter and each environmental variable using bivariate scatterplots with a distance-weighted least squares smoother fitted. These changes in profile shape as temperature warmed were a consequence of the decrease in background chlorophyll concentration, a smaller height parameter, a general decrease in the width of the peak, and a deepening of the chlorophyll maximum (Fig. 6a).

The shape of the profiles was also strongly related to surface chlorophyll concentration, with a clear migration in the position of maximum frequency from the right side of the SOM output map to the left as chlorophyll increased (Fig. 5b). At low concentration of surface chlorophyll (<2 mg m⁻³), subsurface peaks were most common. At moderate concentration of surface chlorophyll (2–10 mg m⁻³), the chlorophyll peak had increased in size and shifted towards the surface. As the surface chlorophyll concentration increased even further (>10 mg m⁻³), the depth of maximum chlorophyll was even shallower. These changes

in profile shape as surface chlorophyll increased were a consequence of an increase in the background chlorophyll concentration, a substantial increase in the height parameter and a general shallowing of the chlorophyll maximum (Fig. 6b). There was no change in the width of the peak as surface chlorophyll increased.

The shape of the profiles was also related to mixed layer depth, although not as strongly as for SST and surface chlorophyll (Fig. 5c). Profiles with surface and subsurface peaks were common at shallow depths of the mixed layer (<20 m). At intermediate depths of the mixed layer (20–40 m), there was considerable variability in profile shape, from those with large peaks of surface chlorophyll to those with smaller deeper peaks. Deep mixed layers (>40 m) were characterized by chlorophyll profiles with surface and near-surface chlorophyll maximum. These changes in profile shape as the mixed layer deepens were a consequence of a decrease in background chlorophyll concentration, a broadening of the chlorophyll peak and a deeper chlorophyll maximum (Fig. 6c). The height parameter appeared to have a dome-shaped relationship, with higher values at intermediate depths of the mixed layer (these profiles mapped to the top left corner of the frequency map for intermediate, i.e. 20–40 m, mixed layer depths). This suggests that the total chlorophyll in the peak was smaller for very shallow and deep mixed layers.

The shape of the chlorophyll profile was strongly related to euphotic depth, with a clear migration from the right to the left side of the SOM map (Fig. 5d). At shallow and deeper euphotic depths (<15 and >30 m), small and relatively deep peaks were common. Large peaks in surface chlorophyll were frequent at intermediate depths (15–30 m). These changes in profile shape as the euphotic layer deepens were a consequence of a general decrease in both the background chlorophyll concentration and the height parameter, and an increase in the depth of the chlorophyll maximum layer (Fig. 6d). There was no consistent change in the width of the peak (σ) as the euphotic layer deepened.

DISCUSSION

To our knowledge, this is the first time that an artificial neural network such as a self-organizing map has been used to identify characteristic vertical chlorophyll profiles in the ocean. This technique is more impartial than subjectively grouping profiles into classes. A major advantage of the technique is that the output is easy to interpret: the profile shapes can be visualized

