A novel quantitative, sub-provincial approach to characterizing the shape of chlorophyll profiles

Nonkqubela F. Silulwane

Thesis submitted to the Faculty of Science, University of Cape Town, in fulfilment of the requirements for the degree of Master of Science

September 2001

Abstract

To estimate primary production from satellite, knowledge of the vertical distribution of phytoplankton is needed. As subsurface vertical structure cannot be measured by satellite, representative seasonal profiles have been determined for regions of the world's ocean (biogeochemical provinces) with distinct physical and biological characteristics. Data on vertical chlorophyll profiles from 1993-1998 were obtained from research cruises in the southern Benguela upwelling system. Sampling was conducted from inshore (~20 m depth) to offshore (~300 m depth) over the continental shelf. An exponential form of a shifted Gaussian function was fitted to 592 profiles to estimate four parameters that described the shape of the curve: the background chlorophyll concentration, the depth of the chlorophyll peak, the spread of the peak and the total chlorophyll concentration within the peak. A type of artificial neural network called a self-organizing map (SOM) was used on the four parameters to identify characteristic profiles, and then classify existing profiles into these representative classes. The change in profile shape was also investigated seasonally, within different subregions (West Coast and Agulhas Bank), and with respect to environmental variables such as sea surface temperature, surface chlorophyll, and water column depth. The SOM technique identified a continuum of patterns, ranging from those with small deep peaks to those with large near-surface peaks. Although profile shape varied seasonally, profiles were variable within each season, making seasonally-averaged chlorophyll profiles, as has been used for this and other biogeochemical provinces, inappropriate. The self-organizing map can be used semi-quantitatively to predict the relative frequency of each characteristic profile class under different environmental conditions. To enable a truly quantitative approach, a combination of generalized additive modelling (GAM) and generalized linear modelling (GLM) was used to model the combined effects of environmental variables on profile parameters, regionally and seasonally. Models explain between 15 and 70% of the variance in the shape of chlorophyll profiles. A canonical succession in profile shape following upwelling of cold water was identified, with large surface chlorophyll maxima in cool water and small deep peaks in warm water. The GLM was then used to predict the shape of chlorophyll profiles from environmental variables. Good relationships between observed and predicted values were obtained for the depth of the peak and total chlorophyll concentration within the peak, indicating good predictions for these parameters. In this study, novel approaches such as artificial neural networks and generalized modelling have highlighted the variability in profile shape and enabled its improved prediction. This will lead to superior regional estimates of primary production.

Acknowledgements

I would first like to thank the Department of Environmental Affairs and Tourism, Marine and Coastal Management branch for providing data used in this project. This research would not be possible without funding from the European research programme, ENVIFISH (contract number IC18-CT98-0329), and the National Research Foundation. Special acknowledgement goes to Prof. Trevor Hastie, Professor of Statistics, Stanford University, USA for helping with generalized modelling, his suggestions in solving some of the problems that I encountered in modelling are greatly appreciated. I thank my supervisors, Drs Anthony J. Richardson and Betty A. Mitchell-Innes, and Prof. Frank Shillington for their help, guidance and advice in this study. This work is specially dedicated to my mother "Hlongie" and my late father "Benz", who unfortunately did not see the final product. Special thanks to my wonderful family, especially my sisters, for their support and encouragement throughout my academic career. To my best friend and "soul mate", Rhodes Kgotso Thuwe, for being there for me when I needed him most. My special friends in the Department of Oceanography and Zoology at UCT, Beau M. Tjizoo, Aina Iita, Siphokazi Ndudane, Benedict Dundee, Tembela Mapukata and Nandipha Twatwa, are also thanked for their support and keeping me company when I was away from home. I thank my great friend, Carolina Parada for the wonderful time we had especially the ice-skating trip, "Carly, that was a wonderful experience". I am also grateful to my friends, Ewa Stachlewska and Maya Pfaff for their help and support. Thank you very much to the department of Oceanography, especially Stewart Bernard for his suggestions. A big thank you to Penny Krohn for finding all the research articles that I needed and for good "lobola" conversations. To Leslie Alley for her assistance and motherly comfort in difficult times. Finally, to my wonderful friends, Zimasa Dziba and Nondumiso Sikiti, thanks for being so understanding and considerate; honestly, I would not have finished without your "technological" help.

PUBLICATIONS FROM THIS THESIS

Publications in scientific journals

B. A. Mitchell-Innes, N. F. Silulwane and M. I. Lucas 2001 – Variability of Chlorophyll Profiles on the West Coast of Southern Africa in June/July 1999. *South Afracan Journal of Science* **97**: 246–250.

Silulwane, N. F., Richardson, A. J., Shillington, F. A. and B. A. Mitchell-Innes 2001 - Identification and classification of vertical chlorophyll patterns in the Benguela upwelling system and Angola-Benguela front using an artificial neural network. In A Decade of Namibian Fisheries Science. Payne, A.I.L., Pillar, S.C. and R.J.M. Crawford (Eds). *South African Journal of marine Science* **23**: 37-51.

Richardson, A.J., Pfaff, M.C., Field, J.G., Silulwane, N.F. and F.A. Shillington (in press) – Identifying characteristic chlorophyll *a* profiles in the coastal domain using an artificial neural network. *Journal of Plankton Research.*

Publications in popular newsletter

Richardson, A. J., Shillington, F. A, Stachlewska, I. S., Silulwane N. F.and M. Pfaff. *South African Network for Coastal and Oceanic Research Newsletter* 164, August 2000. Artificial neural networks in marine science.

PUBLISHED ABSTRACTS BASED ON THIS RESEARCH

Silulwane, N. F., Richardson, A. J., Mitchell-Innes, B. A. and F. A. Shillington. Identification and classification of vertical chlorophyll patterns in the southern Benguela upwelling system using a self-organizing map. *ENVIFISH/VIBES Workshop on Environmental Conditions and Pelagic Fish in the Benguela and Angolan Systems*, 18-20 October 2000, Cape Town, South Africa.

Silulwane, N. F., Richardson, A. J., Mitchell-Innes, B. A. and F. A. Shillington. Identification and classification of vertical chlorophyll patterns in the northern Benguela upwelling system using a self-organizing map. *Symposium 2000: A Decade of Namibian Fisheries Science*, 10-12 October 2000, Swakopmund, Namibia.

Silulwane, N. F., Mitchell-Innes, B. A., Richardson, A. J. and F. A. Shillington. Estimation of integrated chlorophyll *a* from SST: an index of productivity that may be related to pelagic fish recruitment. 10th Southern African Marine Science Symposium, 22-26 October 1999, Wilderness, South Africa.

PRESENTATIONS BASED ON THIS RESEARCH

Oral presentations at international conferences

Silulwane, N. F., Richardson, A. J., Mitchell-Innes, B. A. and F. A. Shillington. Predicting chlorophyll profile shapes in the Benguela upwelling system using a self-organizing map. SPACC/IDYLE/ENVIFISH conference, Globec Special Report, 6-8 September 2001, Cape Town, South Africa.

Silulwane, N. F., Richardson, A. J., Mitchell-Innes, B. A. and F. A. Shillington. Identification and classification of vertical chlorophyll patterns in the southern Benguela upwelling system using a self-organizing map. *ENVIFISH/VIBES Workshop on Environmental Conditions and Pelagic Fish in the Benguela and Angolan Systems*, 18-20 October 2000, Cape Town, South Africa.

Silulwane, N. F., Richardson, A. J., Mitchell-Innes, B. A. and F. A. Shillington. Identification and classification of vertical chlorophyll patterns in the northern Benguela upwelling system using a self-organizing map. *Symposium 2000: A Decade of Namibian Fisheries Science*, 10-12 October 2000, Swakopmund, Namibia.

Oral presentations at international organizations and Universities

Richardson, A. J., Silulwane, N. F, M. Pfaff, I. S. Stachlewska and F. A. Shillington. Predicting the shape of sub-surface chlorophyll profiles in the Benguela upwelling system using some novel quantitative approaches. 10 May, 2001, Marine Biological Association of the United Kingdom, Plymouth, United Kingdom.

Richardson, A. J., Silulwane, N. F, M. Pfaff, I. S. Stachlewska and F. A. Shillington. Using artificial neural networks to identify patterns in data: examples using sea surface temperature and chlorophyll profiles. Presented at:
University of Flinders, Adelaide, Australia. 29 November, 2000.

University of Queensland, Brisbane, Australia. 4 December 2000.

University of Griffith, Gold Coast, Australia. 6 December 2000.

Poster presentation at an international conference

Silulwane, N. F., Mitchell-Innes, B. A., Richardson, A. J. and F. A. Shillington. Estimation of integrated chlorophyll *a* from SST: an index of productivity that may be related to pelagic fish recruitment. 10th Southern African Marine Science Symposium, 22-26 October 1999, Wilderness, South Africa.

TABLE OF CONTENTS

Chapter Three: Results

eastern Agulhas Bank, (b) western Agulhas Bank, and (c) West Coast sub-provinces. The relative frequencies are calculated separately for each sub-region of the Agulhas bank. All regional levels add up to 100%.

Figure 3.15. Relative frequencies of chlorophyll patterns at different seasons: (a) autumn, (b) spring and (c) summer. The relative frequencies are calculated separately for each season with seasonal levels adding up to 100%.

Figure 3.16. Relative frequency maps for different categories of SST: (a) Low SST (10-14°C), (b) medium (14-18 °C), and (c) high (> 18°C) with levels adding up to 100%.

Figure 3.17. Relative frequency maps for different categories of surface chlorophyll concentration: (a) Low (< 1 mgm^{-3}), (b) medium (1-7 $mg.m^{-3}$), and (c) high (> 7 mgm^{-3}) with levels add up to 100%.

Figure 3.18. Relative frequency maps for different categories of water column depths: (a) Shallow (< 80 m), (b) medium (80-170 m) and (c) deep (> 170 m) with levels adding up to 100%.

Figure 3.19. Relationships between background chlorophyll concentration (B_0) and (a) SST (°C), (b) surface chlorophyll concentration ($mg.m^{-3}$), and (c) water column depth (m).

Figure 3.20. Relationships between width of the peak (σ) and (a) SST, (b) surface chlorophyll concentration, and (c) water column depth.

Figure 3.21 Relationships between total chlorophyll concentration within the peak (h) and (a) SST, (b) surface chlorophyll concentration, and (c) water column depth.

Figure 3.22. Relationships between the depth of the peak (z_m) and (a) SST, (b) surface chlorophyll concentration, and (c) water column depth.

Figure 3.23. A generalized additive model of the background chlorophyll concentration (B_0) modelled with surface chlorophyll concentration and water column depth.

List of Tables

Chapter 1: General Introduction

1.1 Estimation of ocean primary production

The survival of organisms at higher trophic levels depends on the supply of phytoplankton (primary production) energy and organic matter, therefore studies central to understanding phytoplankton biomass distribution and primary production estimates are important for most biological oceanographic processes (Kuring *et al.* 1990). The most important elements required to expand our understanding of phytoplankton biomass and primary production, are the surface and vertical chlorophyll structures and photosynthetic parameters, and their distribution on seasonal and regional scales (Platt *et al.* 1991).

The distribution and variability of phytoplankton biomass, and computation of primary production, have used *in situ* chlorophyll data from ship and satellite measurements. Using-remotely sensed ocean colour data from satellites, such as SeaWiFs and Coastal Zone Colour Scanner (CZCS), has improved the understanding of biological variability in the ocean and also proved the viability of using satellite data for estimating ocean primary production (Platt *et al.* 1991). However, there are a number of problems with satellite derived ocean colour data. First, areas with significant cloud cover tend to bias estimates (Longhurst *et al.* 1995). Second, in some studies, small errors in atmospheric correction of remotely sensed data have resulted to large errors of primary production (Gordon 1993). Last, in coastal areas, high concentrations of organic matter as well as chlorophyll degradation products are sometimes misinterpreted as chlorophyll (Sathyendranath and Morel 1983).

In situ data from research cruises provide vertical chlorophyll structure that cannot be obtained from remote sensing by satellite sensors. These *in situ* data are by no means perfect, meaning that difficulties are experienced, which include high light levels in the surface layers. This leads to inaccuracies in estimates of surface chlorophyll, and thus primary production (Cullen and Lewis 1995). Computation of primary production requires both the vertical and surface chlorophyll fields, with *in situ* data extrapolated to match the satellite data temporally and spatially. For extrapolation to areas with no measurements, the ocean can be considered either as a biological continuum or having spatial discontinuities between ecological entities (Platt *et al.* 1995, Sathyendranath *et al.* 1995). In future, remote sensing is likely to be the tool used to understand phytoplankton productivity and the synoptic state of pelagic ecosystems (Platt *et al.* 1995).

Empirical and analytical models for estimating global primary production require *in situ* data with the subsurface chlorophyll structure and remotely sensed surface chlorophyll as inputs for the estimations (Platt *et al.* 1988, Platt *et al.* 1995, Sathyendranath *et al.* 1995, Millán-Núñez *et al.* 1997). Increased estimates of phytoplankton biomass and primary production have been evident when a consistent deep chlorophyll maximum is included in the estimation, compared to underestimation of primary production when the homogeneous biomass distribution is present (Sathyendranath *et al.* 1995, Millán-Núñez *et al.* 1997).

To estimate global primary production, Longhurst *et al.* (1995) and Sathyendranath *et al.* (1995) partitioned the ocean into four primary domains, which were further subdivided into 57 secondary biogeochemical provinces. Each biogeochemical

province was characterized by a single seasonal profile. Boundaries between provinces varied to accommodate seasonal, annual and decadal-scale changes in ocean circulation. These domains were constructed based on the assumptions that the primary forcing of algal blooms is linked to wind mixing, stratification and sea surface irradiance (Longhurst *et al.* 1995), so that each domain has unique physical characteristics that play a major role in the ecology of phytoplankton. The four domains of the global pelagic ecosystem, Polar, West-Wind, Trade-Wind and Coastal-Boundary domains, are briefly discussed below. These four domains are further divided into biogeochemical provinces. Biogeochemical provinces are defined as compartments with seasonal uniformity of the major physical processes that govern their biological processes (Sathyendranath *et al.* 1995).

1.1.1 Global domains of pelagic ecosystems

The most important feature of the high-latitude Polar domain, is the presence of a brackish surface layer caused by melting of the winter ice cover, mainly in spring and early summer, leading to active algal bloom development. This fresher layer induces stability, initiating a bloom as soon as high surface light is available. This feature has been clearly seen in ocean colour satellite images (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). The West-Wind domain constitutes the mid-latitude oceanic regions where strong westerly winds induce winter mixing, resulting in deep mixed layers (>500 m) in winter. Phytoplankton dynamics are controlled by local physical forcing. For instance, spring blooms occur as the wind stress decreases and more light penetrates to subsurface layers (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). The Trade-Wind domain comprises the low-latitude provinces, and the feature of this domain is the shallow tropical pycnocline with very high stability

caused by positive heat flux across the sea surface. Unlike in the West-Wind domain, blooms in mixed layers are generally not light limited (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). As this study focuses on a coastal shelf system, the Coastal-Boundary domain will be discussed in some detail.

1.1.2 Coastal-Boundary domain

The coastal domain comprises provinces where the general ocean circulation is strongly modified by the interaction between the coastal topography and coastal wind regime. This domain has the largest number of provinces. Features of this domain are coastal upwelling, anticyclonic eddy fields and shelf-break fronts. Processes that account for algal bloom developments are more diverse than the other domains because of extreme wind stress, shelf-break and coastal upwelling, tidal fronts and mixing, and bathymetric features (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). Difficulties in allocating boundaries between provinces were encountered because of these processes as well as the complex coastal and shelf topography (Mittelstaedt 1991, Longhurst *et al.* 1995, Sathyendranath *et al.* 1995).

1.1.2.1 The Benguela Current Coastal province

One province within the coastal domain is the Benguela Current Coastal province (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). This province is on the West Coast of southern Africa and covers the region between 14°S and 37°S (Shannon and Nelson 1996) characterized by episodic wind-driven coastal upwelling. Seven upwelling cells as well as offshore filaments and eddy-fields are found within this province (Nelson and Hutchings 1983, Shannon and Nelson 1996). The Benguela

upwelling system has well defined boundaries in the north, the Angola-Benguela frontal region (14-17°S), and in the south, the Agulhas retroflection area (36-37°S) (Shannon and Nelson 1996).

1.2 Study Areas

Study areas include the West Coast of South Africa from Cape Point (34°S) northwards (29°S), and the Agulhas Bank. The West Coast is a sub-province of the Benguela Current Coastal province, which experiences maximum wind-driven upwelling in summer (Shannon 1985). The Agulhas Bank covers the continental shelf area between Cape Point (18°E) in the west and Port Alfred (27°E) in the east. The Bank is mainly subdivided into two parts, the western and the eastern Agulhas Bank (Probyn *et al.* 1994). The western Agulhas Bank, from Cape Point to Cape Agulhas is the most southerly region of the Benguela upwelling system (Shannon 1985). The eastern part with shelf-break upwelling, forms the outer shelf of the Agulhas Bank (Largier *et al.* 1992). These two Agulhas sectors have different temperature regimes related primary to the Benguela coastal upwelling in the west and the Agulhas current in the east (Swart and Largier 1987).

1.3 The use of SST to estimate ocean primary production

Satellite-derived sea surface temperature (SST) has been used to estimate indirectly surface nitrate concentration useful for the calculation of new production in the North Atlantic Ocean (Dugdale *et al.* 1989, Sathyendranath *et al.* 1991). The use of SST to predict other environmental parameters had been successfully applied in a variety of oceanographic regimes (Platt *et al.* 1995). The reason for using SST for estimating

other parameters in upwelling areas is easy to understand. For example, SST is an indicator of the age of upwelled water; SST of recently upwelled water is cool and as upwelling water matures, SST warms (Platt *et al.* 1995). In the southern Benguela upwelling region, predictions of nitrate content of the euphotic zone from the temperature-nitrate relationships used SST from satellites, and these predictions were then used to derive estimates of potential new production (Waldron and Probyn 1992). It is believed that remotely-sensed SST can be used to predict surface chlorophyll and phytoplankton pigment structure also useful for estimating primary production.

1.4 Importance of upwelling areas

The four main wind-driven upwelling areas in the world are the Benguela Current system, the California Current, the Humboldt Current and the Canary Current (Lluch-Belda *et al.* 1989, Hill *et al.* 1998, Shillington 1998). These upwelling areas have high chlorophyll concentrations, and thus high phytoplankton biomass, which make them suitable for pelagic fish spawning. Upwelling areas are also nursery and recruitment grounds for pelagic larvae and fish with pelagic fish representing one-third of the world's total marine fish catch (Lluch-Belda *et al.* 1989). The Benguela upwelling system has a wider continental shelf than other upwelling areas, although it is less productive in fish biomass (Hutchings 1992). This suggests that the size of the continental shelf has less importance for productivity, whereas environmental and biological processes play a vital role in pelagic fish production in upwelling areas (Cury *et al.* 1998).

1.5 Phytoplankton community structure of the southern Benguela region

Phytoplankton are responsible for more than 95% of ocean carbon fixation (Falkowski 1994), which accounts for ~40% of the global carbon fixation (Falkowski and Kolber 1995). Systematic trends of phytoplankton community structure in the southern Benguela region exist in taxa above the species level, mainly in diatoms and flagellates (Pitcher *et al.* 1991, Mitchell-Innes and Walker 1991, Pitcher *et al.* 1992); and their dominance can be predicted qualitatively once changes in vertical stability occur (Mitchell-Innes and Walker 1991). Diatoms are commonly found in nutrient-rich water inshore with the bloom development associated with pulses of upwelling (Brown and Field 1986, Brown and Hutchings 1987), resulting in new production (Probyn 1992). Diatoms are able to sink and therefore remain in deep nutrient-rich water columns close to upwelling centres (Pitcher 1990). Flagellates dominate offshore with stable low phytoplankton biomass in the water column basically supported by recycling of nutrients (Probyn *et al.* 1990, Probyn 1992, Mitchell-Innes and Pitcher 1992). Flagellates, particularly dinoflagellates, congregate at offshore fronts during upwelling, then move shorewards as offshore winds relax. Such relaxation of equatorward winds results in cessation of upwelling. An advantage dinoflagellates have over diatoms is their ability to move vertically in the water column and remain within a band of suitable conditions in order to balance light and nutrient uptake requirements (Pitcher *et al.* 1996).

1.6 Importance of phytoplankton and chlorophyll *a*

Upwelling areas do not only serve as spawning grounds for pelagic fish, but they also provide pelagic fish with a good food environment. The main food for sardine is

phytoplankton (Van Der Lingen 1994) which increases in biomass after upwelling in the surface layers because of high light availability, increased nutrients and warming of surface water (Pitcher *et al.* 1992). High variability in spatial and temporal distribution of phytoplankton is mainly controlled by the interaction of physical and biological factors, which affect the upper mixed layers and the photosynthetically active radiation distribution (Falkowski and Woodhead 1992).

The spawning seasons of sardine and anchovy both coincide with the spring phytoplankton bloom in September or October on the Agulhas Bank. During the spring bloom, enhanced phytoplankton growth occurs, particularly over the inner shelf. The spring bloom seems to provide a good food environment for copepods and adult anchovy and sardine (Mitchell-Innes *et al.* 1999). Successful spawning by anchovy also depends on a stable food supply, consisting mainly of zooplankton (James 1987, Hutchings 1992), which feed mainly on phytoplankton, and then support anchovy serial spawning (Peterson *et al.* 1992, Richardson *et al.* 1998).

Chlorophyll *a* concentration is a measure of the amount of food available to zooplankton and fish in pelagic ecosystems (Armstrong *et al.* 1991, Mitchell-Innes and Pitcher 1992). Chlorophyll *a* is used as an index of phytoplankton biomass (Cullen 1982, Sathyendranath *et al.* 1995) especially in highly productive upwelling areas. The depth-integrated chlorophyll *a* in the upper 30 m sometimes is computed and used as an estimate of phytoplankton biomass (Brown 1992). Remotely-sensed surface chlorophyll concentrations have been used to predict the total chlorophyll in the euphotic zone, and also the chlorophyll distribution as a function of depth (Morel and Berthon 1989).

Spatial and temporal variability in chlorophyll distribution is central to our understanding of phytoplankton biomass and ocean primary production. Regional differences in the vertical distribution of chlorophyll cause major difficulties in deriving precise estimates of phytoplankton biomass and oceanic productivity from surface measurements (Falkowski and Kolber 1995). Techniques, such as artificial neural networks, have been recently applied to estimate oceanic chlorophyll concentrations from satellite-derived surface chlorophyll (Keiner and Brown 1999). Neural networks have also been used to highlight phytoplanton biomass variability in the Benguela upwelling system (Silulwane *et al.* 2001, Richardson *et al.* in press).

1.7 Aims

Aims of this study are:

1. To characterize the shape of vertical chlorophyll profiles of the continental shelf region of the West Coast and Agulhas Bank sub-provinces.
2. To investigate the temporal and spatial variability of *in situ* measured chlorophyll profiles in relation to a range of environmental parameters (e.g. sea surface temperature, surface chlorophyll concentration and water column depth).
3. To predict chlorophyll profile shape from pertinent environmental parameters that are known or can be easily measured (from satellites).

1.8 Thesis outline

The study parameterizes the shape of chlorophyll profiles of the southern Benguela upwelling region and the Agulhas Bank. A type of a neural network is used to highlight variability in vertical chlorophyll structure, based on the environment of these regions. Quantitative techniques such as generalized additive and generalized

these regions. Quantitative techniques such as generalized additive and generalized linear models (GAM and GLM) are also used to model and predict the shape of chlorophyll profiles from environmental information. The methodology outlined in this thesis will allow improved regional primary production estimates in the Agulhas Bank and Benguela upwelling system. Moreover, it provides a framework that can be used in other coastal domains for estimating subsurface chlorophyll structure. This is a first study that uses novel quantitative techniques to identify and highlight variability in phytoplankton pigment structure as well as predicting profile shapes from environmental variables within the context of global primary production (see Platt *et al.* 1988, Longhurst *et al.* 1995 and Sathyendranath *et al.* 1995).

Chapter 2 is the methods chapter, which presents profiles and environmental data collected from cruises, and techniques used to model profiles in order to obtain parameters (that describe the shape of these profiles) important for estimating primary production. This chapter also covers the background and use of a semi-quantitative technique, the self-organizing map (SOM), to identify typical chlorophyll patterns, and to characterize and classify these identified patterns into classes based on different environmental categories.

A self-organizing map (SOM) is a type of an artificial neural network, which is particularly adept at pattern recognition (Kohonen *et al.* 1995). In this study, a SOM is used to characterize chlorophyll profiles of the southern Benguela and Agulhas Bank sub-provinces in relation to the physical environment. This study does not try to describe the SOM technique, as this has been extensively covered in the literature (see Kohonen 1989, Kohonen 1990, Kohonen 1997), but uses the technique as a useful

tool in characterizing the variability in phytoplankton pigment structure for use in primary production in future.

Chapter 2 also deals with modelling the combined effects of environmental variables on profile parameters using quantitative techniques such as generalized additive and generalized linear modelling (GAMs and GLMs), as well as predicting the shape of profiles from these environmental variables using GLMs. The relationship between profile parameters and environmental variables (SST, surface chlorophyll concentration and depth of the water column) at regional and seasonal scales was investigated using statistical models such as the generalized linear model (GLM) and generalized additive model (GAM).

In the current study, GAM and GLM are used to model (simultaneously) the effect of each environmental variable on vertical chlorophyll distribution because physical and biological processes are interrelated (Morel and Berthon 1989, Brown and Hutchings 1987, Pitcher *et al.* 1992). GAMs have been particularly used to capture the nonlinear trends of the parameter values given a set of environmental variables. These GAMs have also been used to determine the most significant environmental variables for modelling profile parameters. GLMs have been applied to obtain an equation for predicting the parameter values from an independent set of environmental variables.

Results are presented in Chapter 3, which covers the semi-quantitative ability of the SOM in relating phytoplankton pigment structure (in terms of chlorophyll profile parameters) to environmental variables. The possibility of using a single typical profile for each season in future primary production estimates in the Benguela

province as suggested and conducted in other studies (Platt *et al.* 1991, Longhurst *et al.* 1995, Sathyendranath *et al.* 1995) is reported. Chapter 3 also reports on variability in phytoplankton pigment structure highlighted by the SOM technique, and characteristic profiles of the southern Benguela and Agulhas Bank sub-provinces at seasonal scales. The results' chapter also focuses on the quantitative approach of GAMs and GLMs in modelling profile parameters as functions of environmental variables, and the capability of GLMs in predicting profile parameters from environmental variables.

Discussion of the results using different quantitative techniques (used in Chapter 3) is given in Chapter 4. The advantages and disadvantages of using the SOM technique in characterizing chlorophyll profiles are discussed. The use of profile parameters in the SOM analysis is justified. The adequacy of using mean seasonal profiles in primary production estimates is mentioned in this chapter. The impact of the physical environmental factors, such as upwelling, light intensity and coastal morphology, in spatial and temporal distribution of phytoplankton pigment structure of the southern Benguela and Agulhas Bank sub-provinces is presented. Modelling of profile parameters from environmental variables using GAM and GLM are also discussed in Chapter 4. The predictive success of profile parameters using GLM from these environmental variables is also mentioned. Chapter 4 also presents the conclusion and future research that can be conducted to improve primary production estimates in the Benguela Current Coastal province, and other biogeochemical provinces in the dynamics of coastal domains. For future profile parameter predictions in the Benguela, the use of other environmental variables, such as satellite-derived chlorophyll and sea surface temperature are suggested.

Chapter 2: Material and Methods

2.1 Data collection

Oceanographic data were collected on research cruises off the west and south coasts of South Africa (Fig. 2.1) by Marine and Coastal Management as part of both the South African Sardine and Anchovy Recruitment Programme, and the Spawner Biomass surveys. Most oceanographic data were from the continental shelf, although some data were collected in deeper water waters. The analysis was restricted to water columns ≤ 300 m deep to focus on shelf processes. Table 2.1 shows the continental shelf profiles collected from different cruises over the west and south coasts of South Africa. Generally, each transect started two miles from the coast and extended to beyond the 200 m isobath, with stations 10 miles apart. Transects were completed within a 12-hour period. Continuous fluorescence profiles down to 100 m or to within 10 m of the seabed were measured at each station (Fig. 2.1) by a thermistor and profiling fluorometer (Chelsea Instruments AquaTracka MKIII). Chlorophyll *a* samples were collected in Niskin bottles as single samples at the surface and at the maximum fluorescence depth. These water samples were filtered using Whatmann GF/F filters and then extracted in 90% acetone at 20°C for 24 hours. These samples were then measured fluorometrically before and after addition of hydrochloric acid to adjust for phaeopigments using a Turner designs Model 10-000R fluorometer (Parsons *et. al.* 1984).

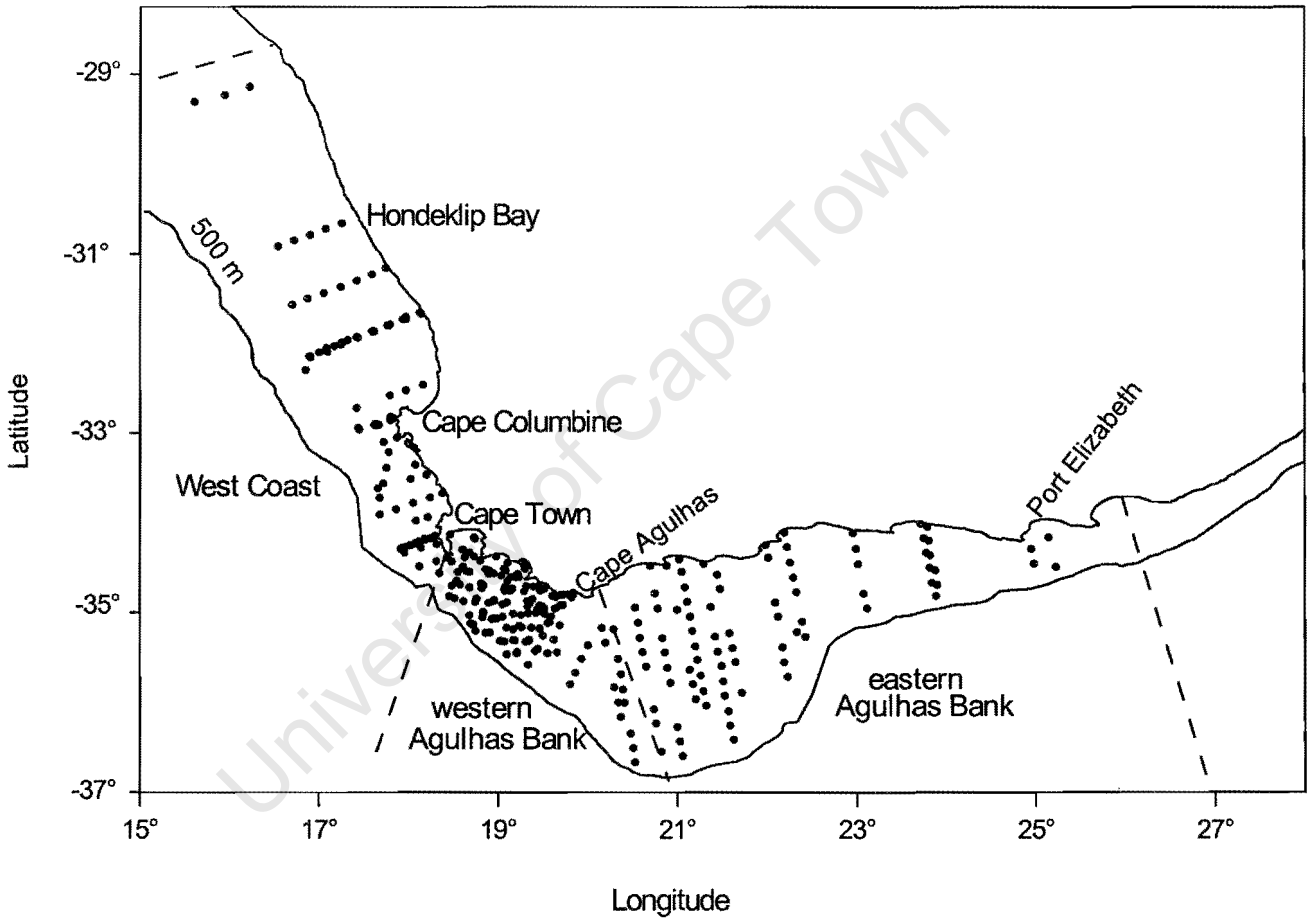


Figure 2.1. Positions of chlorophyll profiles collected from different research cruises over the Agulhas Bank and West Coast of South Africa.

Table 2.1. A summary of the continental shelf chlorophyll profiles collected from different research cruises. Areas for data sampling are EAB (eastern Agulhas Bank), WAB (western Agulhas Bank) and WC (West Coast of South Africa). SARP is the Sardine and Anchovy Recruitment Programme

Cruises	Area sampled	Years	Months	Number of profiles
SARP	WAB and WC	1993/1994 1994/1995	Sept, Oct, Dec, Jan, Feb, and March	309
Spawner Biomass	EAB, WAB and WC	1993-1996, 1998	November	283

Photoinhibition occurs when high surface light causes a reduction in fluorescence measurements, which result to a bias in *in situ* estimates of chlorophyll and primary production. Photoinhibition basically causes changes in the optical characteristics of phytoplankton (Cullen and Lewis 1995) and a rapid decrease in fluorescence per unit of chlorophyll *a* (Falkowski and Kolber 1995). To address photoinhibition, calibrations of fluorescence data for each cruise were done by linearly regressing fluorescence and extracted surface chlorophyll *a* at the surface and maximum flourescence depth for each of the western Agulhas Bank (WAB), eastern Agulhas Bank (EAB) and West Coast (WC) of South Africa. High extracted surface chlorophyll *a* values corresponding to low *in situ* fluorescence measurements were

considered photoinhibited (as indicated in the linear regression), and were excluded from the regression. This led to a further exclusion from all subsequent analyses of profiles corresponding to photoinhibited values. For areas with a very poor linear regression (low r^2), each cruise was separated into different transects (lines) and these were then calibrated to compensate for apparent photoinhibition. The transect with the highest r^2 value from the linear regression was then used to calibrate the whole region. After excluding photoinhibited profiles, 592 profiles (from a total of 689) were left for use in further analysis.

Underwater irradiance was measured concurrently with 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 surface irradiance. The mixed layer depth was calculated as the depth where the difference in temperature from the surface was 0.5 °C. At each station, pertinent environmental variables were collected, including sea surface temperature (10.1-22.4°C) and depth of the water column (≤300 m).

The study area covers the West Coast of South Africa, which is a sub-province of the Benguela Current Coastal province, and the Agulhas Bank. The Agulhas Bank is further subdivided into the western Agulhas Bank, which is the southern-most region of the Benguela upwelling system, and the eastern Agulhas Bank, which experiences little upwelling compared to the western part. Although the Agulhas Bank is regarded as a sub-province on its own, owing to the paucity of chlorophyll profile data (only 592) available for conducting this study, the Agulhas Bank sub-province data were combined with the West Coast sub-province for the purposes of this study. In

addition, the use of semi- and quantitative techniques, such as the SOM, requires large data sets because they can be easily handled by these techniques. In addition, it is believed that analyses conducted with large data sets should give more meaningful results compared to those from small data sets.

2.2 Parameterizing vertical chlorophyll profiles

Fluorescence profiles were not collected at standard depths because of ocean currents, differing cable speeds and rolling of the ship during sampling making it difficult to make direct profile comparisons. Another problem that was experienced was that of variability in inshore and offshore profiles with depth. To enable comparison and classification of profiles, it was then decided that profiles should be standardized. Two standardization methods were possible. One approach would be to interpolate profiles to standard depths and then adjust for water column depth. A second, simpler approach is that of parameterizing chlorophyll profiles by the shifted Gaussian model (Platt *et al.* 1988, Longhurst *et al.* 1995, Platt and Sathyendranath 1995, Sathyendranath *et al.* 1995). This simpler approach was preferred, and therefore vertical chlorophyll profiles (592) from different sub-regions and seasons (Table 2.1) were parameterized by fitting a shifted Gaussian function to each profile. This four-parameter function is expressed as (Platt *et al.* 1988, Platt and Sathyendranath 1995):

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

where $B(z)$ = chlorophyll biomass as a function of depth ($mg.m^{-3}$), B_0 = background chlorophyll concentration ($mg.m^{-3}$), h = total chlorophyll beneath the curve ($mg.m^{-2}$), σ = standard deviation of the peak (m) and z_m = depth of the chlorophyll peak (m). Figure 2.2 depicts the shifted Gaussian curve used for phytoplankton biomass parameterization. The effect on profile shape of changing each of these parameters is shown in Figure 2.3.

2.2.1 Non-linear curve fitting

A non-linear estimation procedure was used to get parameter values that best define the shape of each vertical chlorophyll profile. This non-linear estimation algorithm is an iterative procedure that tries to find best values of the Gaussian curve. This iterative procedure stops when the loss function, which is the least-squares estimate of the fit of the model, is small. The loss function is calculated as: $L = \sum(Obs - Pred)^2$. A small value for the loss function means that the model fits the data well and a large loss function means the model fits the data poorly. If the loss function is still large, then the algorithm changes the parameter values and recalculates the loss function until it is within the convergence criterion. The convergence criterion specifies how small the loss function needs to be when fitting the function in order to provide a good fit to the data.

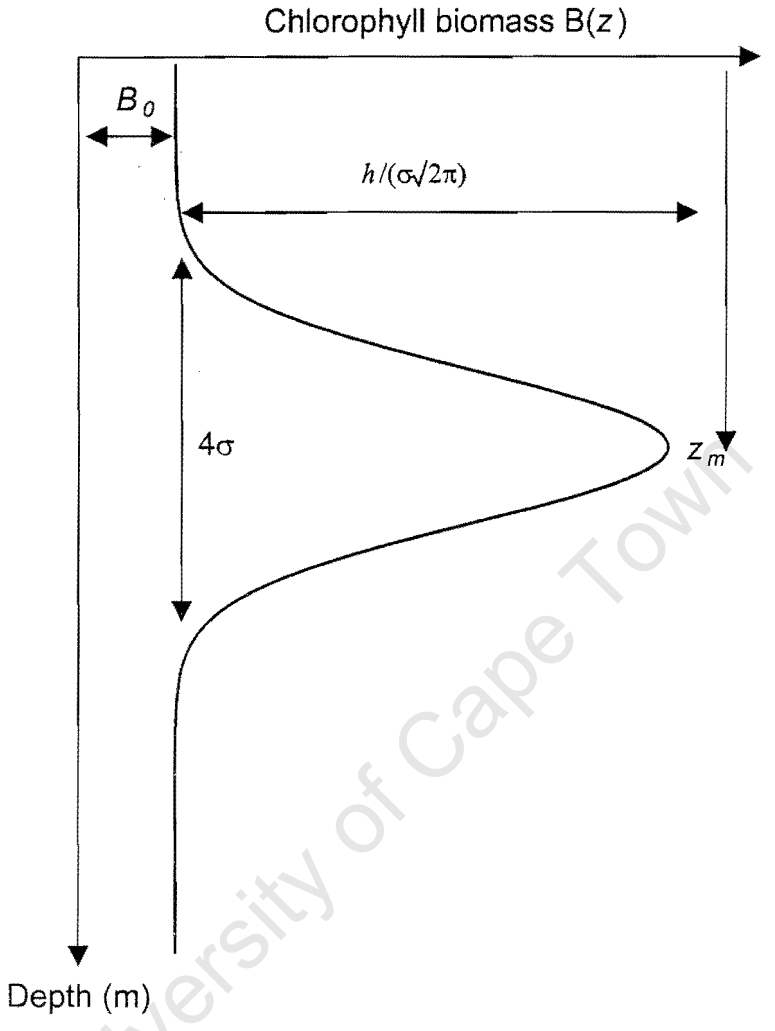


Figure 2.2. A shifted Gaussian curve showing the four parameters (B_0, σ, h and z_m) used to describe *in situ* vertical chlorophyll profiles.

