Relating an archive of *in situ* vertical chlorophyll-a profiles to concurrent remotely sensed surface data

Robert I. Williamson

Submitted in partial fulfilment of the requirements of the degree of Master of Science in Applied Marine Science, Department of Oceanography, Faculty of Science, University of Cape Town, South Africa. Degree by coursework and dissertation.

November 2007

Supervisor: Prof. J.G. Field

Co-supervisor: Prof. C.J. Reason

TABLE OF CONTENTS

Abstract

Knowledge of the vertical distribution of phytoplankton in the upper ocean is essential for accurate estimates of primary production. Satellite remote sensing has given scientists an unprecedented view of near-surface chlorophyll distribution and other surface conditions, including sea surface temperature and wind data, from regional to global scales but little information on the dynamics below the surface. As a result estimates of global production tend to use regional profile averages but these methods oversimplify the smaller scale dynamics, particularly in coastal regions where productivity is highly variable on time scales of weeks. A pilot study by computer science honours students in 2006 showed the viability of using a Dynamic Bayesian Network (DBN) in predicting a representative profile per pixel of a satellite map based on a database of time series satellite surface data.

In this study, 5813 *in situ* profiles were obtained from the highly dynamic upwelling region around the southwestern coastline of southern Africa. The samples were collected between 1988 and 2006 between the coast and the continental slope. The region was divided into three sub-regions according to biophysical processes: the west Coast; the west Agulhas Bank; and the east Agulhas Bank. Of the 5813 profiles, 5557 were included in the sub-regions. Two consecutive processes were then applied to the profile database. First, the profiles were clustered using a k-means clustering program which produced 16 representative clusters. The centroids of the clusters showed a variety of shapes that could depict various upwelling sequences as well as offshore and open water characteristics. The integrated chlorophyll concentration for the average profile of each cluster ranged from 21.5 $mg.m^{-2}$ to 358.8 $mg.m^{-2}$. The number of profiles in each cluster ranged from 148 to 604. Second, the individual profiles were allocated satellite surface data of sea surface temperature, surface chlorophyll and surface wind according to the date and position of the profile. For processing and storage efficiency 5-day averaged SST and surface chlorophyll, and weekly averaged wind data were used. The viability of using this database in a DBN was then analysed by determining the relationships between the profile and the satellite surface data. Seasonal wind patterns and regional variability were evident in the profile frequency distributions and profile shapes related well to SST and surface chlorophyll. Sequences of profile development were also corroborated by SST and surface chlorophyll suggesting a good relationship. The relationship between wind data and profile shape however, was not as robust. This most likely due its variability on time scales of less than a

week. Although the temporal relationships between the surface data and profiles could not be tested by the DBN within the time span of this study the database of profiles and their related remotely sensed surface data has provided an encouraging basis for its development and future potential as an accurate predictive tool.

Acknowledgements

I would like to thank the Department of Environmental Affairs and Tourism, Marine and Coastal Management branch for providing the chlorophyll data used in this project. I would also like to acknowledge the funding provided by the National Research Foundation Benguela Ecology Program and thank Prof. Chris Reason as an NRF grantholder. I thank my supervisor Prof. John Field for his advice and generosity with his time. Thank-you to my colleague Andrew Symington for his help in solving the problems encountered in the code. A final and special thanks goes to my family and friends who generously accommodated me when needed and wined and dined me when my pockets were empty.

Chapter 1: Introduction

Biological oceanographers have in the past tended to focus on the smaller scales of living marine organisms and are primarily concerned with the fertility of the oceans and the production from these organisms in terms of energy that is made available to higher trophic groups. These concerns have long been established in the field of marine science (see Chambers, 1912 for a review of works between 1841 and 1912) with the combined efforts that discovered photosynthesis and those by Hooker (1874) and Ørsted (1844) who independently discovered in the mid 19th century that unicellular planktonic plants support oceanic food webs. The lower energy transfer processes through these webs that incorporate organisms ranging from nanoflagellates to fish larvae have vital implications for fisheries yields and it was this fundamental association that formed the foundation for early research. At the very base of the energy transfer pathway are the phytoplankton, microscopic organisms that obtain their energy autotrophically through photosynthesis. Besides the obvious bottom-up implications for marine food webs they more recently have become a major focal point in the recycling of nutrients, the flux of carbon between the ocean and atmosphere and the transfer of organic matter from the ocean surface to the ocean floor.

1.1 A brief history of phytoplankton research

Essentially the major goal of biological oceanography is the determination of temporal and regional changes in plankton productivity in the world's oceans. Steemann Nielsen (1952) developed the ^{14}C method that gave biological oceanographers a break-through tool to accurately calculate primary productivity with meaningful precision, accuracy and efficiency. Another major leap forward followed the launch of the Coastal Zone Colour Scanner (CZCS) by NASA in 1978 which provided a large scale view of surface phytoplankton distribution across the global oceans.

An understanding of the concept of photosynthesis was evident from the early 1800s and although it was applied to aquatic environments from about 1850, a sound description was only made near the end of the 19th century. The source of inorganic carbon for aquatic plants proved much more problematic for scientists than it did for terrestrial plants. During this period Victor Hensen (1887) was advocating the idea of Hooker (1874), Ørsted (1844)

that oceanic food webs were based on material from macroalgae, aquatic plants and terrestrial derived plant debris (this argument was disputed by Ernst Haekel). It was believed that the determination of the consumption or reproduction or microalgae could indicate the amount of animal biomass that could be supported. Ironically Hensen (1887) did not realize just how small the microalgae were in his research and missed the most abundant phytoplankton, the nanoplankton. Furthermore he made the erroneous assumption that plankton abundance was the same everywhere due to universal horizontal mixing. Gran (1912) hypothesized about the regional, latitudinal, vertical and seasonal variations in primary production. His ideas on the photic zone were based on observations made by Whipple (1899) on phytoplankton growth in bottles at various depths. Light, however, could not explain the latitudinal variability when considering the biomass poverty of the tropical oceans. Nutrients had been proposed as major regulators earlier by Brandt (1899) and Nathansohn (1906) who highlighted the influence of upwelling and mixing through the supply of subsurface nutrients. Gran (1912) reasoned that the physics of ocean gyres with their ascending motion in cyclonic systems and the opposite in anti-cyclonic systems could explain latitudinal variations on the macro scale. Variations in open-ocean and coastal waters he hypothesized as the result of increased upwelling and mixing although he allowed for terrestrial runoff to play some role. Gaarder and Gran (1927) began investigating the seasonal pattern in an Oslo fjord system. Together they documented the spring bloom and summer decreases in phytoplankton abundance. Similar studies by Atkins (1928) and Marshall and Orr (1928) indicated the importance of solar heating and stable conditions that led to a well defined thermocline during a spring time bloom and the subsequent depletion of nutrients from the isolated layer. The outcome of these studies highlighted the necessity for convective winter mixing to restore nutrients to the surface layer and the need for the thermocline to retain phytoplankton in the sunlit and initially nutrient-rich upper layer. Throughout these hypotheses there had been no method of actually calculating plant production. Nets had been used for sampling but were completely inadequate for capturing nanoplankton and frequently clogged.

Between 1910 and 1940 there was a concentration of effort towards developing quantitative methods to measure planktonic productivity. Two methods were already at hand: measuring oxygen production with the Winkler method and measuring carbon dioxide uptake by determining pH changes with indicator dyes. A third method which had

developed earlier was measuring the uptake of phosphate and using the C to P ratio to calculate carbon production. Using these methods led to questions regarding the depth of light at which respiration is balanced by production. This Gran and Gaarder (1916) tested by the procedure now known as the light and dark bottle method. Concurrently efforts were also focused on productivity determinations based on the uptake of inorganic nutrients nitrate, silicic acid or phosphate. Although the methods used prior to World War II may have been labour-intensive with poor precision and unknown accuracy they were able to provide the annual cycle in temperate-latitude, coastal ocean environments if not quantitative measurements of productivity.

Steemann Nielsen's (1952) ^{14}C method, made possible through the development of radioisotopes for the use in biological tracer experiments from around 1937, made its debut on the 1950 *Galathea* Expedition. He had the insight that the radiotracer methodology could be used to quantify a bulk ecological process such as primary production. After WWII biological oceanography became a government supported science in many countries. Oceanography tended towards international collaboration providing a much higher spatial and temporal resolution of the oceans (such as International Geophysical Year, 1957 and International Indian Ocean Expedition, 1964). By the 1970s the ^{14}C method had been used in most parts of the world enabling scientists to estimate a more robust global ocean primary productivity and the characteristics of productivity in the various regions.

A general consensus in 1958 was that nutrient supply in most cases determines the productivity of a region. However there were reservations to the nutrient uptake method because recycling and complex hydrography precluded the rigorous nutrient budgeting necessary to estimate productivity. The oxygen production method too was rendered obsolete due to bacterial growth in the incubators. In 1957 ICES recommended the use of the ^{14}C method for measuring production of organic matter in the oceans. Ryther (1963) used measurements obtained by the ^{14}C uptake method and analyzed the regulation of productivity geographically by looking at all aspects of a given region: water colour; temperature; biomass of higher trophic organisms; wind and seasonality. Later Ryther (1969) estimated fish yields in a bottom-up process of energy transfer between trophic levels from primary production in each of the generalized regions: oceanic; coastal and upwelling.