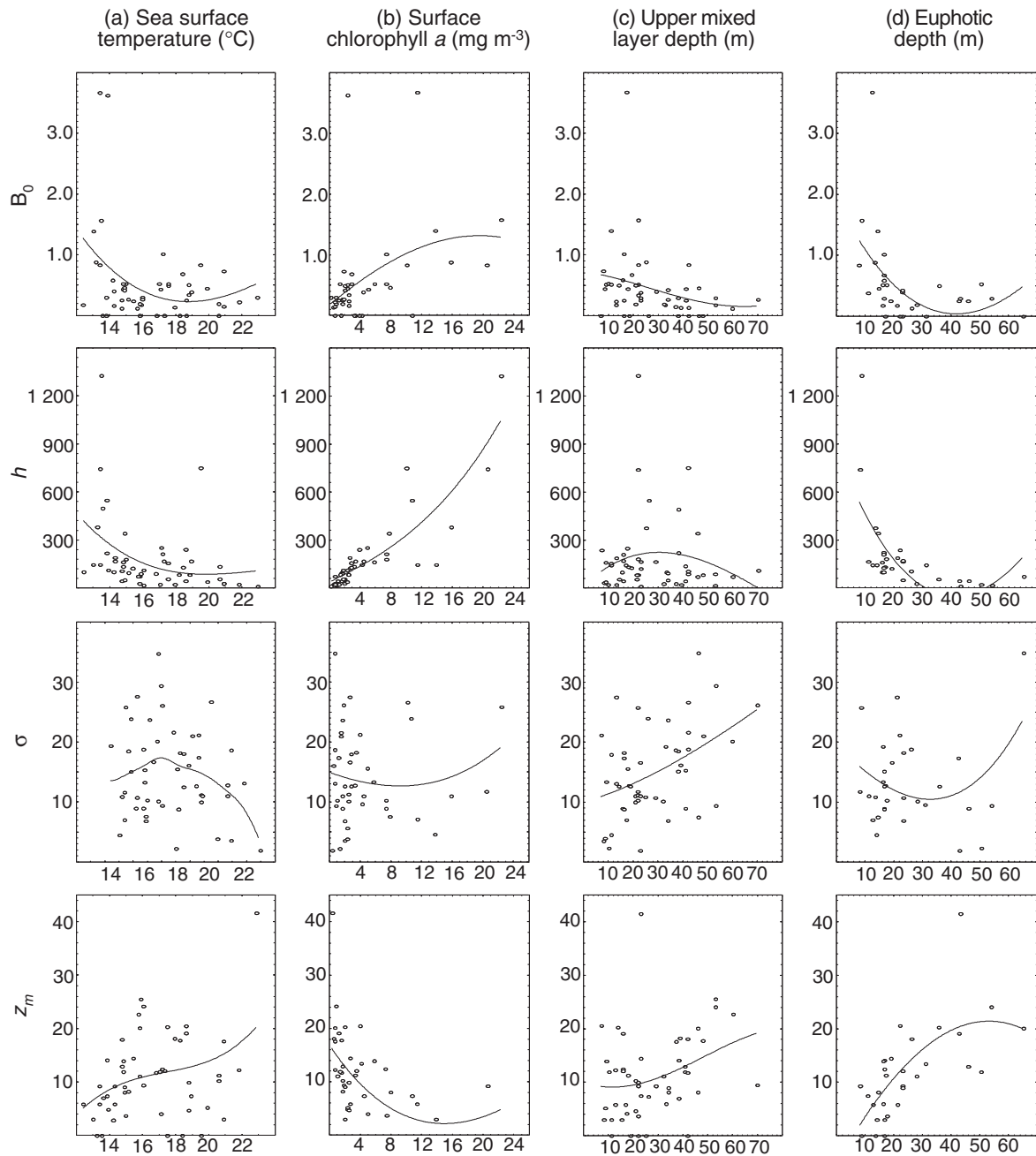


Fig. 6: Relationships between the four parameters of the shifted Gaussian curve (B_0 , h , σ and z_m) fitted to each profile and (a) SST, (b) surface chlorophyll concentration, (c) upper mixed layer depth and (d) euphotic depth. Lines were fitted using the least squares distance-weighted smoothing procedure

directly and the relationship among classes is clear. This is in contrast with the output from conventional multivariate techniques such as principal components, multidimensional scaling and cluster analysis, that use various abstractions of the data. In addition, the SOM makes it possible to visualize large datasets with potentially thousands of profiles.

Characteristic profiles of the Benguela system

The SOM analysis identified a continuum of characteristic chlorophyll profiles, from those with large surface peaks to those with smaller subsurface ones. Although there were no chlorophyll profiles with very deep subsurface peaks, there was still considerable spatial variability in the shape of chlorophyll profiles in the Benguela system during winter. The surface and subsurface peaks identified by the SOM technique resemble those in many eutrophic and mesotrophic waters (Morel and Berthon 1989). Moreover, it is likely that the diversity of profile shapes has been underestimated, for two main reasons. First, the dataset was limited to only 52 profiles from only one winter cruise, and deep (>40 m) chlorophyll maxima were not observed, even though they are common midshelf and on the shelf-edge of the southern Benguela system (Brown and Hutchings 1987). Second, eight profiles were not used in the SOM analysis because the shifted Gaussian curve did not fit well, so from the outset some profiles that were different were excluded. The diversity of profiles identified in the Benguela system suggests that a single seasonal profile is insufficient to capture the variability in chlorophyll patterns in the Benguela Current biogeochemical province. If this variability in profile shape is related to environmental variables, then improved regional estimates of primary production may follow.

Relationship of profile shapes to environmental variables

Despite the relatively small dataset, some strong relationships between profile shape and environmental variables were identified. The shape of vertical chlorophyll profiles was related to SST, surface chlorophyll, mixed layer depth and euphotic layer depth. Profiles with large peaks in surface chlorophyll dominated cool water with high concentrations of surface chlorophyll, shallow-to-deep depths of the mixed layer and intermediate euphotic depths. By contrast, profiles with smaller subsurface peaks dominated warm water, with low concentration of surface chlorophyll, intermediate mixed layers and deep euphotic layers.