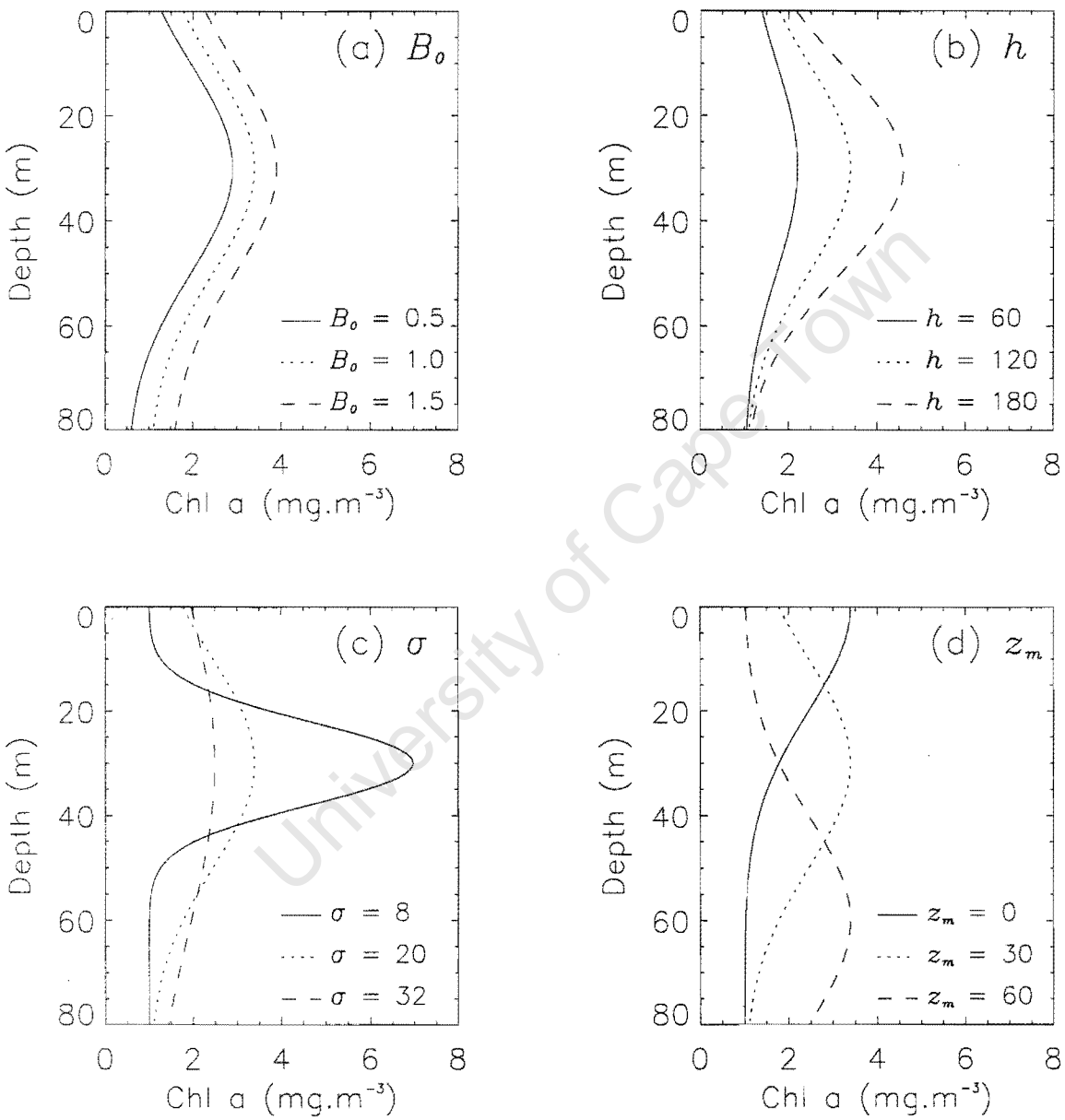


Figure 2.3. The effect of changing parameter values on the shape of the Gaussian model. Constant parameter values are B_0 = 1mg.m^{-3}, h = 120 mg.m^{-2}, σ = 20 m and z_m = 30 m, except for the parameter that is being changed.

By recalculating the loss function, the algorithm is trying to find a point where the model best fits the data. When the convergence criterion is met, the procedure stops (estimation converges) and the parameter values at that point are the best ones for the model. The amount that the parameter values are changed in each model step is the step size.

Fitting curves with non-linear estimation does not always result to converged algorithms, which means parameter estimates are meaningless or they balloon to infinity. When this happens, the initial values of the parameters, which are the initial step size and a convergence criterion, can be changed and the fitting procedure is restarted. In this study, the main algorithm used for curve fitting was *Quasi Newton*. This algorithm has (for all parameters) the start value of 0.1 and the initial step size of 0.5. The convergence criterion was chosen to be 0.0001. When the algorithm did not converge to meaningful estimates, the *Hooke-Jeeves patterns* algorithm was used, which required different start values and step sizes for parameters to be estimated. Although the *Hooke-Jeeves patterns* algorithm was rather slow compared to the *Quasi-Newton*, it managed to fit all the curves satisfactorily. Profiles were fitted in STATISTICA 5.5, using the non-linear estimation module, and the user-specified regression was chosen for curve fitting.

2.2.2 Fitting vertical chlorophyll profiles using the exponential Gaussian function

Fitting the shifted Gaussian function to chlorophyll profiles as in Platt *et al.* (1988), Longhurst *et al.* (1995), Platt and Sathyendranath (1995) resulted in occasional negative values for parameters. However, B_0, h and σ are not defined for negative values: B_0 (mg.m^{-3}) is a chlorophyll *a* concentration; h (mg.m^{-2}) is the total biomass

above the background chlorophyll concentration; and σ (m) is the width of the peak. A negative z_m parameter for the Gaussian function is interpreted as z_m being above the surface of the water. In reality, the peak of the biomass is at the surface, with the imaginary chlorophyll maximum of the generalized profile located above the surface (Platt *et al.* 1991). For our data, when z_m was negative, the h values became unrealistically large (> 10000). By constraining the value of z_m to be positive, more realistic h values were obtained. This had very little effect on the proportion of variance explained by the curve (< 3% difference). To constrain all parameters to be positive, an exponential form of a Gaussian function was used (see Platt and Sathyendranath (1995) for its derivation). The formula is:

$$B(z) = e^{c} + \frac{e^{a}}{e^{b}\sqrt{2\pi}} e^{-\frac{(z-z_m)^2}{2(e^{b})^2}}$$

where $B_0 = e^c$, $h = e^a$, $\sigma = e^b$ and $z_m = e^d$. The exponential shifted Gaussian function converged more easily than the standard shifted Gaussian curve. Although more profiles could be satisfactorily fitted with the exponential curve, some of the curves fitted the profiles poorly, and these were eliminated. Profiles with two or more peaks or with a lot of scatter were not included in further analysis. Profiles were also rejected that had chlorophyll spikes caused by gelatinous masses of the colonial diatom *Thallasiosira* sp. (Mitchell-Innes *et al.* 1999). After curve fitting, only profiles with variance explained (r^2 value) ≥ 80% were considered, leading to 435 profiles (from a total of 592 profiles) to be used in further analyses. The latest Spawner Biomass survey (November 1998) was chosen as an independent data set, which was used to predict profile parameters in further analysis. Chlorophyll profiles from this

survey were from the spring season, but collected at different sub-regions (eastern Agulhas Bank, western Agulhas Bank and West Coast of South Africa), and these represented ~20% of profiles with $r^2 \geq 80\%$ after curve fitting.

2.3 Pattern recognition using artificial neural networks

To identify and classify typical *in situ* chlorophyll profiles of the West Coast and Agulhas Bank sub-provinces, an artificial neural network was used.

2.3.1 Artificial neural networks

Artificial neural networks (ANNs) are computer algorithms used for analysing nonlinear data, with nonlinear data referring to nonlinear relationships between dependent and independent variables. ANNs simulate the functioning and operation of their biological counterpart, the brain. Unlike traditional computer algorithms, ANNs have the ability to learn and generalize from the given information, and this makes them powerful mathematical tools (Hewitson and Crane 1994). Generally, they classify data more accurately than conventional statistical methods (Hewitson and Crane 1994). A neural network can identify patterns provided to it, and even previously unseen cases correctly (Foody 1999). Two particular advantages of neural networks are the ability to filter noise in the data as well as working with large data sets (Gross *et al.* 2000). Unlike nonlinear regression methods, neural networks do not require a *priori* knowledge of the nature of the data set. Statistical assumptions such as normal distribution, linearity and equal variances are not a prerequisite when using neural networks (Chen and Ware 1999). The primary disadvantage of neural networks

is the lack of statistical rigour such as absence of significance levels. Neural networks have been widely used in diverse areas, including pelagic fish biomass and recruitment forecasting (Jarre-Teichmann *et al.* 1995, Chen and Ware 1999), stock market prediction, synoptic climatology studies of the Southern Hemisphere (Crane and Hewitson 1994, Main 1997). Relatively few applications of neural networks have been conducted in studies of phytoplankton biomass and primary production. Most applications to primary production studies have used a network called a multi-layer perceptron (MLP) for prediction of near-surface phytoplankton pigment concentration and primary production using satellite-derived surface chlorophyll (Foody 1999, Keiner and Brown 1999, Gross *et al.* 2000, Belgrano *et al.* 2001, Buckton *et al.* 2001). The MLP is the most popular network that performs supervised learning, and is mainly used for predictions (Jarre-Teichmann *et al.* 1995).

In this study, a different type of neural network is used, called a self-organizing map (SOM) (Fig. 2.4) that is particularly adept at pattern recognition. A SOM, which is also used for data exploration approaches, performs unsupervised learning (Hewitson and Crane 1994). A SOM consists of a regular array of nodes (patterns), with each node having a reference vector that is equal in dimension to the input data (Hewitson and Crane in press). In this study, each reference vector for each node has four-elements which correspond to the four parameters from the Gaussian curve.

Artificial neural networks, such as a SOM, can generally handle large and more noisy data than standard statistical methods used for data exploration (Cluster analysis and multidimensional scaling). The SOM has the ability to capture non-linear patterns and to span breakpoints in the data. For identifying groups in a data set, the primary

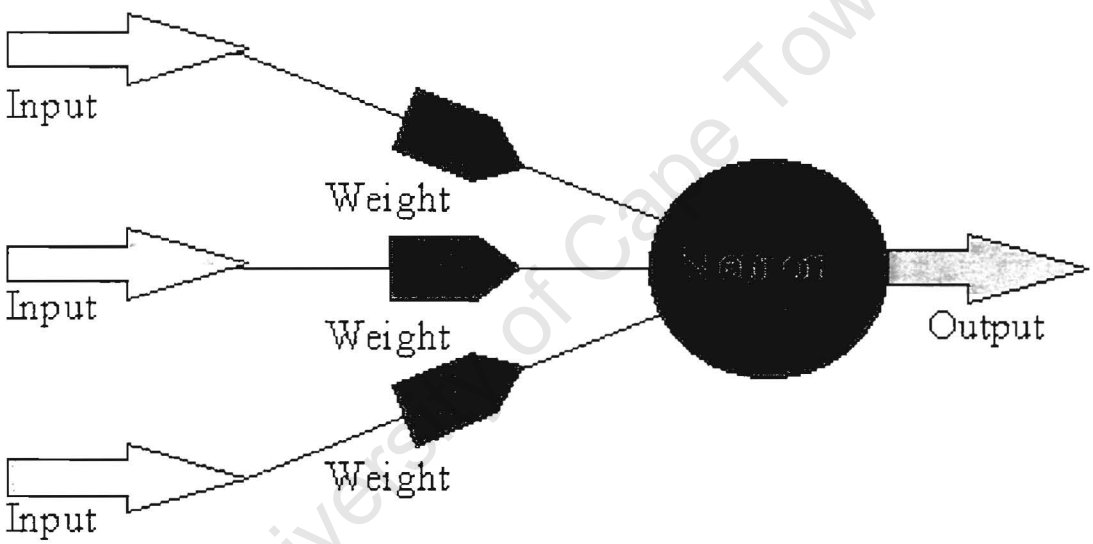


Figure 2.4. A structure of the Self-organizing map (SOM) showing the input layer, and the output layer connected by connection weights.

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).

A SOM is primarily used for recognizing and classifying patterns from a data set. This technique basically maps patterns from a multi-dimensional array (4-dimensional in this study) into a two-dimensional space (Fig. 2.5). Patterns are presented as a continuum on the two-dimensional output map based on the underlying information of the input data (Dayhoff 1990, Hewitson and Crane 1994). In the output map, the SOM concentrates patterns in regions of the data space that have more information. The selection of the size (dimensions) of the output map is arbitrary, but mostly depends on the degree of generalization required. Thus, a larger SOM output map (many patterns) will reveal more detailed structure of the input data, whereas a smaller output map will not give details, but produces more generalized patterns compared to a larger map. For the analysis, the data set is treated as continuous, and the technique mainly captures the non-linear properties in the data set (Hewitson and Crane in press). A SOM has been recently applied in phytoplankton chlorophyll studies in the Benguela upwelling region (Silulwane *et al.* 2001, Richardson *et al.* in press).

2.3.2 Identifying characteristic shapes of chlorophyll profiles using an artificial neural network, the Self-Organizing Map (SOM)

The four model-derived profile parameters were used as input to the self-organizing map. Input data consisted of the four Gaussian parameter values, which form

columns, by the number of profiles (rows). Prior to the SOM analysis, a column standardization of parameters was performed, to give each parameter equal weighting because of the large difference in magnitude of the different parameters. The SOM_Pak software version 3.1 (Kohonen *et al.* 1995) was used to perform the SOM analysis, and is freely available with its guideline applications from the Neural Network Research Centre at the Helsinki University of Technology (http://www.cis.hut.fi/research/som_1vq_pak.shtml).

2.3.3 The structure of the SOM

A SOM consists of two layers (Fig. 2.4), the input (first) layer, which contains the input data, and the output (second) layer with the identified SOM patterns. In this study, the input layer consists of the profile parameters from the Gaussian curve for each profile, and the output layer consists of the identified parameter values for the vertical chlorophyll patterns. These layers are interconnected by connection weights, where each node (pattern) of the input layer is connected to every single node in the output layer (Dayhoff, 1990)

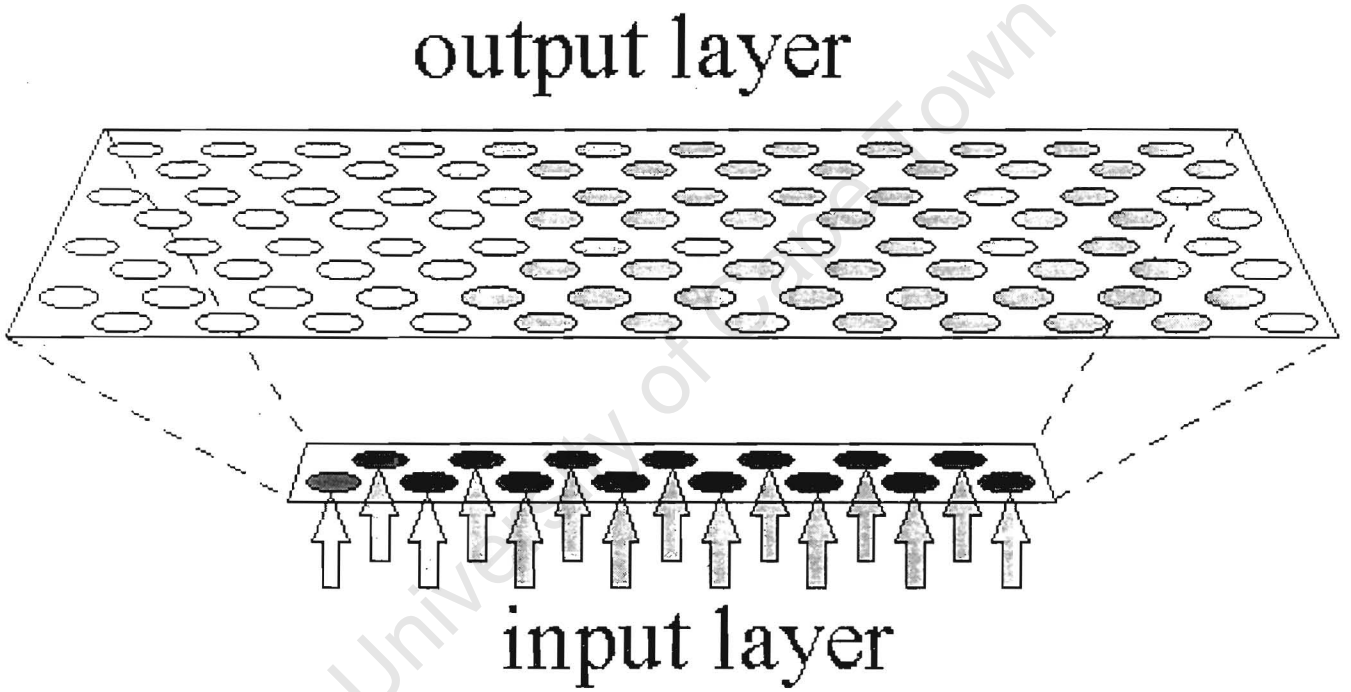


Figure 2.5. A general diagram of a multi-dimensional array of a self-organizing map (SOM) reduced to a two dimensional array.

2.3.4 The SOM algorithm

Defining the shape and dimensions of the SOM output

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 hexagonal topology produces more continuous output patterns compared to the rectangular shape, whereas the rectangular topology is easier to plot. The size of the output map, defined by the x- and y-dimensions, determines the number of output patterns. Too many output patterns do not adequately reduce the data to characteristic patterns, whereas too few output patterns do not allow differentiation of underlying patterns. To choose the SOM output that best represents all 347 vertical chlorophyll profiles, SOM output maps with different dimensions were considered. A number of

output maps were explored because there is no *a priori* fixed size for the output patterns. Dimensions of these maps varied from 7 columns by 5 rows (7x5), 4 columns by 3 rows (4x3), and to a map with 6 columns by 4 rows (6x4). It should be noted that a single dimension map (1x1) would approximate the mean of the data. All these SOM outputs had a rectangular topology.

Initialization

Prior to the training process, connection weights for each node need to be initialized. These were the initial values of the four Gaussian parameters in the present analysis. These initial values 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. For the analysis, the SOM was initialized by performing a

principal components analysis (PCA) on the input data. A PCA placed axes of most variation along the diagonals on the SOM output map.

Training

During the training stage, the SOM adapts itself iteratively to identify patterns in the input data (Fig. 2.6). Input data are presented onto the SOM for training in a consecutive manner. The first input vector consists of profile parameters from the first profile, is compared using Euclidean distance with the reference vector on each node (profile pattern) on the output map. The node that is most similar to the input chlorophyll 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 profile 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 was repeated for all input data. Figure 2.6 shows an example of the SOM under different stages of training, starting from zero to 100 000 iterations. In the present analysis, the training consisted of 100 000 iterations with an initial learning

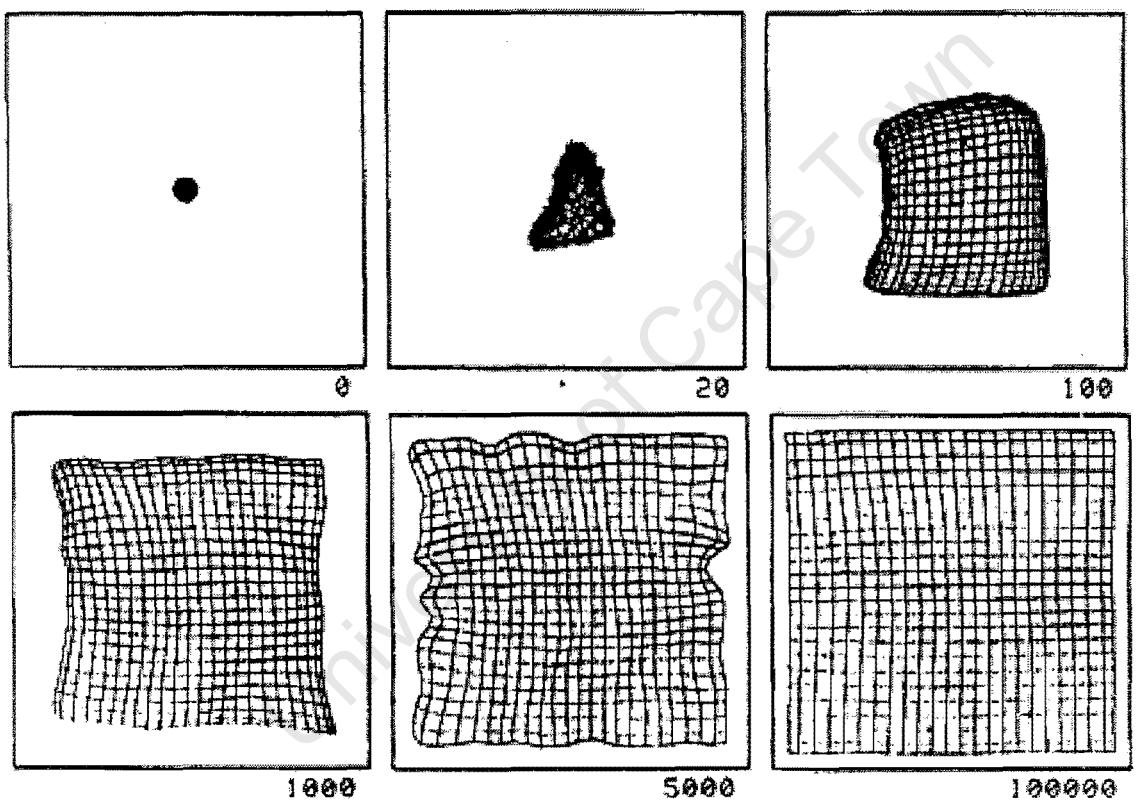


Figure 2.6. An unravelling procedure of a SOM map during different stages of training.

rate (α) of 0.01, a bubble neighbourhood function and an initial update neighbourhood of 1.5.

Convergence

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

2.4 Visualization

2.4.1 Classifying chlorophyll profiles into characteristic shapes

Once the underlying patterns (classes) have been identified, the SOM can be used to classify the input data, which are profiles in the present study, 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. Large errors identify input data that are not well represented by the output map.

Frequency maps of the chlorophyll output patterns from the SOM for each different SOM output map (7x5, 4x3, and 6x4) were then constructed to find out the relative

frequency of each of the patterns in our data set. Frequency maps have the same dimensions as the SOM output map and show the relative frequency (in percentages) of each pattern in the input data. All relative frequencies on a self-organizing map sum to 100%.

2.5 Relating the shape of chlorophyll profiles to environmental variables: a semi-quantitative approach using a SOM

Vertical chlorophyll patterns identified in the SOM were then related to environmental variables that may influence the profile shape. This was achieved by constructing frequency maps based on different levels of the environmental variable in question. This identifies those patterns that are most frequent for each level of an environmental variable. Environmental variables used were sea surface temperature (SST; taken from 3 m), surface chlorophyll concentration, water column depth (sounding), season and geographic area of the data. Seasons were autumn (March-May), spring (September-November) and summer (December-February). Areas were the Agulhas sub-province, which was further sub-divided into eastern and western regions, and West Coast sub-province, and the description of these study areas is presented in Chapter One (section 1.2). SST was partitioned into low (<14°C), medium (14-18°C) and high (>18°C). Surface chlorophyll *a* concentrations were calculated from the shifted Gaussian curve for z~0 m for each profile, and surface concentration ranged from 0.09-37.02 $mg.m^{-3}$ with levels low (<1 $mg.m^{-3}$), medium (1-7 $mg.m^{-3}$) and high (>7 $mg.m^{-3}$). Water column depth levels were; shallow (<70 m), medium (70-140 m) and deep (>140 m).

2.6 Relating the shape of chlorophyll profiles to environmental variables: a quantitative approach using generalized additive and generalized linear models

Relationships of the four profile parameters (B_0, h, σ, and z_m) were established with surface chlorophyll concentration and easily measured environmental variables such as SST, depth of the water column over spatial and temporal scales. No data on vertical chlorophyll profiles for the winter season were available for the analyses. Scatterplots were used as an exploratory data analysis tool to identify relationships between each parameter and environmental variables. These scatterplots indicated non-linear relationships between each parameter and each environmental variable. Then generalized additive models (GAMs) were used to model the nonlinear relationships of each parameter with all environmental variables simultaneously, and this was impossible to model using simple regression methods. Generalized linear models (GLMs) were then used to get an equation that can be used to predict each parameter from a set of environmental variables. Only those environmental variables that could be estimated (depth of the water column) or measured by satellites (SST and surface chlorophyll concentration) over space and time scales were used.

GLM and GAM are used to model a response variable from a set of independent variables. Unlike standard linear regression techniques, the dependent variable is not usually modelled as a continuous real-valued variable (Hastie and Tibshirani 1990). Although GAMs were run first in the model procedure of this study, the GLM will be described first because the GAM is a generalization of the GLM. GAMs and GLMs were performed with the S-plus software (a statistical module).

2.6.1 Generalized Linear Model (GLM)

A generalized linear model (GLM) is a generalization of the linear regression model:

$$Y = \alpha + \Sigma_j X_j \beta_j$$

A GLM fits a linear function to estimate the relationship between the response variable (Y) and predictor variables (X_j). The equivalent number of parameters fitted by the model is defined in terms of their *degrees of freedom* (df). For linear terms, the model uses one *degree of freedom* to fit each term, and if the model uses three predictor variables to fit a response, the total number of parameters/coefficients used in the model will be equal to 3. A GLM can model data sets having different underlying distributions: including normal, binomial, Poisson, gamma, and inverse normal (Hastie and Tibshirani, 1990).

2.6.2 Generalized Additive Model (GAM)

An additive model extends the notion of a linear model by allowing some or all linear functions of the predictors to be replaced by arbitrary smooth functions of the predictors. Thus, the standard linear regression form is replaced by the additive form, given by:

$$Y = \alpha + \Sigma_j f_j(X_j)$$

Generalized additive models are nonlinear models, which require minimum equivalent number of parameters, that is, the quantities estimated by the model coefficients, to be defined. Like a GLM, a GAM also uses *degrees of freedom* to

define the equivalent number of parameters used in the model. For an additive model with a linear term, and therefore a single coefficient, such as a model with an intercept (null model), one *degree of freedom* is required for modelling. Models with nonlinear terms use more than one *degree of freedom*. For instance, a quadratic polynomial, which is a predictor variable fitted as a nonlinear term, contributes two coefficients to the model and therefore two *degrees of freedom* will be required for modelling this term. The number of *df* is a function of the smoother of the predictor variable in the data set, and not a function of the response variable (Hastie and Tibshirani 1990)

A generalized additive model gives the estimated relationship between the individual fitted terms and each of the corresponding predictors. The response is adjusted as a smooth function for each predictor using smoothers. Therefore, GAM plots basically show the response as the adjusted relationship of each predictor against each predictor variable. These plots allow the examination of each fitted term in the model formula, giving an assessment of the goodness of fit of the model to the data.

2.6.3 Smooth functions in generalized additive model

Smooth functions or smoothers are tools that summarize the trend of values of the response (Y) as a function of predictors (X_j). Smooth functions produce an estimate of the trend of a response that is less variable than the response itself. In nature, smoothers have nonlinear properties, which means that they do not assume a fixed dependence of a response on a predictor; hence they are referred to as nonlinear regression tools. They are mainly used for describing plots by identifying trends, and for estimating the dependence of the mean of the response on predictors. Thus,

smoothers serve as important tools for estimating additive models (Hastie and Tibshirani 1990).

Smoothers basically average the response values in each neighbourhood in the data set and decide the size of this neighbourhood. How to average within the neighbourhood depends on the type of smoother that is used. Large neighbourhoods do a lot of smoothing resulting in model estimates with low variances; conversely small neighbourhoods give high variances in models (Fig. 2.7) (Hastie and Tibshirani 1990).

Different smooth functions exist, and the two smooth functions useful for transforming the predictors are the loess (lo) and spline (s) smoothers. The loess smoother fits a locally weighted least-squares regression to estimate the smooth function. The cubic B-spline smoother does not use a local averaging (Hastie and Tibshirani 1990). These smooth functions basically introduce nonparametric fitting into the model. Models involving smoothed terms use both parametric and nonparametric degrees of freedom. Parametric degrees of freedom result from fitting a linear component (trend) for each smooth term, and nonparametric degrees of freedom result from fitting the smooth function to the non-linear part after considering the modelled linear part (Hastie and Tibshirani 1990).

2.6.4 Assumptions in GLM and GAM

Generalized models have less statistical restrictions (assumptions) than standard linear regression method, but nonetheless, certain assumptions do apply. For instance, these models assume that the observed response variable is independent of the predictor

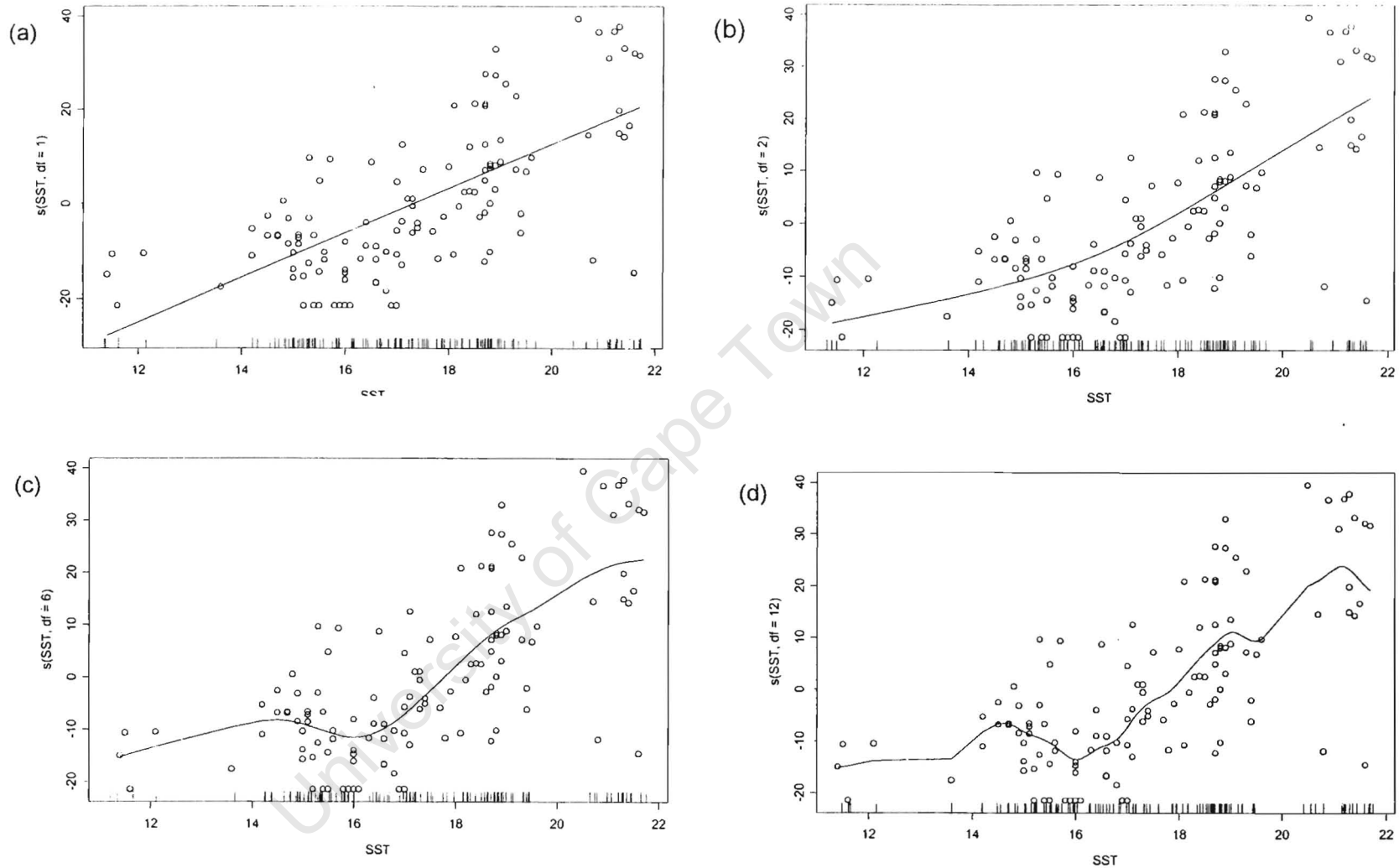


Figure 2.7. Generalized additive models showing the effect of different degrees of freedom on the smooth function: (a) $df = 1$, (b) $df = 2$, (c) $df = 6$, and (d) $df = 12$.

variable. For generalized models, errors should be distributed as Gaussian (normal), binomial, Poisson, gamma and inverse normal variables when developing models These error distributions depend on the type of data set used (Hastie and Tibshirani 1990, Morisette *et al.* 1999). In the current analysis, all parameters were considered to have a Gaussian distribution for modelling, and therefore their errors were expected to have a normal distribution. Plots of residuals against quantiles of the standard normal distribution were created for each parameter to test for normality. These standard normal plots showed errors that were not normally distributed for σ and h. The skewed distributions for σ and h were greatly improved by transforming these parameters (response variables) to get normal relationships and better fitted errors. Standard normal plots for these parameters suggested a log-transformation for σ and square root-transformation for h, which were explored to improve their error distribution. Standard normal plots for B_0 and z_m showed normal distributions and therefore were not transformed.

2.6.5 Building generalized additive models

Criteria for building a GAM depend on the type of data, whether the predictor and response variables are continuous, factors (categorical) or a combination of both. When developing a GAM, both continuous and discrete factors for response and predictors can be used. Similar to GLM, a GAM can model data with different distributions (Hastie and Tibshirani 1990).

Separate additive models were constructed for each profile parameter. These models were built by the forward stepwise approach, which involves sequential addition of each predictor variable to an existing model. Initially, a model with an intercept only

(null model) was built for a profile parameter (response variable). Then models with each of the predictor variables (one-factor model) were built, with the intercept included in all models. Each of these one-factor models was then compared with the null model, using an F-test, in order to choose the best one-factor model. The one-factor model with the lowest residual deviance and the most significant predictor variable (indicated by the lowest *p-value*, which should be less than 0.05) was regarded as the best model. The residual deviance is the variance not explained by the model, so that the smaller the residual deviance the better is the model. By adding one predictor variable to a single factor model, a two-factor model (with two predictor variables) was then built which was then compared to a single factor model. An analysis of deviance for the sequential addition of each variable to an existing model was obtained with the ANOVA function, which enabled the best model to be chosen. To build subsequent models, one variable was added to each previous model to get the complex multiple-factor model. The model with lowest residual deviance (error), highest total variance, and with all *p-values* of the predictor variables significant at the 5 % level ($p < 0.05$) was considered as the best additive model for that profile parameter. The cubic spline smoother was used in all additive models with *degrees of freedom* equal to three and these were set by default.

In building models by the stepwise forward approach, some single-factor models were not significant when compared to the null model. But, the factors of these models were then included in building more advanced models because their interaction with other factors (partial effects) resulted in more variance explained by these models (higher r^2 values) than when they are run as single factor models (marginal effects). The stepwise forward technique is simple and performs better in building models than

the stepwise backward approach, although it is rather slow. A summary table of GAMs of profile parameters including their r^2 values is given in Table 3.2.

2.6.6 Use of Analysis of variance (ANOVA) in GAM and GLM

The ANOVA function has been used to produce an analysis of deviance for the stepwise addition of each predictor to an existing model, specifying the F-test to test for differences between two models. F-test was used because we assume a normal regression, which is set by default when running the test, instead of using a Chi-squared test normally used for logistic regressions which deal with binary or two-valued response variables where under- or over- dispersions are specified. In addition, the Chi-squared test is more conservative compared to the F-test (Hastie and Tibshirani 1990).

2.6.7 Generalized additive models for the shifted Gaussian profile parameters

Generalized additive models for profile parameters were created using environmental variables. The background chlorophyll concentration (B_0) was modelled as a function of water column depth and surface chlorophyll concentration. The width (σ) and depth of the peak (z_m) were modelled as a function of all environmental variables used in this study, which are water column depth, season, surface chlorophyll, area and SST. The total chlorophyll concentration within the peak (h) was modelled with surface chlorophyll concentration, water column depth and season as predictor variables.

2.6.8 Building generalized linear models from generalized additive models

To obtain predictive equations for profile parameters, GLMs were then used. As in GAM, both continuous and categorical response and predictor variables were used. To develop a GLM for each response (profile parameter), the form of the relationship between each response and each environmental variable was identified visually from each GAM plot. GLMs were developed using significant predictor variables identified by the GAM. This was achieved by introducing a number of parametric relationships including piecewise linear regression, quadratics, log and exponential fits (by looking at GAM plots).

2.7 Predicting profile parameters using developed GLM equations

Best generalized linear models were established for each parameter, and then their equations were used to make predictions of profile parameters from environmental variables. To assess the capability of GLMs to predict the shapes of chlorophyll profiles, parameter estimates were predicted from environmental variables using an independent data set. The independent data set of chlorophyll profiles and environmental variables was collected on a cruise in November 1998, with profiles for the spring season but collected at different sub-provinces (Agulhas Bank and West Coast). This independent data set was ~20% of the whole data set of chlorophyll profiles used in this study. Predictions of the response, which are parameter values, were also performed using S-plus.

S-PLUS commands used for creating GAMs and GLMs for profile parameters are included in appendix II.

Chapter 3: Results

3.1 Mean seasonal chlorophyll profiles

Mean seasonal chlorophyll profiles for all seasons (autumn, spring and summer) were developed by calculating average values of each profile parameter for each season, which were then used to plot mean seasonal profiles. The mean profiles for each season with their mean profile parameters are shown in Figure 3.1. The chlorophyll maximum was between 11 m and 23 m deep for all seasons (autumn, summer and spring), indicating near-surface to subsurface chlorophyll peaks. The change in shape and size of the pigment structure was indicated by changes in profile parameters. The mean autumn profile had a moderate chlorophyll peak (~5 $mg.m^{-3}$), which was relatively narrow (σ~9 m) and close to the surface (z_m~11 m). The average profile for summer had a narrow, larger (>7 $mg.m^{-3}$), subsurface peak (z_m~18 m) with high total chlorophyll concentration beneath the peak (h~ 186 $mg.m^{-2}$). The mean spring profile was different from autumn and summer profiles, having relatively broader (σ~14 m) and deeper peak (z_m~22 m) with lowest chlorophyll concentration beneath the curve (h~66 $mg.m^{-2}$). The spring profile was situated deeper in the water column with low chlorophyll concentration (~2 $mg.m^{-3}$) than the autumn profile. The background chlorophyll concentration (B_0) did not change much from one season to the other, with summer having a lowest background concentration (B_0~0.20 $mg.m^{-3}$), and a highest concentration in autumn (B_0~0.27 $mg.m^{-3}$).