In summary, pre-WWII thorough descriptions of the dynamics of plankton productivity were achieved particularly in the vertical dimension. Thereafter the international expansion of biological oceanography and the developments that led to Steemann Nielsen's ^{14}C method enabled a greater understanding of the seasonal and regional variations and estimates of global productivity. The addition of satellite information at much higher spatial and temporal resolution than was possible from ships was a second milestone in understanding phytoplankton dynamics.

1.2 Vertical Distribution of Phytoplankton

1.2.1 Vertical Dynamics

Sverdrup (1953), quantifying the earlier work of Gran (1931), incorporated the interactions between light, nutrients, mixing and stability to explain the onset of the North Atlantic spring bloom. Nutrients, light and temperature control the individual algal cell growth whereas environmental conditions that control water column stability, advection, grazing and sinking control population growth through the control of concentrated layers. Models based on these dynamics simulate growth and loss for phytoplankton biomass through photosynthesis, photorespiration, predation and sinking. Of great consequence to production is the mixed layer depth relative to the critical depth (where the total photosynthetic gain is equal to respiratory loss). During winter months and high surface winds the mixed layer is deep and cells will find themselves below the critical depth and will experience a negative net growth because on average they respire more than they can photosynthesize. At the same time deep mixing introduces new nutrients into the upper layer. With the onset of spring and increasing sun angle and incident radiation, the upper layer warms and stratifies and between wind-mixing events a spring bloom can develop. At lower latitudes where seasonal changes are not as pronounced wind mixing and irradiance forcing is less significant. Here the mixed-layer depth is consistently shallower than the critical depth so that there is a net gain in carbon fixation.

1.2.2 Deep Chlorophyll Maximum

The understanding of the horizontal and vertical patterns of phytoplankton was improved through technical advances in proxies for phytoplankton biomass such as the use of chlorophyll fluorescence (Yentsch and Menzel, 1963; Holm-Hansen *et al.*, 1965) and the development of continuous monitoring of *in vivo* chlorophyll (Lorenzen, 1966). Many of

the vertical profiles of chlorophyll observed during the early stages from diverse regions featured a subsurface or deep chlorophyll maximum (DCM) layer. During the late spring and summer months in coastal upwelling systems the pycnocline separates the phytoplankton in the nutrient-depleted upper layer from the deeper nutrient-rich water. The phytoplankton are however able to survive for some time on the nutrients excreted by the predatory zooplankton and other heterotrophs in the mixed layer. Typically this period is marked by moderate productivity but low chlorophyll as the phytoplankton biomass is consumed as rapidly as it grows. In many instances there is a community change among the phytoplankton that coincides with the nutrient depletion from one dominated by diatoms to one dominated by dinoflagellates. The dinoflagellates are able to out-compete the diatoms at low nutrient concentration due to their better surface to volume ratio and motility that enables them to move out of their low nutrient micro-environments. At this stage a chlorophyll maximum can be found at the pycnocline as phytoplankton tolerable to the low light conditions exploit the nutrients that diffuse across the pycnocline. It should be noted that diffusion of nitrate across the density barrier can result in deep chlorophyll maxima but this does not necessarily indicate a biomass maximum. Species under low light but higher nitrate concentrations can produce more chlorophyll than species higher in the water column (Cullen, 1982). The DCM has implications for primary production estimation because their location and size are not constant or easily predictable. Assuming a constant profile, as used in the models of global production by Platt (1986), generally can lead to an underestimation of total primary production (Platt *et al.*, 1991) by as much as 30% (Millan-Nunez *et al.*, 1997). Sathyendranath *et al.* (1995) point out that the DCM is often below the surface mixed layer and that this can be regarded as new production. This new production is what interests many ecologists and climatologists because when it is averaged over a long period it is the proportion of total production that is exported from the euphotic zone and is an important component of the biological pump that removes carbon from the upper layers.

1.3 Remote Sensing and Primary Production

The physical forcing of the upper ocean and the associated response of phytoplankton was well documented by the late 1970s. This was particularly due to the small spatial scales and accessibility of the relevant data from ships. In contrast to this was the lack of concurrent knowledge of basin-scale horizontal distribution of phytoplankton due to the much larger spatial scales and inadequate sampling ability of ships. Without doubt this

problem would have persisted today were it not for the development of satellites. As Longhurst (1998) points out in his book *Ecological Geography of the Sea*, sensors mounted on earth-orbiting satellites have provided scientists the means to shift their focus from understanding the physiological processes governing the pelagic biota to a better understanding of the horizontal distribution, abundance and biomass of the organisms themselves. Images obtained from 1978 by the Coastal Zone Colour Scanner (CZCS) sensors on board the NIMBUS satellite provided biological oceanographers with an unprecedented look at regional and seasonal variations in ocean colour representing near-surface phytoplankton chlorophyll. Due to the optical complexities of Case 2 waters (coastal waters where the presence of other substances could influence the water-leaving radiance) the initial mission of the CZCS was to retrieve data mainly from open-ocean waters (Case 1 waters). However new sensors and better algorithms for the retrieval of surface chlorophyll information from satellite data as well as scientific endeavours to improve the interpretation of ocean colour have much improved the capabilities of remotely sensed ocean colour. A major source of the difficulty still endemic to remote sensing of Case 2 waters is the non-linear nature of the problem and the similarities between the characteristic optical properties of individual components to be retrieved (Neumann *et al.*, 2000 in Sathyendranath (ed.), 2000).

1.3.1 ***Modelling Primary Production*** Previous models, for example Smith (1981) and Eppley *et al.* (1985) have estimated time and depth-integrated primary productivity as a function of surface chlorophyll values (Behrenfeld and Falkowski, 1997). Good correlations have been found between surface chlorophyll and integrated chlorophyll. For example surface chlorophyll accounted for 82% and 95% of integrated chlorophyll variation respectively (Lorenzen, 1970 Smith and Baker, 1978). More complex models that incorporate surface irradiance produce depth integrated production that is the product of depth integrated chlorophyll, daily surface irradiance and a constant for average photosynthetic yield (Morel, 1978; Falkowski, 1981; Platt, 1986; Morel, 1991). A more mechanistic approach has been the use of complex bio-optical models (Platt and Sathyendranath, 1988; Morel and Berthon, 1989; Morel, 1991; Platt *et al.*, 1991). These models account for the attenuation of photosynthetically active radiance through the water column and include photosynthesis vs. irradiance parameters.

Empirical and semi-analytical models which apply to the entire euphotic zone have been compared (Platt and Sathyendranath 1993; Behrenfeld and Falkowski, 1997) and ideally require chlorophyll biomass with depth profiles. Morel and Berthon (1989) suggested that surface chlorophyll can be used to predict not only the depth-integrated biomass but also the vertical structure. They were able to reproduce categorized mean profiles (that could be reproduced by a generalized Gaussian profile) for Case 1 waters from surface chlorophyll values and computed production using photosynthesis vs. irradiance parameters. They pointed out however, that relationships based on statistical averages do not account for unusual vertical distributions. Other authors have used a Gaussian distribution function to represent the vertical profile of chlorophyll biomass (Lewis *et al.*, 1983; Platt and Sathyendranath, 1988; Sathyendranath *et al.*, 1995). This method uses the four parameters that describe the curve fitted to a particular profile (interpreted as background chlorophyll, depth of the chlorophyll maximum, breadth of the peak and total biomass above the baseline concentration) in estimations of primary production.

Scaling regionally-developed models of primary production to global scales presented two problems. Due to the paucity of the *in situ* database two options were available to oceanographers in modelling global primary production: they could assume the ocean is a biological continuum or they could partition the ocean into geographically characteristic regions. Regions do not only differ according to their physical dynamics but as a consequence, the vertical structure of phytoplankton. Sathyendranath *et al.* (1995) found that the relationship between surface chlorophyll and chlorophyll as a function of depth varied significantly to warrant the use of biogeochemical provinces that describe the reality of seasonal phytoplankton growth as initially proposed by Platt and Sathyendranath (1988). Surface chlorophyll values do to some extent predict the subsurface profile but in terms of integrated chlorophyll it is more useful for analytical models to have depth related concentrations. Sathyendranath *et al.* (1995) classified the North Atlantic Basin according to the differences in the physical environment that would most likely have an influence on regional algal dynamics. Within each of these provinces the authors established seasonally differentiated photosynthesis-light curve parameters and parameters that determine the vertical structure in the chlorophyll profile. Thus they advocate the need for both satellite data to resolve the temporal and spatial issues and *in situ* data to resolve the dynamics below the surface.