These relationships are a consequence of the strong influence of physical processes such as upwelling, stratification and turbulence (Pitcher *et al.* 1992, Ediger and Yilmaz 1996, Mitchell-Innes *et al.* 1999, 2001), which limit the availability of nutrients and light in the water column. Phytoplankton biomass is initially low in very cold, nutrient-rich, newly upwelled water. As upwelled water warms, the high level of available light and warm conditions in the surface layers lead to high phytoplankton growth and the development of surface phytoplankton blooms. Bloom development leads to high concentrations of chlorophyll in the euphotic layer (Morel and Berthon 1989) and also causes less light penetration (shading) at depth, resulting in low subsurface phytoplankton biomass. As nutrients become depleted in the surface layers and the water column stabilizes, phytoplankton biomass shifts deeper, where more nutrients are available, leading to subsurface chlorophyll peaks (Cullen 1982, Brown and Hutchings 1987, Morel and Berthon 1989, Pitcher *et al.* 1992). Although high concentrations of nutrients are present deep in the water column, light limitation at these depths leads to low subsurface chlorophyll concentration.

The variability in profile shape is also evident within a particular range of an environmental variable: the SOM did not identify a single typical profile, but rather a variety of profile shapes. This is because blooms do not necessarily develop and decline in discrete water parcels moving offshore. Because upwelling is pulsed, there are strong shearing motions between water parcels, and turbulence and horizontal advection causes considerable mixing of new and old water (Pitcher *et al.* 1992). Also, considerable variation in bloom development and hence profile shape is caused by different phytoplankton seed communities in newly upwelled water (Pitcher 1990).

Frequency maps from the SOM output can be used semi-quantitatively to predict the probability of different profile shapes at a particular level of an environmental variable. This is an improvement in parameterization of vertical chlorophyll structure over assuming a single typical seasonal profile. This study has been a "proof of concept" (only 60 profiles were available) for using the SOM technique to identify characteristic vertical chlorophyll profiles. The SOM has proven to be a powerful tool for classifying the shapes of chlorophyll profiles and relating them to environmental variables. This technique is now being applied to a larger dataset (~500 profiles) from the southern Benguela. The current study suggests that estimates of vertical chlorophyll structure and hence the regional primary production estimated from ocean colour data could be improved by use of critical environmental variables to predict the shape of the chlorophyll

biomass profile. The rapid sampling and synoptic coverage of SST and surface chlorophyll measured by satellite could permit prediction of the shape of chlorophyll profiles over appropriate scales of time and space in a dynamic upwelling area. The use of other techniques, such as generalized linear modelling and generalized additive modelling, to predict profile parameters directly from SST, surface chlorophyll and a suite of other key environmental variables, is now being investigated.

ACKNOWLEDGEMENTS

We thank Dr B. C. Hewitson, Mr J. Main and Prof. G. B. Brundrit (all University of Cape Town) for useful discussions on the use of self-organizing maps. Comments by Dr J-F. Berthon and an anonymous reviewer are also appreciated. This research was carried out under the auspices of the ENVIFISH program, contract number IC18-CT98-0329, which provided funding for NFS and AJR.