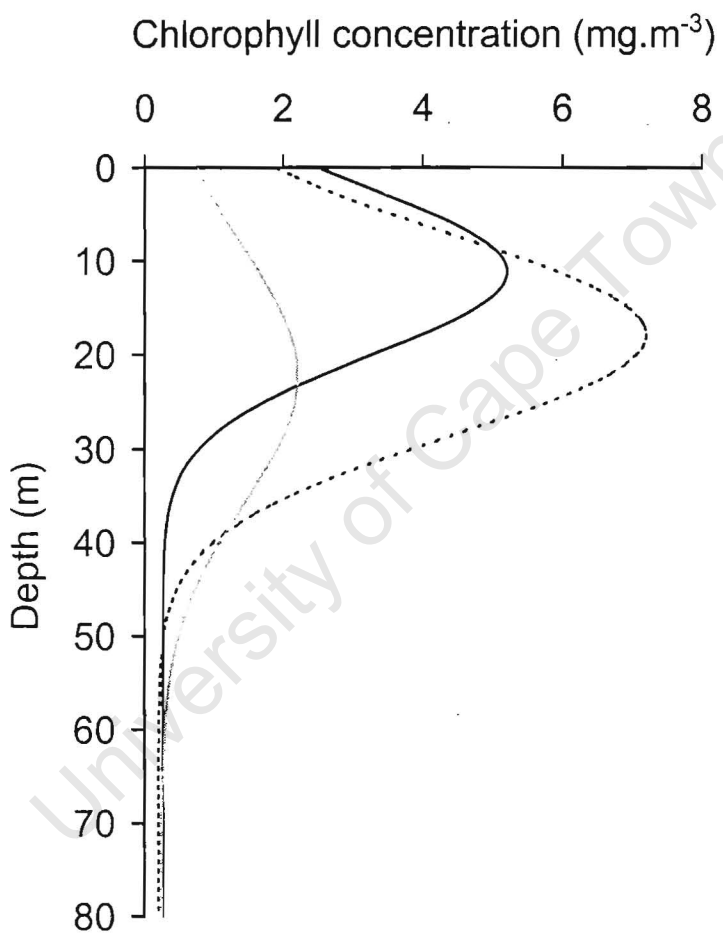


Figure 3.1. Mean chlorophyll profiles for autumn (black solid line), spring (grey solid line) and summer (dotted line). Average parameter values are: autumn (B_0 = 0.27, h = 109.6, σ = 8.9, z_m = 11.2), spring (B_0 = 0.25, h = 66.4, σ = 13.5, z_m = 22.3) and summer (B_0 = 0.20, h = 185.9, σ = 17.8, z_m = 17.8).

The adequacy of mean seasonal profiles were assessed by looking at frequency histograms of the profile parameters within each season. Generally, there was a considerable spread of parameter values within each season, suggesting that seasonally-averaged profiles are insufficient for characterizing profile shapes in each period. In autummn, almost all parameters are skewed to the left with the depth of the peak (z_m) showing bimodal distribution (Fig. 3.2). The depth of the peak (z_m) varied from 0 to >45 m, and the spread of the peak (σ) from 0 to >35 m. Except for z_m and σ, there was less variation in parameter values during the spring period. Although relative frequencies for z_m were considerably low in spring, there was most variation (from 0 to >50 m) in z_m values, indicating the ocurrence of different profiles shapes with surface and subsurface chlorophyll maxima (Fig. 3.3). Spring profiles had relatively low background chlorophyll concentration (B_0~0.5) and total chlorophyll concentration within the peak (h~ 100). In summer, the magnitude of profile shapes changed considerably and the depth of the peak also showed most variation as in spring with values ranging from 0 to >60 m. Two different distributions of z_m were evident, indicating profiles with surface and subsurface peaks (Fig. 3.4).

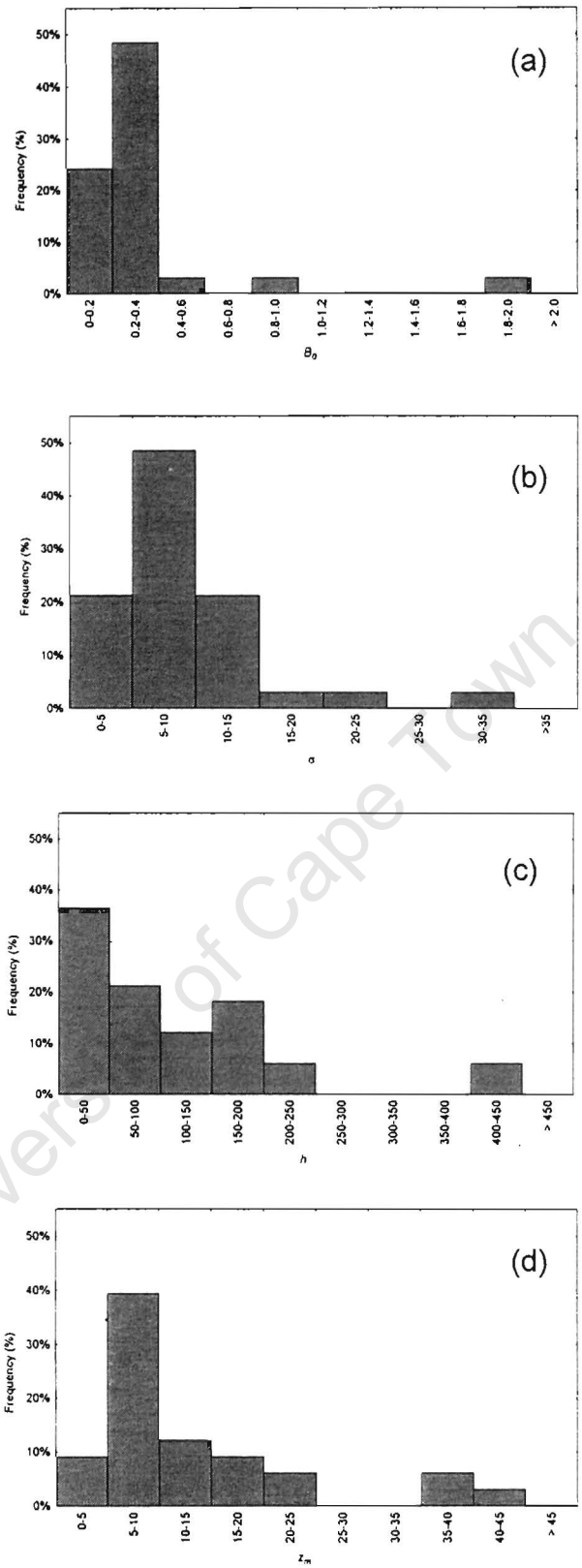


Figure 3.2. Relative frequency histograms of (a) B_0, (b) σ, (c) h and (d) z_m for autumn (n =33).

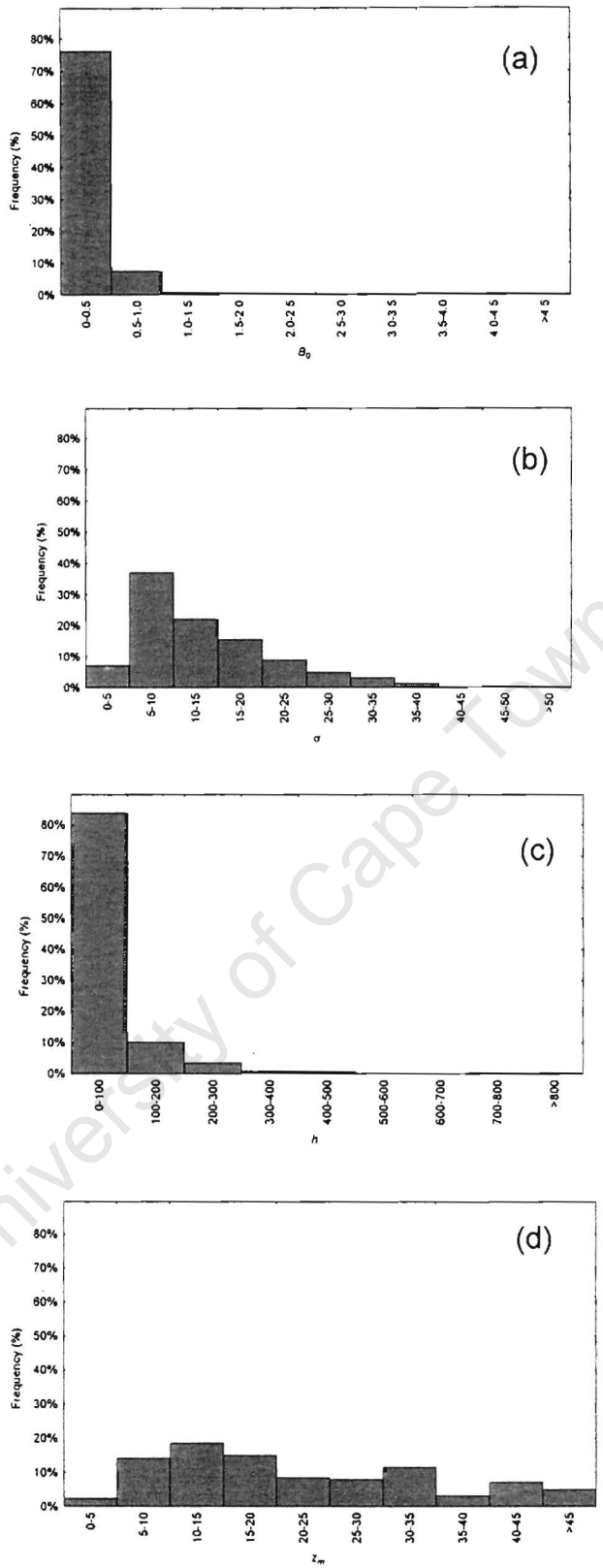


Figure 3.3. Relative frequency histograms of (a) B_0, (b) σ, (c) h and (d) z_m for spring (n=227).

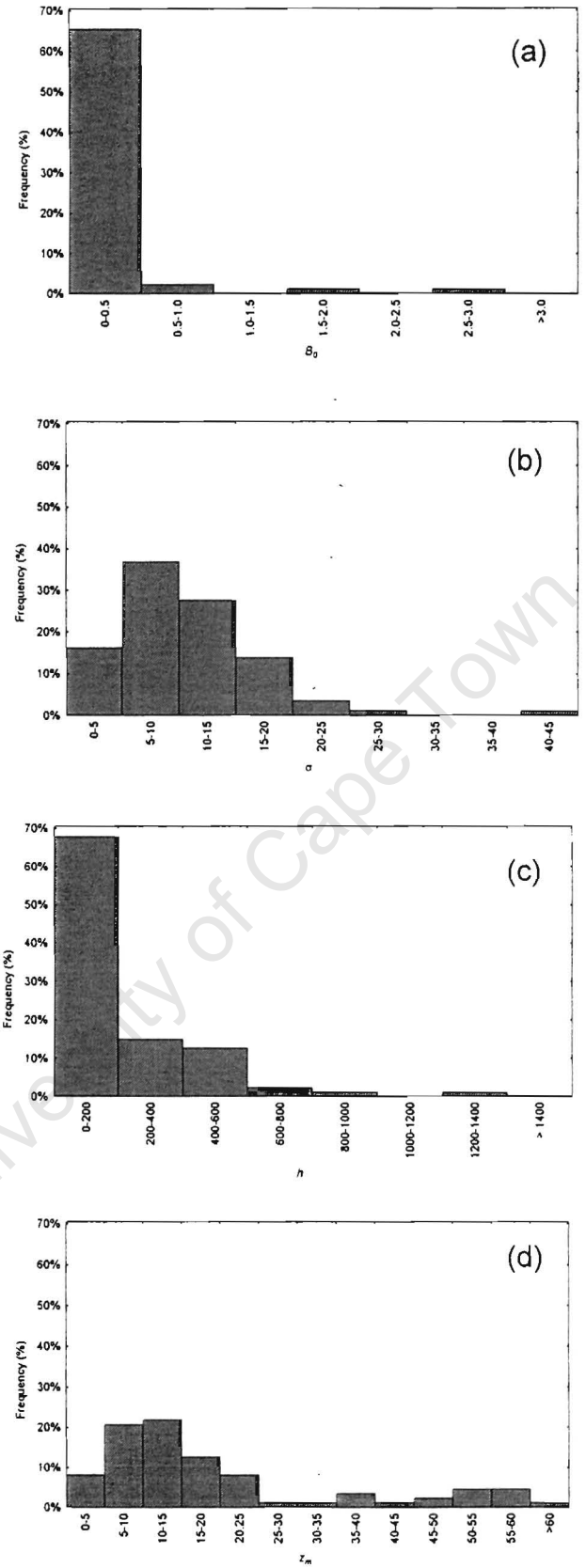


Figure 3.4. Relative frequency histograms of (a) B_0, (b) σ, (c) h and (d) z_m summer (n = 87).

3.2 Identified characteristic shapes of chlorophyll profiles from the SOM analysis

Different sizes of SOM were explored to find a SOM output with patterns that could summarize the data but still capture sufficient detail of the 347 *in situ* chlorophyll profiles used as input in the SOM analysis. The SOM technique produced a continuum of chlorophyll patterns mapped from a multi-dimensional array into a two-dimensional space, with similar patterns mapped close together and dissimilar patterns further apart. Different initializations were used for different sizes of SOM outputs (5x3, 6x4 and 7x5), which resulted to a shift in position of chlorophyll patterns with surface and subsurface peaks on a 2-dimensional SOM output.

3.2.1 The 5x3 SOM output

The smallest SOM output map (5x3, Fig. 3.5) showed chlorophyll patterns on the bottom left side of the map with high surface chlorophyll concentration (>10 $mg.m^{-3}$) (e.g. #11) and those toward the right side with small (~1 $mg.m^{-3}$), subsurface peaks (#5). The gradual change in patterns was a result of change of parameter values across the output map. The depth of the peak (z_m) changed from the bottom left corner of the map with near-surface peaks (#11, 12) and these shifted to deep chlorophyll peaks (~40 m) on the top right corner of the SOM output (#5; z_m increasing). The width of the peak changed across the SOM output, with narrow peaks at the bottom right corner (#15) and these becoming broader towards the top left corner (#1; σ increasing). The total chlorophyll concentration beneath the peak (h) increased from subsurface peaks on the top right corner (#5) to surface peaks on the bottom left

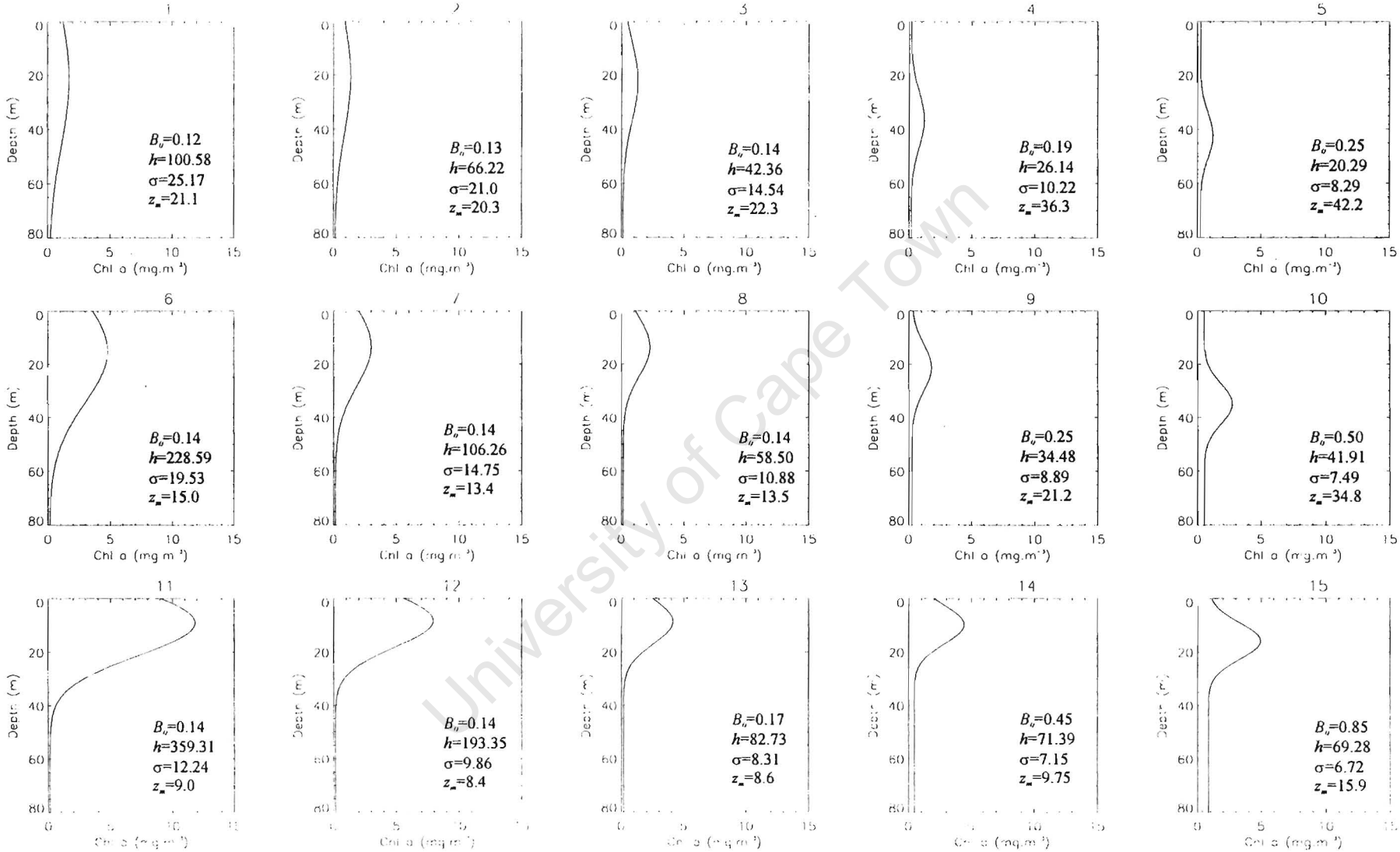


Figure 3.5. A 5x3 SOM output map for the 347 input chlorophyll profiles.

corner (#11). The background chlorophyll concentration (B_0) increased from a minimum in the top left corner (#1) to a maximum in the bottom right corner (#15). Considering these changes in profile parameters across the SOM output, it was clear that a continuous change in chlorophyll patterns was shown by the SOM technique. The SOM technique could identify only a few generalized patterns because of the small SOM output map. The analysis tried to map as many similar profiles as it could to each pattern, but some were very different within each pattern, and this was highlighted by the error histograms.

3.2.2 The 5x3 error histogram

One of the advantages of the SOM is its ability to produce an error that is associated with each profile mapped to a particular pattern in the output map. The error indicates how well each profile was represented by the pattern it gets mapped to in the SOM output with a larger error indicating a poorer fit. The error map of the 5x3 SOM (Fig. 3.6) had most patterns with relatively small errors (≤1), indicating profiles that were represented by particular patterns. Exceptions were patterns #1, 11 and 15 and these had large errors (>1), indicating poor representation of profiles within those output patterns. The SOM output showed pattern #1 with broader peak at intermediate depths (~21 m) and high B_0 value (0.12). Some profiles that mapped to this pattern (#1) had a large range of B_0 values (0-0.5 $mg.m^{-3}$) and high variability in total chlorophyll within the peak (h, 30-208 $mg.m^{-2}$). Pattern #11 had a near-surface peak, maximum B_0 (0.14 $mg.m^{-3}$) and maximum h value (~360 $mg.m^{-2}$), and profiles that mapped to this pattern had surface and near-surface peaks (0-14.6 m), and higher h values (>400 $mg.m^{-2}$).

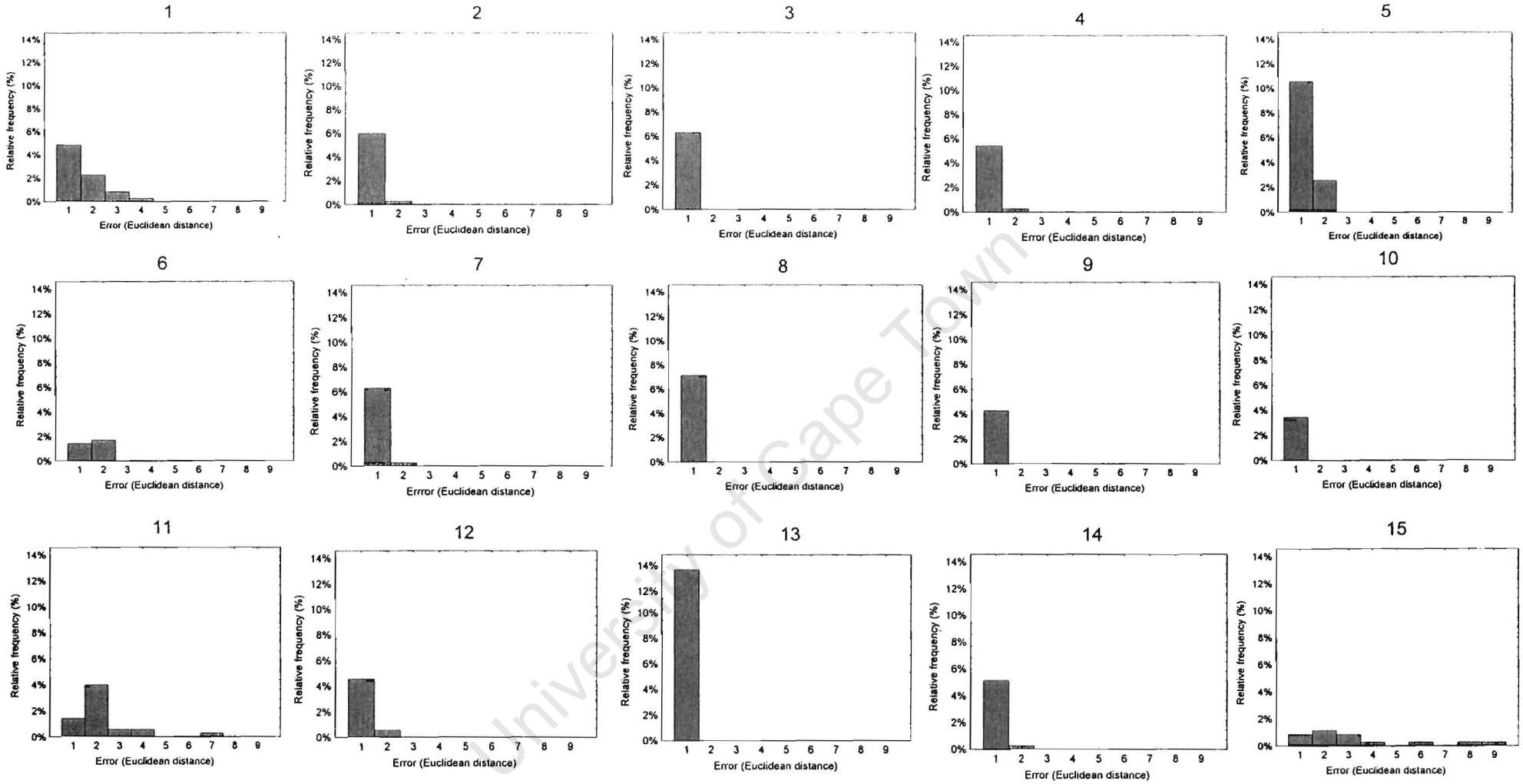


Figure 3.6. Frequency histograms of errors of the 5x3 SOM for the input profiles that mapped to each output pattern. The error is the Euclidean distance between a profile and its associated pattern.

Pattern #15 had a pronounced, near-surface peak (~16 m) and profiles that belonged to this pattern had surface peaks (0-5 m).

3.2.3 The 7x5 SOM output

A SOM output with many (35) patterns was presented and showed a continuum of patterns ranging from profiles with small (~1 $mg.m^{-3}$) surface peaks at the bottom left corner to large (>15 $mg.m^{-3}$) surface peaks (typical blooms) at the top right corner of the output map (Fig. 3.7). Changes in patterns were a consequence of the change in the following parameters in all directions of the output map. The depth of the chlorophyll maximum (z_m) decreased from a maximum (~48 m) in the bottom left side (#29) to a minimum (~6 m) at the bottom right side of the map (#35), and across the map, from the top left to the bottom right side. The total chlorophyll beneath the peak (h) showed an opposite trend, increasing across the map, from the bottom left corner (#29) to the top right corner (#7). The width of the peak narrowed (σ decreasing), while the background chlorophyll (B_0) increased across the map, from the top left side (#1) to the bottom right corner (#35) of the SOM. The 7x5 SOM output map could identify a large number of patterns, so that more subtle differences in profiles can be identified.

3.2.4 The 7x5 error histogram

The error histogram of the 7x5 SOM (Fig. 3.8) showed most patterns characterized with small errors (≤1), suggesting that a group of similar profiles were mapped to a particular pattern, and therefore there was less variability in chlorophyll profiles within each pattern. Patterns with large errors were also present (e.g. #1, 7 and 35).

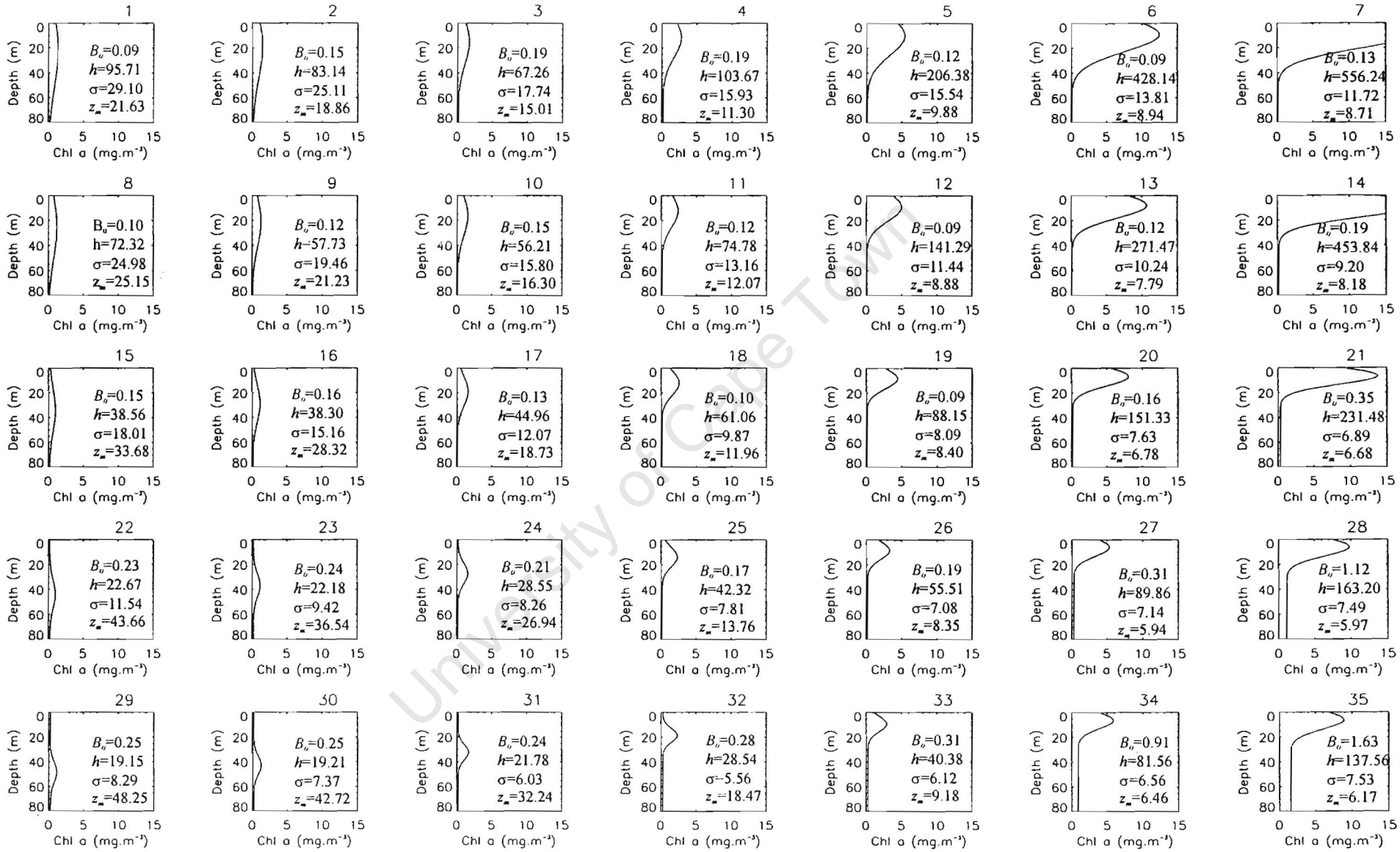


Figure 3.7. A 7x5 SOM output map for the 347 input chlorophyll profiles.

For example, pattern #1 had low B_0 value and a broader peak; largest surface peak and high total chlorophyll beneath the peak (556 $mg.m^{-2}$) were characteristic features of pattern #7; and pattern #35 had a highest B_0 concentration (1.63 $mg.m^{-3}$). Looking at the data set, typical profiles that mapped to pattern #1 had surface chlorophyll peaks and low B_0 concentrations; input profiles that mapped to pattern #7 had higher h values (>556 $mg.m^{-2}$ for the output pattern); and input profiles for pattern #35 had higher B_0 concentrations, which ranged from 1.52-4.48 $mg.m^{-3}$, compared to the background concentration for the output pattern. Input profiles that were different (in terms of parameter values) from the SOM output patterns were the ones that were identified as outliers (patterns with high error values) in the error histogram (Fig. 3.8).

3.2.5 The 6X4 SOM output

Generally, chlorophyll patterns of the 6x4 SOM output map changed across the map. Patterns at the bottom left of the map (e.g. #19) had low subsurface chlorophyll concentrations (~1$mg.m^{-3}$) and these became high surface chlorophyll concentration (>10 $mg.m^{-3}$) on the top right of the map (e.g. #6) (Fig. 3.9). In addition to changes in chlorophyll concentration across the SOM, the change in profile parameters was also shown by the SOM analysis. The depth of the chlorophyll maximum (z_m) changed across the map, with small (< 1 $mg.m^{-3}$), deep chlorophyll maximum (~47 m) at the bottom left corner (#19), and these shifted to surface layers to be large (~15 $mg.m^{-3}$) near-surface chlorophyll peaks (~ 8 m) at the top right corner (#6). The subsurface chlorophyll peaks at the bottom right corner (#24) became near-surface peaks at the top left corner (#1). The width of the peak (σ) also changed across the map, with

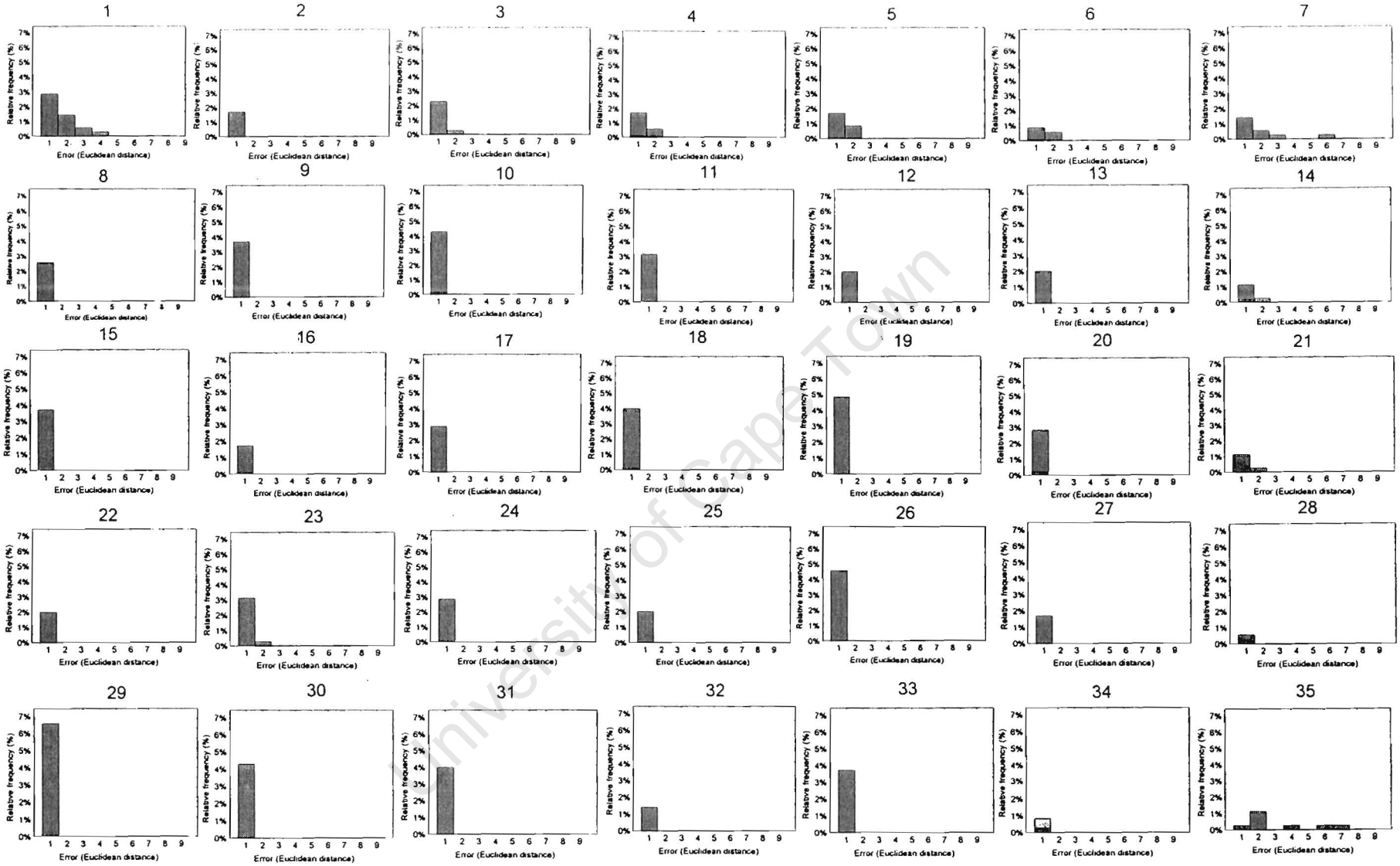


Figure 3.8. Frequency histograms of errors of the 7x5 SOM for the input profiles that mapped to each output pattern.

narrower near-surface peaks situated at the top left corner and then became broader as the chlorophyll maxima shifted deeper at the bottom right corner of the map (#24). The total chlorophyll concentration within the peak (h) increased from a minimum (18.3 mg.m^{-2}) in the bottom left (#19) to a maximum (461.4 mg.m^{-2}) at the top right corner (#6). The background chlorophyll concentration (B_0) decreased from a maximum (1.4 mg.m^{-3}) at the top left (#1) to a minimum (0.1 mg.m^{-3}) at the bottom right part of the SOM (#24). Surface and subsurface chlorophyll maxima were characterized by low background chlorophyll concentration (Fig. 3.9). The SOM technique produced a continuum of patterns in all directions of the SOM output, with every pattern representing profiles from input data with no discontinuities in profile parameter values. However, the analysis highlighted the variability in chlorophyll profiles, in terms of parameter values, across all directions of the map.

3.2.6 The 6x4 error histogram

Generally, frequency error histograms for the 6x4 map (Fig. 3.10) showed most patterns had relatively small errors (≤1), which means profiles were appropriately represented. Profiles not well represented by pattern #1 had B_0 values ranging from 1.99-4.48 mg.m^{-3}, which were greater than 1.43 mg.m^{-3} for this pattern, as well as high values of h (183-351 mg.m^{-2}) that were much higher than 108.6 mg.m^{-2} for this SOM output pattern. Outliers that mapped to pattern #6 had exceptionally large h, ranging from 567-1362 mg.m^{-2}, considerably larger than 461.4 mg.m^{-2} for this pattern. Outliers were easily identified from the SOM output patterns by considering their errors.

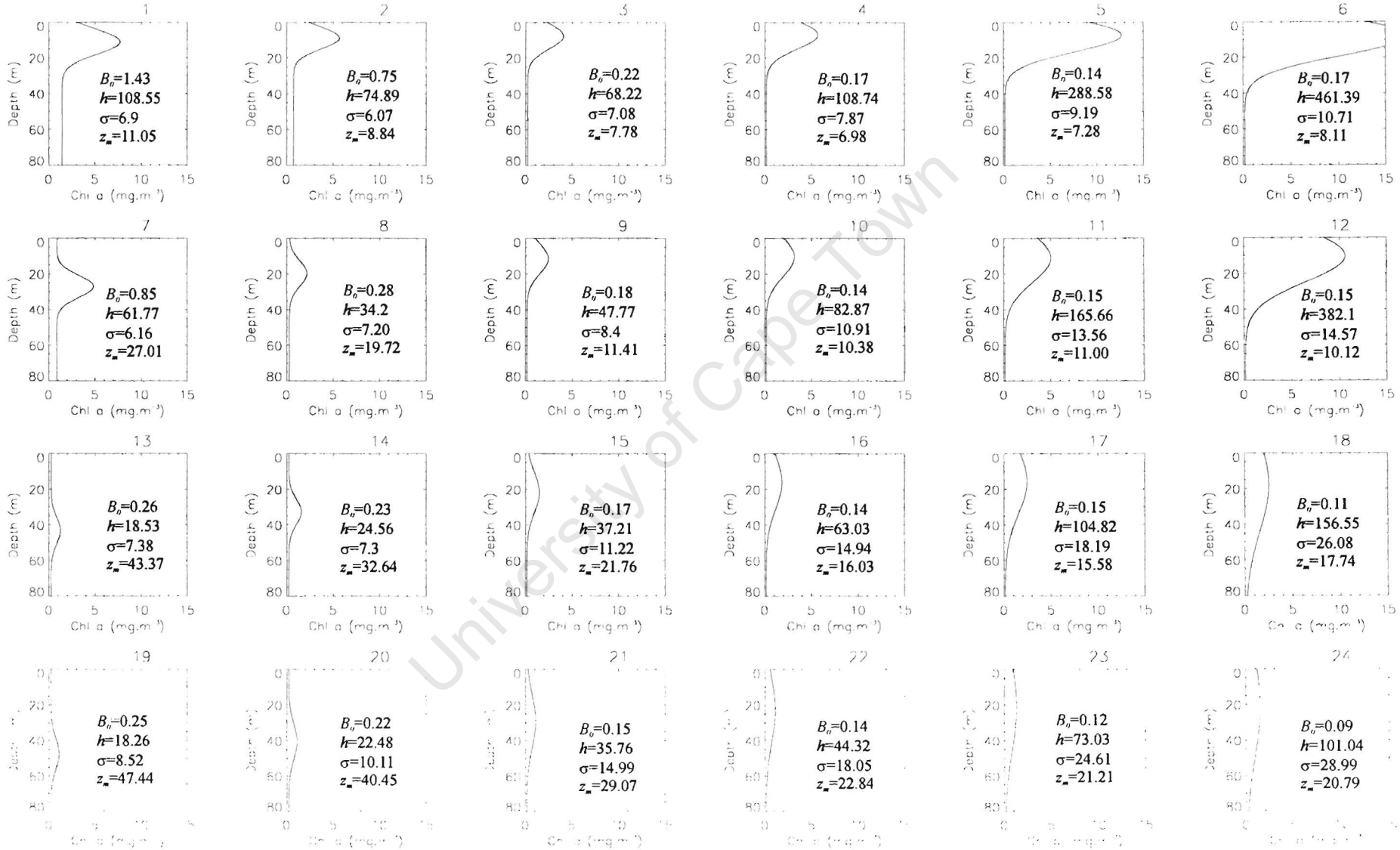


Figure 3.9. A 6x4 SOM output map for the 347 input chlorophyll profiles.