1.3.2 Ocean Variability

Afanasyev *et al.* (2000) noted studies by Platt and Sathyendranath (1988) and Platt *et al.* (1991) that suggested that regions with characteristic seasonal physical variations that effect the depth of the mixed layer could describe biological dynamics. Numerical modelling studies (e.g. Dutkiewicz *et al.*, 2000) that calculate the spatial pattern of chlorophyll concentrations support this. Longhurst *et al.* (1995) created a biogeography of the global ocean based on how the ecology of plankton responds to regional oceanography. The authors initially used the forcing described by Sverdrup (1953) *viz.* forcing of the mixed layer depth by local wind and irradiance at the surface, but found this insufficient to describe the dynamics in all regions. They recognised four primary domains by their characteristic forcing, some of which required additional factors: the seasonal cycle of sea ice in the *Polar Domain*; the westerlies wind stress from quasi-permanent low-pressure cells in the *Westerlies Domain*; the frictional wind stress by the seasonal trade winds in the *Trade Winds Domain*; and the modified oceanic circulation from interaction with the coastal topography and coastal wind regime in the *Coastal Domain*.

In coastal waters (which are described as extending from the edge of the continental shelf to the high water mark) the processes effecting biological production are complicated in particular by shallow depths, tidal currents and the proximity to the coast. Shallow water means that the mixed layer may frequently extend to the bottom and entrain accumulated nutrients into the water column making them available for photosynthetic production. Related to the shallowness are tidal currents that cause turbulence as they flow over the bottom, which may reach to the surface. Frequently more shallow regions are mixed all year round where adjacent regions may be seasonal (first suggested by Bigelow, 1927, in Mann & Lazier, 1996 p.219). The region of transition between deeper stratified water and mixed shallower water is a tidal front and sometimes occurs over the continental shelves. It has been suggested that the position of the tidal front varies according to the lunar cycle and the strength of the tidal currents. Simpson (1981) described the circulation and structure of a front as having downwelling at the front and upwelling on the well-mixed side. There was a strong along-front mean flow and fontal eddies. The newly stratified water behind the advancing front will contain nutrients characteristics of the mixed water just ahead of it. In addition areas of weak tidal mixing will be stratified before those with stronger tidal mixing and coincide with the development of a bloom (Pingree *et al.*, 1976). The third

characteristic acts as a barrier to advection during along-shore winds and can result in either upwelling or downwelling. Upwelling frequently replenishes nutrient concentrations in the upper layers and fuels some of the most biologically productive regions in our oceans. Fresh water input from land can also play a role in stratification due to its lower density, sometimes creating a temperature inversion. The more buoyant fresh water leads to a surface height anomaly and buoyancy-driven currents. These processes, simplified as those enhancing stability (freshwater input) and those creating turbulence to break down stability (wind-driven and tidal currents), create much stronger horizontal gradients than in the open ocean and water properties tend to be distributed in the horizontal.

1.4 Study Area

Coastal upwelling areas contribute about 20% of the global fish production although constituting only around 1% of the ocean surface area (Ryther, 1969; Cushing, 1971; Mann, 2000). The disproportionate productivity is a result of the localized upwelling winds. The concentrated food environment provides good spawning grounds for pelagic fish and for fish production. Productivity can vary between these upwelling systems as well as spatially within each region in response to variations in wind strength, bathymetry, latitude and the hydrographic properties of the water column. The variation in productivity, particularly on inter-annual time scales is of interest as the availability of phytoplankton may limit the occurrence and productivity of organisms at higher trophic levels (Borchers and Hutchings, 1986). The region of interest for this paper falls into such an upwelling coastal zone that is described by Longhurst *et al.* (1995) as the Benguela Current Coastal Province.

The map area in this study extends from 5°S, 5°E to 40°S, 40°E which includes the most northern extent of the Benguela Current system which has been argued to extend as far north as 12-13°S (Moroshkin *et al.*, 1970 in Shannon, 1985) to the Angola-Benguela frontal region, and the most southern extent as far as the Agulhas retroflection area at approximately 40°S (Shannon *et al.*, 1981). The separation of the Benguela Current into the westward northern limb of the subtropical gyre is completed in the vicinity of Cape Frio at 18°S where the warm tropical surface water of the Angola Current meets the cool water of the Benguela. The Angola-Benguela Front is the result of the convergence of the two oppositely moving fronts (Shannon *et al.*, 1987). Upwelling in the Benguela system can extend as far east as Cape Agulhas (35°S, 20°E) (Shannon, 1985). Upwelling also occurs east of Cape Agulhas (Shannon & Nelson, 1996; Lutjeharms *et al.*, 2000). Across the

Agulhas Bank Ekman transport is predominantly onshore due to the prevailing Westerlies however upwelling can occur where the Agulhas Current flows against shelf edge (Bakun, 1993; Lutjeharms *et. al.*, 2000).

The continental shelf is wider off southwestern Africa than it is off Angola. Off the Orange River mouth the shelf extends 150km offshore. There are anomalously deep zones where double shelf breaks are common at about 150 to 200m and 300 to 500m respectively. The shelf topography plays a significant role in determining the locations of surface features (Longhurst, 1998). Nelson and Hutchings (1983) suggest that upwelling occurs preferentially where the continental shelf is narrowest such as off salient capes. Equatorward flow is carried in a series of jet currents and eddies associated with the coastal and shelf topography. A prominent front is frequently observed near the shelf break where upwelled water sinks on the shoreward side. Beyond the shelf break front an eddy field marks the edge of the clear waters of the subtropical gyre of the South Atlantic. Cold core eddies appear as isolated areas of chlorophyll enhancement along this frontal zone.

1.4.1 Local Forcing and Processes that effect Vertical Structure

The Benguela is one of the four major boundary current regions of the global ocean (Hill *et al.*, 1998) and is similar to those off California, Peru and North West Africa in the domination of coastal upwelling (Shannon, 1985). The oceanic conditions off the west coast of southern Africa are largely controlled by basin-scale, ocean-atmosphere interactions over the South Atlantic. In particular the seasonal migration of the South Atlantic Anticyclone interacting with the equatorial low-pressure belt of the intertropical convergence zone (ITCZ) in the north and the continental low-pressure cells over southern Africa in the east to control the southeasterly trade winds along the west coast of southern Africa. These winds drive the basin–scale circulation and the constituent Benguela Current regime. The upwelling is induced by equatorward winds that flow near-parallel to the coast causing offshore Ekman divergence of surface waters. The local wind is closely related to upwelling particularly in the correspondence between peak wind stress and major upwelling cells (Cape Frio, Lüderitz, Hondeklip Bay and Cape Columbine) (Hardman-Mountford *et al.*, 2003). Over these areas the seasonal variation is evident as the trade winds migrate north and south. Elevated levels of chlorophyll biomass are noted downstream of these cells. Between these upwelling cells there tends to be a local minimum in wind stress and offshore Ekman transport during the winter leading to strong

stabilization of the water column. These irregularities in associated currents (including upwelling filaments that extend offshore) are attributed to variations in the alongshore windstress, coastal geometry, and topography of the continental shelf. These filaments can merge neighbouring upwelling centres making it impossible to accurately estimate production in each individual cell from satellite imagery.

The biological response to upwelling differs significantly between the northern and southern Benguela regions. The predominant reason suggested by Chapman and Shannon (1985) is the short-pulsed forcing in the south relative to the more continuous forcing in the north. In the southern region chlorophyll-enhanced water follows the outline of the shelf edge where a well-defined thermal front separates the coastal waters with their high biomass from smaller stocks of phytoplankton typical of open ocean waters. A chlorophyll maximum band is found about mid-shelf (15-25 km offshore).

The upwelling cells of the Southern Benguela coincide with the summer winds when they are at their peak and furthest south. Strub *et al.* (1998) showed that north of 32°S winds are upwelling favourable and currents equatorward all year round but strongest in summer, whereas from Cape Columbine to the Cape Peninsula winds and currents are more seasonal. Hardman-Mountford *et al.* (2003) showed that between 16° and 22°S wind forcing is bimodal with the first peak in October-November and the second peak in March-April as noted by Shannon and Nelson (1996). Chlorophyll biomass is greatest during the summer upwelling season in the Southern Benguela but occurs during winter along the central Namibian coast (Hardman-Mountford *et al.*, 2003).

Pulses of upwelling (as indicated by sea-level change) propagate southeast along the west coast and continue along the south coast, beyond Port Elizabeth, at speeds of 5 to 8 $m.s^{-1}$. These wave like features are generated by winds that alternately increase and decrease upwelling. Because the winds are associated with atmospheric high- and low-pressure systems moving eastward past southern Africa the regions of enhanced upwelling also move around the Cape in concert with the wind systems (Mann & Lazier, 1996 see Nelson, 1992 and Jury & Brundrit, 1992). The structure and speed of the upwelling features are similar to coastally trapped waves (Gill, 1992). These are internal Kelvin waves on the horizontal scales of 30-60 km rather than 1000 km or so of waves associated with the tides.