LITERATURE CITED

- ANDRÉ, J. M. 1992 — Ocean color remote-sensing and the sub-surface vertical structure of phytoplankton pigments. *Deep-Sea Res.* **39**(5A): 763–779.
- AOKI, I. and T. KOMATSU 1997 — Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanologica Acta* **20**(1): 81–88.
- BROWN, P. C. and L. HUTCHINGS 1987 — The development and decline of phytoplankton blooms in the southern Benguela upwelling system. 1. Drogue movements, hydrography and bloom development. In *The Benguela and Comparable Ecosystems*. Payne, A. I. L., Gulland, J. A. and K. H. Brink (Eds). *S. Afr. J. mar. Sci.* **5**: 357–391.
- CHEN, D. G. and D. M. WARE 1999 — A neural network model for forecasting fish stock recruitment. *Can. J. Fish. aquat. Sci.* **56**: 2385–2396.
- CULLEN, J. J. 1982 — The deep chlorophyll maximum: comparing vertical profiles of chlorophyll *a*. *Can. J. Fish. aquat. Sci.* **39**: 791–803.
- DAYHOFF, J. E. 1990 — *Neural Network Architectures. An Introduction*. New York; Van Nostrand Reinhold: 259 pp.
- EDIGER, D. and A. YILMAZ 1996 — Characteristics of deep chlorophyll maximum in the northeastern Mediterranean with respect to environmental conditions. *J. mar. Syst.* **9**: 291–303.
- HEWITSON, B. C. and R. G. CRANE (in press) — Self organizing maps: applications to synoptic climatology. *Climate Res.*
- JARRE-TEICHMANN, A., BREY, T. and H. HALTOF 1995 — Exploring the use of neural networks for biomass forecasts in the Peruvian upwelling system. *Naga* **18**(4): 38–40.
- KAMEDA, T. and S. MATSUMURA 1998 — Chlorophyll biomass off Sanriku, northwestern Pacific, estimated by ocean color and temperature scanner (OCTS) and a vertical distribution model. *J. Oceanogr.* **54**: 509–516.
- KOHONEN, T. 1997 — *Self-Organizing Maps*. Berlin; Springer: 426 pp.
- KOHONEN, T., HYNINEN, J., KANGAS, J. and J. LAAKSONEN 1995 — SOM-PAK: the self-organizing map program package. Version 3.1.
- LEWIS, M. R., CULLEN, J. J. and T. PLATT 1983 — Phytoplankton and thermal structure in the upper ocean: consequences of nonuniformity in chlorophyll profile. *J. geophys. Res.* **88**(C4): 2565–2570.
- LONGHURST, A. R. 1995 — Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.* **36**: 77–167.
- LONGHURST, A. R., SATHYENDRANATH, S., PLATT, T. and C. M. CAVERHILL 1995 — An estimate of global primary production in the ocean from satellite radiometer data. *J. Plankt. Res.* **17**(6): 1245–1271.
- MANN, K. H. and J. R. N. LAZIER 1991 — *Dynamics of Marine Ecosystems. Biological Physical Interactions in the Oceans*. Boston; Blackwell Scientific: 466 pp.
- MITCHELL-INNES, B. A., RICHARDSON, A. J. and S. J. PAINTING 1999 — Seasonal changes in phytoplankton biomass on the western Agulhas Bank, South Africa. *S. Afr. J. mar. Sci.* **21**: 217–233.
- MITCHELL-INNES, B. A., SILULWANE, N. F. and M. I. LUCAS 2001 — Variability of chlorophyll profiles on the west coast of southern Africa in June/July 1999. *S. Afr. J. Sci.* **97**: 246–250.
- MOREL, A. and J. F. BERTHON 1989 — Surface pigments, algal biomass profiles, and potential production of the euphotic layer: relationships reinvestigated in view of remote-sensing applications. *Limnol. Oceanogr.* **34**(8): 1545–1562.
- PITCHER, G. C. 1990 — Phytoplankton seed populations of the Cape Peninsula upwelling plume, with particular reference to resting spores of *Chaetoceros* (Bacillariophyceae) and their role in seeding upwelling waters. *Estuar. coast. Shelf Sci.* **31**(3): 283–301.
- PITCHER, G. C., BROWN, P. C. and B. A. MITCHELL-INNES 1992 — Spatio-temporal variability of phytoplankton in the southern Benguela upwelling system. In *Benguela Trophic Functioning*. Payne, A. I. L., Brink, K. H., Mann, K. H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 439–456.
- PITCHER, G. C., RICHARDSON, A. J. and J. L. KORRÜBEL 1996 — The use of sea temperature in characterizing the mesoscale heterogeneity of phytoplankton in an embayment of the southern Benguela upwelling system. *J. Plankt. Res.* **18**(5): 643–657.
- PLATT, T., CAVERHILL, C. M. and S. SATHYENDRANATH 1991 — Basin-scale estimates of oceanic primary production by remote sensing: the North Atlantic. *J. geophys. Res.* **96**(C8): 15147–15159.
- PLATT, T., SATHYENDRANATH, S., CAVERHILL, C. M. and M. R. LEWIS 1988 — Ocean primary production and available light: further algorithms for remote sensing. *Deep-Sea Res.* **35**(6A): 855–879.
- SATHYENDRANATH, S., LONGHURST, A. R., CAVERHILL, C. M. and T. PLATT 1995 — Regionally and seasonally differentiated primary production in the North Atlantic. *Deep-Sea Res.* **42**: 1773–1802.
- STORBECK, F. and B. DAAN 2001 — Fish species recognition using computer vision and a neural network. *Fish. Res.* **51**: 11–15.