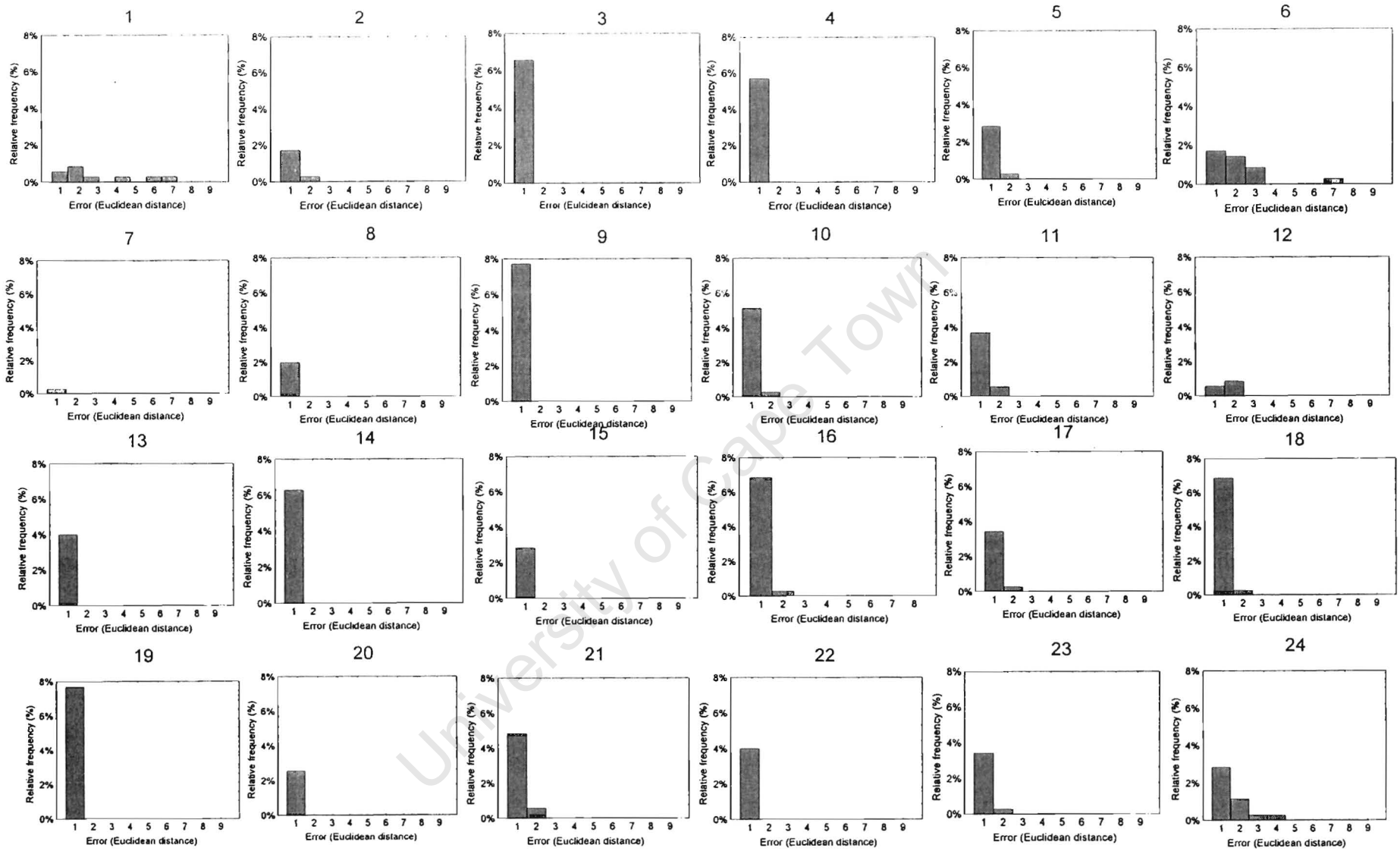


Figure 3.10. Frequency histograms of errors of the 6x4 SOM for the input profiles that mapped to each output pattern.

The SOM analysis: a semi-quantitative approach

To identify the frequency of occurrence of different patterns identified by the SOM analysis, a relative frequency map of each of the SOM output maps was produced. These frequency maps have the same dimensions as those of the SOM output and can be visualized as being superimposed on the SOM output map, so that their coordinates correspond to those of the output map.

3.3.1 The 5x3 SOM output frequency map

The overall frequency map for the 5x3 SOM output (Fig. 3.11) showed patterns that are dominant and rare from the input data set. Patterns that dominated were #5 and 13, with these patterns representing approximately 13% and 14% of profiles from the input data. Characteristic profiles that mapped to pattern #5 had a narrow (σ~8 m), deep chlorophyll maximum (z_m~42 m). The other frequent pattern (#13) had a near-surface peak (z_m~9 m) with moderate chlorophyll concentration (5mg.m^{-3}) and low background chlorophyll concentration (B_0~2 mg.m^{-3}). Pattern #6 with moderate, subsurface chlorophyll maximum (z_m~15 m) and high total chlorophyll within the peak (h~229 mg.m^{-2}), was the least common pattern, and represented ~3% of input profiles. A narrow, deep chlorophyll maximum with moderate chlorophyll concentration and high B_0 (~0.5 mg.m^{-3}) were features of profiles that mapped to pattern #10, a least common pattern representing ~3% of profiles.

Figure 3.11. Overall relative frequency map of the 5x3 SOM output of the Agulhas Bank and West Coast sub-provinces. Note the squares correspond to patterns in the same position in the 5x3 SOM output map in Figure 3.5. Relative frequencies for all patterns add up to 100%.

3.3.2 The 7x5 SOM output frequency map

Generally, patterns with very high chlorophyll concentrations and surface peaks were not commonly occurring in the 7x5 overall frequency map (Fig. 3.12). This 7x5 overall frequency map had very low relative frequencies compared to the 5x3 frequencies. The chlorophyll pattern (#29) with the deepest chlorophyll maximum (z_m~48 m) and minimum chlorophyll concentration within the peak (h~19 mg.m^{-2}) was the most frequent occurring pattern, with 7% of profiles mapping to it. Other frequent patterns, with 5% of profiles mapping to each pattern, were #1, 19 and 26. Pattern #1 had almost uniform chlorophyll concentration throughout the water column and lowest background chlorophyll concentration (B_0 = 0.09 mg.m^{-3}). A lowest background concentration, narrow, near-surface peak (σ~8 m; z_m~8 m) and low chlorophyll within the peak were characteristic features of pattern #19. Pattern #26 was similar to pattern #19, and had a narrow peak (σ~7 m), except that this pattern had a higher B_0 (~0.2 mg.m^{-3}) and a smaller (~3 mg.m^{-3}), near-surface peak (z_m~7 m) compared to #19. A pattern with highest chlorophyll concentration (>15 mg.m^{-3}), near-surface peak (z_m~9 m) and highest chlorophyll concentration within the peak (# 7) was the least occurring pattern. This rare pattern represented only 1% of profiles from the input data. Other infrequent patterns, each representing 1% of profiles, were patterns #32 and 34.

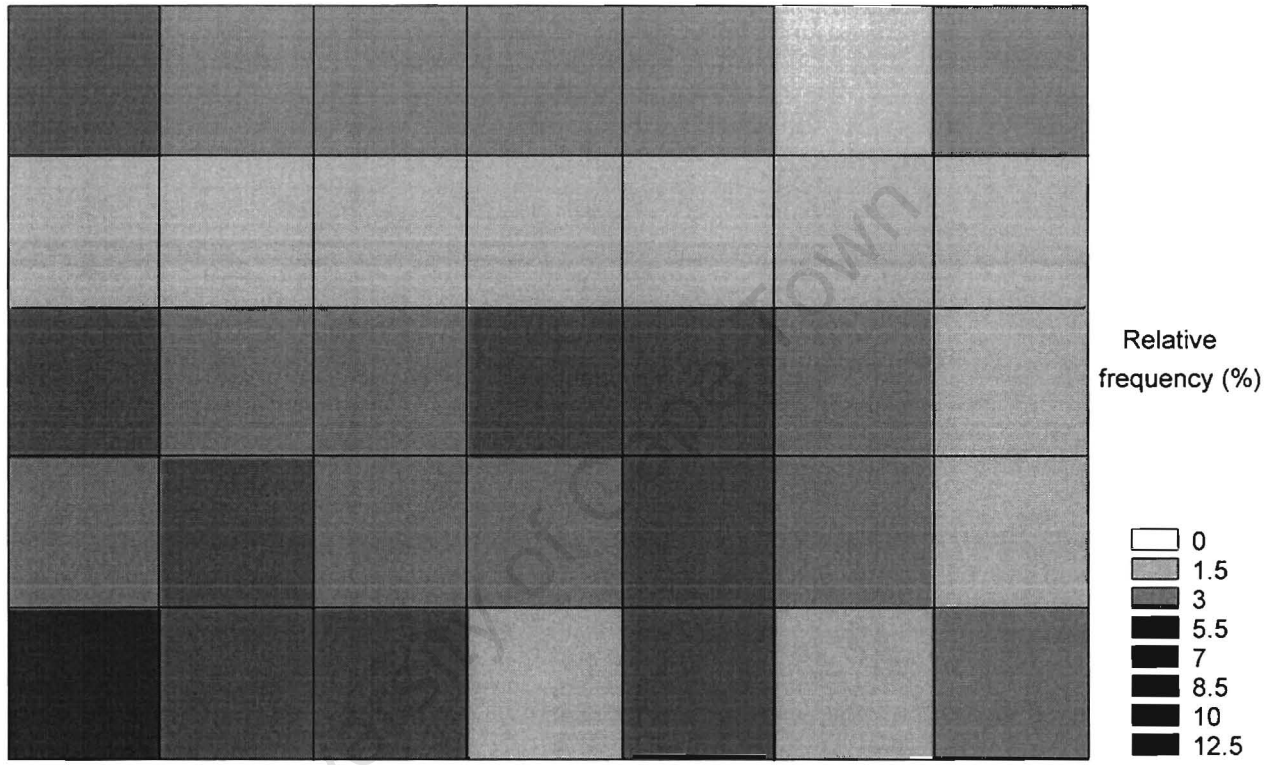


Figure 3.12. Overall relative frequency map of the 7x5 SOM output of the Agulhas Bank and West Coast sub-provinces. Note the squares correspond to patterns in the same position in the 7x5 SOM map in Figure 3.7, and frequencies sum up to 100%.

3.3.3 The 6x4 SOM output frequency map

The 6x4 overall frequency map (Fig. 3.13) showed patterns #9 and 19 as most dominant patterns in the input profile data with each pattern representing 8% of input chlorophyll profiles. Pattern #9 had a low background chlorophyll concentration, and a narrow chlorophyll peak (σ~8 m) situated at near-surface layers (~11 m). Pattern 19 had a narrow (σ~8 m), deepest chlorophyll maximum (z_m~47 m) and lowest chlorophyll concentration within the peak (h~18 mg.m^{-2}). Other frequent patterns (#3 and 16), representing 14% in total of profiles, with moderate near-surface peak for pattern #3, and broad, subsurface maximum for pattern #16. The least frequent pattern (#7), with relative frequency ~0.3%, had a moderate (~5 mg.m^{-3}), narrow (σ~6 m), subsurface chlorophyll maximum (z_m~27 m).

After exploring different sizes of SOM output maps (5x3 with few patterns, 6x4 with intermediate patterns, and 7x5 map with many patterns), and looking at their error and frequency histograms, the selection of a best output map for the 347 input profiles was considered. The 7x5 map had too few profiles mapped to each node, and this was indicated by very low relative frequencies in the overall map (Fig. 3.12), while the very compact 5x3 map did not show a sufficient range of chlorophyll profiles, and therefore low chlorophyll variability was highlighted (Fig. 3.11). The 6x4 SOM gave the required details of chlorophyll profile shapes of the sub-provinces, and had showed the variation in these profiles without too much generalization and over detailed description of chlorophyll variability (Fig. 3.13). Therefore it was considered as the best SOM output map with sufficient details for the purposes of this study, and was used in further analyses.

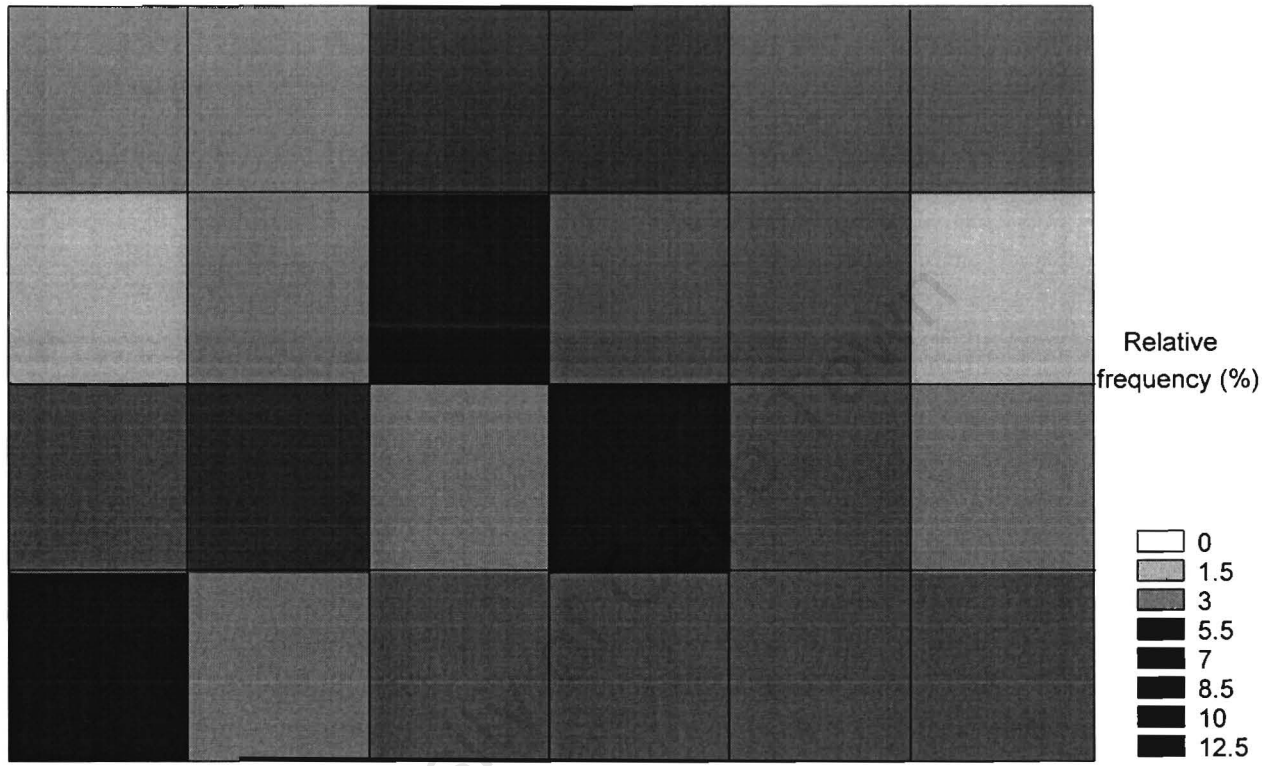


Figure 3.13. Overall relative frequency map of the 6x4 SOM output of the Agulhas Bank and West Coast sub-provinces. Note the squares correspond to patterns in the same position in the 6x4 SOM map in Figure 3.9, and frequencies sum up to 100%.

3.4 Relating the shape of chlorophyll profiles of the Agulhas Bank and West Coast sub-provinces to environmental variables: a semi-quantitative approach using a SOM

Owing to the relatively small number of profiles available in this study, it was not possible to investigate the effect of environmental variables in each sub-province and season. Thus, all data were combined to elucidate relationships with environmental variables. Categories (low, medium and high) were created for SST, surface chlorophyll and water column depths to find chlorophyll variability at these levels.

3.4.1 Spatial variability in chlorophyll profiles

Variability in phytoplankton pigment structure at spawning and recruitment grounds (Agulhas Bank and West Coast regions) was explained by looking at different chlorophyll profile shapes at these regional scales. The relative frequencies of occurrence of profiles (for the 6x4 SOM) showed variability in chlorophyll patterns in these different sub-regions (EAB, WAB and WC).

The eastern Agulhas Bank sub-region (EAB)

The eastern Agulhas Bank, with the lowest number (70) of chlorophyll profiles showed patterns with small, deep, chlorophyll peaks (>30 m), located at the bottom left of the output map, as most common patterns (#14 and 19), with pattern #14 accounting for ~4%, and pattern #19 representing ~3% of chlorophyll profiles (Fig.

3.14a). This region lacked profiles with high surface peaks (>10 $mg.m^{-3}$, <10 m) and only few patterns with pronounced near-surface peaks were found. The majority of patterns were characterized by small (<2 $mg.m^{-3}$), relatively deep chlorophyll peaks (z_m> 20 m).

The western Agulhas Bank sub-region (WAB)

The western Agulhas Bank, with 202 profiles, had a greater variety of chlorophyll patterns than either the EAB or WC sub-regions, with almost all patterns from the SOM output represented in this region (Fig. 3.14b). The majority of chlorophyll patterns had small (~1 $mg.m^{-3}$), subsurface peaks (e.g. #19) and large, surface peaks with high chlorophyll concentration (>5 $mg.m^{-3}$), (e.g. #6). Relatively dominant patterns were characterized by small, deep chlorophyll peaks (#19 and 21), and each pattern had ~5% of profiles mapped to it. Another frequent pattern with relative frequency ~4% had highest surface chlorophyll maximum and total chlorophyll within the peak (#36). Moderate to high, narrow near-surface peaks were features of other common patterns (#9, 10 and 11). Patterns with almost uniform chlorophyll throughout the water column (#23 and 24) were also common. Least common patterns had a moderate, narrow surface peak (#2); and a small, subsurface peak (#8), with each pattern representing ~1% of profiles. Patterns with pronounced subsurface chlorophyll maximum (~27 m; e.g. #7) were not occurring in this sub-region.

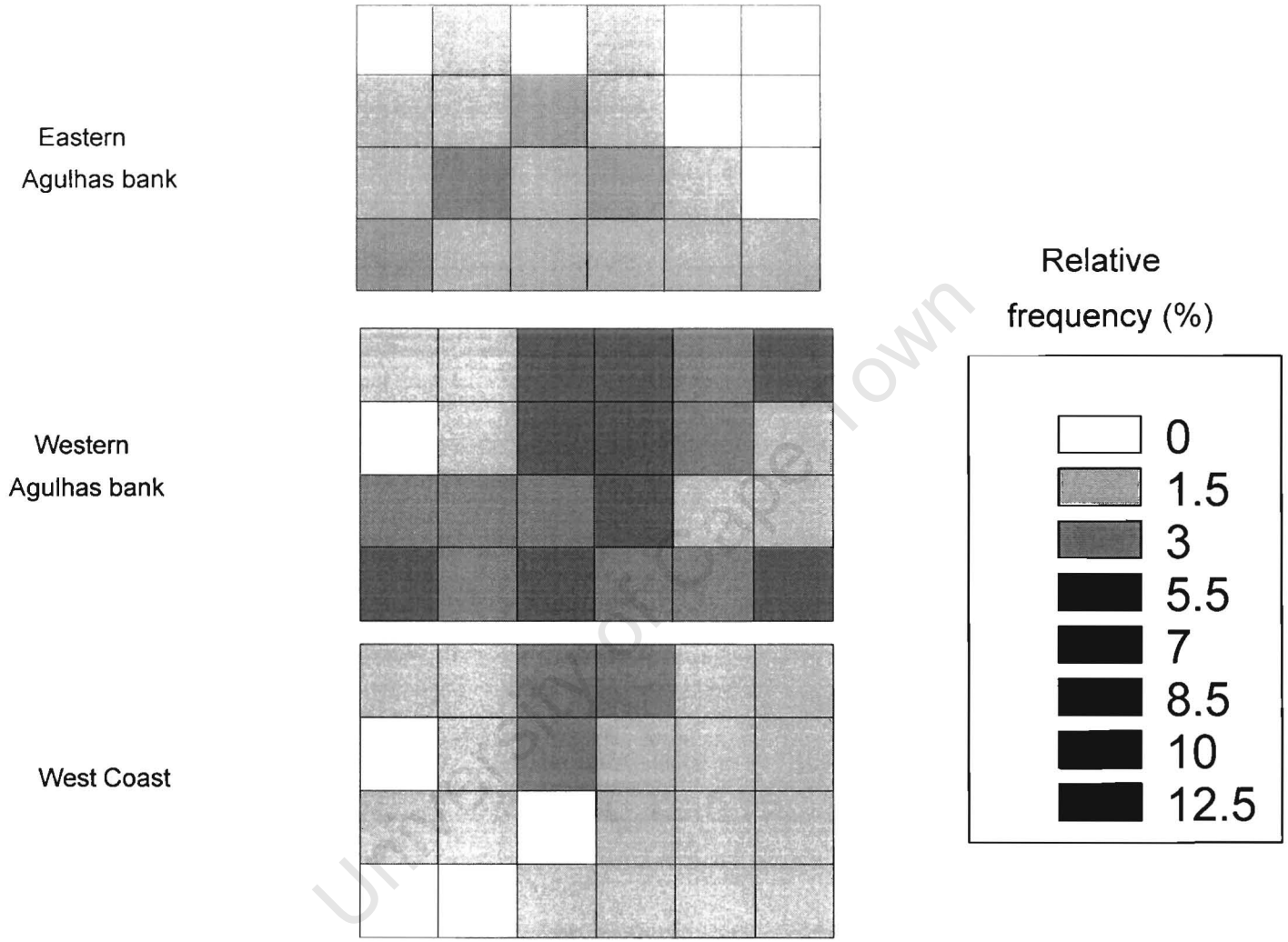


Figure 3.14. Relative frequencies of chlorophyll patterns at different sub-provinces: (a) Eastern Agulhas Bank, (b) Western Agulhas Bank, and (c) West Coast sub-provinces. All regional levels add up to 100%.

The West Coast sub-province (WC)

Few profiles (75) were from the West Coast region of South Africa. The majority of patterns from this sub-province had surface and near-surface peaks with high and moderate chlorophyll concentrations (3-12 $mg.m^{-3}$). The most common pattern (#3) had a moderate (~5 $mg.m^{-3}$) surface peak and represented ~3% of profiles (Fig. 3.14c). Some patterns with very deep and narrow chlorophyll maxima (z_m> 40 m) were not found in this region (#19 and 20). Patterns that were almost uniform throughout the water column (#22, 23 and 24) were few, and these represented an overall frequency of ~2%.

3.4.2 Seasonal variability in chlorophyll profiles

Considering changes in phytoplankton biomass distribution and hence of pigment structure with time, variability in chlorophyll profile shapes was identified by looking at the frequency of occurrence of 6x4 SOM output patterns at different seasons (autumn, spring and summer) (Fig. 3.15).

Autumn

Only 33 out of 347 profiles were from this period, therefore there was less confidence in relative frequencies of patterns in this period compared to other seasons. Generally, autumn profiles had narrow surface and near-surface peaks (Fig. 3.15a). Few profiles with a deep chlorophyll maximum (>30 m) occurred in this period. The dominant patterns were those with near-surface peaks and moderate chlorophyll concentrations

(4-6 $mg.m^{-3}$; #3 and 4), and these represented ~4% of chlorophyll profiles from the data set. Less than 1% of profiles with high (>15 $mg.m^{-3}$), near-surface chlorophyll concentration maximum were found in this season (e.g. #6).

Spring

Most profiles from the data set were collected during spring (227), with higher variability in profiles found in this period than in summer and autumn (Fig. 3.15b). A mixture of chlorophyll profiles with near-surface and subsurface maxima, and those that were almost uniform throughout the water column were found. Patterns with more pronounced, but moderate near-surface maximum were the most dominant (#9 and 16), and these represented ~12% of chlorophyll profiles. Patterns with very low chlorophyll concentrations (1-3 $mg.m^{-3}$) and deep subsurface peaks (>40 m) were more common in spring than in autumn and summer. The least frequent patterns with high surface chlorophyll concentration (>10 $mg.m^{-3}$) were in the top right corner of the output map, with ~2% of profiles mapping to patterns 5 and 12, and less than 1% of profiles mapping to pattern #6.

Summer

The summer season, with 87 profiles, showed two different patterns dominating, and these were profiles with high (>15 $mg.m^{-3}$), surface chlorophyll maxima in the top right corner (#6), and small (~1 $mg.m^{-3}$) deep subsurface maxima (~47 m) in the bottom left corner (#19). These dominant patterns represented ~7% of profiles from the summer period (Fig. 3.15c). Other common patterns included those with

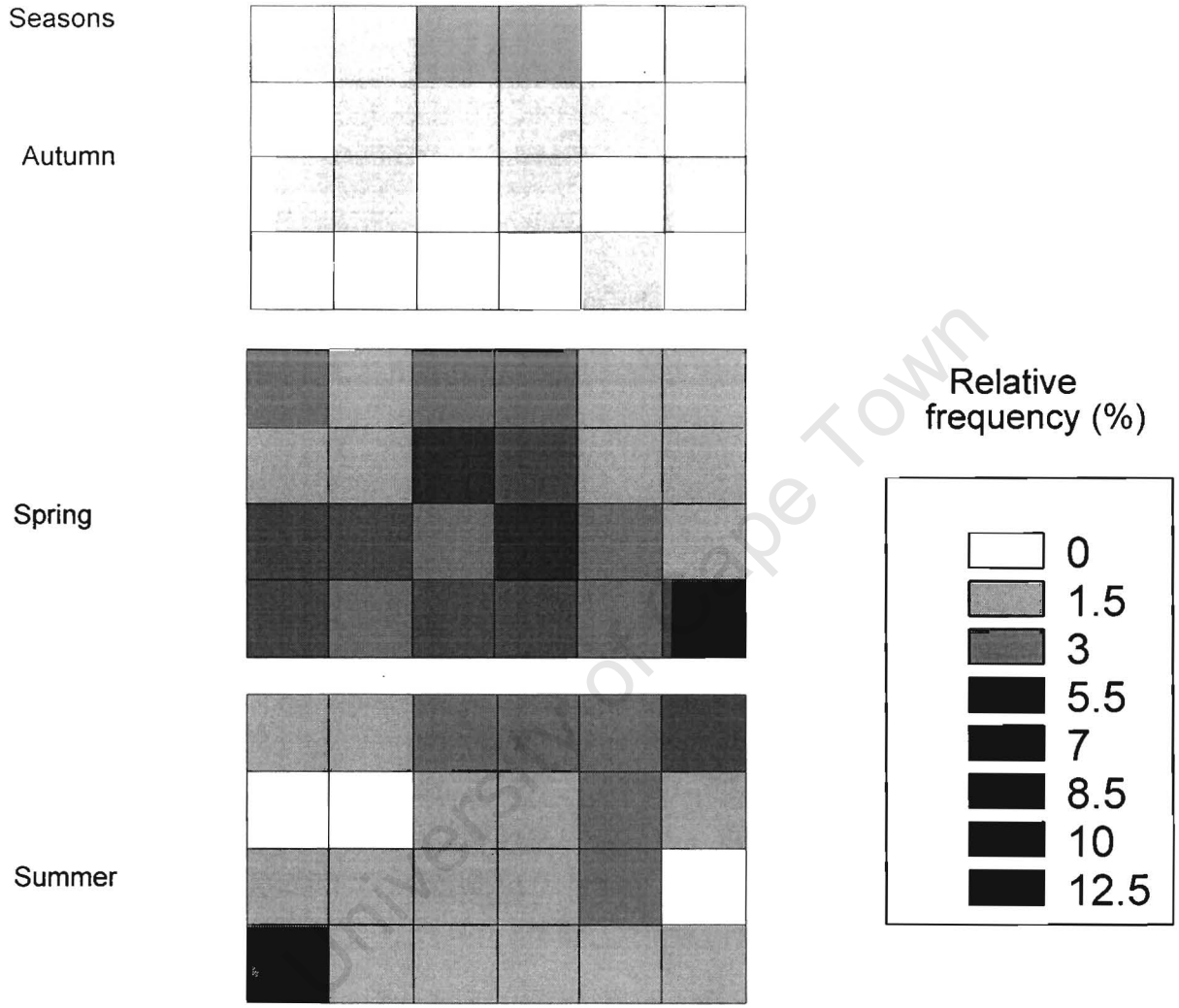


Figure 3.15. Relative frequencies of chlorophyll patterns at different seasons: (a) autumn, (b) spring, and (c) summer. All seasonal levels add up to 100%.

moderate and high (4-12 $mg.m^{-3}$), narrow surface and near-surface peaks (patterns 3, 4, 5, 10, 11 and 17). Profiles with almost uniform chlorophyll throughout the water column and subsurface peaks were less common in this season than in spring.

3.4.3 Relating seasonal and spatial patterns to environmental variables

The overall frequency map of the 6x4 SOM was partitioned into categories of SST, surface chlorophyll concentration and the depth of the water column to get chlorophyll distribution at these different categories. A shift in chlorophyll peak was evident as the temperature of the water warmed (Fig. 3.16). Surface and near-surface peaks with high chlorophyll concentration were dominant at low SST (<14°C). A profile with moderate surface peak (#3) dominated (frequency ~2%) at cool SST. As the sea surface water warmed, surface peaks with high total chlorophyll concentration within the peak (h) and narrow peaks (σ) developed (Fig. 3.16). Profiles with deep chlorophyll maximum were frequently found at warm SST (e.g. #19 had a frequency ~8%) with a mixture of profile shapes at intermediate SST (14-18°C). A shift in the depth of the peak with increasing concentrations of chlorophyll *a* at the sea surface is shown in Figure 3.17. Profiles with subsurface peaks were common at low surface chlorophyll concentrations (<1 $mg.m^{-3}$). At intermediate concentrations (1-8 $mg.m^{-3}$), profiles with near-surface and surface peaks dominated. As high surface chlorophyll concentration (>8 $mg.m^{-3}$) prevailed, large surface maxima were evident, and these represented ~6% of profiles. The chlorophyll maximum also occurred deep in the water column (Fig. 3.18). Inshore (<80 m), surface and near-surface peaks were frequent, and in midshore a mixture of profiles with surface and subsurface peaks predominated. In deeper waters, subsurface chlorophyll maxima dominated.

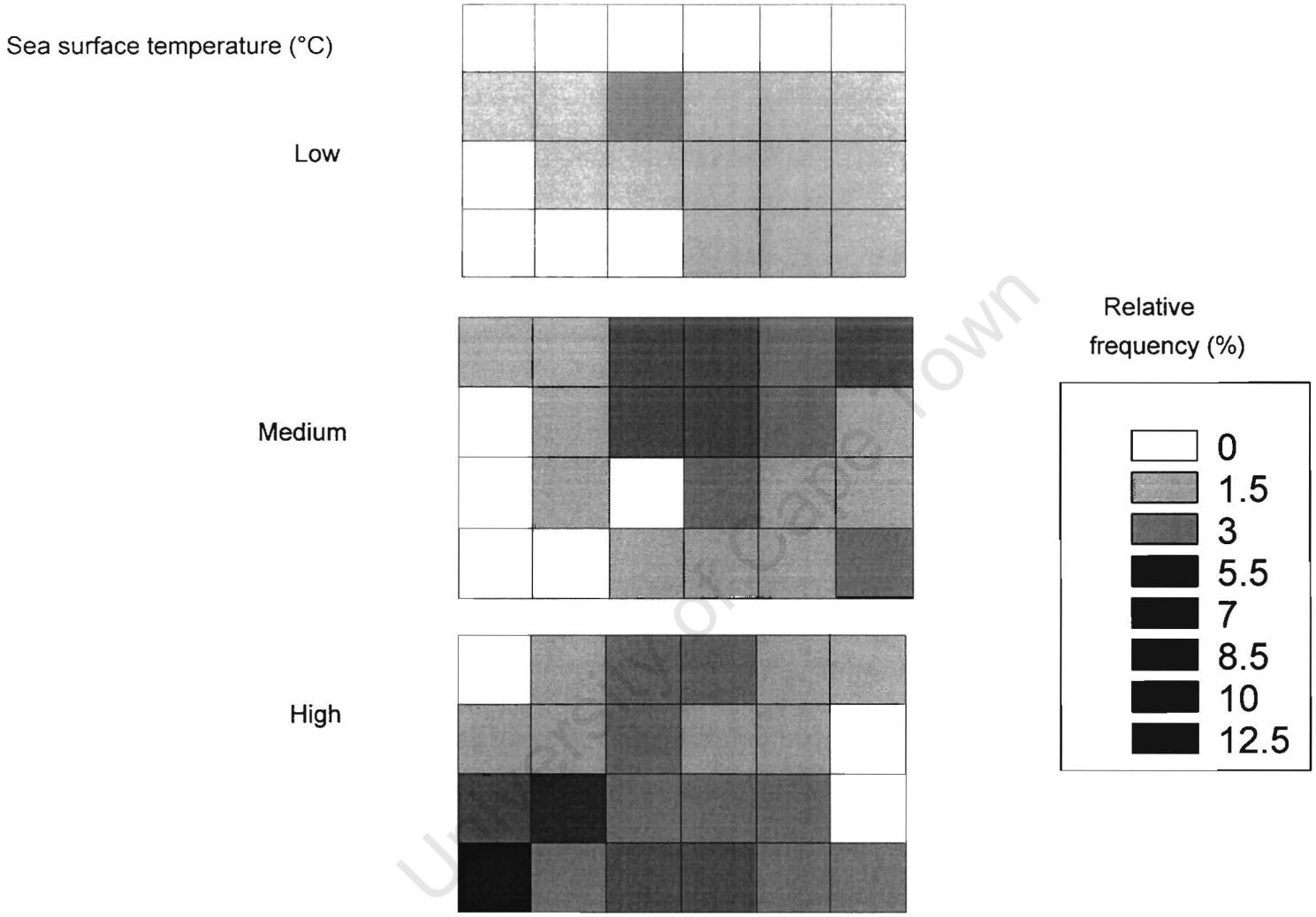


Figure 3.16. Relative frequencies of chlorophyll patterns at different sea surface temperatures: (a) low <14°C, (b) medium = 14-18°C, and (c) high >18°C. Levels add up to 100%.

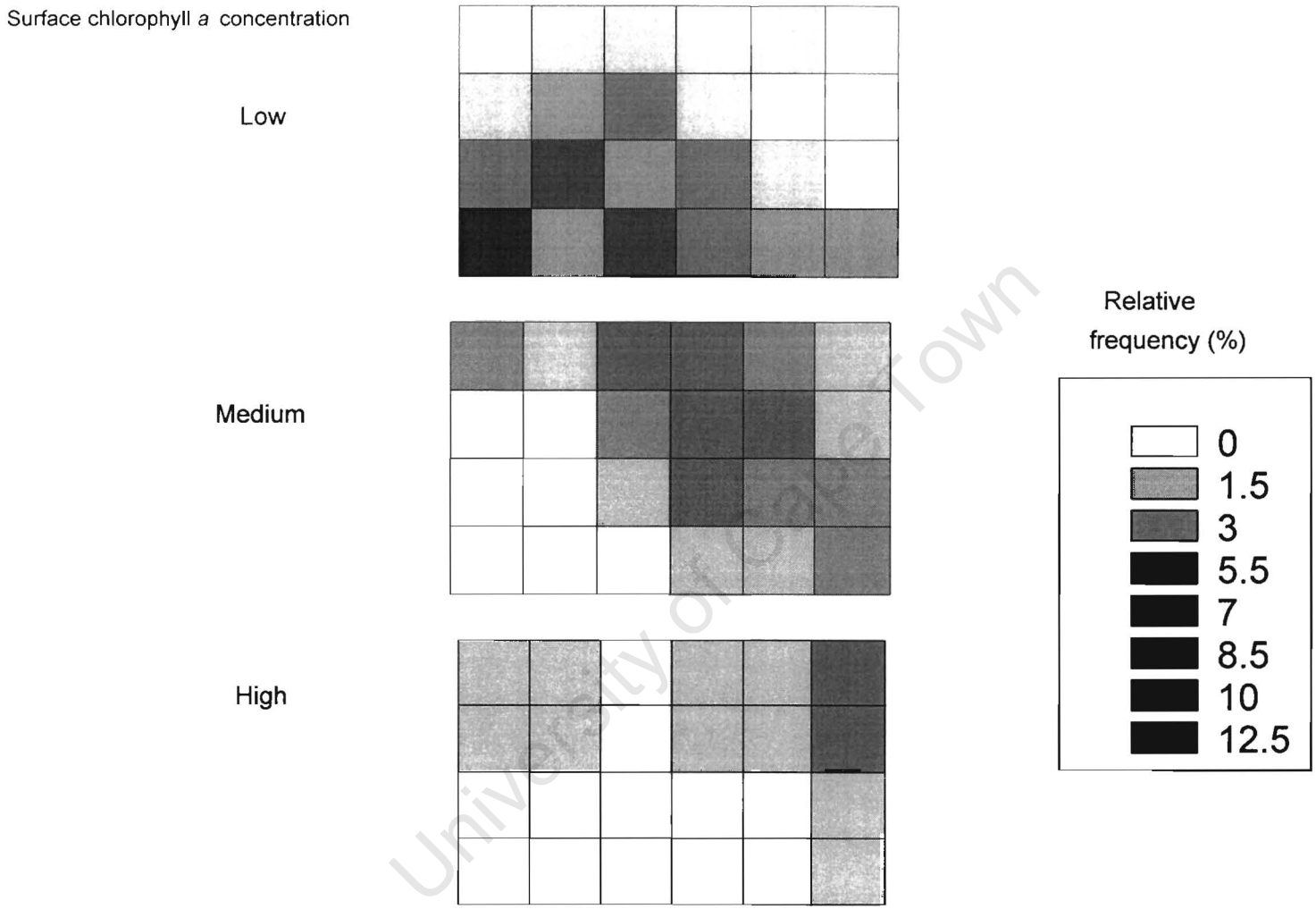


Figure 3.17. Relative frequencies of chlorophyll patterns at different surface chlorophyll concentration: (a) low <1 mg.m^{-3}, (b) medium 1-8 mg.m^{-3}, and (c) high >8 mg.m^{-3}. Levels add up to 100%.