These coastally trapped waves propagating south along the west coast of southern Africa can enhance or diminish the upwelling events. For example, at the crests of the waves the deep, nutrient rich water is raised nearer than normal to the surface, which leads to more productive upwelling. At the wave troughs the opposite occurs as the nutrient rich water is pushed deeper that normal. The net result of this complex system of interactions is that upwelling events with duration close to the optimal 10-day period of the coastal trapped waves occur with some regularity in the Benguela upwelling system.

1.4.2 Horizontal Chlorophyll Distribution

Barlow *et al.* (2001) investigated the distribution of phytoplankton and the associated absorption characteristics in the southern Benguela ecosystem in the winter of 1999. Chlorophyll-a concentrations ranged from 0.3 to 18.5 $mg.m^{-3}$ at the surface and diatoms were the most abundant phytoplankton inshore. Offshore localities indicated nanoflagellates and cyanobacteria to be more dominant. A mixed population of diatoms and small flagellates was prevalent on the shelf in South African waters, while diatoms dominated the Namibian shelf zone. There was a poor relationship between measured chlorophyll and SeaWiFs chlorophyll with only 55% of the variance being explained. The regression line indicated that SeaWiFs overestimated chlorophyll-a for concentrations <2 $mg.m^{-3}$ and underestimated concentrations that were >2 $mg.m^{-3}$. This was probably attributed to the poor atmospheric conditions and the failure of the satellite derived atmospheric correction algorithm. In the southern Benguela chlorophyll concentrations typically reach 5-30 $mg.m^{-3}$ in aged upwelled water (Barlow *et al.*, 2001).

Mitchell-Innes *et al.* (2001) described the characteristic winter profiles of four regions within the Benguela system using a shifted-Gaussian curve. Each region was defined hydrologically. The region can be subdivided as the shelf area <200m depth and the slope 200 – 500m. The use of a Gaussian fitted distribution was generally good but poor where chlorophyll concentrations were low (1-2 $mg.m^{-3}$). The profiles generally changed from inshore well-defined surface or near-surface peaks to progressively weaker and deeper offshore. The height of the peak declined offshore as well as getting deeper whereas the spread of the peak increased. The Namaqua region had a mean peak at about 10 m and a value of approximately 4.5 $mg.m^{-3}$ on the shelf and 25 m and 1.5 $mg.m^{-3}$ offshore.

1.4.3 Recent Predictions of Chlorophyll Profiles in the Benguela

A review of the biogeography by Platt and Sathyendranath (1999) suggests that the variables in each region for the prediction of primary production should be either fixed seasonally or continuous and predictable. This is most certainly not appropriate in the highly dynamic pulsed upwelling of the Benguela (Silulwane, 2001: Richardson *et al.*, 2002). Silulwane *et al.* (2001) and Richardson *et al.* (2002) used Self Organising Maps (SOMs), a type of artificial neural network, to determine characteristic profile classes from a database of Gaussian curve parameters. The database was created by fitting the curve to an archive of *in situ* profiles. The profile class could then be predicted from their relationship between the Gaussian curve parameters and a suite of easily measurable environmental variables including satellite data (SST and surface chlorophyll). The approach however, is only semi-quantitative as it does not indicate the strength of the relationships and it is still not a continuous parameterisation of profiles (Richardson *et. al.*, 2002). Furthermore this method requires that profiles in the raw chlorophyll data that do not fit a Gaussian model be removed prior to analysis (e.g. ~15% in Silulwane *et al.*, 2001; Longhurst *et al.*, 1995).

Vertical profiles have been studied and frequently explained in terms of a 4 parameter Gaussian curve (Platt *et al* 1998) with Taguchi *et al.* (1994) explaining more than 85% of the variability in the vertical profile. Their model was successfully able to predict three parameters of the Gaussian curve: background chl-a concentration; depth of the subsurface maximum; and the total biomass above the background concentration. They noted that low integrated chlorophyll (<300 $mg.m^{-2}$) produced much variability in the profiles. The non-linear relationships between the Gaussian model and environmental conditions were then parameterised using generalised linear models (GLMs) to obtain a predictive equation for each profile parameter.

Richardson *et al.* (2003) took the process of dynamic profiles further for the region by relating the four parameters of a shifted Gaussian model to real-time satellite-derived information (surface chlorophyll-a concentration and SST) and known inputs such as season, area and depth of the water column. The authors were able to predict well the depth of the chlorophyll maximum and the height of the peak above the background chlorophyll but were not successful in determining significant relationships for the background

chlorophyll and width of the peak. An important forcing variable that was omitted was surface winds which play an important role in the definition of many of the biomes presented by Longhurst *et al.* (1995) and is the primary forcing of upwelling in the Benguela system (Shannon, 1985; Pitcher *et al.*, 1992). Demarcq *et al.* (*in press*) used a series of 15 SOM-derived profile categories based on raw chlorophyll profile data and a simple light model to estimate the primary production in the Benguela and Agulhas systems. The authors used satellite derived surface chlorophyll concentration and SST and known variables (season, locality and depth of the water column) to predict the profile number (1 to 15). This was performed using a two-step generalized modelling approach. The first step involved ascertaining the relationship between the surface observations and each of the 15 profiles using Generalized Additive Models (GAMs) and the predictor variables of surface chlorophyll, SST, depth, season and locality. Locality was categorized as follows: West Coast; western Agulhas Bank; and eastern Agulhas Bank. Depth was used as a proxy for distance from the shore. The second step involved assessing the relationship between profile number and each environmental predictor by inspecting the GAM plots and then parameterizing them using piecewise linear regression or exponential transformation. These parameterizations were implemented as a Generalized Linear Model (GLM) to predict the profile type from the satellite data.

The aim of this paper is to take the process of quantitatively predicting the subsurface chlorophyll distribution further by including wind data in the suite of remotely accessed environmental data. The inclusion of wind data is evaluated in conjunction with SST and surface chlorophyll concentration data by relating them to an archive of *in situ* chlorophyll-a profiles. The data will be integrated into a database for training a Dynamic Bayesian Network. The original aim was to train the DBN and test its predicting ability using surface data not used in the training process but this was not possible owing to time constraints on the developers of the DBN software.

Chapter 2: Methods

This section summarises the chlorophyll profile prediction stage of a prototype system currently named the Plankton Prediction System (PPS). The goal of this system is to predict the vertical distribution of phytoplankton biomass for each pixel in a regional map for the calculation of real-time primary production. A Dynamic Bayesian Network (DBN) will be used to predict sub-surface chlorophyll profiles from remotely sensed data. The DBN is trained on available archives of co-incident satellite data and *in situ* profile data. The PPS has the following objectives:

To provide;

- A means of processing satellite data for display and inference purposes
- A method of clustering profile data into groups
- A Dynamic Bayesian Network to infer chlorophyll profiles over time, incorporating temporal relationships between profiles
- A method of visualizing the satellite data as well as the inferred profiles
- A method of visualizing the Dynamic Bayesian Network

The PPS can be separated into three preliminary modules; *pre-processing*; *training* and *testing*. The first stage is responsible for the integration of all the raw data from a number of different sources. After integration, the data is further discretized through a clustering process, which reduces complexity. The second stage, *training*, is responsible for training a DBN with the clustered data produced in the *pre-processing* stage. The *testing* will evaluate the predictions of the system using data not used in the *training* stage. As the DBN has not been constructed at the time of writing only the *pre-processing* stage will be evaluated in this paper.

2.1 Prediction System

2.1.1 Bayesian Networks

Bayesian networks are probabilistic graphical models that represent a set of variables and their causal influences or probabilistic dependencies using Bayes rule. In the model the variables are represented by nodes. Arcs represent the causal or probabilistic dependencies between the nodes given the value of its parents (if there is an arc from node A to another node B then A is the parent of B as B depends directly on A). The strength of the

dependencies is expressed in a Conditional Probability Table (CPT) attached to each node. This gives the probability distribution over the values of the node. Both the parameters and the structure of the network are obtained using learning algorithms. In this case the nodes are not all observed (SST, surface chlorophyll, surface wind, temperature and chlorophyll profiles, area and season) but the structure is specified intuitively due to the causal relationships between nodes. Because of this the Expectation Maximization algorithm can be used as opposed to the Maximum Likelihood Estimation algorithm (when the structure all the nodes is known). Dynamic Bayesian Networks model how the state or probability distribution of the variables changes at each time step although the structure of the network does not change. In short, they are used to predict the future based on time-series data from the past and the present.

2.1.2 Dynamic Bayesian Networks

Dynamic Bayesian Networks (DBNs), a close relative of Bayesian Networks (BNs), have a number of advantages over BNs. In addition to being well suited to handling time series data, this framework can handle missing data in a principled way as well as model stochasticity, prior knowledge and hidden variables (Murphy and Mian, 1999). Hidden Markov Models (HMMs) and Kalman Filter Models (KFMs) are two other techniques which have been used with some success in areas ranging from speech and gesture recognition, to tracking planes and predicting the economy. Both of these lack expressive power. An HMM is a temporal probabilistic model in which a single discrete random variable describes the state of the process, with the possible values of the variable being the possible states of the system. The DBN overcomes this limitation in expressive power by allowing for multiple random variables to describe the possible states of the system. The KFM is a way of estimating the state of a dynamic system based on a series of noisy measurements, which is the case in most real world applications. KFMs use a simple unimodal linear-Gaussian probability distribution, whereas DBNs can make use of an arbitrary probability distribution.