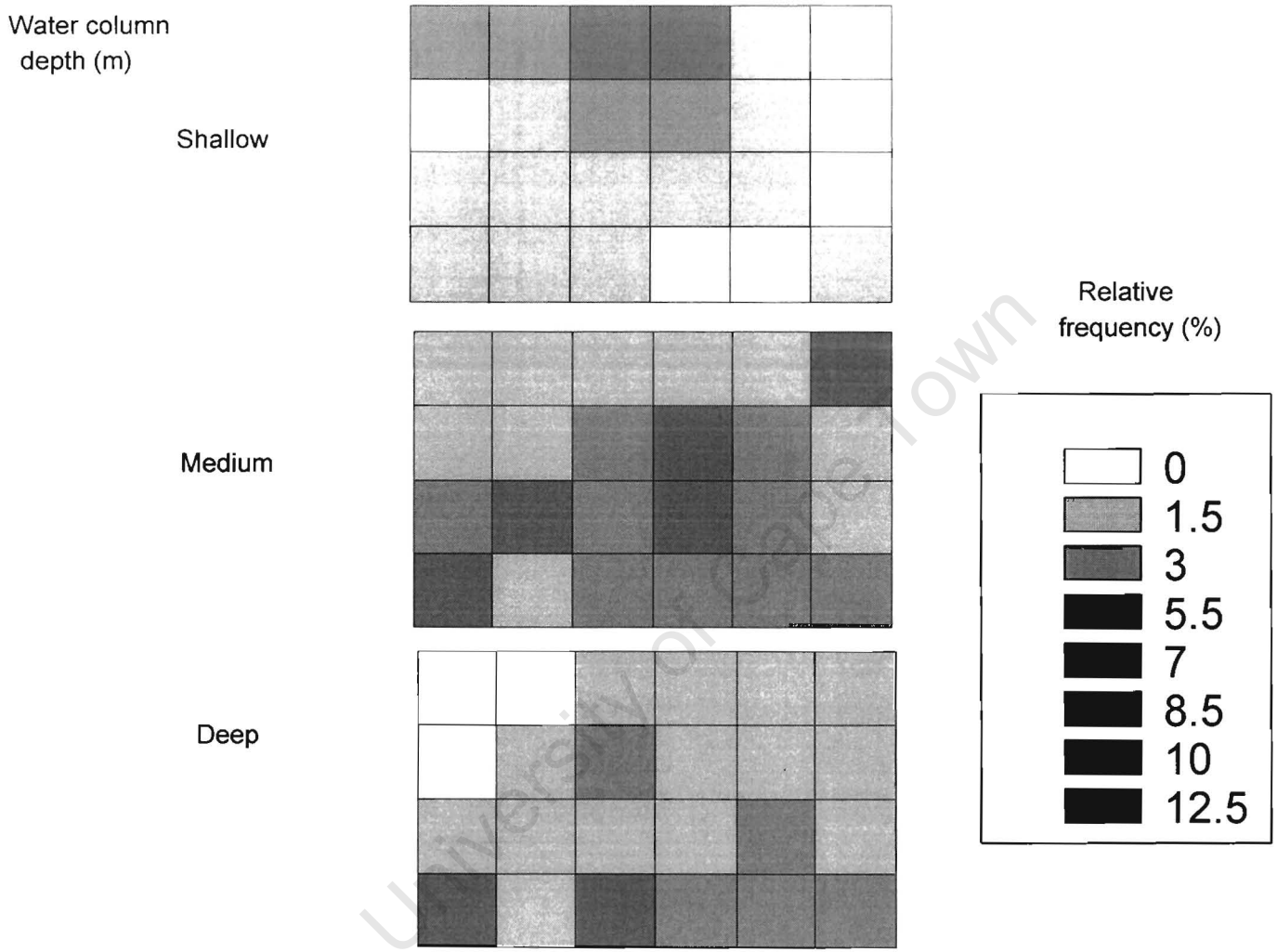


Figure 3.18. Relative frequencies of chlorophyll patterns at different water column depths: (a) shallow <80 m, (b) intermediate = 80-170 m, and (c) deep >170 m. Levels add up to 100%.

3.5 General trends of profile parameters with environmental variables

As an initial attempt at finding the relationship between the parameters (B_0, σ, h and z_m) and environmental variables, scatterplots with Least-squared weighted smoothers were created. These scatterplots were made to visualize possible trends, and not to test statistical significance of parameters with environmental variables. Generally, plots indicated non-linear relationships between parameters and environmental variables.

Background chlorophyll concentration (B_0) scatterplots (Fig. 3.19) showed non-linear relationships between B_0 and environmental variables. At shallow water column depths, cool SST and low surface chlorophyll concentrations, relatively high background chlorophyll concentrations were found. As SST warmed and high surface chlorophyll concentrations prevailed, low background chlorophyll concentrations occurred, particularly offshore.

The width of the peak (σ) plots (Fig. 3.20) showed non-linear trends of this parameter in response to environmental variables. At cool (<14 °C) and very warm SST (>20 °C), narrow peaks were found particularly at shallow water column depths (<80 m) and high surface chlorophyll concentrations (>8 $mg.m^{-3}$). Broader peaks were found offshore at intermediate SST and low chlorophyll concentrations (Fig. 3.20).

The total chlorophyll concentration within the peak (h) was higher at intermediate SST (14-18 °C), while at cool SST (<14 °C) and very warm SST (>18 °C) low concentrations of total chlorophyll within the peak existed, especially at shallow water

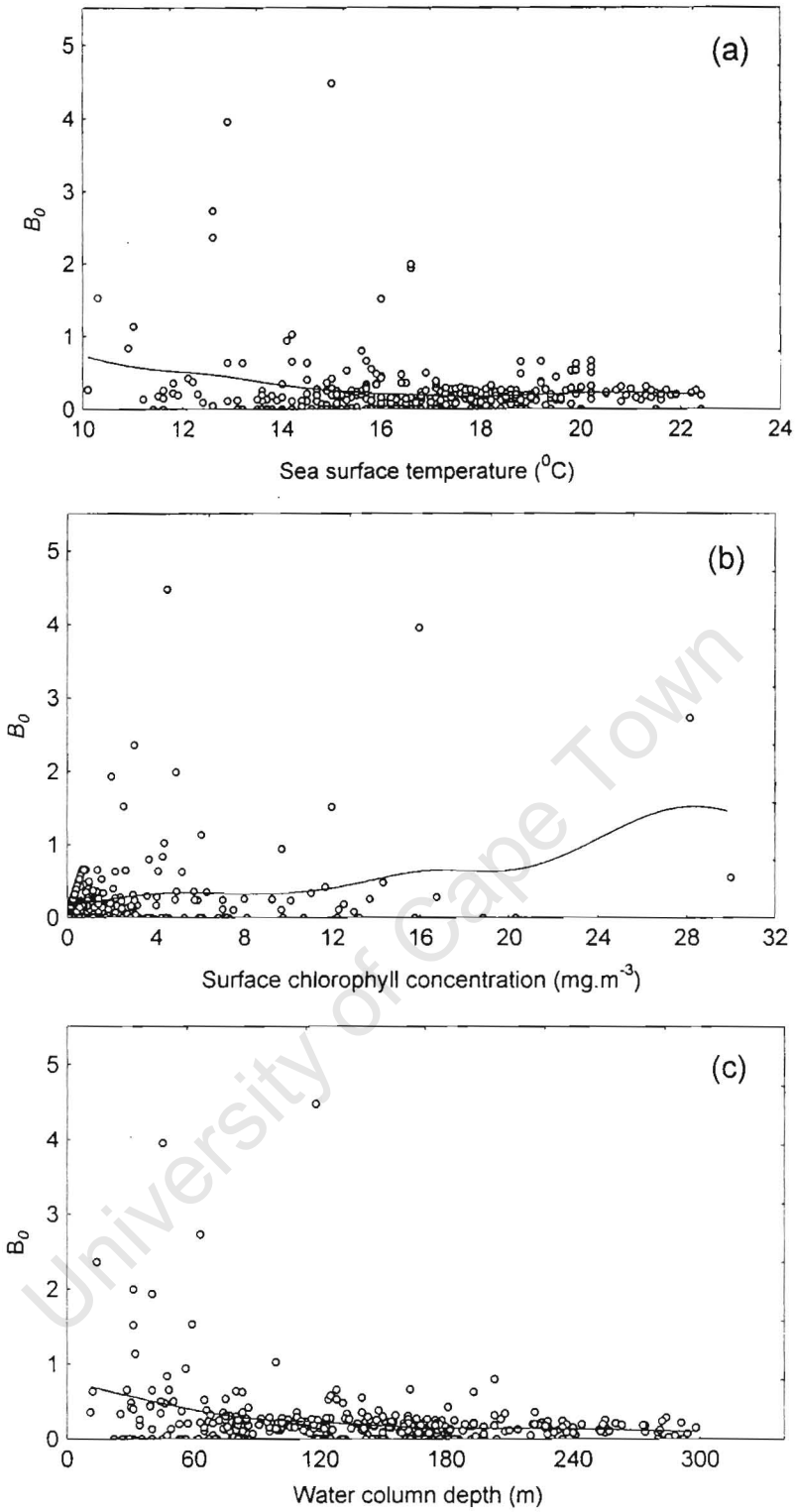


Figure 3.19. Relationships between background chlorophyll concentration (B_0) and (a) SST (°C), (b) surface chlorophyll concentration (mg.m^{-3}), and (c) water column depth (m).

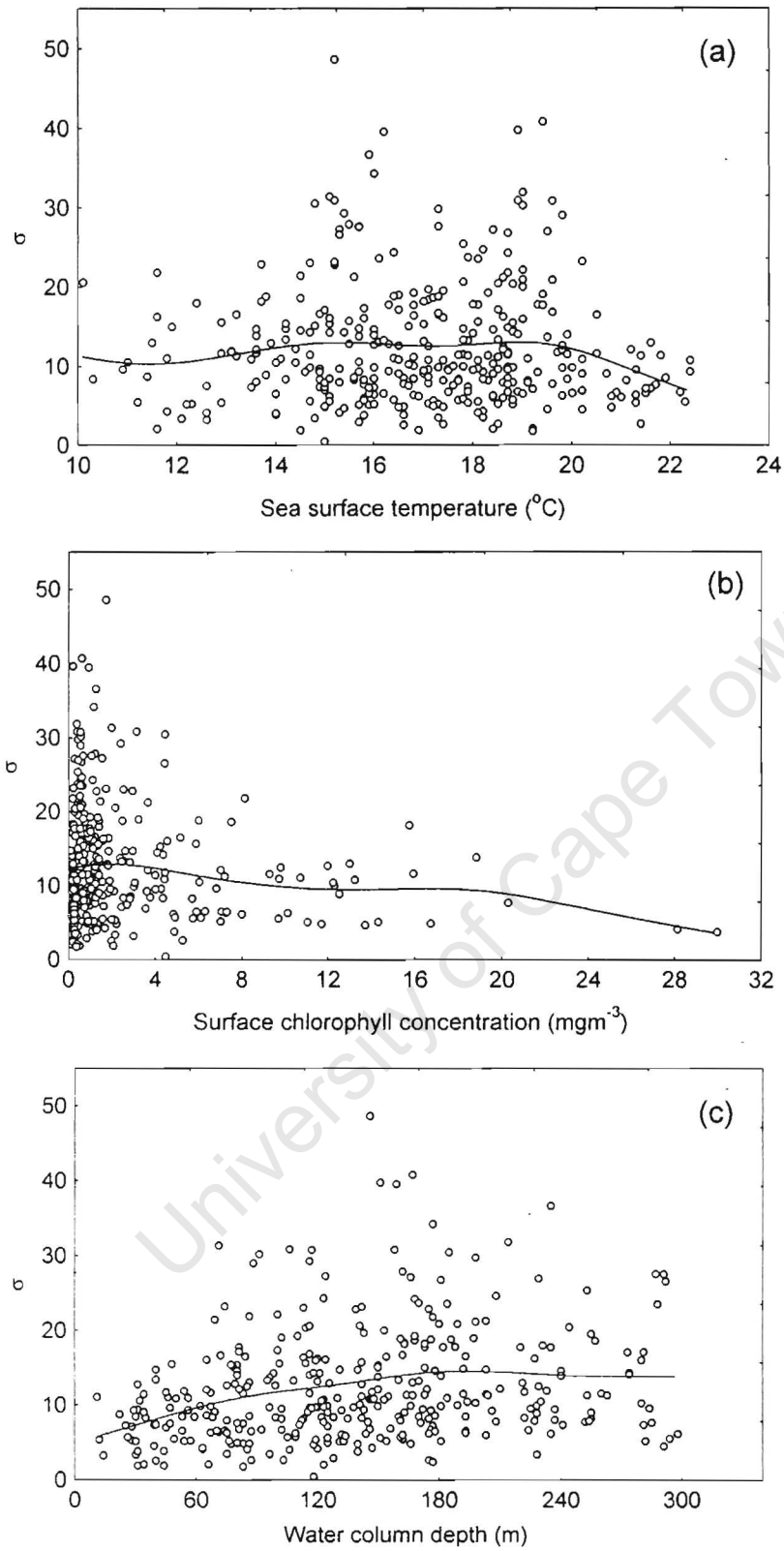


Figure 3.20. Relationships between the width of the peak (σ) and (a) SST (°C), (b) surface chlorophyll concentration ($mg.m^{-3}$), and (c) water column depth (m).

column depths (<80 m). As the surface chlorophyll concentration increased, the total chlorophyll concentration within the peak (h) also increased (Fig. 3.21).

The inshore region, where upwelling occurs, was characterized by cool SST (10-14 °C), indicating the onset of upwelling. Plots of the depth of the peak (z_m) (Fig. 3.22) showed near surface and surface peaks (z_m<10 m) with high surface chlorophyll concentrations (~35 mg.m^{-3}) at cool SST (10-14 °C), a possible indication of bloom development at these temperatures. Deep chlorophyll maxima (z_m~40 m) with low chlorophyll concentrations existed at warm SST and deeper water columns (offshore).

3.6 Relating the shape of chlorophyll profiles to environmental variables: a quantitative approach

3.6.1 Generalized additive models for the shifted Gaussian parameters

Relationships of environmental variables and profile parameters from the shifted Gaussian curve were investigated using GAMs. The scale of the y-axis from the GAM plots gives zero (0) as the mean effect of the adjusted predictor variable on each response variable; positive y-axis indicates a positive impact of adjusted predictor variable on the response, with negative values implying a negative effect of the predictor on the response. For this study, the y-axis of the graph indicates the relative magnitude of the effect of smoothed environmental variable on the profile parameter. High positive y-values show a greater effect of the environmental variable on the response, while low values indicate a lesser effect. For example, if SST has y-values (-2, 0, 2, 4) and surface chlorophyll has values (-1, 0, 1), then SST has a greater influence on the modelled parameter than surface chlorophyll. The 95% confidence intervals are shown in each plot (dotted lines), and a rug plot is included on the x-axis.

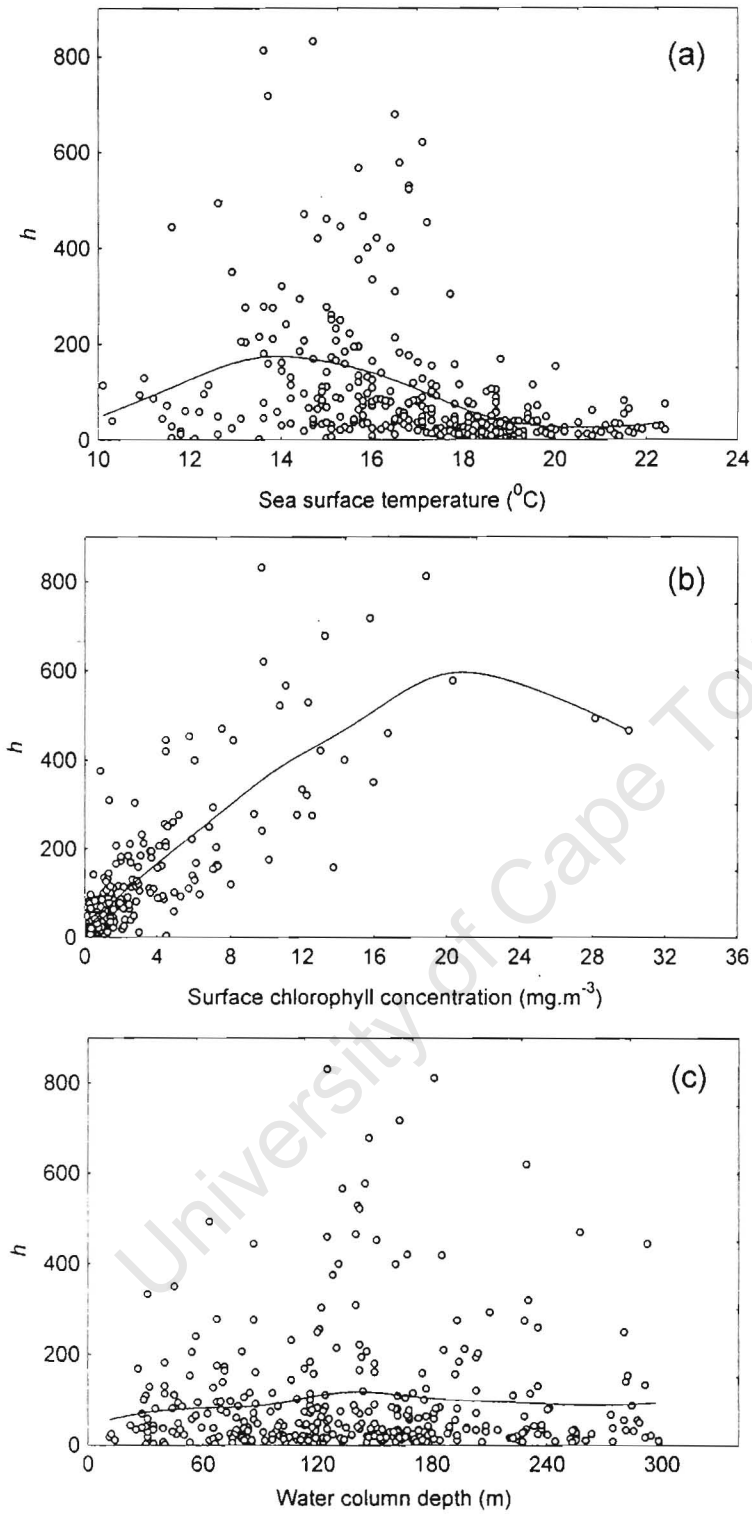


Figure 3.21. Relationships between the total chlorophyll concentration within the peak (*h*) (b) surface chlorophyll concentration ($mg.m^{-3}$), and (c) water column depth (m).

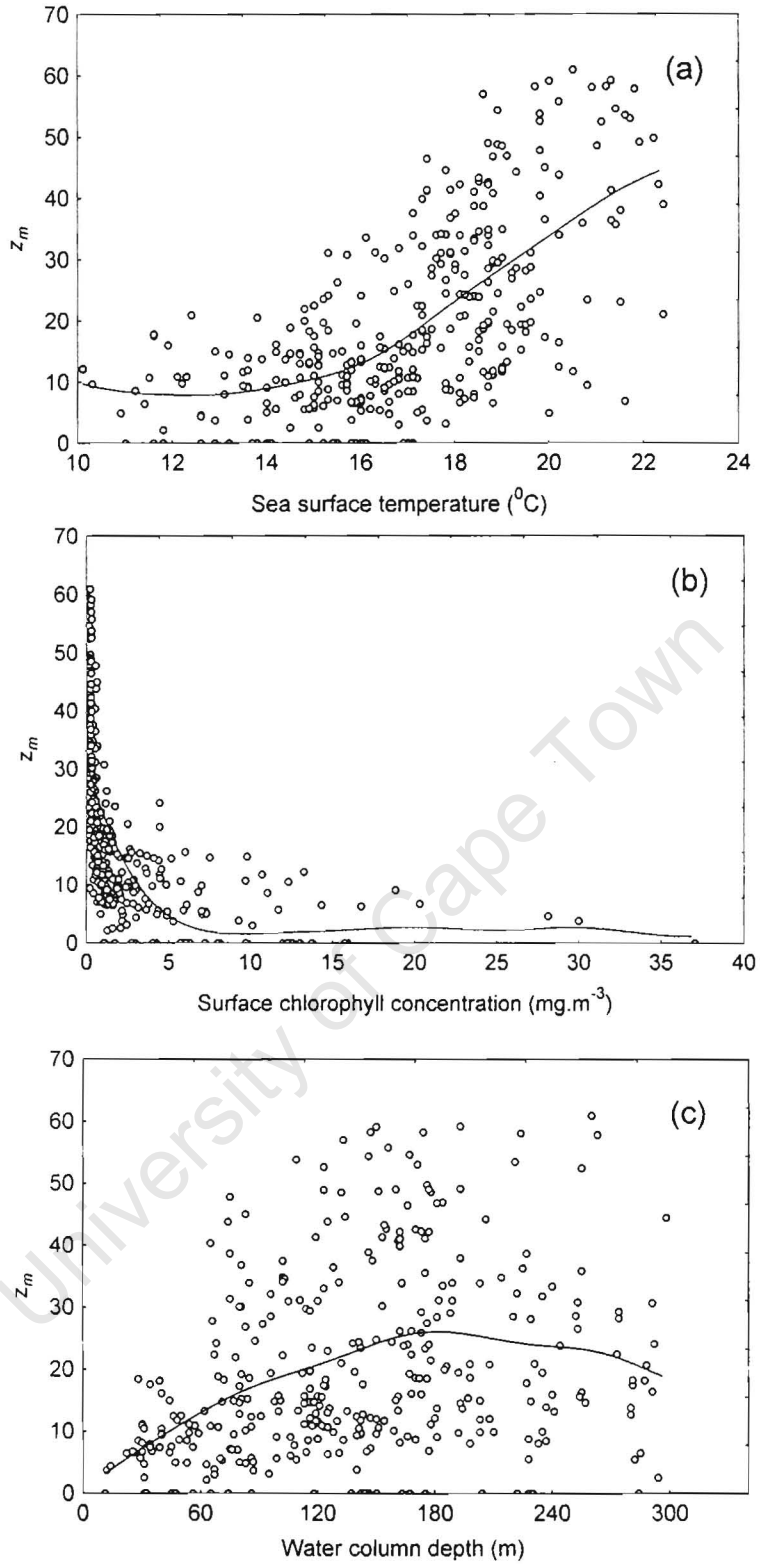


Figure 3.22. Relationships between the depth of the peak (z_m) and (a) SST(°C), (b) surface chlorophyll concentration (mg.m^{-3}), and (c) water column depth (m).

The most important parameters that could significantly model the chlorophyll distribution were the total chlorophyll within the peak (h) and the depth of the chlorophyll peak (z_m).

The background chlorophyll concentration (B_0)

The background chlorophyll concentration was modelled as a function of two variables, sounding (water column depth) and surface chlorophyll concentration (Fig. 3.23) with a greater effect of surface chlorophyll than water column depth. The background chlorophyll concentration (B_0) was higher inshore, and then decreased as the water column deepened until B_0 remained fairly constant at depths greater than 150 m. Low concentrations of B_0 occurred at low surface chlorophyll concentrations. There was a general increase in B_0 as surface chlorophyll concentrations increased. Modelling B_0 with these variables only explained ~18 % of the variance in the model (Table 3.1), indicating poor predictability of this parameter from surface chlorophyll concentration and water column depth.

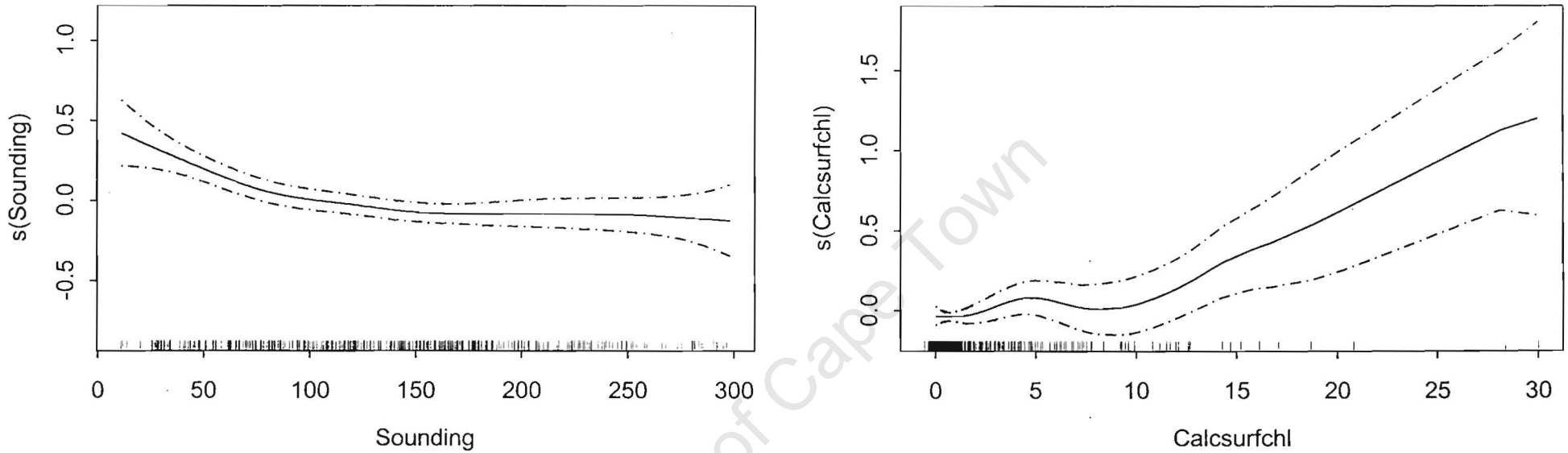


Figure 3.23. A generalized additive model of the background chlorophyll concentration (B_0) modelled with surface chlorophyll and water column depth. Note the y-axis is the B_0 given as a smoother of each environmental variable. Dotted lines are 95% confidence intervals.

Table 3.1: A summary table of GAMs and GLMs for profile parameters. Variables used in modelling are indicated by √, and variance explained (r^2) is included for each model. Square-root and log transformations were used for h and σ respectively.

Parameter	Surface chlorophyll	SST	Water column depth	Area	Season	r^2 -GAM (%)	r^2 –GLM (%)
B_0	√	-	√	-	-	17.7	15.1
Sqrt h	√	-	√	-	√	80.0	73.6
Log σ	√	√	√	√	√	23.4	20.8
z_m	√	√	√	√	√	66.7	69.8

The width of the peak (σ)

From the model fitting procedure, the transformed width of the peak (log σ) was modelled as a function of water column depth, season, surface chlorophyll concentration and area (Fig. 3.24). The water column depth and surface chlorophyll concentration had a greater effect on σ than season and area. In shallow water columns, the width of the peak was narrow, and σ became broader offshore. Broader peaks were found mainly in spring with narrow peaks occurring at other seasons. Narrower peaks were commonly found at high surface chlorophyll concentrations

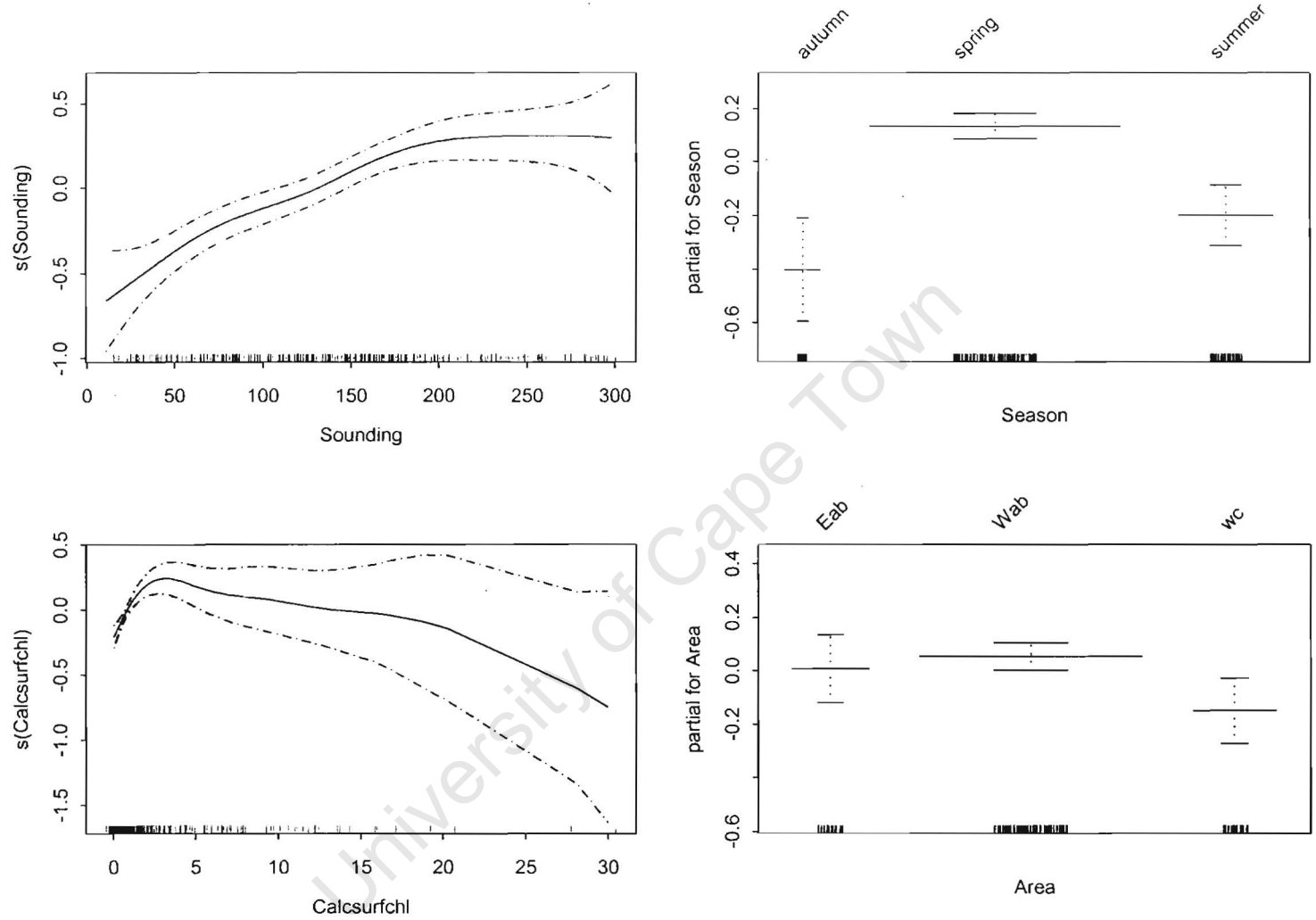


Fig. 3.24. A generalized additive model of the width of the peak (log σ) modelled as a function of water column depth, season, surface chlorophyll concentration and areas. Note the y-axis is log σ given as a smoother of each environmental variable. Dotted lines are 95% confidence intervals.

(>10 $mg.m^{-3}$) and these became broader as surface chlorophyll concentrations decreased. The West Coast sub-province was characterized by narrower peaks, and eastern and western Agulhas Bank sub-provinces had intermediate peaks. Approximately 23% of the variance was explained by the additive model for this parameter (Table 3.1).

The total chlorophyll concentration within the peak (h)

The total chlorophyll concentration within the peak was transformed (sqrt *h*) and modelled as a function of surface chlorophyll concentration, sounding, season and SST (Fig. 3.25). These parameters explained the seasonal distribution of the total chlorophyll concentration within the peak (*h*) well. Surface chlorophyll concentration had a greater effect on this parameter than other parameters. An increasing chlorophyll concentration within the peak coincided with an increase in surface chlorophyll concentrations. In summer, peaks had high chlorophyll concentrations, and intermediate chlorophyll concentrations within the peak occurred in spring, and autumn had low chlorophyll within the peak. The total concentration within the peak increased as surface water warmed, and low concentrations of *h* developed at high SST (>18 °C). The GAM explained 80% of the variance for the *h* model (Table 3.1), indicating high predictability of this parameter from these environmental variables.

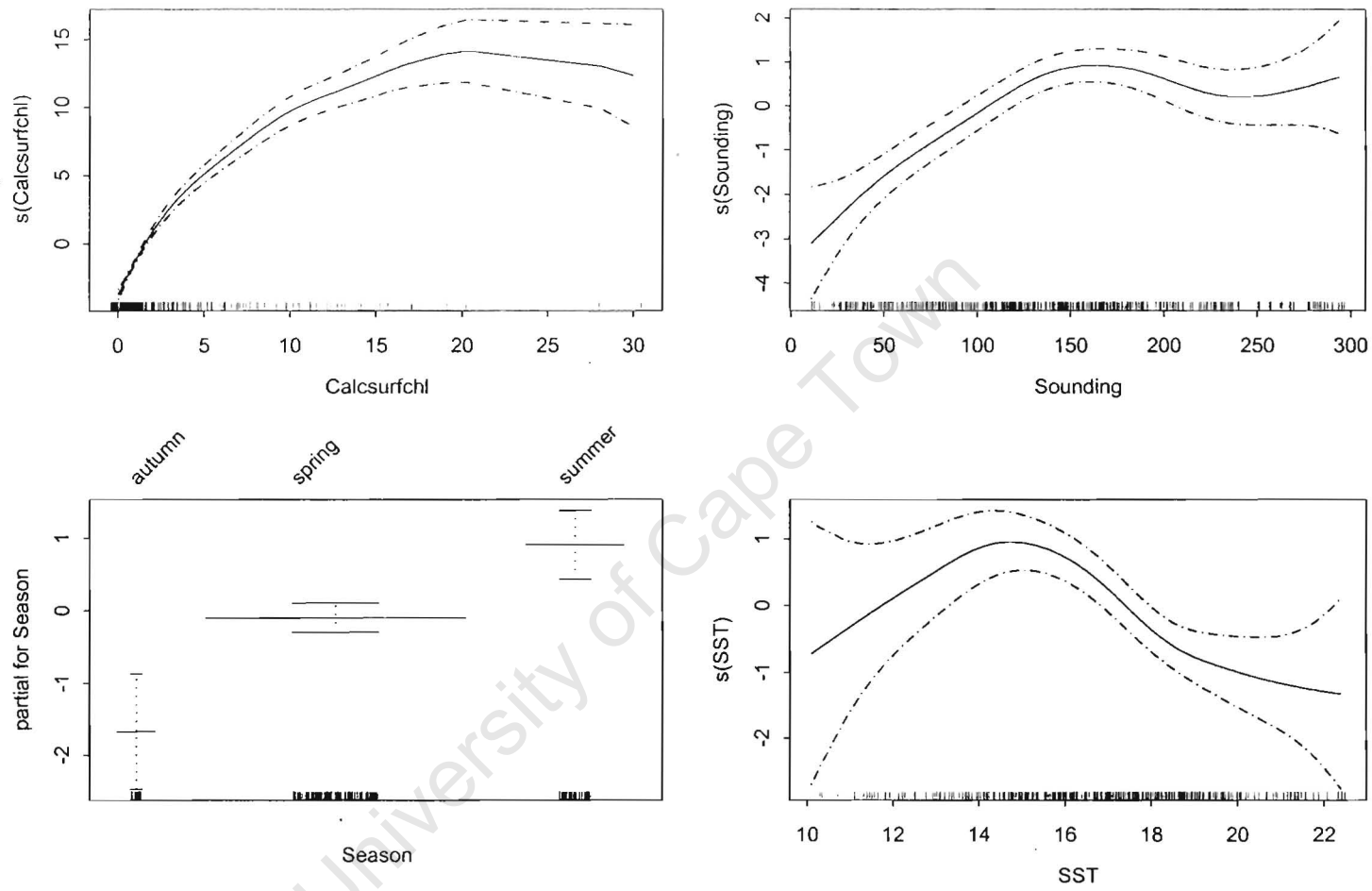


Figure 3.25. A generalized additive model of the total chlorophyll concentration within the peak (sqrt *h*) modelled as a function of surface chlorophyll water column depth, season and SST. The y-axis is the sqrt *h* given as a smooth function of environmental variables.
Dotted lines are 95% confidence intervals.

The depth of the chlorophyll peak (z_m)

A GAM modelled the depth of the chlorophyll maximum (z_m) as a function of surface chlorophyll concentration, SST, water column depth, season and area (Fig 3.26). Plots showed SST as the variable with a greater magnitude effect on this parameter. Surface chlorophyll maxima dominated at high surface chlorophyll concentrations (>10 $mg.m^{-3}$). The depth of the maximum shifted subsurface at low surface chlorophyll concentrations. Surface peaks developed at cold and intermediate SST (10-16 °C) and as surface temperature warmed, the chlorophyll maximum deepened. Surface peaks were features of shallow depths, and subsurface chlorophyll maxima dominated offshore. In summer and autumn, surface chlorophyll maxima were found, and these maxima shifted to subsurface layers in spring. Subsurface chlorophyll peaks were frequent on the Agulhas bank sub-province, with West Coast sub-province dominated by surface peaks. The model explained ~67% of the total variance for z_m (Table 3.1).

3.6.2 Summary of GAMs for profile parameters

A summary of the GAM for each profile parameter is given below, showing all the significant environmental variables used in modelling.

$B_0 = \mu + s_1$ (sounding) + s_2 (surface chlorophyll).

Log (σ) = $\mu + s_1$ (sounding) + (season) + s_2 (surface chlorophyll) + (area) + s_3 (SST).

Sqrt $h = \mu + s_1$ (surface chlorophyll) + s_2 (sounding) + (season).

$z_m = \mu + s_1$ (surface chlorophyll) + s_2(sounding) + s_3(SST)+ (season) + (Area).

Where μ is the overall mean of the profile parameter and the s_i are smooth functions (cubic spline smoothers). Note that the categorical factors do not have a smoothing spline.

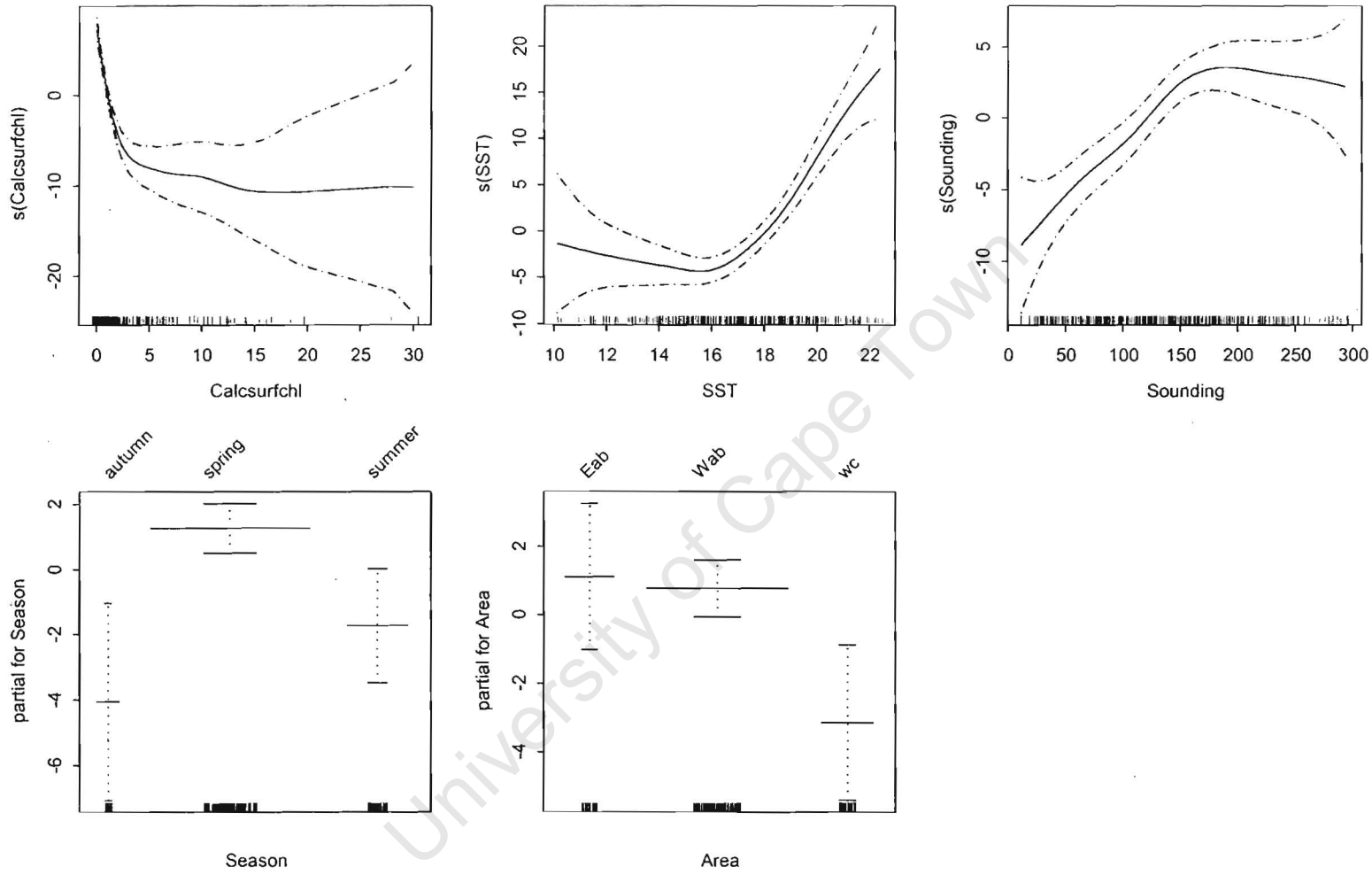


Figure 3.26. A generalized additive model of the depth of the chlorophyll maximum (z_m) modelled as a smooth function of surface chlorophyll, SST, water column depth, season and area. The y-axis is the z_m given as a smoother of each environmental variable. Dotted lines are 95% confidence intervals.

3.6.3 Advantages of GLM over GAM

Generalized linear models are more parsimonious than generalized additive models because they often use fewer parameters. Moreover, a linear model produces an analytical expression (equation) that can be used for predictions, which cannot be obtained when fitting using GAM (GAM does provide predictions but not an analytical expression). The linear equation is a result of the estimation of the coefficients of the linear relationship whereas for the additive model with smoothers, the smooth functions are nonparametric estimates of the relationship. These nonlinear estimates do not give an equation because they have no analytical form and are based on an iterative algorithm. The explicit equation from a GLM makes the results more portable (equation can be used in other studies).

3.6.4 Generalized linear models for the shifted Gaussian parameters

GAM plots highlighted possible linear relationships of each parameter with environmental variables; therefore GLMs were created to model these relationships. Although generalized additive models gave meaningful results of parameters with environmental variables and could be used for predictions, GLMs were preferred over GAMs because they give an equation that can be used for predictions. Except for the z_m linear model, the variance explained by the GLMs was lower than for the GAMs because some of the information is lost in constant breakpoint regressions. Table 3.2 shows the ANOVA results for the GLMs of the Gaussian parameters with their F-values; and p-values of the variables were considered significant at 5% level.

Table 3.2: ANOVA Table for the GLMs of the Gaussian parameters. Terms are significant at 5% level, and $^{+}$ indicates terms that are marginally significant with $p<0.1$

Response (parameter)	Terms used in GLM	*df*	Deviance	Residual *df*	Residual deviance	F-value	*p*-value
B_0	Sounding	1	5.4	344	57.5	34.6	0.000
	Surface chlorophyll	1	14.5	343	53.0	28.9	0.000
Log σ	Sounding	1	12.3	344	137.3	38.1	0.000
	Season	2	9.0	342	116.0	14.0	0.000
	Surface chlorophyll	2	5.8	340	110.2	9.0	0.000
	Area	2	1.5	338	108.7	2.4	0.096^{+}
Sqrt h	Surface chlorophyll	1	6211.3	342	3102.4	853.7	0.000
	Sounding	1	141.0	341	2961.4	19.4	0.000
	Season	2	97.3	339	2864.1	6.7	0.001
	SST	1	405.1	338	2459.1	55.7	0.000
Z_m	Surface chlorophyll	1	45519.3	342	29670.3	672.8	0.000

SST	1	4419.2	341	25251.6	65.3	0.000
Sounding	1	1654.8	340	23596.4	24.5	0.000
Season	2	472.6	338	23123.8	3.5	0.031
Area	2	390.2	336	22733.5	2.9	0.057^{+}

The background chlorophyll concentration (B_0)

As in GAM, for B_0 two important predictors were sounding and surface chlorophyll. Piecewise linear fit was considered for both predictors with breakpoint at 90 m for sounding (indicating shallow depths) and at 10 $mg.m^{-3}$ for surface chlorophyll concentration (Fig. 3.27). Exponential fit for sounding and piecewise linear fit for surface chlorophyll were also tried, but resulted to a lower proportion of variance explained (r^2 value) and a higher residual deviance (*Error*) when compared to the previous model, and therefore were discarded. Approximately 15% of the proportion of variance was explained by the GLM for B_0, and this is slightly lower than for the B_0 GAM.

The width of the peak (σ)

The width of the peak (log σ) was modelled as a function of sounding, season, surface chlorophyll and area. Season and area were categorical variables and these gave the same results as in GAMs. Surface chlorophyll concentration was modelled as a polynomial with 3 *degrees of freedom*; and a breakpoint regression at 170 m was used

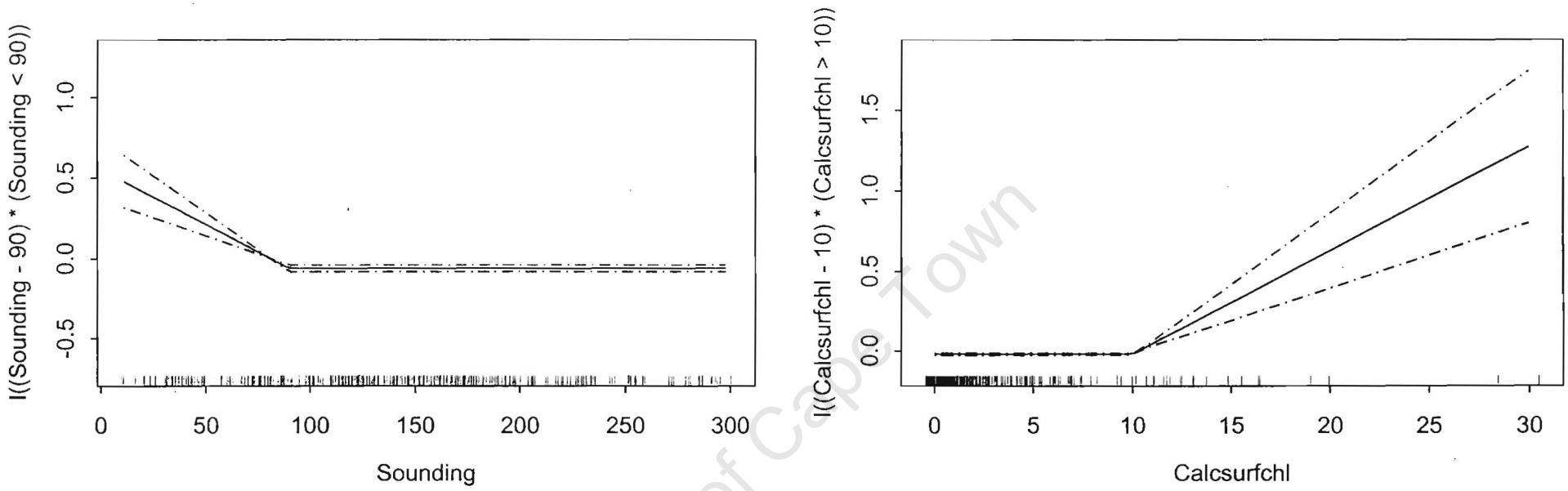


Fig. 3.27. A generalized linear model of the background chlorophyll concentration (B_0) modelled with water column depth and surface chlorophyll, y-axis is B_0 given as a breakpoint regression of the variables. Dotted lines are 95% confidence intervals.

to model sounding (Fig. 3.28). The GLM explained ~21% of the variance for this parameter (Table 3.1).

The total chlorophyll concentration within the peak (h)

The transformed *h* parameter (square-root *h*) was modelled as a function of three significant predictor variables, surface chlorophyll, sounding, season and SST (Fig 3.29). Surface chlorophyll suggested a breakpoint at 20 $mg.m^{-3}$; sounding was modelled with piecewise linear fit (breakpoint at 140 m). Season was fitted as a categorical variable, therefore same results as in GAMs were obtained; and SST had a breakpoint at 15°C (Fig. 3.29). Other possible relationships included log fit for both surface chlorophyll and sounding. These relationships were also tried in order to build a better model, but it was found that they were not better compared to the initial model for this parameter (piecewise linear fit for chlorophyll, sounding and SST). The model explained 74% of the variance for the *h* parameter (Table 3.1).

The depth of the peak (z_m)

The depth of the peak (z_m) used all the predictor variables to build the GLM. Surface chlorophyll suggested an exponential fit; SST and sounding were modelled using piecewise linear fits, with breakpoints at 17°C and 170 m respectively. Season and area were treated as categorical variables (Fig. 3.30). Table 3.1 also gives a summary of GLMs constructed for each parameter. Surprisingly, the linear model for this parameter explained even a higher proportion of the variance (70%) than the additive model (67%) (Table 3.1).

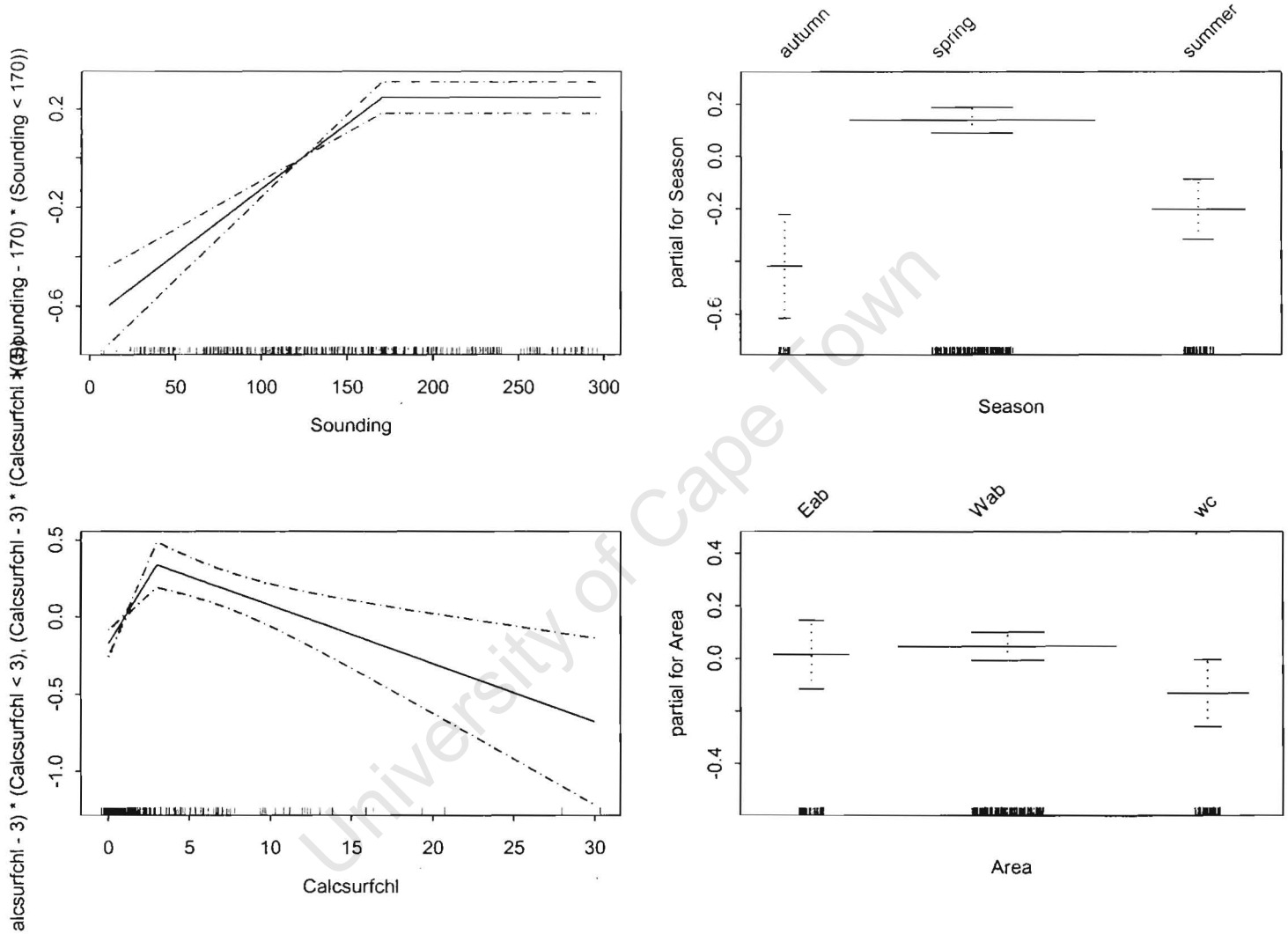


Fig. 3.28. A generalized linear model of the width of the peak (σ) (log σ) modelled as a function of water column depth, season, surface chlorophyll concentration and area. Note the y-axis is the log (σ) shown as breakpoint regression for depth and chlorophyll. Dotted lines are 95% confidence intervals.

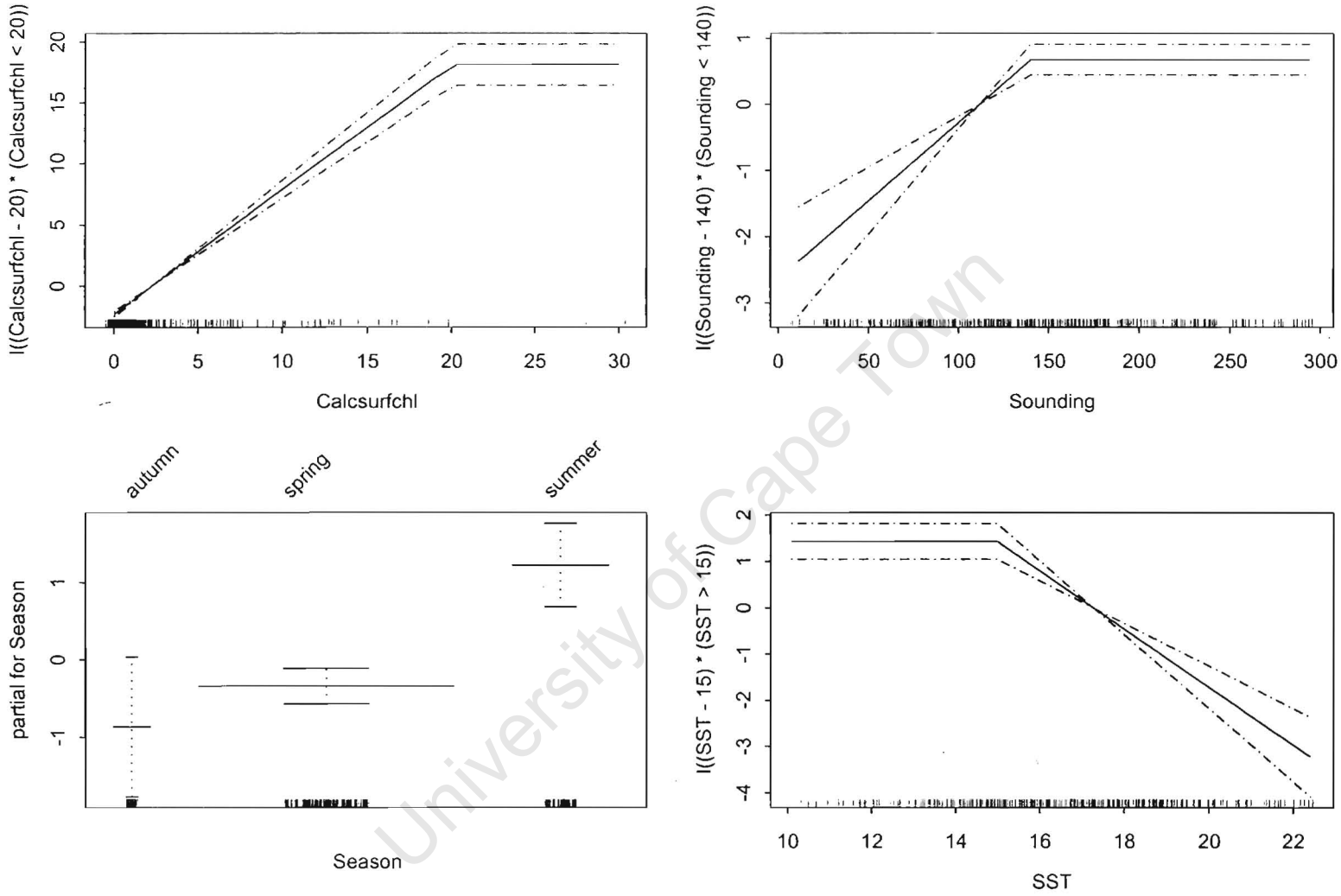


Figure 3.29. A generalized linear model of the total chlorophyll concentration within the peak (sqrt *h*) modelled as a function of surface chlorophyll, water column depth, season and SST. The y-axis is sqrt *h* given as a regression of environmental variables. Dotted lines are 95% confidence interval

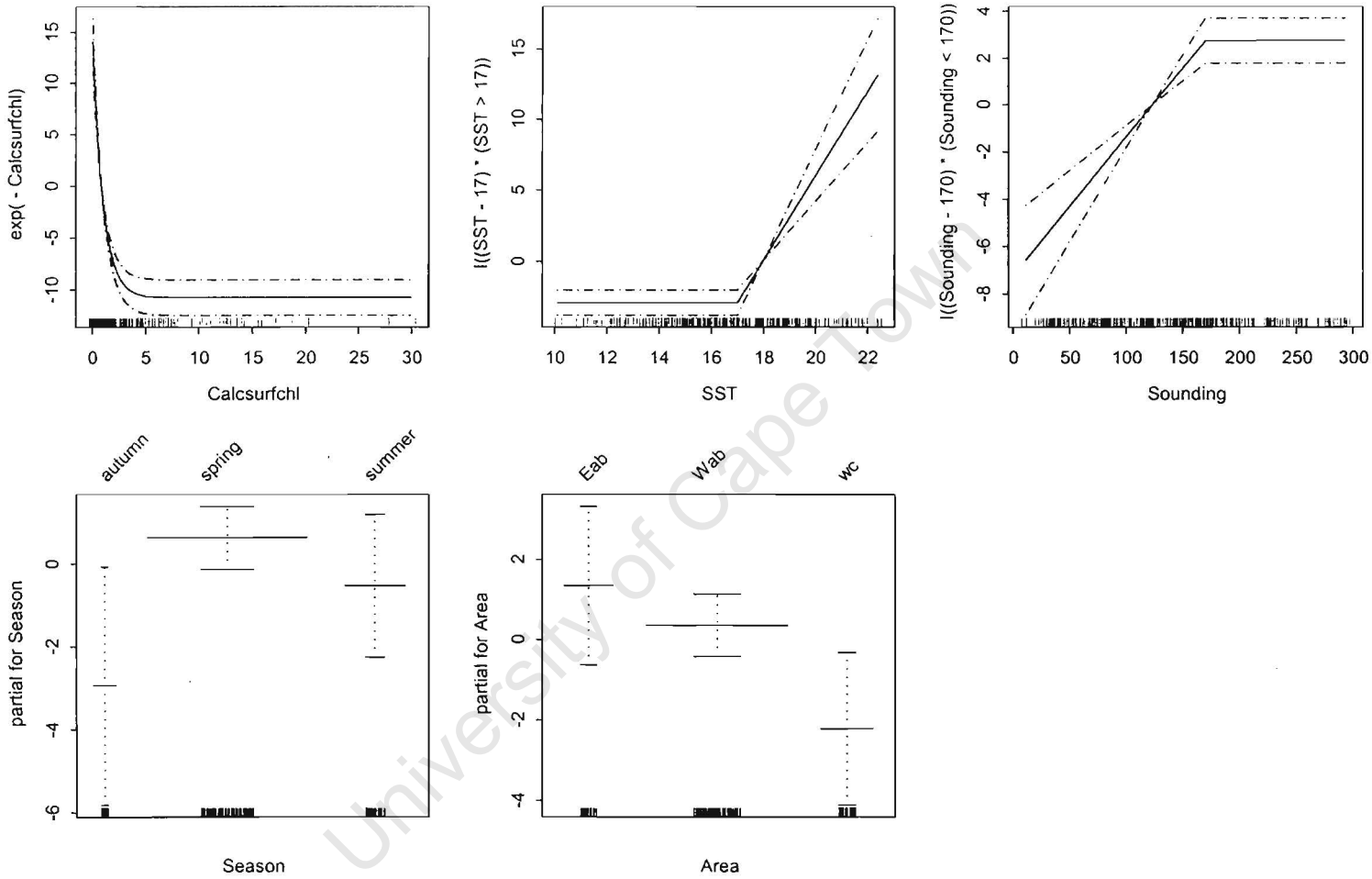


Figure 3.30. A generalized linear model of the depth of the chlorophyll maximum (z_m) modelled as a function of surface chlorophyll, SST, water column depth, season and area. The y-axis is z_m given as an exponential and regression for environmetal variables. Dotted lines are 95% confidence intervals.

3.7 GLM equations for the profile parameters

Equations for predictions were obtained from generalized linear modelling of profile parameters, and these are shown in Table 3.3 with the intercept and terms of the model included.

3.8 Predicting profile parameters using the generalized linear model

Predictions of profile parameters, which describe the shape of chlorophyll profiles, from environmental variables using generalized linear models (GLMs), are shown in Table 3.4. Except for z_m and h, correlations between the observed and predicted values were not significant (weak).

Table 3.3: Predictive equations of profile parameters from generalized linear modelling

$$B_0 = 0.163 - 0.0068 \times \begin{Bmatrix} (Dep - 90) & \text{, if } Dep < 90 \\ 0 & \text{, if } Dep > 90 \end{Bmatrix} + 0.064 \times \begin{Bmatrix} 0 & \text{, if } Chl < 10 \\ (Chl - 10) & \text{, if } Chl > 10 \end{Bmatrix}$$

$$\log \sigma = 2.51 + 0.005 \times \begin{Bmatrix} 0 & \text{, if } Dep < 170 \\ (Dep - 170) & \text{, if } Dep > 170 \end{Bmatrix} + \begin{Bmatrix} 0 & \text{, } Aut \\ 0.557 & \text{, } Spr \\ 0.216 & \text{, } Sum \end{Bmatrix} - 0.038 \times (Chl - 3) + \begin{Bmatrix} 0 & \text{, } EAB \\ 0.031 & \text{, } WAB \\ -0.147 & \text{, } WCO \end{Bmatrix}$$

$$\sqrt{h} = 27.73 + 1.020 \times \begin{Bmatrix} (Chl - 20) & \text{, if } Chl < 20 \\ 0 & \text{, if } Chl > 20 \end{Bmatrix} + 0.024 \times \begin{Bmatrix} (Dep - 140) & \text{, if } Dep < 140 \\ 0 & \text{if } Dep > 140 \end{Bmatrix} + \begin{Bmatrix} 0 & \text{, } Aut \\ 0.524 & \text{, } Spr \\ 2.085 & \text{, } Sum \end{Bmatrix} - 0.629 \times \begin{Bmatrix} 0 & \text{, if } SST < 15 \\ (SST - 15) & \text{if } SST > 15 \end{Bmatrix}$$

$$z_m = 7.487 + 25.33 \times e^{-Chl} + 2.975 \times \begin{Bmatrix} 0 & \text{, if } SST < 17 \\ (SST - 17) & \text{if } SST > 17 \end{Bmatrix} + 0.059 \times \begin{Bmatrix} 0 & \text{, if } Dep < 170 \\ (Dep - 170) & \text{, if } Dep > 170 \end{Bmatrix} + \begin{Bmatrix} 0 & \text{, } Aut \\ 3.573 & \text{, } Spr \\ 2.416 & \text{, } Sum \end{Bmatrix} + \begin{Bmatrix} 0 & \text{, } EAB \\ -0.994 & \text{, } WAB \\ -3.571 & \text{, } WCO \end{Bmatrix}$$

Where:

B_0 = Background chlorophyll concentration, σ = width of the peak, h = total chlorophyll concentration beneath the peak,

z_m = depth of the chlorophyll maximum.

Dep = Depth of the water column, Chl = surface chlorophyll concentration, Aut = autumn, Spr = spring, Sum = summer

EAB = eastern Agulhas Bank, WAB = western Agulhas Bank, WCO = West Coast of South Africa and SST = sea surface temperature.

Table 3.4: Summary of the statistical correlation between observed and predicted parameter values for the test data set of November 1998 (n = 88). GLM equations were used to predict each parameter. The r^2 is the variance explained and p-values are significant at 5% level

Parameter	Mean	Std dev.	r^2 (%)	p-value
B_0				
Observed	0.5	0.3		
Predicted	0.2	0.1	1.9	0.2051
σ				
Observed	13.2	6.2		
Predicted	12.5	2.8	15.9	0.0017
h				
Observed	74.4	51.3		
Predicted	59.6	43.0	43.5	0.0000
z_m				
Observed	18.7	51.3		
Predicted	18.6	43.0	61.0	0.0000

Predicted background chlorophyll concentration (B_0)

The relationship of parameter values from the shifted Gaussian curve with those predicted using the GLM for B_0 is presented in Figure 3.31a. Too much scatter was evident from the graph with no clear trend between the observed and predicted B_0 values. Low variance was explained by the model (r^2~2%), indicating a weak relationship and a possible poor prediction of B_0 from the depth of the water column

and surface chlorophyll concentration. The mean and standard deviation of predicted values were less than half of observed values (Table 3.4), suggesting an under-estimation of modelled B_0 by the generalized linear model.

Predicted width of the peak (σ)

Prediction of this parameter (σ) from surface chlorophyll, water column depth, season and area was fairly poor. The GLM explained less variance (r^2~16%) between observed (from the Gaussian curve) and predicted values from the GLM (Table 3.4). Many data points were above and below the 95% confidence interval (dotted line), showing a weak relationship between observed and predicted values. No clear relationship could be observed from the graph (Fig. 3.31b). Low variability in predicted parameter values compared to observed values is shown by the standard deviation in Table 3.4. The linear model resulted in under-estimations of predicted parameter values from environmental variables.

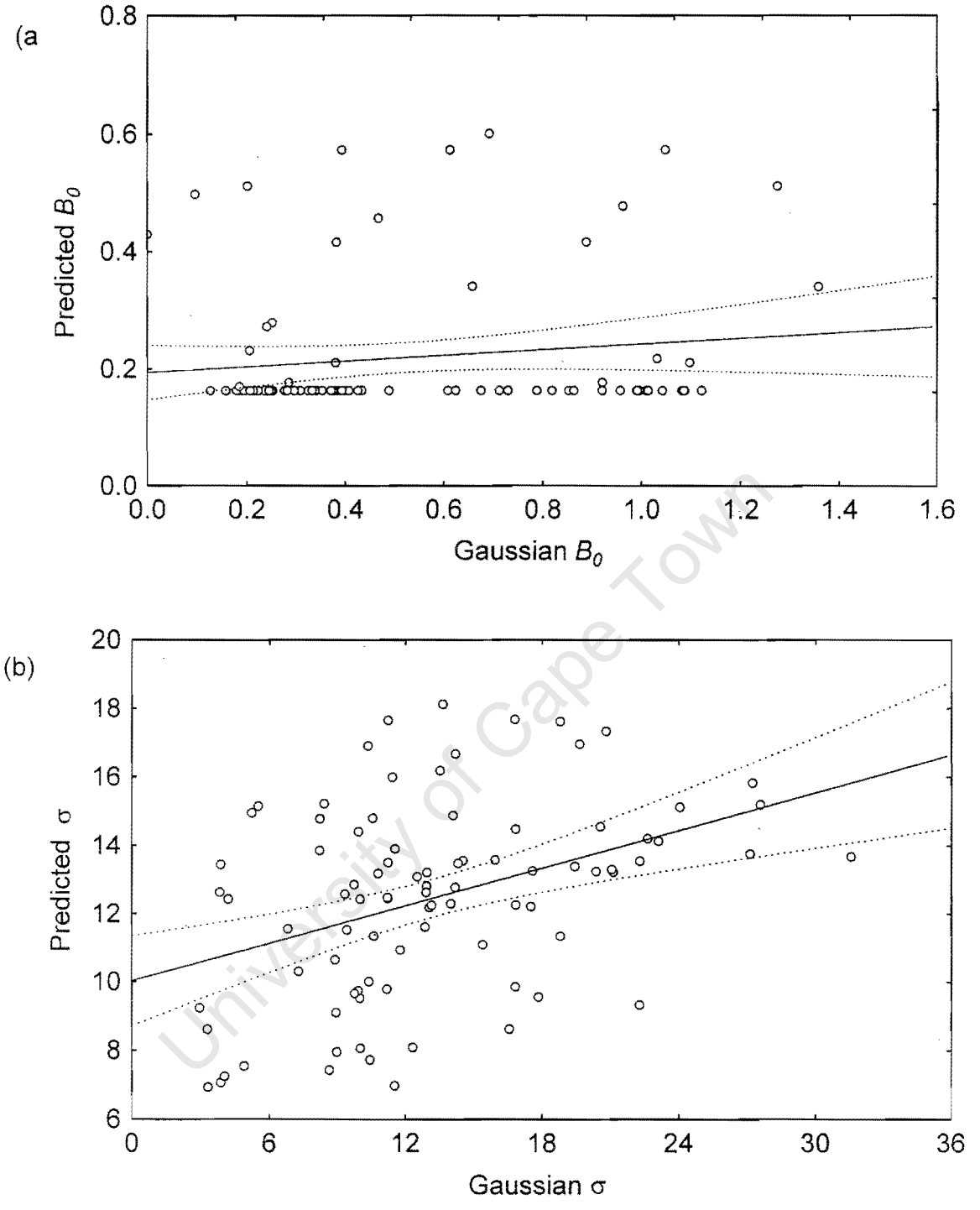


Figure 3.31. A relationship between a Gaussian-derived and predicted: (a) background chlorophyll concentration (B_0), and (b) width of the peak (σ) from a generalized linear modelling. 95% confidence intervals are included (dotted lines).

Predicted total chlorophyll within the peak (h)

Predicting h from surface chlorophyll, SST, water column depth and season using the GLM gave better r^2 value (~43%) as indicated in Table 3.4, and confidence in predicting this parameter from these environmental variables. A significant relationship between observed parameter values (from the shifted Gaussian curve) and predicted values (from the GLM) is shown in Figure 3.32a, with less scatter in data points compared to B_0 and σ. The relationship presented a high correlation coefficient of 0.66, suggesting a good prediction of this parameter from significant environmental variables used in the GLM.

Predicted depth of the peak (z_m)

A good prediction of the depth of the chlorophyll peak (z_m) from surface chlorophyll, SST, depth of the water column, season and area was shown (Fig. 3.32b). The model explained high total variance (r^2~61%), indicating a high correlation between observed and predicted z_m values as shown in Table 3.4. An increase in observed values corresponded to increasing predicted values, highlighting a significant linear relationship (Fig. 3.32b). The mean and standard deviation of the observed and predicted values were similar (Table 3.4), and this indicated a lack of under estimation of predicted parameter values

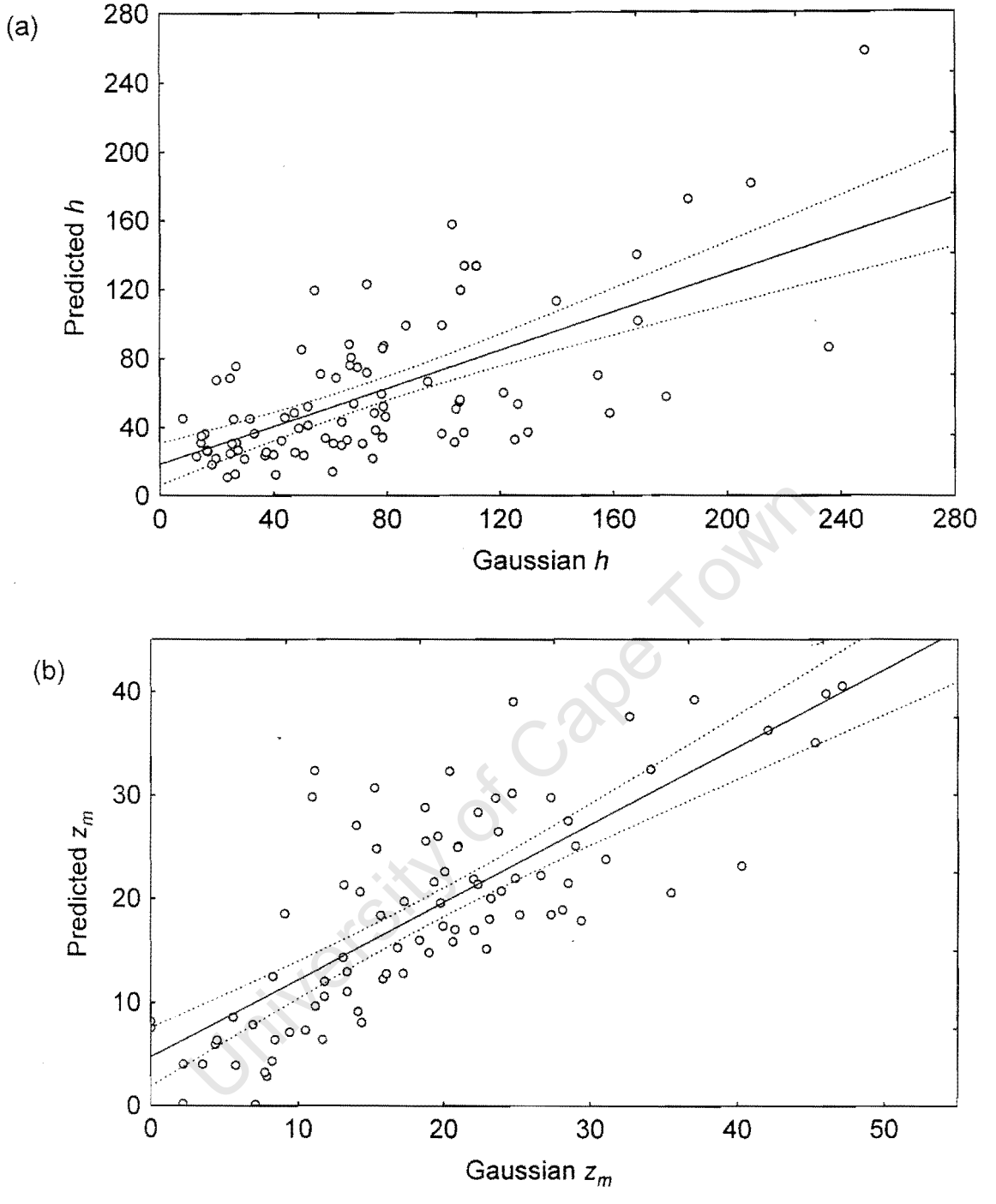


Figure 3.32. A relationship between a Gaussian-derived and predicted: (a) total chlorophyll concentration within the peak (h), and (b) depth of the chlorophyll maximum (z_m) from a generalized linear modelling. 95% confidence intervals are included as dotted lines.

Chapter 4: Discussion

4.1 The self-organizing map (SOM) analysis

The SOM technique identified a continuum of chlorophyll patterns, which represent the chlorophyll profiles used as input in the SOM analysis for the current study. The analysis also highlighted different chlorophyll patterns with considerable variability in profile parameters and therefore of chlorophyll distribution at seasonal and sub-provincial scales. The approach of fitting parameters was preferred over the one that uses raw profiles interpolated to standard depths and adjusted for the water column depth, because it is a typical approach that has been used globally in other studies (Platt 1988, Longhurst *et al.* 1995, Sathyendranath *et al.* 1995) as a first step in estimating primary production from satellites. For computation of primary production in the Benguela upwelling system, a productivity algorithm is needed to retrieve the phytoplankton photosynthetic rate, and thus primary production. Profile parameters have been used as inputs in analytical models for deriving the algorithm needed to estimate primary production (Platt 1988, Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). Although the approach adopted in this study is subjective, it strongly constitutes a semi-quantitative way of identifying changes in chlorophyll profile shapes (based on profile parameters) at different environmental levels, which would be impossible with raw profiles. This was achieved easily by using the four parameter values (B_0, σ, h and z_m) given by the SOM technique for each identified output pattern, which where then used in further analyses. Furthermore, profile parameters used in this study could be useful for conducting research in other studies that would

require profile shapes. The use of raw profiles, interpolated to standards depths, in the SOM analysis to look at the variability in profile shapes and phytoplankton distribution is currently underway as part of an on-going project that aims to improve primary production of the Benguela upwelling system".

Self-organizing maps have only recently been applied to chlorophyll profiles for pattern recognition, and have been used as semi-quantitative tools with considerable potential in classifying chlorophyll profiles into characteristic shapes based on environmental information of the Benguela upwelling system (Silulwane *et al.* 2001, Richardson *et al.* in press). The same study is carried out for the sub-provinces of the Benguela Current and Agulhas Bank, but using a bigger data set of chlorophyll profiles. To my knowledge, the SOM technique has not been used in other biogeochemical provinces except for the Benguela upwelling system, and its potential in highlighting variability in chlorophyll patterns is one of its advantages, and it looks promising.

This technique addresses the same problems of organizing data in order to investigate pattern groupings, which is possible with other statistical tools such as cluster analysis, multidimensional scaling and principal components analysis. An advantage of a SOM analysis over these techniques is that it does not need statistical assumptions about the data set. In addition, the output from the SOM can be easily visualized in two-dimensional space as the input, which are chlorophyll profiles in this study, and easily interpreted. Furthermore, large complex data sets can be easily handled with the SOM analysis, and this is very difficult if not impossible when using other statistical techniques mentioned earlier (Openshaw 1994). The size of the output

map is chosen by the user depending on the output details required and can include whatever is being considered as important, making this technique a flexible tool.