A Dynamic Bayesian Network is defined as a pair of Bayes nets, $(B_1, B_{\rightarrow})$, where the first member of the pair defines the prior $P(Z_1)$, and the second member is a two slice temporal Bayes net which defines $P(Z_t|Z_{t-1})$ by means of a directed acyclic graph (DAG) as follows:

$$P(Z_t|Z_{t-1}) = \prod_{i=1}^{N} P(Z_t^i|Pa(Z_t^i))$$

where Z_t^i is the i'th node at time t, and Pa(Z_t^i) are the parents of Z_t^i in the graph. The nodes in the first slice do not have any parameters associated with them, but each node in the second slice has an associated conditional probability distribution (CPD) which defines $P(Z_t^i|Pa(Z_t^i))$ for all t > 1. The parameters of the CPDs are assumed to be unchanging over time (Murphy, 2002). To determine changing probabilities over time the resulting DBN can be "unrolled" until there are the desired number of time slices, T. The resulting joint probability distribution is then given by:

$$P(Z_{1:T}) = \prod_{t=1}^{T}\prod_{i=1}^{N} P(Z_t^i|Pa(Z_t^i))$$

The Expectation Maximization (EM) algorithm was to be used to learn the parameters of the network as not all nodes in the network are fully observable. However, due to time constraints on the developers of the software code for the DBN was not constructed in time to analyse the present data.

2.2 Pre-processing

2.2.1 In situ *data*

Shipboard oceanographic data were collected between 1988 and 2006 during fisheries surveys off the west and south coasts of South Africa by Marine and Coastal Management, Department of Environmental Affairs and Tourism, South Africa. Generally, each transect was perpendicular to the coast and had stations 10 miles apart, starting two miles from the coast and extended to the shelf edge, although some transects were extended further offshore (Figure 2.1). Each transect was completed within a 12 hour period. Fluorescence and temperature profiles were measured at each station down to 100m or within 10m of the seabed by a thermistor and profiling fluorometer (Chelsea Instruments Aqua Tracka MKIII). The total number of usable fluorescence profiles was 5813 from 49 cruises ranging from approximately 14°S 10°E off the west coast of southern Africa to 30°S 33°E along the southeast coast. The vast majority of profiles however were collected between the Orange River mouth and Port Elizabeth.

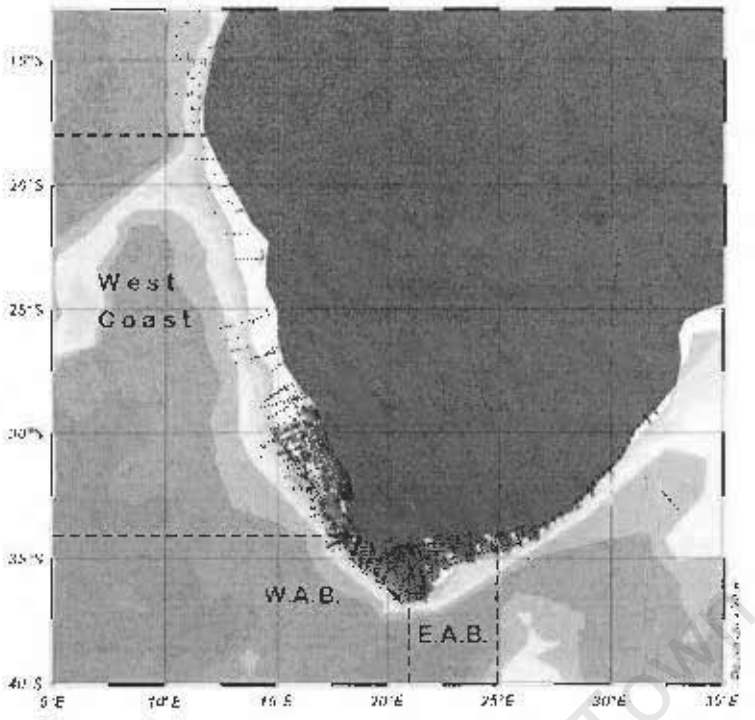

Fig. 2.1. Map of the study area showing station positions where the 5813 profiles were collected. The three sub-regions are demarcated by the 18°S and 34°S latitudes and the 21°E and 25°E meridians.

Chlorophyll-a samples were collected in Niskin bottles on a CTD rosette beginning a few meters below the surface and at intervals of about 10m down to between 30 and 50m depending on the profiling signature. The collected samples were filtered onto Whatmann GF/F filters and the chlorophyll extracted by placing the filters in 90% acetone at 20°C for 24 hours. The discrete chlorophyll samples were measured fluorometrically using a Turner design model 10-000R fluorometer before and after addition of hydrochloric acid to adjust for phaeopigments according to Parsons *et al.* (1984). The reverse-phase HPLC (high performance liquid chromatography) method described by Barlow (1997) was used for the HPLC samples.

Cullen (1982) discusses the inherent problems with using *in vivo* fluorescence as a measure of chlorophyll. Ideally the relationship between fluorescence and chlorophyll concentration should be linear but due to factors such as species composition, nutrition and ambient light this is not the case. Although little could be done regarding the first two factors in formulating a relationship between fluorescence and chlorophyll concentration, the effect of changes in light could be addressed. This was done by selecting data collected between late

evening and early morning only. Although this eliminated the need to remove data that exhibited photo-inhibition, applying the conversion formula derived from night-time data to day-time data can lead to an underestimation of chlorophyll concentration in the high light surface layer (Cullen and Lewis, 1995). A total of 1053 chlorophyll concentrations samples, comprising of 651 fluorometrically determined and 402 HPLC determined concentrations were regressed against their corresponding fluorescence voltages. The fluorescence voltage profiles were edited and binned into 1m intervals using the standard Sea-Bird software, SBE Data Processing Version 5.37m. The data were grouped according to season and then region in order to determine any significant differences among the mean regional and seasonal regression slopes.

2.2.2 Satellite Data

SST and Surface Chlorophyll Biomass

The use of SST in the prediction of other environmental parameters has been successfully applied in a variety of oceanographic regimes (Platt *et al.*, 1995). The reason for using SST for estimating these parameters in upwelling areas is easy to understand. For example SST can be used as an indicator of the age of upwelled water since upwelled water is cool and as upwelling water matures SST warms (Dugdale *et al.*, 1997). SST has also been correlated to nutrient concentrations (Waldron and Probyn, 1992). Temperature has also been closely related to P-I parameters, in part through the influence of physical forcing on phytoplankton community structure (Harrison and Platt, 1985; Macedo *et al.*, 2001).

There has been some correlation between surface chlorophyll concentration and P-I parameters although the relationship tends to be weak (Harrison and Platt, 1985). This can be explained by the poor relationship between chlorophyll concentration and the environmental conditions that regulate species succession and photo-physiology. Low biomass can occur during winter when winds are strong and light is limiting. It can also result during summer when nutrient levels are low and conditions are quiescent thus both SST and surface chlorophyll should be used to assign the photosynthetic parameters. The biomass in the near surface water which is accessible from remote sensors provides information on the state of the upwelling when used in conjunction with temperature.

Archives of SST and sea surface chlorophyll data from satellites dating from 2002 to the beginning of 2007 were used. The SST and surface chlorophyll data were retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor launched onboard

AQUA in May 2002, via the http://www.rsmarinesa.org website. The website generates pentad products from level 2 data at a 1 km resolution, as a 5 day-to-day average using default NASA ocean-colour processing.

Wind

Averaged longshore wind displacement is often invoked as a metric linearly related to productivity on timescales of months to seasons through its connection with upwelling, (Nelson, 1992). Nelson (1992) points out that upwelling is never uniform in the longshore direction but is enhanced by features such as coastal and submarine topography, such as upwelling capes and canyons where upwelling cells form. Upwelling dynamics respond to wind-forcing which is variable typically over 2-5 days producing the characteristic pulsed upwelling of the Benguela system. Level 3 mean wind-field data were retrieved from the NASA SeaWinds scatterometer onboard QuikSCAT via CERSAT, at IFREMER, Plouzané (France). The data were collected from July 1999 and comprised of weekly averaged wind fields over global 0.5° x 0.5° resolution geographical grids. The wind values were quantized to values of precision 1.4° to reduce the amount of space required for storage. The precision was judged to be sufficient for the PPS.

Season

The vertical distribution of chlorophyll is a seasonally varying structure that is closely related to the stratification of the upper layers. Stratification is closely related to solar heating and upwelling favourable winds, each seasonally dependent. Furthermore as mentioned above, low biomass in the surface waters may result from different environmental conditions according to season. Solar heating is also in phase according to latitude, and the regions in the northern Benguela will experience different heating to the southern region. Hence, all the satellite derived data were grouped according to season.

Region

Longhurst (1998, p182) delimits the *Benguela Current Coastal Province* between Cape Frio (18°S) and the Cape of Good Hope. In their paper which estimated phytoplankton biomass in the Benguela ecosystem, Brown *et al.* (1991) included the south coast of South Africa (and consequently the Agulhas Bank) extending the Benguela system to 27°E. They separated the South Coast from the West Coast by a line due west of Cape Point. Boyd and Shillington (1994) discuss the physical forcing and circulation patterns on the Agulhas Bank and limit the region between 18°E and 25°E. Probyn *et al.* (1994) review primary production on the Agulhas Bank and describe a central and eastern Agulhas Bank east of

roughly 21°E as apposed to a western sector. These references were used to divide the west and south coast data into: the West Coast between 18°S to 34°S; the Western Agulhas Bank (WAB) south of 34°S and east of 21°E; and the Eastern Agulhas Bank (EAB) between 21°E and 25°E. The West Coast is characterised by southeast trade winds which are responsible for year round coastal upwelling. Relaxation or reversals of these winds are regularly forced by the passage of cyclones to the south of the Cape Peninsula. More seasonal wind reversals are associated with the south coast where the mid-latitude westerlies prevail in winter and cause deep mixing, particularly on the western Agulhas shelf. Different hydrological conditions characterise the Western and Eastern Agulhas Bank primarily wind-driven coastal upwelling on the WAB and shelf-edge upwelling resulting from interactions with the Agulhas Current on the EAB.

*2.2.3 **Clustering***

The chlorophyll profiles were provided as averaged values at 1m intervals from 1m below the surface to 100m depth. In order to be able to cluster the profile data it had to be resampled to produce a continuous curve at precisely 1m depth intervals. This was done using *Akima interpolation* (Akima, 1970). Unlike its counterparts (e.g. linear-, polynomial-or cubic spline interpolation), *Akima interpolation* provides a curve with an unusually natural fit, which through experimentation has been the most akin to that preferred by the human eye. The profiles were each given a value according to the week within which the profile date fell (i.e. 1-52) and linked with their corresponding satellite-derived environmental variable values. Clustering is another method performed on all data to further reduce the overall dimensionality so that the DBN is able to process the data as efficiently as possible. In effect the clustering performed on all satellite data reduced the initial dimensionality of each variable to a set of predetermined intervals. For the profile data the clustering technique aimed to partition the *in situ* profiles into an optimal number of representable clusters. A number of algorithms adapted from the open-source *Cluster 3.0* program that represent each of the three broad categories *viz.* hierarchical, partition relocation (k-clustering) and machine learning (SOMs), were assessed previously using the r^2 goodness of fit method. Essentially this method divides the total variance of the profiles in the cluster from the cluster centroid by the total amount of variation from between the profiles. The overall r^2 value for each algorithm is the arithmetic mean of all the individual cluster values. Of 43 experimental configurations tested on 2412 profiles the k-clustering technique yielded the top 7 best results (r^2 ranging from 0.81 for k-means, to 0.75 for k-medians) with the next best technique, hybrid

clustering, being ranked 8th ($r^2 = 0.74$). SOMs and hierarchical clustering did not feature in the top 25. The k-means technique was chosen as the optimal clustering method.

As each profile now had an associated environmental variable interval and belonged to a specific cluster the relationships could be evaluated graphically. This was done by considering the frequency distributions of the profiles within each of the clusters by season and the interval distribution of each of the environmental variables within each cluster.

Chapter 3: Results

3.1 Ship Observations

3.1.1 Regression of Fluorescence versus Discrete Chlorophyll Samples

Profile and corresponding discrete log-chlorophyll data from a total of 651 night-time stations collected between January 2004 and May 2006 on 15 cruises were used in the regression of discrete log-chlorophyll concentrations against fluorescence voltages, and 402 samples were used in the HPLC regressions. Although it is reasonable to expect the fluorescence signal to vary according to a number of internal physiological and external environmental factors the resolution of such detail is unnecessary at the current stage of the project. The co-efficient of determination for the discrete cruise data indicates a good relationship ($r^2 = 0.73$, Figure 3.1a). The regression line of the data was compared to the HPLC regression (Figure 3.1b). The slope of the regression lines shows no significant difference at the 5% significance level (using the student T-test). The HPLC regression equation was chosen to convert the fluorescence profile data to chlorophyll concentrations rather than a pooled data set due to the better fit r^2 value ($r^2 = 0.82$) and the more reliable results obtained when using the HPLC methodology. The conversion algorithm was applied to the archive of *in situ* profile data and produced a total of 5813 usable profiles. These profiles were entered into the data base and clustered by the *Cluster 3.0* program.

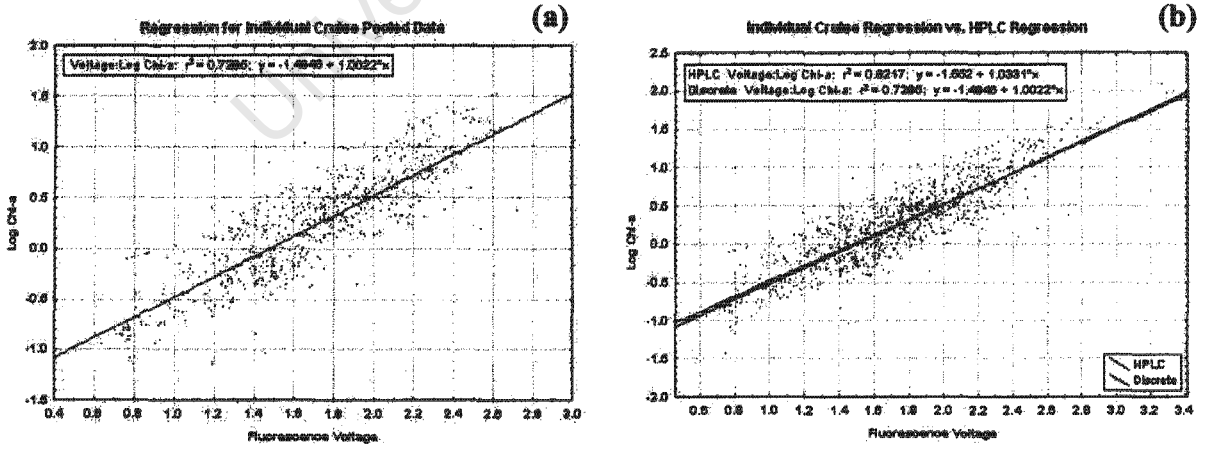

Fig. 3.1. Regression for (a) the pooled discrete chlorophyll data showing that almost 73% of the variance in log chlorophyll concentrations is explained by the fluorescence voltages and (b) the regressions of the pooled data and HPLC data. The HPLC data has a slightly better co-efficient of determination. There was no significant difference between the two slopes at the 5% significance level.

3.1.2 Profile Clusters

The cluster program produced 16 clusters which are arranged in order of increasing integrated chlorophyll-a concentrations ranging from 21.5 $mg.m^{-2}$ to 358.8 $mg.m^{-2}$ (Figure 3.2). The distributions of the profiles range from 148 profiles in cluster 16 to 604 in cluster 11. Although the clusters are arranged according to integrated chlorophyll concentrations, there is no clear order for any other profile attribute such as chlorophyll peak depth or surface concentration.

3.2 Satellite-Derived Surface Data

Satellite-derived surface chlorophyll values were associated with each profile in the data base. Because of the limitations of current satellite technology not every *in situ* sample could have concurrent satellite coverage and not all missing data could be interpolated. As a result there were fewer available profiles to be analysed in the sections dealing with the satellite-derived surface values.

3.2.1 Wind Speed and Direction (Figure 3.3)

Southeasterly winds blow along the West Coast throughout the year with the strongest average wind speed during summer and winter. In autumn these winds are at their lowest speed and during spring they are at there most variable. More westerly winds begin in autumn and reach a maximum occurrence in winter. Northwesterly winds rival the southeasterlies in wind speed during this season whereas the southwesterly winds are notable weaker. Comparatively the Agulhas Bank region indicates much more variability. Southeasterly winds are seasonal in this region and are associated with spring and summer. Westerly to northwesterly winds dominate the autumn and winter seasons over the Agulhas Bank but are stronger on the eastern side.

3.2.2 SST (Figure 3.4)

The West Coast has much cooler surface water for all seasons compared to the West and East Agulhas Bank whereas the EAB has the warmest. Summer temperatures on the WAB and EAB are almost consistently >18°C and only on the EAB do temperatures not drop <16°C. The WAB and EAB show coolest temperatures in winter while the West Coast has coolest temperatures in autumn with no profiles having surface temperatures >18°C.