The SOM technique is by no means perfect, and this means there are disadvantages when using this technique. A single disadvantage of particular importance is the lack of statistical information such as significance level, and variance explained by the model (Openshaw 1992). These disadvantages hold true for cluster analysis and multidimensional scaling, which are primarily used for pattern groupings with no necessary significance tests (Richardson *et al.* in press).

However, the present approach of the SOM constitutes an objective way of conducting a semi-quantitative analysis of the phytoplankton pigment structure in the southern Benguela sub-province. It has been shown that the results obtained with this approach are meaningful and consistent with previous studies that looked at changes in phytoplankton pigment distribution with environmental variables (Mitchell-Innes *et al.* 1999, Mitchell-Innes *et al.* 2001), particularly those studies that used the same analysis (Silulwane *et al.* 2001, Richardson *et al.* in press).

It should be noted that some profiles were excluded and only 347 profiles (from a total of 592) were used in the SOM analysis. Profiles used as input in the SOM analysis were from three seasons, but mainly representing spring and summer, with very few (33) from the autumn period, and with no profiles from winter. In addition, the number of chlorophyll profiles from each sub-region was substantially different. The inclusion of all available profiles in the SOM analysis would result in a higher variability in vertical chlorophyll distribution compared to the one highlighted in the

present SOM analysis. Moreover, the use of a large data set of vertical chlorophyll profiles from all seasons (with reasonable number of profiles from each season and sub-province) would characterize a higher seasonal variability in phytoplankton pigment structure within each sub-province that is not represented in the present study. The large data set would also highlight the variation in inter-annual changes of phytoplankton biomass, which would be useful in other studies that consider phytoplankton biomass as an index of primary production.

Previous studies on profile parameterization (Sathyendranath *et al.* 1995) considered some other useful parameters that can be derived from the four parameters of the Gaussian curve, thereby reducing the four parameters to three for primary production estimates. In this study, it was decided that the four initial parameters from the Gaussian curve should be kept for simplicity and be used in the SOM technique. The SOM technique easily handles multi-dimensionality, which is reduced to two-dimensional space as seen in this study. For consistency, the four parameters applied in the SOM algorithm were also used in other subsequent analyses.

This study forms part of the incomplete primary project that tries to estimate primary production from satellite derived chlorophyll data in the Benguela upwelling system. Thus, once a large data set of chlorophyll profiles and satellite-derived surface chlorophyll and SST has been gathered, use would be made of other parameters derived from other Gaussian parameters for primary production estimates, as has been in other studies (Longhurst *et al.* 1995, Sathyendranath *et al.* 1995). For future studies in the Benguela, the reduction in profile parameters will be possible in modelling with GAMs and GLMs, but not with the SOM technique.

4.2 Mean seasonal chlorophyll profiles

Mean chlorophyll profiles give a misleading representation of the characteristic shape of profiles in the current sub-provinces. For example, in summer, the mean profile (Fig. 3.1), has a large, broad subsurface peak (z_m~18 m), and characteristic profiles for summer as derived from the SOM analysis have large surface and near-surface peaks and are narrower compared with the mean summer profile (Fig. 3.15). The average spring profile had a relatively broad subsurface chlorophyll maximum compared to the summer and autumn mean profiles (Fig. 3.1). The SOM technique identified patterns with more pronounced, moderate near-surface maximum as characteristic profiles for the spring period, but also profiles with surface and subsurface chlorophyll maximum occurred (Fig. 3.15). This showed that mean chlorophyll profiles are inappropriate to define typical chlorophyll profile shapes for different seasons because the parameter values are averaged, and therefore some of the information is not expressed in the averaged profiles.

4.3 Variability in phytoplankton pigment structure of the Benguela and Agulhas Bank continental shelf sub-provinces

4.3.1 Spatial variability in phytoplankton distribution

Dynamics of phytoplankton growth differ strongly between provinces of the ocean (Sathyendranath *et al.* 1995). A shift in chlorophyll maximum (z_m) occurs from one trophic state to the other (Morel and Berthon 1989). The width of the chlorophyll peak (σ) also varies in different trophic states, with mesotrophic coastal waters characterized by narrow peaks (σ~8 m) (Hoepffner *et al.* 1999). This variation in

profile parameters also occurs at fine scales in the West Coast and Agulhas Bank sub-provinces. The depth of the chlorophyll maximum and the thickness of the peak vary from one sub-province to the other in the southern Benguela region. The West Coast sub-province has mainly profiles with narrow, surface chlorophyll peaks compared to the Agulhas Bank, where profiles with broader subsurface chlorophyll peaks occur.

Apart from physical and biological processes that influence phytoplankton pigment structure, the coastal morphology, such as the width of the continental shelf, also affects the spatial phytoplankton pigment structure (Hoepffner *et al.* 1999, Peliz and Fiúza 1999). The broad continental shelf of the West Coast of South Africa, with narrow belt of coastal upwelling water in summer (Mitchell-Innes *et al.* 2000), is characterized by high surface (>15 $mg.m^{-3}$) and moderate to high near-surface chlorophyll maxima (4-12 $mg.m^{-3}$), because of nutrient-rich upwelled water and increased light, which favour phytoplankton growth. In areas where the shelf narrows and steepens with little upwelling, particularly in the eastern part of the Agulhas Bank, low chlorophyll maxima occur subsurface.

The southern-most part of the Benguela upwelling system, the western Agulhas Bank, has a narrow shelf characterized by high chlorophyll concentrations. The SOM approach has revealed that the western Agulhas Bank has a mixture of profiles with surface and near-surface, as well as sub-surface chlorophyll maxima with moderate to high surface chlorophyll concentrations. This leads to a greater variation in phytoplankton biomass distribution on the Agulhas Bank compared to the one highlighted in the West Coast sub-province. Recent studies on chlorophyll distribution on the western Agulhas Bank using the SOM approach have also revealed

this greater variability in pigment structure (Richardson *et al.* in press). Previous studies on plankton biomass have shown that the Agulhas Bank is in fact a highly productive sub-province in comparison to other east coast shelf regions (Hutchings 1994).

The eastern sub-region of the Agulhas Bank sub-province has warm water of Agulhas current origin; therefore very low pigment concentrations (<1 $mg.m^{-3}$) prevail in this region. The phytoplankton pigment distribution of this eastern part is mainly subsurface (Carter *et al.* 1986) and represents that of the oligotrophic state with warm slope waters. These warm, slope waters have very low chlorophyll concentrations that are separated from the enhanced inshore surface pigment concentrations of the cool shelf waters by a strong frontal structure (Peliz and Fiúza 1999).

The well developed subsurface maximum on the Agulhas Bank is associated with strong thermal stratification in the shelf waters (Carter *et al.* 1986). Although this subsurface maximum is well maintained (Carter *et al.* 1987) especially on the eastern side of the Agulhas Bank as highlighted in this study, it is believed that variation in this subsurface phytoplankton structure exists. Deep chlorophyll maximum occurs as a result of less light penetrated to deeper depths due to self-shading, which is caused by the phytoplankton bloom development at the surface layers. Phytoplankton blooms in the euphotic layer (Morel and Berthon 1989) limit light penetration at depth, and therefore low subsurface chlorophyll maximum develops (Brown and Hutchings 1987, Carter *et al.* 1987, Pitcher *et al.* 1992).

Considerable variability in pigment biomass is found in areas more frequently affected by upwelling events (Peliz and Fiúza 1999) particularly the West Coast sub-province of South Africa. In this sub-province, high concentrations of chlorophyll *a* (>10 $mg.m^{-3}$) were frequent, particularly near- and at the sea surface. In addition to spatial changes in phytoplankton biomass in an upwelling region, species composition of phytoplankton are quite variable spatially and temporally (Cullen 1982, Mitchell-Innes *et al.* 2000) because dynamics of phytoplankton growth differ strongly (Sathyendranath *et al.* 1995) from inshore to offshore between sub-regions. Inshore, diatoms tend to dominate in cool (<13°C) upwelled water with high nutrient concentrations causing surface chlorophyll peaks, while dinoflagellates predominate during quiescent periods with moderate water temperatures, and as water warms (>16°C) micro-flagellates become common (Pitcher *et al.* 1998). Mobile dinoflagellates develop strategies and behavioural patterns that enable them to position themselves within deep layers with light and rich nutrients when surface upwelled nutrients have been depleted (Pitcher *et al.* 1998). Their motility enables them to dominate farther offshore, resulting in a subsurface chlorophyll maximum, particularly in late summer and autumn in the midshelf region (Mitchell-Innes *et al.* 2000).

4.3.2 Seasonal variability in phytoplankton distribution

Seasonality of profile parameters within the Agulhas Bank and West Coast sub-provinces was investigated. The depth and the thickness of the chlorophyll maximum (z_m and σ) differ from one season to the other. The depth of the chlorophyll maximum becomes progressively deeper, and the width of the peak becomes broader during

spring compared with summer and autumn. The progressively narrow chlorophyll peaks for summer and autumn for the sub-province of the Benguela Current province are similar to those for the East Atlantic Canary Current province during the same seasons. Furthermore, this similarity in width of the chlorophyll peak may be explained by the common seasonal and perennial upwelling in some subregions of the Canary Current (Sathyendranath *et al.* 1995) and the Benguela Current Coastal province (Shannon 1985, Shannon and Nelson 1996).

During wind-forcing events, the inshore region experiences large temporal variability in thermal and biological characteristics within short (hours) time scales (Pitcher *et al.* 1998, Mitchell-Innes *et al.* 2000). Previous studies looking at the subsequent path of newly-upwelled water in the Benguela region showed temporal changes in vertical phytoplankton pigment distribution that occurred within a few days (Brown and Hutchings 1987, Pitcher *et al.* 1996). Well-mixed water columns with low chlorophyll concentrations exist during the onset of upwelling. Surface chlorophyll maxima are then established as the water column becomes stratified, and within 6-8 days these surface maxima sink becoming subsurface chlorophyll peaks (Brown and Hutchings 1987, Pitcher *et al.* 1996). Such rapid changes in vertical chlorophyll distribution within finer-scales in an upwelling region can lead to seasonal variability in phytoplankton pigment structure as indicated by the SOM analysis for both Agulhas Bank and West Coast sub-provinces. In view of the seasonal variability in profile shapes identified by the SOM, it would be inappropriate to consider the pattern that predominates in each season as a characteristic profile for the sub-provinces, as has been the approach in other previous studies (see Kameda and Matsumura 1998, Longhurst *et al.* 1995, Sathyendranath *et al.* 1995).

In spring, a greater diversity of chlorophyll patterns occurs, with a mixture of profiles from near-surface and subsurface chlorophyll maxima, and those that are almost uniform throughout the water column. These different profile shapes indicate a high variability in phytoplankton biomass distribution over the Agulhas Bank and West Coast sub-provinces. A high variability in phytoplankton pigment structure was also shown in a recent study on the western Agulhas Bank in the same season (Richardson *et al.* submitted). This is because spring is a transition period with characteristic profiles representing both winter and summer phytoplankton pigment structures. This is shown, for example, by dominant near-surface chlorophyll peaks typical of summer profiles, and those that are almost uniform throughout the water column, which are representatives of winter profiles. Furthermore, deep subsurface peaks were more common in spring than in summer and autumn, suggesting the existence of deep mixed layers and low light levels, which could not allow rapid development of surface blooms (Richardson *et al.* 1998).

In summer, highly variable and enhanced concentrations of phytoplankton in the coastal zone (Brown and Hutchings 1987) are a result of changes in nutrient status of upwelled water (Brown 1984), which covers the broad continental shelf of the West Coast of South African sub-province (Probyn *et al.* 2000). In the current study, maximum chlorophyll concentrations are evident in summer in the upwelling inshore zone. The existence of maximum productivity that coincides with high mean values of satellite-derived surface chlorophyll concentration is a spring feature of the coastal region of the north-east Atlantic ocean (Hoepffner *et al.* 1999). The Benguela and Agulhas Bank sub-provinces have highest chlorophyll concentrations at intermediate SST (14-18°C) in summer and autumn in aged upwelled waters. In addition, the

lowest phytoplankton pigment concentrations are evident in spring in the present study. Lowest integrated production at the onset of upwelling in summer in the Benguela shelf region results from strong winds, which induce upwelling associated with decreased surface productivity because of higher turbulence, leading to deep mixing and homogenous water column. By contrast, calm days with decreased wind strength are associated with high primary production, as the upwelled water becomes stratified (Hoepffner *et al.* 1999).

4.3.3 Variability in vertical chlorophyll distribution with environmental variables

Characteristic chlorophyll profiles of the Benguela upwelling system include profiles with surface or near-surface peaks, subsurface peaks, and those with uniform chlorophyll distribution (typical of the winter period). Profiles with deep chlorophyll maxima were common in the Agulhas Bank sub-province, and these may have arisen from different causes (Cullen 1982). Usually, strong relationships between mixing, light, nutrients and vertical pigment distribution exist throughout the water column. Profiles with near-surface and surface chlorophyll maximum exist when high nutrients and increased solar irradiance exist at surface layers with intensified stratification, thereby leading to increased primary productivity in these layers (Kameda and Matsumura 1998). Patterns with subsurface peaks appear when the near-surface nutrients are depleted and therefore phytoplankton sink to deeper layers causing subsurface maximum primary production.

Strong relationships between profile shapes, in terms of profile parameters, and environmental variables have been identified. It was found that surface chlorophyll

peaks dominate at cool temperatures and high surface chlorophyll concentrations with warm waters characterized by low surface chlorophyll concentrations and subsurface chlorophyll maxima. The same trends of profile shapes with SST and surface chlorophyll concentration were found in a recent study on chlorophyll profiles of the Benguela upwelling system (Silulwane *et al.* 2001).

An inverse relationship of surface chlorophyll concentration, from the Gaussian curve, and depth of the chlorophyll maximum (z_m) was evident. The same relationship was also obtained in the northwestern Pacific, but using remotely sensed surface chlorophyll concentration (Kameda and Matsumura 1998).

4.4 Modelled profile parameters using generalized modelling

The generalized models (GAM and GLM) can model profile parameters as functions of combined environmental variables. The most significant modelled parameters are the depth of the chlorophyll maximum (z_m) with models explaining 67-70% of the variance, and the total chlorophyll within the peak (h) with the variance range of 74-80%. The depth of the chlorophyll maximum is very important in computations of primary production, where increased production estimates resulted in the inclusion of the vertical chlorophyll structure (Sathyendranath *et al.* 1995). Moreover, the vertical pigment structure in the euphotic zone, which is mostly indicated by the chlorophyll maximum at depth, is required to estimate primary production in the water column from satellites (Morel and Berthon 1989, Sathyendranath and Platt 1995). The total chlorophyll concentration within the peak (h) is also important and it is equivalent to the integrated chlorophyll concentration within certain depths of the water column.

An advantage of generalized modelling is that the effect of many environmental variables on each profile parameter could be modelled, and this is impossible to model using simple (linear) regression methods. Previous studies of profile parameters with surface chlorophyll (the background chlorophyll concentration), used statistical correlation, where a single profile parameter was correlated to other Gaussian profile parameters (Sathyendranath *et al.* 1995, Longhurst *et al.* 1995). GAMs model the effects of many variables with each parameter, but each variable is adjusted in terms of others, therefore each response is modelled as an adjusted function of these environmental variables.

The generalized modelling approach conducted in this study can also be expanded to other biogeochemical provinces of the primary domains, where profile parameters can be modelled and predicted from environmental information. The relationships obtained between profile parameters and environmental variables of the continental shelf region for the current sub-provinces might be applicable to the Benguela province. Changes in offshore chlorophyll distribution can also be incorporated to study extensively the phytoplankton pigment structure and its prediction, and primary production in future for this sub-province and other provinces. For other provinces, the relationships might differ because of the hydrography and biology of such provinces. It is important to note that the current data set used needs to be expanded to include more profiles particularly winter profiles, which have not been used in this study. Although the models could be improved by including other environmental variables from the research cruises, such as euphotic depth, upper mixed layer depth, and nutrient concentrations, they cannot be directly measured synoptically from satellites and thus were not included in the present study. Once larger data sets are

used, not only changes in finer scales (within season and sub-province variability) as considered in the current study, but also the inter-annual variation in primary production in large spatial scales (biogeochemical provinces and primary domains) can be determined. Moreover, satellite-derived surface chlorophyll and sea surface temperature can be sampled synoptically in short time periods over large spatial scales, therefore a larger data set can be available to conduct primary production studies using satellite data in the Benguela.

4.5 Predicted profile parameters using generalized linear modelling

Profile shapes could be predicted, in terms of profile parameters, with good predictions for the depth of the peak (z_m) and total chlorophyll concentration within the peak (h). These parameters (z_m and h) could be well predicted because their generalized models explain more variance than those for the background chlorophyll concentration (B_0) and the width of the peak (σ). The suggestion of background surface chlorophyll (B_0) as a predictor of the chlorophyll profile structure by Morel and Berthon (1989) was proven to be insufficient for such purposes. The relationship between (B_0) and the depth of the chlorophyll maximum (z_m) was weak for all provinces and seasons of the four primary domains, including the Benguela Current Coastal province. To improve GLM equations for B_0 and hence predictability of this parameter, it would be useful to include the depth of the peak (z_m) concurrently with environmental variables as predictors in modelling. The prediction of the width of the peak from (B_0) was also impossible (Sathyendranath *et al.* 1995). For this study, the depth of the peak is well predicted but using a set of significant environmental variables. A large degree of uncertainty exists for the prediction of the width of the

peak, which is likely to be impossible to predict from environmental variables used in the current modelling.

It should be noted that the independent dataset used for predictions was from a single cruise with profiles mainly for spring within sub-regions of the Agulhas bank, and from the West Coast sub-province, therefore poor predictions for B_0 and σ could be the result of these data restrictions. Meanwhile, predictions for z_m and h are considered as good predictions because this is the first study (that is known) that has taken this approach of predicting profile parameters using generalized modelling. In addition, the approach of predicting parameters from surface chlorophyll in the context of biogeochemical provinces and estimation of primary production has not introduce the techniques used in this study, therefore no comparisons with other studies can be done at this stage.

Although the current models seem to under-estimate the variability in profile parameters, they are considered as good predictive models, particularly for the depth of the peak (z_m) and total chlorophyll concentration within the peak (h) with more variance explained in these models. This under estimation might be due to biasing in our sampling procedure, where other cruises were conducted in spring (Spawner Biomass surveys) with most profiles from the Agulhas Bank sub-province. To improve these models in future studies, cruise data should be from the entire Benguela province and Agulhas Bank sub-province, with almost equal profiles from all seasons.

4.6 Conclusion and future recommendations

The basic assumption of slow varying changes in water column phytoplankton biomass, and therefore of a typical shape of the biomass profile for each season for the Benguela Coastal Current province (Platt and Sathyendranath 1988) has been proven to be inapplicable in the Benguela Current sub-provinces. The semi-quantitative approach of the SOM technique has highlighted the variability in phytoplankton pigment structure in the Benguela and Agulhas Bank sub-provinces, which suggests that seasonal profiles for these sub-provinces are not applicable. This study focuses on fine spatial and temporal scales, which are different sub-regions and seasons of the Agulhas bank, and seasons of the West Coast sub-province. It is believed that variability in chlorophyll profiles as indicated by the SOM analysis will not only be highlighted on smaller scales, but also on larger scales (annually in the Benguela Current Coastal) because of fast changes in hydrography, which lead to changes in biology of this dynamic province.

Generalized additive and generalized linear modelling used to model profile parameters with environmental variables have highlighted the combined effect of these environmental variables on profile parameters, as this was impossible using standard (simple) regression methods. Moreover, the GLM showed meaningful predictions of profile parameters from significant environmental variables, particularly for the depth of the peak (z_m) and total chlorophyll concentration within the peak (h). To my knowledge, this was the first study that applied the GLMs (statistical models) for predicting profile shapes in the Benguela sub-province, as well as in other biogeochemical provinces and primary domains for primary production purposes. The framework of this study can be applied to the entire Benguela Current

Coastal province and other biogeochemical provinces, and can also include the use of remotely sensed SST and surface chlorophyll concentrations in predicting the water column pigment structure.

Ecological limitations of satellite data to only the near-surface information make estimates of primary production to be impossible without the parameters from vertical chlorophyll profiles. For computations of primary production in the Benguela upwelling system, a productivity algorithm is needed to retrieve the photosynthetic rate of phytoplankton from satellite images, especially from images that capture distinct upwelling events in this system. The computation of production estimates using satellite information can be extended to the entire Benguela Current Coastal province once variability in profile shapes and hence of phytoplankton pigment structure in all seasonal scales have been considered.

The methodological approach of using quantitative techniques to quantify phytoplankton pigment distribution can also be implemented in other biogeochemical provinces and primary domains. The SOM technique can be used for recognizing chlorophyll patterns, highlight any variability in chlorophyll profiles, and show characteristic profiles in larger scales. In addition, the phytoplankton pigment structure of each province and domain may be predicted using generalized modelling. A high degree of confidence in prediction of profile parameters should be expected for tropical biogeochemical provinces, where less variable physical, chemical and biological processes prevail. For a dynamic upwelling province, such as the Benguela Current Coastal where changes in physical and biological structure occur within few days, predictions should be done and analysed with caution.

REFERENCES

AMSTRONG, M. J., JAMES, A. G. and E. S. VALDÉS SZEIN-FELD 1991 - Estimates of annual consumption of food by anchovy and other pelagic fish species off South Africa during the period 1984-1988. *S. Afr. J. mar. Sci.* **11**: 251-266.

ANDRÈ, J. M. - 1992. Ocean color remote-sensing and the subsurface vertical structure of phytoplankton pigments. *Deep-Sea Res.* **39** (5): 763-779.

AOKI, I. And T. KOMATSU 1997 - Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanolog. Acta.* **20** (1): 81-88.

BELGRANO, A., MALMGREN, B. A. and O. LINDAHL 2001 – Application of artificial neural networks (ANN) to primary production time-series data. *J. Plankton Res.* **23** (6): 651-658.

BROWN, P. C. 1984 - Primary production at two contrasting near-shore sites in the southern Benguela upwelling region, 1977-1979. *S. Afr. J. mar. Sci.* **2**: 205-215.

BROWN, P. C. 1992 - Spatial and seasonal variation in chlorophyll distribution in the upper 30 m of the photic zone in the southern Benguela/Agulhas ecosystem. In *The Benguela Trophic Functioning*. Payne. A.I.L., Brink, K.H., Mann, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 515-525.

BROWN, P. C. and J. G., FIELD 1986 – Factors limiting phyto-plankton production in a nearshore upwelling area. *J. Plankton Res.* **8**: 55-68.

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.

BUCKTON, D., O'MONGAIN, E. and DANAHER, S. 2001 – The use of Neural Networks for the estimation of oceanic constituents based on the MERIS instrument. *Int. J. Rem. Sens.* **20** (9): 1841-1851.

CARTER, R. A., BARTLETT, P. D. and V. P. SWART 1986 – Estimates of the nitrogen flux required for the maintainance of subsurface chlorophyll maxima on the Agulhas Bank. *In Marine Interfaces Ecohydrodynamics*. Nihoul, J. C. J. (Ed.). Amsterdam; *Elsevier Oceanogr. Series* **42**: 331-339.

CARTER, R. A., McMURRAY, H. F. and J. L. LARGIER 1987 – Thermocline characteristics and phytoplankton dynamics in Agulhas Bank waters. In *The Benguela and Comparable Ecosystems*. Payne, A.I.L., Gulland, J.A. and K.H. Brink (Eds). *S. Afr. J. mar. Sci.* **5**: 327-336.

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.

CULLEN, J. J. and M. R. LEWIS 1995 – Biological processes and optical measurements near the sea surface: Some issues relevant to remote sensing. *J. Geophys. Res.* **100** (C7): 13, 255-13,266.

CURY P., ROY C. AND V. FAURE 1998 - Environmental constraints and pelagic fisheries in upwelling areas: The Peruvian puzzle. In *Benguela Dynamics. Impacts on Shelf-Sea Environments and Their Living Resources*. Pillar, S.C., Moloney, C. L., Payne, A.I L. and F.A. Shillington (Eds). S. *Afr. J. mar. Sci.* **19**: 159-167.

CUSHING, D. H. 1990 - Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. *Adv. Mar. Biol.* **26**: 249-293.

DAYHOFF, J. E. 1990 - *Neural network architectures: An introduction.* Van Nostrand Reinhold, New York: 259 pp.

DUGDALE, R. C., MOREL, A., BRICAUD, A. and F. P. WILKERSON 1989 – Modeling new production in upwelling centers: A case study of modeling new production from remotely sensed temperature and color. *J. Geophys. Res.* **94**: 18119-18132.

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.

FALKOWSKI, P. G. and A. D. WOODHEAD (Eds) 1992 – *Primary productivity and biogeochemical cycles in the sea.* Plenum Press, New York: 550 pp.

FALKOWSKI, P. G. 1994 – The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth. Res.*, **39**: 235-258.

FALKOWSKI, P. G. and Z. KOLBER 1995 – Variations in Chlorophyll Fluorescence Yields in Phytoplankton in the World Oceans. *Aust. J. Plant. Physiol.* **22**: 341-355.

FOODY, G. M. 1999 – The significance of border training patterns in classification by a feedforward neural network using a back propagation learning. *Int. J. Rem. Sens.* **20** (18): 3549-3562.

GORDON, H. R. 1993 – Radiative transfer in the atmosphere for correction of ocean color remote sensors. In Barale, V. and P. M. Schlittenhardt (Eds), *Ocean Colour: Theory and Applications in a Decade of CZCS Experience.* Kluwer Academic Publishers, Brussels, pp. 33-78.

GROSS, L., THIRIA, S., FROUIN, R. and B. C. MITCHELL 2000 - Artificial neural networks for modelling the transfer function between marine reflectance and phytoplankton pigment concentration. *J. Geophys. Res.* **105** (C2): 3483-3495.

HASTIE, T. and R. TIBSHIRANI 1990 – *Generalized Additive Models.* Chapman and Hall, London: 325 pp.

HEWITSON, B. C. and R. G. CRANE 1994 – *Neural Nets: Applications in Geography.* Kluwer Academic Publishers, London: 194 pp.

HEWITSON, B. C. and R. G. CRANE (in press) - Self Organizing Maps: Applications to Synoptic Climatology. *Climate Res.*

HILL, A. E, B. M. HICKEY, F. A. SHILLINGTON, P. T. STRUB, K. H. BRINK, E. D. BARTON AND A. C. THOMAS 1998 - Eastern Ocean Boundaries (E). In *The Sea*, Vol 11, *The Global Coastal Ocean; Regional studies and syntheses*, A. R. Robinson and K. H. Brink (Eds). Wiley, New York, Chap **2:** 29-67.

HOEPFFNER, N., STURM, B., FINENKO, Z. and D. LARKIN 1999 - Depth-integrated primary production in the eastern tropical and subtropical North Atlantic basin from ocean colour imagery. *Int. J. Rem. Sens.* **20** (7): 1435-1456.

HUGGETT, J. A., BOYD, A. J., HUTCHINGS, L. and A. D. KEMP 1998 - Weekly variability of clupeiod eggs and larvae in the Benguela jet current: implications for recruitment. In *Benguela Dynamics. Impacts on Shelf-Sea Environments and Their Living Resources*. Pillar, S.C., Moline, CL, Payne, A.I L. and F.A. Shillington (Eds). S. *Afr. J. mar. Sci.* **19**: 197-210.

HUNTER, J. R. and R. LEONG 1981 - The spawning energetics of female northern anchovy, *Engraulis mordax. Fishery Bull.* **79**: 215-230.

HUTCHINGS, L. 1992 - Fish harvesting in a variable, productive environment – searching for rules or searching for exceptions? In *Benguela Trophic Functioning. Payne*, A.I.L., Brink, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 297-318.

HUTCHINGS, L. 1994 – The Agulhas Bank: A synthesis of available information and a brief comparison with other east-coast shelf regions. *S. Afr. J. Sci.* **90** (3): 179-185.

JAMES, A. G. 1987 - Feeding ecology, diet and field-based studies on feeding selectivity of the Cape anchovy *Engraulis capensis* Gilchrist. In *The Benguela and Comparable Ecosystems*. Payne, A. I. L., Gulland, J. A. and K. H. Brink (Eds). *S. Afr. J. mar. Sci.* **5**: 673-692.

JARRE-TEICHMANN, A., BREY, T. and H. HALTOF 1995 - Exploring the use of neural networks for biomass forecasts in the Peruvian upwelling system. Naga. ICLARM Q. Rep. **18**: 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.

KEINER, L. E. and C. W. BROWN 1999 - Estimating oceanic chlorophyll concentrations with neural networks. *Int. J. Rem. Sens.* **20** (1): 189-194.

KOHONEN, T. 1989 – *Self-Organization and Associated Memory*, 3rd edition. Springer, Berlin.

KOHONEN, T. 1990 - The self-organizing map. *Proceedings of the IEEE.* **78** (9): 1464-1480.

KOHONEN, T., HYNNINEN, J., KANGAS, J. and J. LAAKSONEN 1995 - SOM_PAK: The self-organizing map program package. Version 3.1

KOHONEN, T. 1997 – *Self-Organizing Maps*. Springer, Berlin: 426 pp.

KURING, N., LEWIS, M. R., PLATT, T. and J. E. O'REILLY 1990 – Satellite-derived estimates of primary production on the northwest Atlantic continental shelf. *Cont. Shelf Res.* **10** (5): 461-484.

LARGIER, J. L., CHAPMAN, P., PETERSON, W. T. and V. P. SWART 1992 - The western Agulhas Bank: circulation, stratification and ecology. In *Benguela Trophic Functioning. Payne*, A.I.L., Brink, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 319-339.

LLUCH-BELDA, D., CRAWFORD, R. J. M., KAWASAKI, T., MacCALL, A. D., PARRISH, R. H., SCHWARTZLOSE, R. A. and P. E. SMITH 1989 - World-wide fluctuations of sardine and anchovy stocks: The regime problem. *S. Afr. J. mar. Sci.* **8**: 195-205.

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. Plankton Res.* **17** (6): 1245-1271.

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.

MAIN, J. P. L. 1997 – Seasonality of circulation in southern Africa using the Kohonen Self-Organising map. M.Sc. thesis, University of Cape Town: 84pp.

MANN, K. H. and J. R. N. LAZIER 1991 - *Dynamics of marine ecosystems: biological physical interactions in the oceans.* 1st ed.; Boston; Blackwell Scientific Publications: 466 pp.

MILLÁN- NÚÑEZ, R., ALVAREZ-BORREGO, S. and C. C. TREES 1997 - Modelling the vertical distribution of chlorophyll in the California Current system. *J. Geophys. Res.* **102** (C4): 8587-8595.

MITCHELL-INNES, B. A. and WALKER, D. R. 1991 – Short-term variability during an anchor station study in the southern Benguela upwelling system. Phytoplankton production and biomass in relation to species changes. *Prog. Oceanogr.* **28**: 65-89.

MITCHELL-INNES, B. A. and G. C. PITCHER 1992 - Hydrographic parameters as indicators of the suitability of phytoplankton populations as food for herbivorous copepods. In *Benguela Trophic Functioning.* Payne, A.I.L., Brink, K.H., Mann, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 355-365.

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., PITCHER, G. C. and PROBYN, T. A. 2000 - Productivity of dinoflagellate blooms on the West Coast of South Africa, as measured by natural fluorescence. *S. Afr. J. mar. Sci.* **22**: 273-284.

MITCHELL-INNES, B. A., SILULWANE, N. F. and M. A. LUCAS 2001 - Variability of chlorophyll profiles on the West Coast of southern Africa in June/July 1999. *S. Afr. J. mar. Sci.* **97**: 246–250.

MITTELSTAEDT, E. 1991 – The ocean boundary along the northwest African coast: circulation and oceanographic properties at the sea surface. *Prog. Oceanogr.* **26**: 307-355.

MOREL, A. and J. -A. 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**: 1545-1562.

MORISETTE, J. T., KHORRAM, S. and T. MACE 1999 – Land-cover change detection enhanced with generalized linear models. *Int. J. Rem. Sens.* **20** (14): 2703-2721.

NELSON, G. and L. HUTCHINGS 1983 – The Benguela upwelling area. *Prog. Oceanogr.* **12** (3): 333-356.

PARSONS, T. R., MAITA, Y. and C. M. LAILLI 1984 – *A Manual of Chemical and Biological Methods for Seawater Analysis.* Pergamon, New York.

PELIZ, A. J. and F. G. FIÚZA 1999 - Temporal and spatial variability of CZCS-derived phytoplankton pigment concentrations off the western Iberian Peninsula. *Int. J. Rem. Sens.* **20** (7): 1363-1403.

PETERSON, W. T., HUTCHINGS, L., HUGGETT, J. A. and J. L. LARGIER 1992 - Anchovy spawning in relation to the biomass and the replenishment rate of their copepod prey on the western Agulhas Bank. In: *Benguela Trophic Functioning.* Payne, A.I.L., Brink, K.H., Mann, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 487-500.

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., WALKER, D. R., MITCHELL-INNES, B. A. and C. L. MOLONEY 1991 – Short-term variability during an anchor station study in the southern Benguela upwelling system. Phytoplankton dynamics. *Prog. Oceanogr.* **28**: 39-64.

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. KORRUBEL 1996 - The use of sea temperature in characterizing the mesoscale heterogeneity of phytoplankton in an embayment of the southern Benguela upwelling system. *J. Plankton Res.* **18** (5): 643-657.

PITCHER, G. C., BOYD, A. J., HORSTMAN, D. A. and B. A. MITCHELL-INNES - 1998. Subsurface dinoflagellate populations, frontal blooms and the formation of red tide in the southern Benguela upwelling system. *Mar. Ecol. Prog. Ser.* **172**: 253-264.

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** (6): 855-879.

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): 15,147-15,159.

PLATT, T. and S. SATHYENDRANATH 1995 – *Software for Use in Calculation of Primary Production in the Oceanic Water Column.*

PLATT, T., SATHYENDRANATH, S. and A. R. LONGHURST 1995 – Remote sensing of primary production in the ocean: promise and fulfillment. *Phil. Trans. R. Soc. London Ser. A*, **348**: 191-202.

PROBYN, T. A. 1992 – The inorganic nitrogen nutrition of phytoplankton in the southern Benguela: new production, phytoplankton size and implications for pelagic food webs. In *Benguela Trophic Functioning*. Payne, A.I.L., Brink, K.H., Mann, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 411-420.

PROBYN, T. A., WALDRON, H. N. and A. G., JAMES. 1990 – Size-fractioned measurements of nitrogen uptake in aged upwelled waters: implications for pelagic food webs. *Limnol. Oceanogr.* **35**: 202-210.

PROBYN, T. A., PTCHER, G. C., MONTEIRO, P. M. S., BOYD, A. J. and G. NELSON 2000 - Physical processes contributing to harmful algal blooms in Saldanha Bay, South Africa. *S. Afr. J. mar. Sci.* **22**: 285-297.

RICHARDSON, A. J. 1998 - Variability of copepod abundance and growth in the southern Benguela upwelling system and implications for the spawning of the Cape anchovy. Ph.D. thesis, University of Cape Town: 206 pp.

RICHARDSON, A. J., MITCHELL-INNES B. A., FOWLER, J. L., BLOOMER, S. F., VERHEYE, H. M., FIELD, J. G., HUTCHINGS, L. and S. J. PAINTING 1998 - The effect of sea temperature and food availability on the spawning success of cape anchovy *Engraulis capensis* in the southern Benguela. In *Benguela Dynamics: Impacts on Shelf-Sea Environments and Their Living Resources*. Pillar, S.C., Moloney, C.L., Payne, A.I.L. and F.A. Shillington (Eds). *S. Afr. J. mar. Sci.* **19**: 275-290.

RICHARDSON, A.J., PFAFF, M.C., FIELD, J.G., SILULWANE, N.F. and F.A. SHILLINGTON (in press) – Identifying characteristic chlorophyll *a* profiles in the coastal domain using an artificial neural network. *J. Plankton Res.*

SATHYENDRANATH, S. and A. MOREL 1983 – Light emerging from the sea-interpretations and uses in remote sensing. In A. P. Cracknell (Ed.), *Remote Sensing Applications in Marine Science and Technology*. D. Reidel, Dordrecht, pp. 323-357.

SATHYENDRANATH, S. and T. PLATT 1989 - Remote sensing of ocean chlorophyll: consequence of nonuniform pigment profile. *Appl. Opt.* **28** (3): 490-495.

SATHYENDRANATH, S., PLATT, T., HORNE E. P. W., HARRISON, W. G., ULLOA, O., OUTERBRIDGE, R. and N. HOEPFFNER 1991 – Estimation of new production in the ocean by compound remote sensing. *Nature* **353**: 129-133.

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** (10): 1773-1802.

SHANNON, L. V. 1985 – The Benguela ecosystem, Part 1. Evolution of the Benguela, physical features and processes. *Oceanogr. Mar. Biol. A. Rev.* **23**: 105-182.

SHANNON, L. V. and G. NELSON 1996 – The Benguela: large-scale features and processes and system variability. In *The South Atlantic: Present and Past Circulation*. Wefer, G., Berger, W. H., Siedler, G. and D. J. Webb (Eds). Springer-Verlag, Berlin, pp. 163-210.

SHANNON, L. V., HUTCHINGS, L., BAILEY, G. W. and P. A. SHELTON 1984 - Spatial and temporal distribution of chlorophyll in the southern African waters as deduced from ship and satellite measurements and their implications for pelagic fisheries. *S. Afr. J. mar. Sci.* **2**: 109-130.

SHELTON, P. A. and L. HUTCHINGS 1982 - Transport of anchovy, *Engraulis capensis* Gilchrist, eggs and early larvae by a frontal jet current. *J. Cons. perm. int. Explor. Mer.* **40**: 185-198.

SHILLINGTON, F. A. 1998 – Benguela upwelling system off southwestern Africa, coastal segment (16,E). In *The Sea*, Vol 11, *The Global Coastal Ocean; Regional studies and syntheses*, A. R. Robinson and K. H. Brink (Eds). Wiley, New York, Chap. **20**: 583-604.

SILULWANE, N. F., RICHARDSON, A. J., SHILLINGTON, F. A. and B. A. MITCHELL-INNES (in press) - Identification and classification of vertical chlorophyll patterns in the Benguela upwelling system and Angola-Benguela front using an artificial neural networks. *In A Decade of Namibian Fisheries Science.* Payne, A.I.L., Pillar, S.C. and R.J.M. Crawford (Eds). *S. Afr. J. mar. Sci.* **23**: 37-51.

STORBECK, F. and B. DAAN 2001 – Fish species recognition using computer vision and a neural network. *Fish. Res.* **51**: 11–15.

SWART, V. P. and J. L. LARGIER 1987 – Thermal structure of Agulhas Bank waters. *S. Afr. J. mar. Sci.* **5**: 243-253.

VAN DER LINGEN, C. D. 1994 – Effect of particle size and concentration on the feeding behaviour of adult pilchard *Sardinops sagax*. *Mar. Ecol. Prog. Ser.* **109**: 1-13.

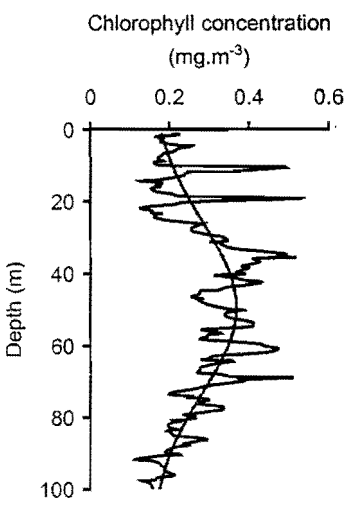
WALDRON, H. N. and T. A. PROBYN 1992 - Nitrate supply and potential new production in the Benguela upwelling system. In *Benguela Trophic Functioning*. Payne, A.I.L., Brink, K.H., Mann, K.H. and R. Hilborn (Eds). *S. Afr. J. mar. Sci.* **12**: 29-39.

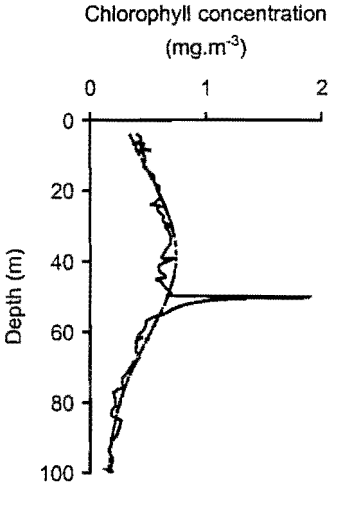
APPENDIX I

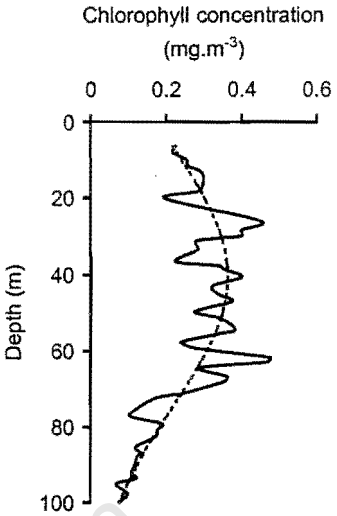
A subset of poorly fitted profiles (r^2<80%) from curve fitting using the Gaussian function.

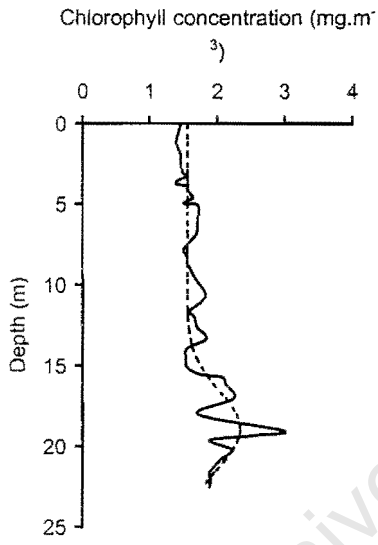
A subset of best fitted profiles (r^2≥80%) from curve fitting using the Gaussian function.

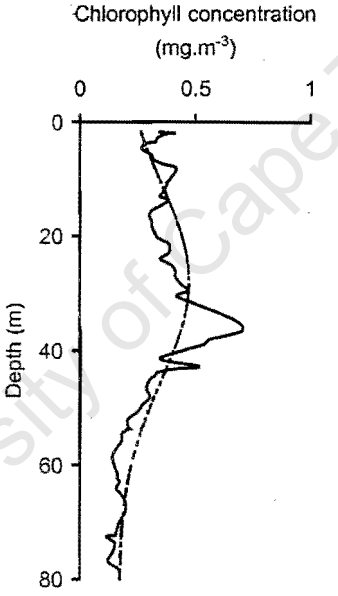
Poor-fitted profiles

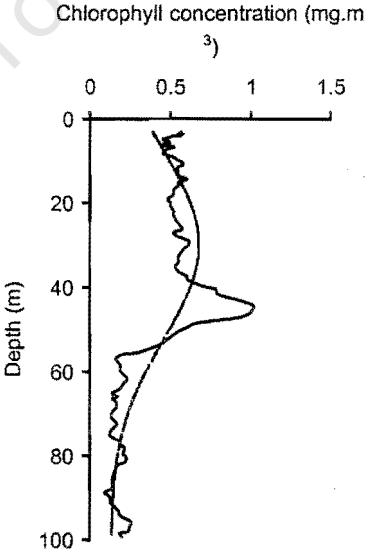


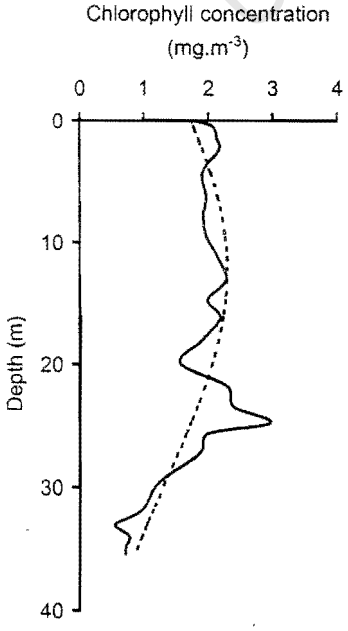


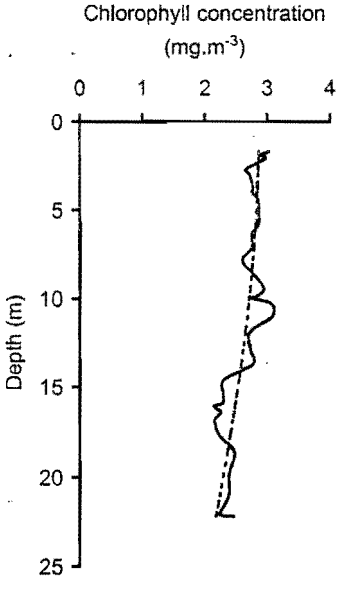


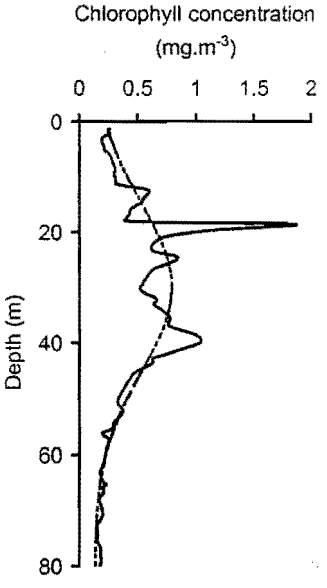




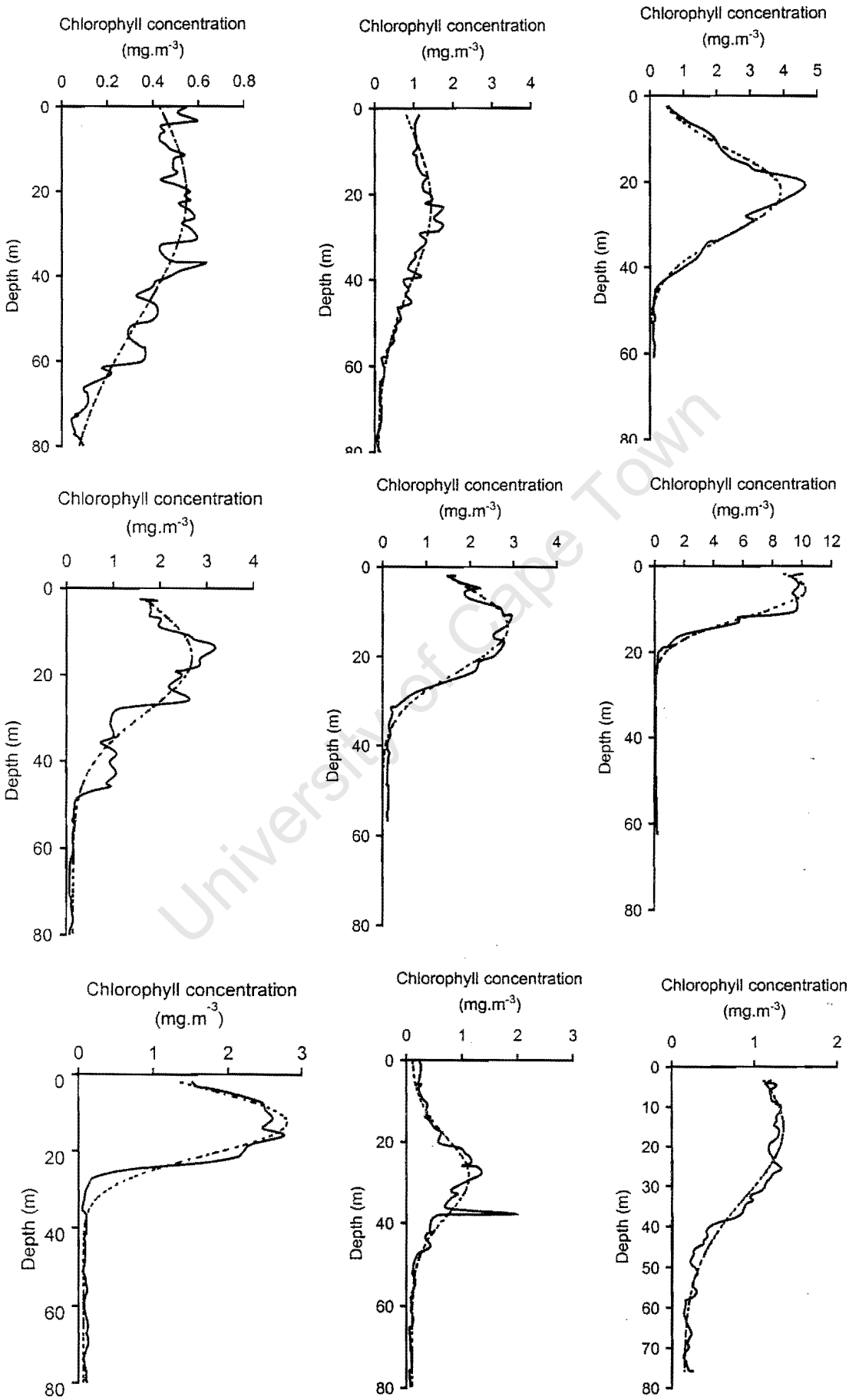


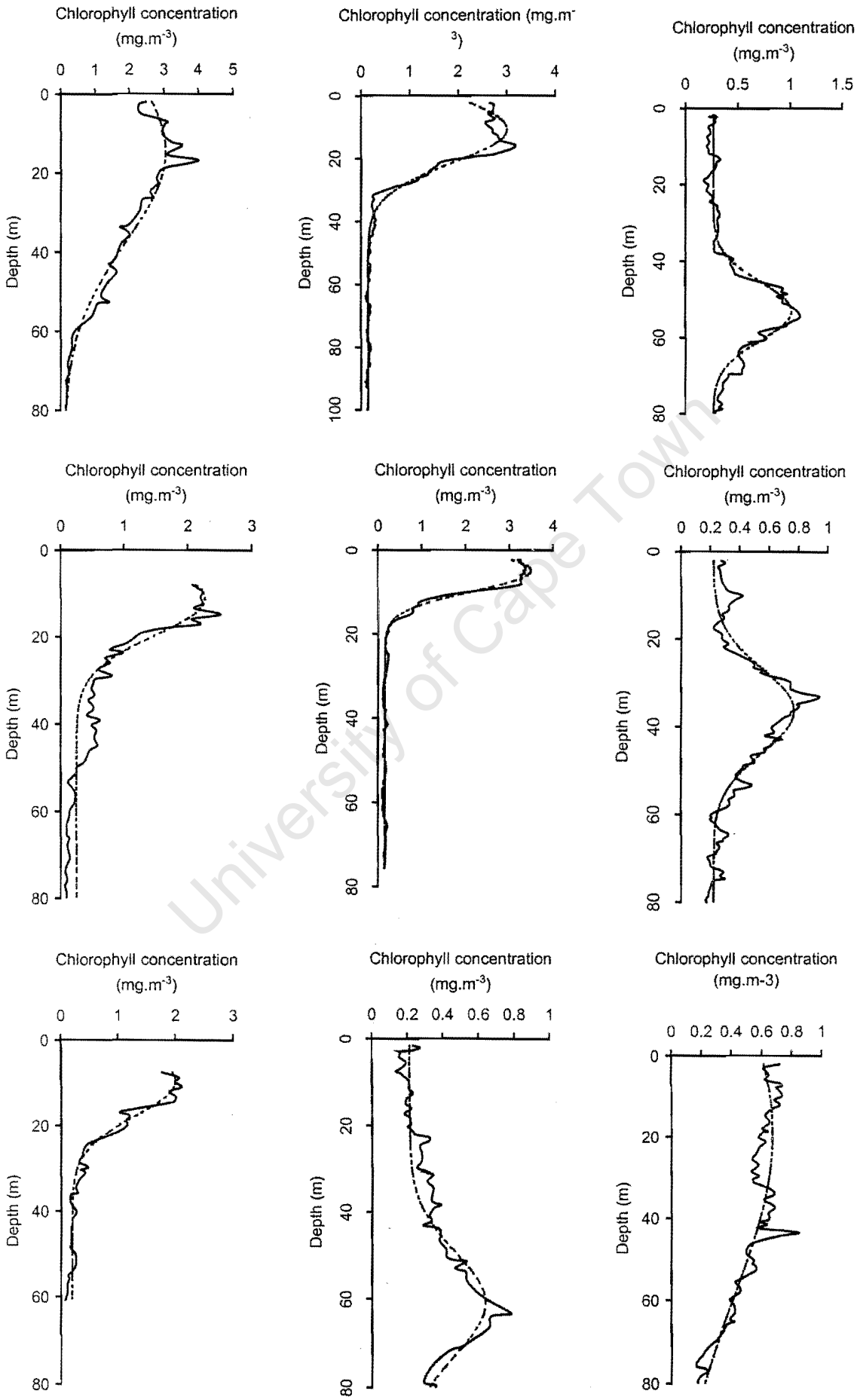


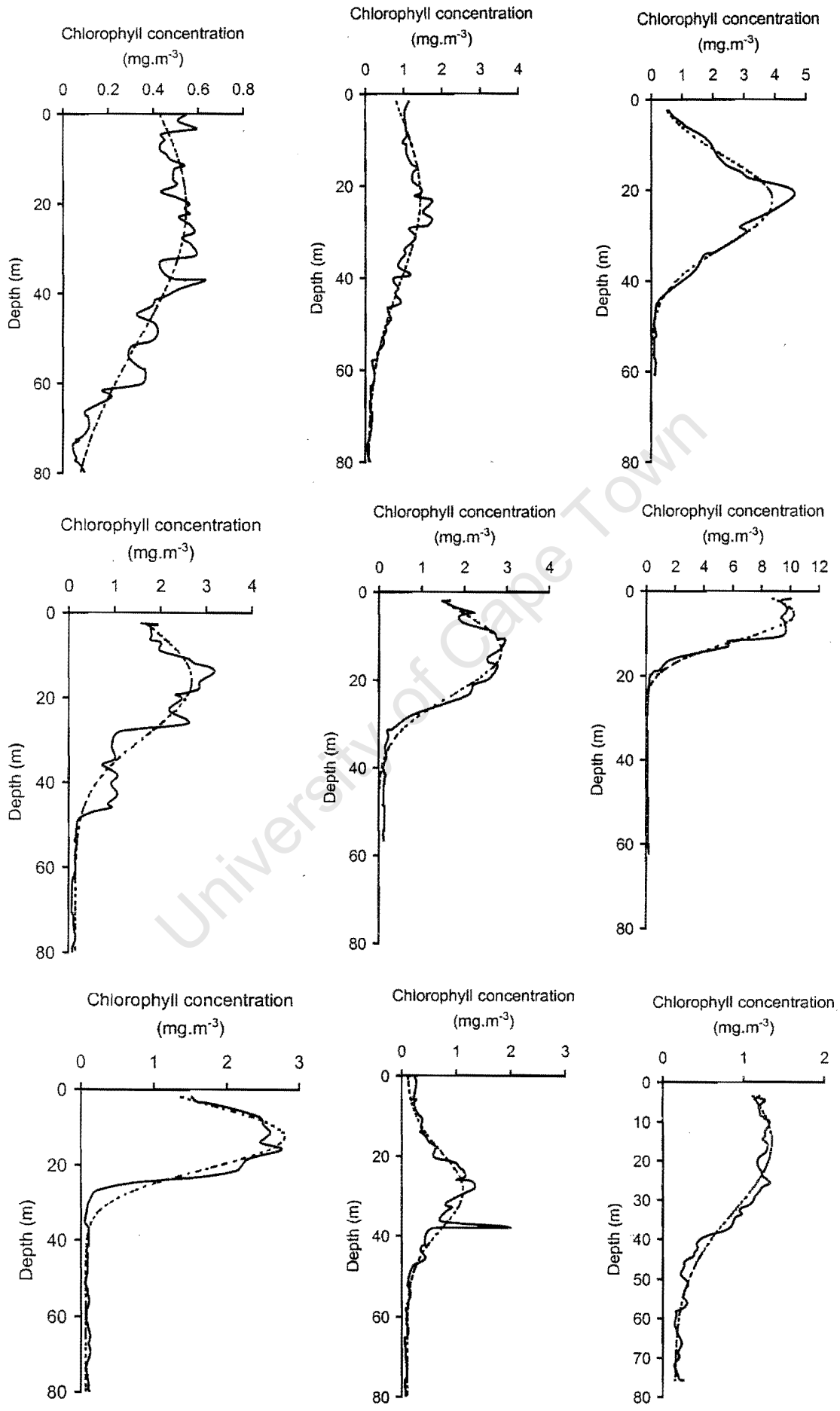




Best-fitted profiles







APPENDIX II

Table of parameter values derived from Gaussian curve fitting, (B_0, h, σ and z_m) for each profile. The variance explained (r^2 value) is included to show how good the model fitted each profile. Only profiles with $r^2 \geq 80\%$ were used in further analyses, and patterns are from the 6x4 SOM analysis.

S-plus GAM and GLM commands used for generalized modelling.

Pattern number	B_0	h	σ	z_m	r^2 value (%)	Pattern mapped to
1	0.0	173.2	31.4	14.8	82	24
2	0.0	207.8	48.6	0.0	88	24
3	0.2	109.9	27.9	26.2	91	24
4	0.3	73.2	16.6	18.3	82	16
5	0.2	184.6	29.3	14.7	83	18
6	0.2	232.6	30.9	6.1	95	18
7	0.1	18.4	17.8	28.6	88	21
8	0.0	33.5	14.6	21.5	87	16
9	0.0	117.1	16.5	15.3	97	17
10	0.0	112.1	16.1	17.3	96	17
11	0.0	26.6	27.2	24.0	81	23
12	0.0	56.9	14.9	19.5	94	16
13	0.5	26.0	11.1	12.5	95	9
14	0.2	14.0	11.4	23.9	85	19
15	0.1	60.2	26.8	28.5	95	23
16	0.1	37.2	27.0	28.2	88	23
17	0.2	13.6	11.2	42.7	86	20
18	0.0	57.5	24.3	26.3	90	23
19	0.2	37.5	20.9	31.1	89	22
20	0.5	8.6	7.4	7.4	92	3
21	0.2	8.0	6.8	54.4	87	19
22	0.2	19.9	8.5	49.0	91	19
23	0.2	11.9	8.5	42.6	85	13
24	0.1	25.7	23.7	18.7	96	23
25	0.1	44.1	24.7	20.9	95	23
26	0.2	39.3	14.4	29.4	97	21
27	0.2	31.3	15.4	12.9	95	16
28	0.1	75.5	16.4	11.2	96	16
29	0.1	36.0	17.8	42.3	91	21
30	0.2	35.3	11.5	34.0	98	20
31	0.2	24.6	14.6	33.4	97	21
32	0.2	31.6	14.3	14.8	92	16
33	0.1	80.1	24.4	17.5	94	3
34	0.1	27.0	17.8	29.2	94	21
35	1.5	334.2	12.8	0.0	97	1
36	0.1	79.8	9.1	10.7	90	10
37	0.1	49.1	19.7	17.7	93	22
38	0.1	29.5	18.4	33.9	87	21
39	0.1	20.1	18.2	26.0	94	22
40	0.0	70.7	5.3	8.5	89	3
41	0.1	23.9	19.6	16.3	95	22
42	0.1	58.4	8.5	12.4	97	9
43	0.2	83.8	13.2	15.7	94	16
44	0.0	40.9	29.8	20.9	93	24
45	0.1	63.7	18.7	22.4	95	22

46	0.0	97.1	9.9	22.4	97	15
47	0.0	71.2	9.9	13.3	98	10
48	0.0	165.1	6.4	5.4	98	4
49	0.1	24.8	9.7	28.6	81	14
50	0.0	114.8	13.4	10.4	94	11
51	0.0	87.8	15.1	22.0	95	16
52	0.7	131.4	14.8	16.1	83	11
53	0.5	86.4	27.3	18.4	89	23
54	0.2	37.3	8.3	9.6	97	9
55	0.3	43.4	8.6	11.2	94	9
56	0.1	170.0	23.1	14.6	96	18
57	0.5	2.3	5.2	11.2	81	2
58	0.2	53.4	17.7	15.3	97	17
59	0.2	12.4	11.6	60.9	83	19
60	0.2	41.0	4.9	4.9	99	3
61	0.1	76.9	10.7	10.7	98	10
62	0.0	68.7	40.8	19.4	91	24
63	0.4	3.2	3.4	10.9	83	3
64	0.2	13.1	9.6	36.4	91	20
65	0.3	278.6	11.7	3.9	96	5
66	0.2	222.3	15.8	0.0	80	11
67	0.3	14.3	7.7	53.0	94	19
68	0.4	276.9	4.9	5.8	96	5
69	0.1	303.9	8.4	15.6	97	5
70	0.2	105.6	9.9	9.2	96	10
71	0.2	82.7	6.6	38.0	96	14
72	0.0	256.4	14.3	14.2	84	11
73	1.9	182.6	2.6	9.6	96	1
74	0.2	21.6	6.7	58.1	93	19
75	0.4	98.2	6.6	0.0	92	3
76	0.1	21.0	12.1	58.2	85	19
77	0.0	45.2	8.7	6.5	84	9
78	0.2	30.5	8.2	52.5	94	19
79	0.2	72.5	13.0	10.8	98	10
80	0.0	250.3	9.7	8.8	96	5
81	0.2	12.6	6.4	59.2	85	19
82	0.0	445.2	21.9	0.0	90	12
83	0.2	48.9	4.2	31.1	98	14
84	0.2	34.3	11.3	54.6	96	19
85	0.0	111.9	7.0	7.6	83	4
86	0.0	162.5	15.4	0.0	98	11
87	0.1	65.5	7.2	6.9	98	3
88	0.2	176.2	6.3	3.1	100	4
89	0.3	120.1	6.2	0.0	97	4
90	0.3	35.8	5.4	41.2	90	13
91	0.3	8.6	2.7	35.6	87	14
92	2.0	59.4	3.8	4.7	93	1
93	0.3	13.4	9.0	35.9	92	14
94	0.3	158.8	4.7	0.0	99	4
95	0.1	421.9	13.0	0.0	94	6

96	0.2	158.0	11.5	9.8	98	11
97	0.2	12.8	20.0	30.2	82	22
98	0.2	21.2	20.4	29.8	93	22
99	0.3	12.3	11.0	46.9	84	19
100	0.1	35.3	31.9	34.9	94	24
101	0.2	10.2	20.5	23.9	81	22
102	0.2	18.6	14.1	29.5	95	21
103	0.0	40.7	39.7	48.8	86	24
104	0.1	18.1	21.8	42.2	90	21
105	0.2	9.8	12.5	18.6	91	15
106	0.1	29.8	30.9	24.5	92	24
107	0.3	9.2	9.2	44.3	81	19
108	0.1	39.7	16.1	0.0	94	10
109	0.2	71.0	12.8	9.8	96	10
110	0.0	110.0	8.5	6.8	90	4
111	0.0	22.9	12.8	7.0	95	10
112	0.0	27.1	14.3	14.7	96	16
113	0.0	33.6	9.7	13.0	94	9
114	0.0	38.7	17.2	30.2	87	21
115	0.2	97.8	21.4	18.9	83	17
116	4.5	4.6	0.5	11.1	93	1
117	0.6	207.6	14.6	14.6	85	11
118	0.0	134.6	27.6	30.8	88	24
119	0.2	17.9	13.0	53.5	91	19
120	0.3	62.4	4.8	9.5	86	3
121	0.2	22.0	10.8	38.9	93	20
122	0.3	31.0	5.5	42.2	92	13
123	0.2	30.4	6.7	49.8	97	19
124	0.0	169.1	5.7	6.6	91	4
125	0.2	26.3	11.3	57.8	97	19
126	0.0	154.5	6.6	5.0	97	4
127	0.0	92.9	2.7	3.7	99	4
128	0.0	54.7	7.2	23.0	89	15
129	0.2	23.2	8.5	49.1	96	19
130	0.4	19.4	1.9	2.6	96	3
131	0.0	140.5	6.6	5.6	97	4
132	0.0	75.6	4.3	7.3	97	3
133	0.1	81.7	5.5	8.2	94	3
134	0.1	9.4	6.2	23.4	88	8
135	0.0	161.7	6.5	5.1	97	4
136	0.0	215.4	11.0	11.7	97	11
137	0.0	116.3	13.3	8.7	92	11
138	0.0	679.3	10.9	12.2	93	6
139	0.1	530.1	10.1	10.5	98	6
140	0.4	309.7	4.8	12.4	98	5
141	0.2	454.2	5.7	10.6	98	6
142	0.0	578.6	7.8	6.8	99	6
143	0.0	1362.0	14.7	0.0	96	6
144	0.5	401.1	5.2	6.6	99	6
145	0.3	461.2	5.0	6.3	99	6

146	0.3	195.4	9.4	12.8	99	11
147	0.3	567.5	5.2	8.7	98	6
148	0.7	376.2	2.9	9.9	96	5
149	0.1	832.2	5.6	10.7	97	6
150	0.6	466.9	3.8	3.8	99	6
151	2.7	494.5	4.2	4.6	98	1
152	0.0	621.2	12.6	14.9	94	6
153	0.0	523.0	11.2	11.8	98	6
154	1.0	87.5	8.4	5.6	90	2
155	0.1	65.4	3.4	5.6	98	3
156	0.1	159.8	22.9	0.0	96	18
157	0.4	18.5	11.0	0.0	95	3
158	0.6	26.1	5.4	3.8	97	2
159	0.1	24.4	15.6	15.0	95	16
160	1.1	129.9	10.6	0.0	95	1
161	0.2	275.5	8.9	0.0	95	5
162	0.0	56.9	27.6	18.3	92	24
163	0.2	260.8	6.2	10.0	96	5
164	0.2	72.4	6.0	12.0	91	3
165	0.4	72.8	5.2	13.7	84	8
166	0.0	82.0	5.7	13.4	82	9
167	0.8	94.6	9.6	5.0	99	2
168	0.8	195.0	21.3	15.0	96	17
169	0.1	87.2	5.5	8.7	89	3
170	0.1	59.3	12.9	13.8	95	16
171	0.1	77.8	14.7	9.4	93	16
172	0.0	180.6	12.1	12.0	95	11
173	0.0	211.3	18.8	20.5	82	17
174	0.1	83.3	7.4	13.3	96	9
175	0.0	141.9	7.4	17.5	87	4
176	4.0	351.0	11.7	0.0	96	1
177	0.3	89.3	9.6	0.0	93	4
178	0.2	30.9	10.0	8.8	97	9
179	0.3	30.8	3.9	6.6	99	3
180	0.3	31.9	7.8	12.1	88	9
181	0.9	241.1	11.0	0.0	93	2
182	0.1	80.4	14.0	16.0	99	16
183	0.1	57.0	10.3	13.9	97	10
184	0.2	12.6	4.3	2.2	98	3
185	0.0	213.4	19.0	15.4	96	17
186	0.0	51.3	23.6	20.7	88	23
187	0.1	44.7	12.0	8.1	98	10
188	1.5	40.3	8.4	9.7	94	1
189	0.6	276.4	16.6	14.6	98	12
190	0.0	445.8	26.6	24.1	90	12
191	2.4	12.7	3.3	4.4	95	1
192	0.4	111.3	8.3	0.0	88	4
193	0.0	250.7	16.1	12.7	97	11
194	0.2	96.3	5.3	10.9	87	3
195	0.0	204.3	11.3	0.0	94	5

196	0.0	400.6	18.9	15.7	91	12
197	0.0	321.3	10.5	0.0	91	5
198	0.0	205.9	11.9	11.1	95	11
199	0.3	121.4	14.8	9.9	99	11
200	0.0	49.9	7.5	0.0	93	4
201	0.3	4.8	2.1	17.6	91	8
202	0.0	813.1	13.9	9.1	98	6
203	0.1	23.0	5.6	2.6	97	3
204	0.1	471.5	18.7	14.7	96	12
205	0.1	294.0	12.2	9.9	99	5
206	0.0	718.5	18.2	0.0	92	6
207	0.2	144.9	4.0	9.1	95	4
208	0.2	28.7	16.3	17.8	97	16
209	0.2	18.6	4.5	16.4	93	8
210	0.4	58.9	5.2	9.8	92	3
211	0.0	185.4	10.5	13.7	99	11
212	0.0	33.1	7.7	6.6	98	9
213	0.2	154.9	5.2	5.5	98	4
214	0.0	420.7	30.5	20.0	98	18
215	0.3	91.2	13.8	9.6	99	10
216	0.2	89.7	22.8	19.7	94	23
217	0.3	37.5	9.9	13.0	93	9
218	0.4	102.4	5.8	5.4	98	3
219	0.1	101.6	18.8	18.5	97	17
220	0.1	36.9	9.5	8.5	99	9
221	0.1	35.0	17.1	18.3	96	22
222	0.0	67.2	9.3	7.5	96	10
223	0.1	40.1	21.2	18.7	96	22
224	0.5	82.4	1.9	15.0	95	2
225	0.1	33.1	7.6	10.1	91	9
226	0.2	17.1	11.8	32.3	93	21
227	0.2	11.5	6.2	44.5	87	13
228	0.1	15.9	7.8	28.7	95	14
229	0.2	128.4	8.6	10.9	97	4
230	0.5	24.5	10.9	12.5	91	14
231	0.1	49.7	29.0	24.6	85	24
232	0.2	44.1	19.1	22.3	93	24
233	0.1	51.2	16.2	11.7	99	16
234	0.2	37.5	10.3	8.1	97	9
235	0.1	40.0	8.2	13.3	98	9
236	0.2	45.2	8.2	15.7	98	9
237	0.1	89.4	11.6	19.9	96	15
238	0.2	115.9	13.7	19.4	98	16
239	0.2	21.2	14.0	17.3	91	16
240	0.1	101.8	9.6	16.9	98	10
241	0.2	72.8	16.9	19.8	95	22
242	0.0	76.0	9.3	21.0	95	15
243	0.1	12.0	9.1	8.6	97	9
244	0.1	8.2	9.5	18.6	90	15
245	0.1	7.8	11.3	30.9	90	21

246	0.1	10.1	14.2	28.3	90	21
247	0.1	13.3	9.3	27.5	91	15
248	0.2	8.5	2.1	33.1	91	14
249	0.1	21.3	8.0	7.2	96	9
250	0.0	30.7	9.4	10.5	91	9
251	0.1	21.5	7.9	19.5	87	8
252	0.1	24.1	6.5	48.6	96	19
253	0.1	15.1	5.0	5.0	98	3
254	0.2	13.6	6.7	34.9	89	14
255	0.1	8.6	7.9	40.8	85	13
256	0.1	15.1	9.9	46.8	82	19
257	0.2	9.9	7.9	38.7	89	20
258	0.0	42.0	8.5	8.1	96	9
259	0.2	7.4	2.7	34.6	90	14
260	0.1	22.6	6.0	41.0	92	19
261	0.2	8.6	2.4	49.0	89	19
262	0.1	2.5	7.4	9.4	86	9
263	0.2	12.6	5.2	43.3	86	13
264	0.3	16.5	7.8	49.0	97	19
265	0.3	9.4	6.0	57.0	90	19
266	0.3	19.6	11.7	58.3	94	19
267	0.3	21.8	9.8	59.1	92	19
268	0.3	18.6	6.9	55.8	94	19
269	0.4	38.5	4.5	18.2	97	8
270	0.1	26.8	23.2	43.8	85	21
271	0.3	13.7	6.3	53.8	95	19
272	0.3	17.8	9.8	52.6	91	19
273	0.3	16.5	6.1	48.6	94	19
274	0.5	15.9	12.1	40.4	80	20
275	0.5	14.4	12.6	47.8	85	19
276	0.6	10.9	11.5	45.0	82	19
277	0.6	17.0	9.0	43.8	80	13
278	0.5	25.2	8.4	36.5	94	14
279	0.7	18.3	9.1	33.9	80	7
280	0.7	40.4	7.1	18.5	96	2
281	0.4	4.5	2.1	27.8	85	14
282	0.4	15.7	1.8	26.9	91	14
283	0.2	30.4	30.2	12.5	88	24
284	0.2	22.1	22.2	15.8	94	23
285	0.1	41.0	30.8	23.5	86	24
286	0.2	36.6	16.5	11.8	97	16
287	0.7	35.5	15.5	11.8	91	16
288	0.2	75.2	13.3	3.2	95	10
289	0.2	49.9	19.3	13.4	97	17
290	0.2	48.0	15.6	6.7	96	16
291	0.2	19.0	16.0	11.7	98	16
292	0.2	28.0	20.9	12.1	96	17
293	0.2	107.1	12.1	19.3	93	15
294	0.2	78.9	9.6	10.9	81	10
295	0.0	101.5	9.3	5.7	90	4

296	0.2	69.4	17.1	22.5	93	22
297	0.2	36.6	7.9	15.7	96	9
298	0.2	46.4	8.0	14.0	95	9
299	0.1	115.1	18.0	21.0	93	17
300	0.2	61.4	15.0	16.0	95	16
301	0.3	114.6	20.6	12.2	96	17
302	0.0	131.9	36.7	19.6	90	24
303	0.0	126.2	34.2	24.1	95	24
304	0.0	166.4	23.2	23.6	89	18
305	0.1	34.9	15.4	14.9	92	16
306	0.2	25.9	17.8	31.8	94	21
307	0.0	85.5	23.6	33.6	86	23
308	0.1	85.4	39.5	10.2	83	24
309	0.1	64.1	10.6	11.0	96	10
310	0.2	7.4	8.5	8.6	80	9
311	0.1	82.6	17.8	31.2	96	21
312	0.1	51.6	17.4	13.3	96	16
313	0.2	24.4	5.2	7.1	90	3
314	0.3	35.7	7.3	6.8	87	3
315	0.1	33.7	25.4	26.6	82	23
316	0.2	42.7	10.0	34.0	94	14
317	0.2	78.3	19.3	16.2	97	17
318	0.1	44.0	12.6	9.1	99	10
319	0.2	44.0	11.5	7.9	90	10
320	0.2	33.4	11.4	29.2	88	21
321	0.2	66.1	16.7	39.9	84	21
322	0.1	49.5	15.1	24.8	90	21
323	0.2	22.9	13.3	7.8	96	10
324	0.0	81.8	8.3	6.8	91	4
325	0.2	12.9	8.2	41.1	89	13
326	0.3	19.7	4.9	46.5	92	13
327	0.2	17.8	10.8	41.4	90	20
328	0.4	38.2	10.8	37.6	86	20
329	0.2	31.1	13.4	15.1	98	16
330	0.2	24.6	9.7	24.3	90	15
331	0.0	49.0	14.7	34.1	87	21
332	0.3	21.9	8.2	34.2	96	14
333	0.3	11.4	6.7	27.4	89	14
334	0.3	17.8	7.6	30.2	83	14
335	0.3	19.0	5.7	31.0	90	14
336	0.3	13.7	3.5	32.1	89	14
337	0.1	42.3	20.7	24.5	89	16
338	0.3	14.0	5.9	44.6	85	13
339	0.3	16.4	5.4	41.4	95	13
340	0.2	13.0	7.4	31.2	89	14
341	0.2	15.3	9.5	37.5	84	20
342	0.3	12.7	4.1	34.0	88	14
343	0.3	12.2	6.8	36.8	89	14
344	0.3	17.0	6.3	38.7	88	13
345	0.3	6.7	3.5	31.4	80	14

346	0.3	19.9	8.4	24.3	87	8
347	0.3	36.1	9.1	7.5	80	9

Examples of GLM and GAM commands used in modelling

```
sqrth.interpt<-
gam(sqrt(h)~1,data=Shefallpar,na.action=na.omit,subset=Calcsurfchl<32)

season<-ordered(c("autumn","spring","summer"))

Levels(season)

area<-ordered(c("eab","wab","wc"))

levels(area)

LogSD.sound.sea.chl.area.glm<-glm(log(sd)~I((Sounding-
170)*(Sounding<170))+Season+cbind((Calcsurfchl-3)*(Calcsurfchl<3),(Calcsurfchl-
3)*(Calcsurfchl>3))+Area,data=Shefallpar,na.action=na.omit,subset=Calcsurfchl<32)

summary(logSD.sound.sea.chl.area.glm)

anova(LogSD.sound.sea.chl.area.glm,test="F")

sqrtH2.Chl.Sound.Season.SST.glm<-glm(sqrt(h)~I((Calcsurfchl - 20) * (Calcsurfchl
< 20)) + I((Sounding - 140) * (Sounding < 140)) + Season+I((SST-
15)*(SST>15)),data=Shefallpar,na.action=na.omit,subset=Calcsurfchl<32)


sqrtH.Chl.Sound.Season.SST.Area<-
gam(sqrt(h)~s(Calcsurfchl)+s(Sounding)+(Season)+s(SST)+(Area),data=Shefallpar,n
a.action=na.omit,subset=Calcsurfchl<32)

summary(sqrtH.Chl.Sound.Season.SST.Area)

plot.gam(sqrtH.Chl.Sound.Season.SST.Area, res=T, scale=2)

plot.gam(sqrtH.Chl.Sound.Season.SST.Area, se=T)

r2. sqrtH.Chl.Sound.Season.SST.Area <-1-( sqrtH.Chl.Sound.Season.SST.Area $deviance/
sqrtH.Chl.Sound.Season.SST.Area $null.deviance)

cat(r2. sqrtH.Chl.Sound.Season.SST.Area)


Zmtest2.glm<-glm(Zm~exp(-Calcsurfchl)+Season+Area,data=Shefallpar,
na.action=na.omit,
Subset=Calcsurfchl<32,ordered=T,contrasts=levels("contr.treatment"))

Zmtest2.glm<-glm(Zm~exp(-Calcsurfchl)+Season+Area,data=Shefallpar,
na.action=na.omit,
Subset=Calcsurfchl<32,labels=as.levels,contrasts=levels("contr.treatment"))
```

```
predict.gam(Zmtest2.glm,newdata=Env98,type="link")

predict.gam(Zmtest2.glm,newdata=Env98,type="terms")

predict.gam(Zmtest2.glm,newdata=Env98,type="response")
```