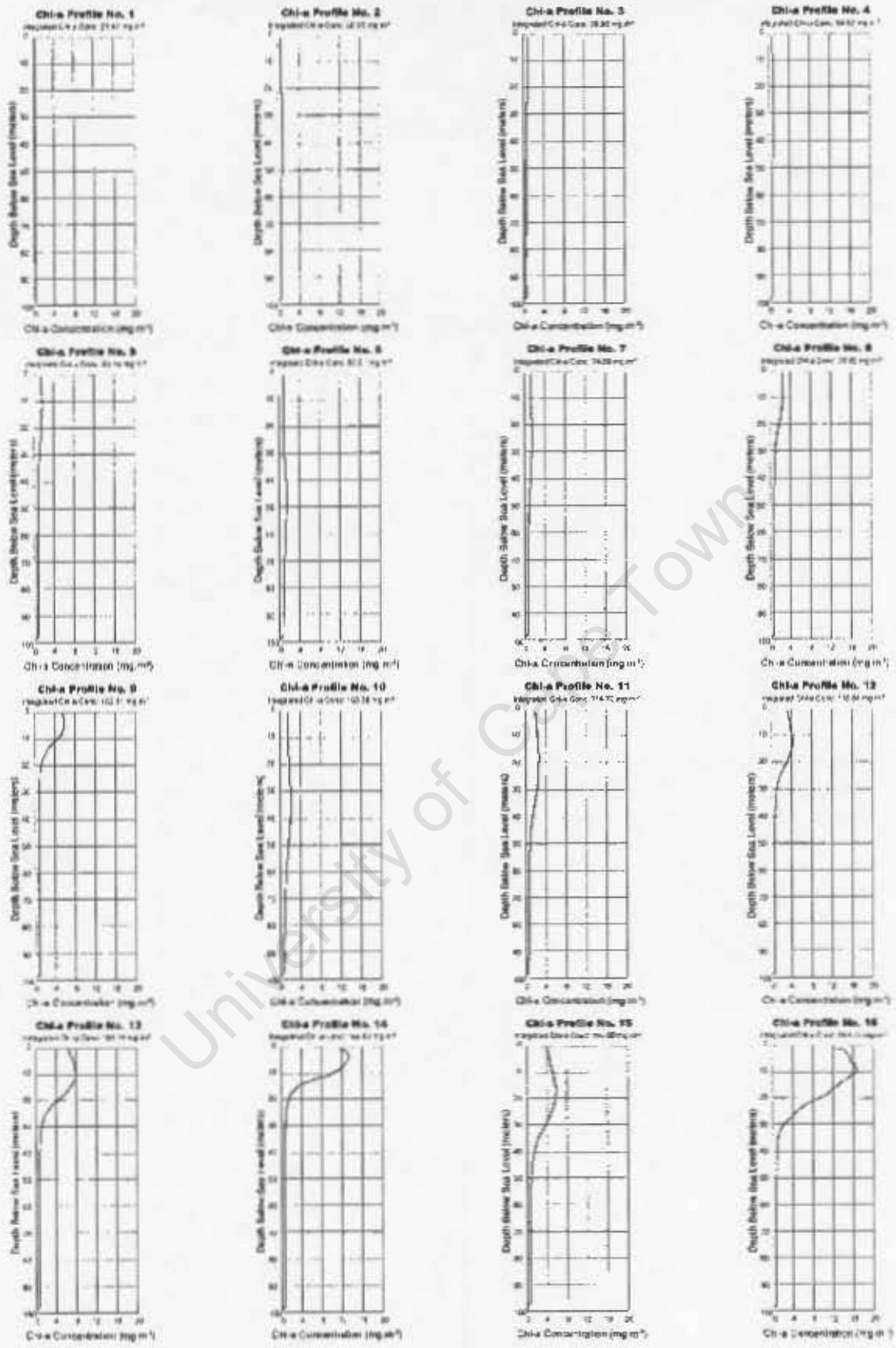

Fig. 3.2. The results of the categorisation of the 5813 profiles into 16 clusters produced by the cluster program. The profiles are arranged according to increasing integrated chlorophyll concentrations only

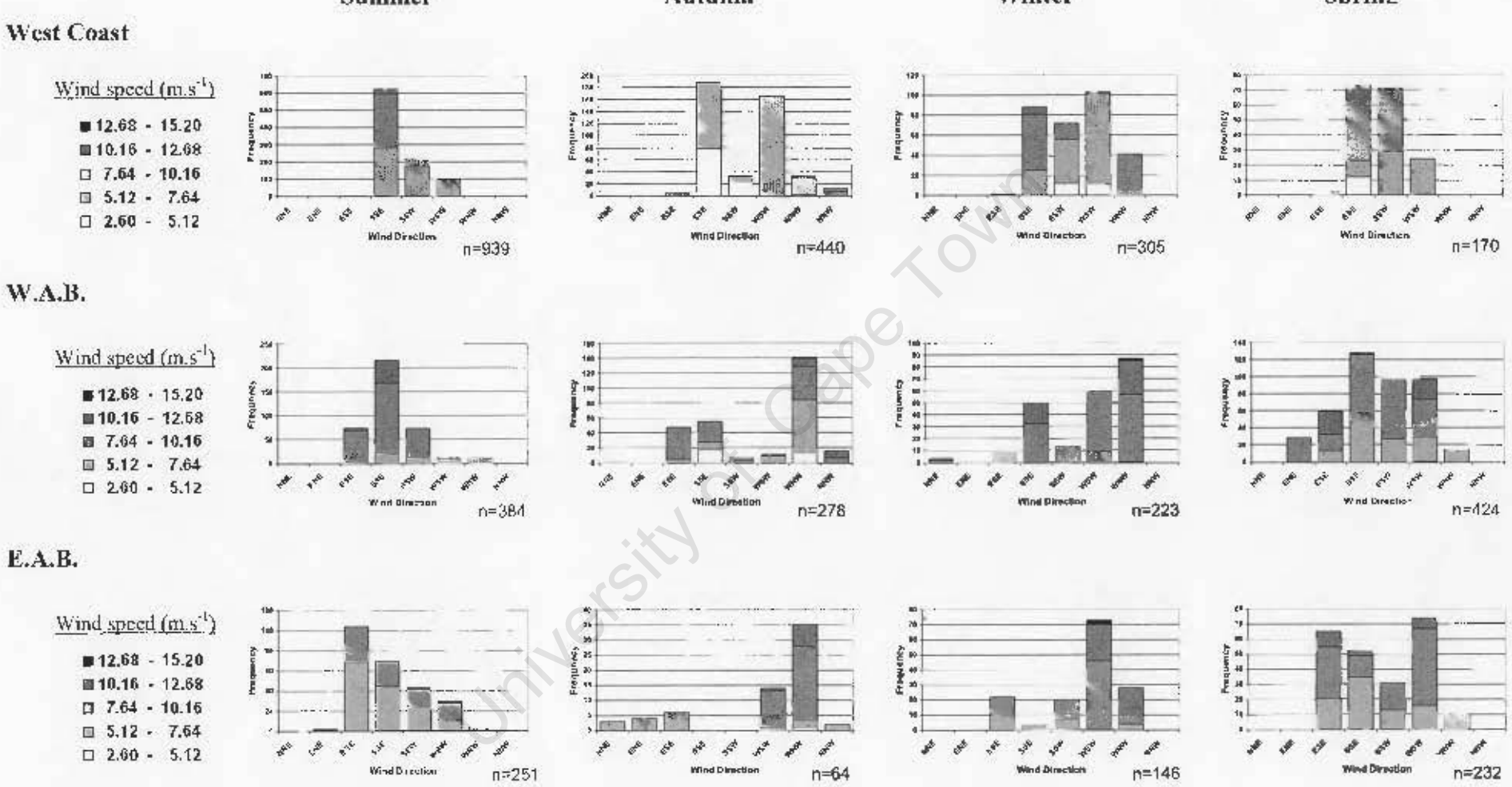

Fig. 3.3. An analysis of seasonal satellite-derived wind data for each region.

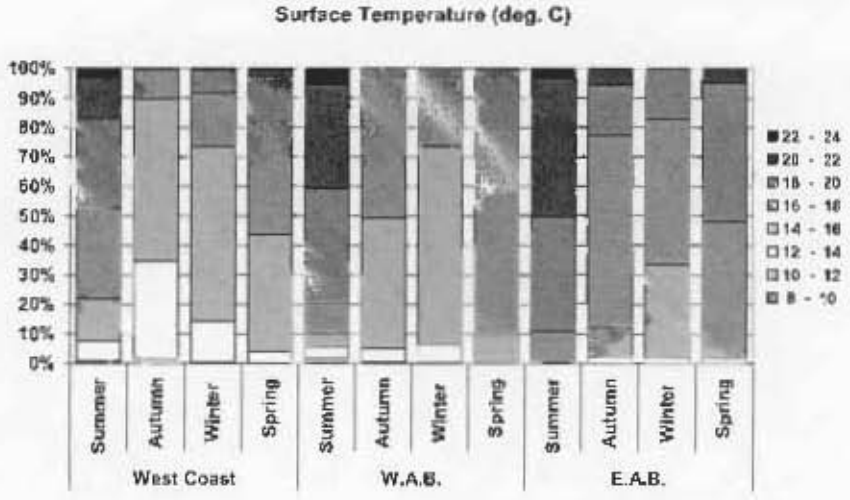

Fig. 3.4. Seasonal proportions of SST intervals for each region (see Figs. 3.8., 3.12. and 3.16. for n counts for each season).

3.2.3 Surface Chlorophyll (Figure 3.5)

For all three regions summer and autumn have higher average surface chlorophyll concentrations than winter and spring. For the West Coast the seasons have marked differences with autumn having the highest average and winter the lowest where as there is less distinction among the seasons of the other two regions. The WAB is dominated by the interval 0.29 to 0.86 $mg.m^{-3}$ whereas the EAB is bimodal with the interval above. The West Coast shows more variation in surface chlorophyll with two modal classes in summer and autumn (0.86-2.60 and 2.60-7.85 $mg.m^{-3}$) and three in winter and spring (including 0.29-0.86 $mg.m^{-3}$).

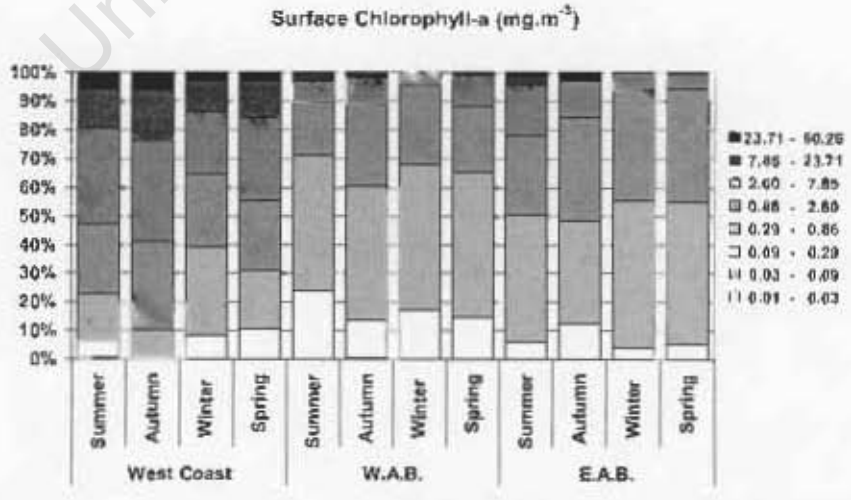

Fig. 3.5. Seasonal proportions of sea surface chl-a concentration intervals data for each region (see Figs. 3.9., 3.13. and 3.17. for regional and seasonal n counts).

3.3 Relating Ships Observations to Satellite Surface Data

3.3.1 West Coast

Wind Direction (Figure 3.6a-d)

In summer when the southeasterly winds prevail almost all clusters are well represented. Although no cluster is associated with a single wind direction category, some patterns are evident. Cluster 7 and 10 have a large proportion of profile associated with SSW wind and 9, 14 and 16 have relative large proportions associated with WSW wind. In autumn the most frequented clusters contain similar proportions of the seasonal wind direction variability with the exception of 7, 10 and 14 which are dominated by westerly winds. There are much fewer samples representing the winter and spring season however cluster 6 profiles rarely occur during westerly winds and cluster 14 and 16 profiles rarely occur during southeasterly winds.

Wind Speed (Figure 3.7a-d)

The stronger summer winds occur most frequently with profiles in cluster 6 and to a lesser extend clusters 4 and 10. Clusters 9 and 14 are mostly associated with intermediate wind speeds. The lower autumn wind speeds constitute the highest proportion of profiles in cluster 5 and cluster 16 in contrast to cluster 4, 7 and 10. During the winter clusters 6, 14 and 15 have no low wind speeds which is also seen in cluster 6, 7, 11, 15 and 16 in spring.

Sea Surface Temperature (Figure 3.8a-d)

Temperatures >18°C are associated with all the summer profiles but only for clusters 4, 6, 7, 10, 11 do they constitute the majority of profiles. Associated SST of profiles in clusters 8, 9, 12, 13 and 14 are mostly <18°C and include the majority of profiles with temperatures in the range 12 to 14°C. In contrast the autumn profiles have few surface temperatures >16°C and hardly any >18°C. Surface temperatures are almost completely confined to the range 12°C to 16°C. In the winter season profiles are largely confined to the 14°C to 16°C range but there is a greater proportion >16°C than in autumn particularly in clusters 7. In spring almost all the clusters have profiles confined to the range 14°C to 20°C. Clusters 11 and 15 have a large proportion of temperatures >16°C.

Surface Chlorophyll (Figure 3.9a-d)

The summer period is dominated by high surface chlorophyll concentrations. Surface chlorophyll concentrations >2.60 $mg.m^{-3}$ are associated with almost all the profiles in

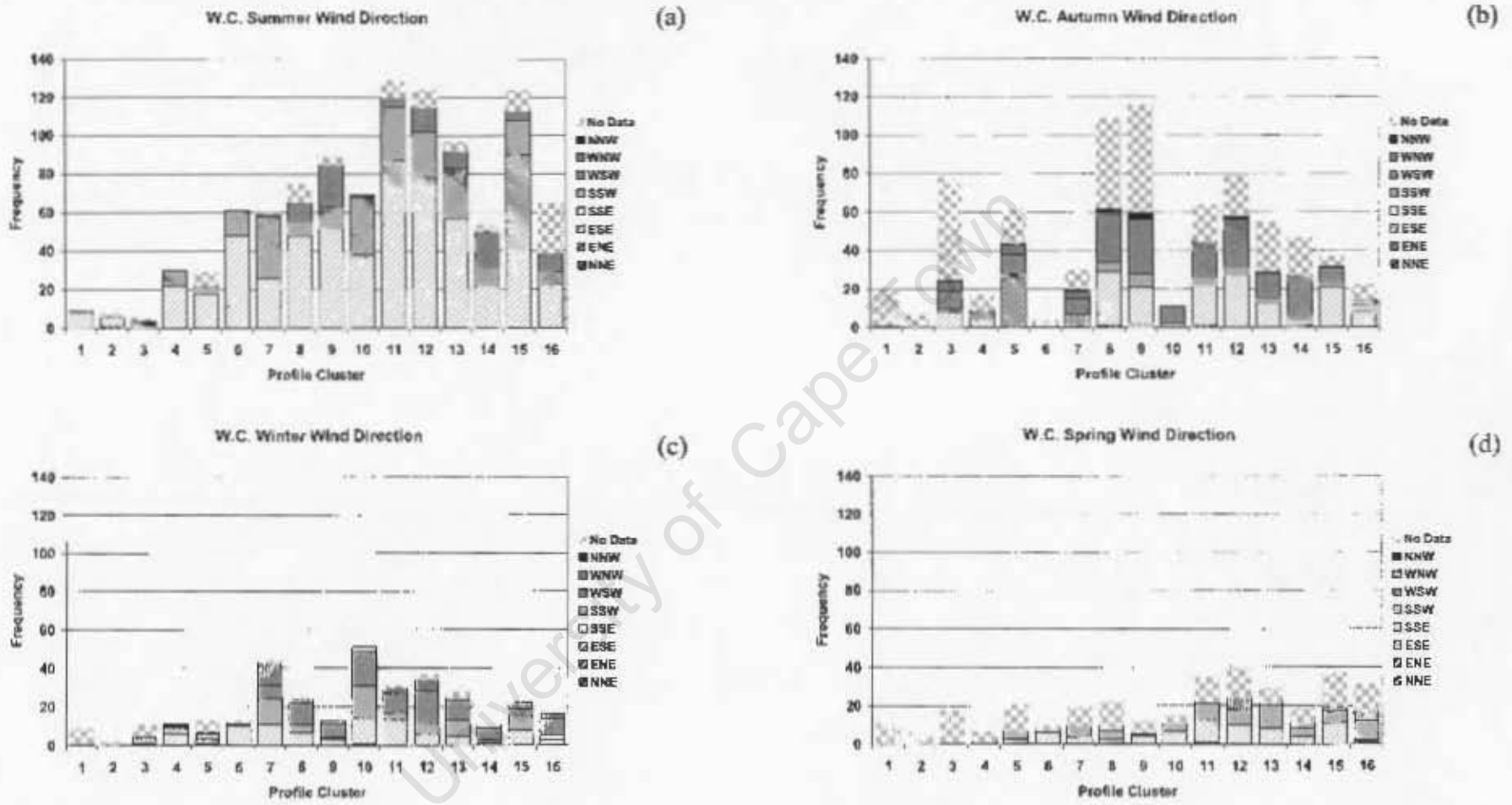

Fig. 3.6. Frequency distributions of profiles and their associated surface wind direction data intervals for each season: (a) summer (n = 939), (b) autumn (n = 440), (c) winter (n = 305) and (d) spring (n = 170). Including profiles with no surface data gives the total number of profiles in each cluster for the season.

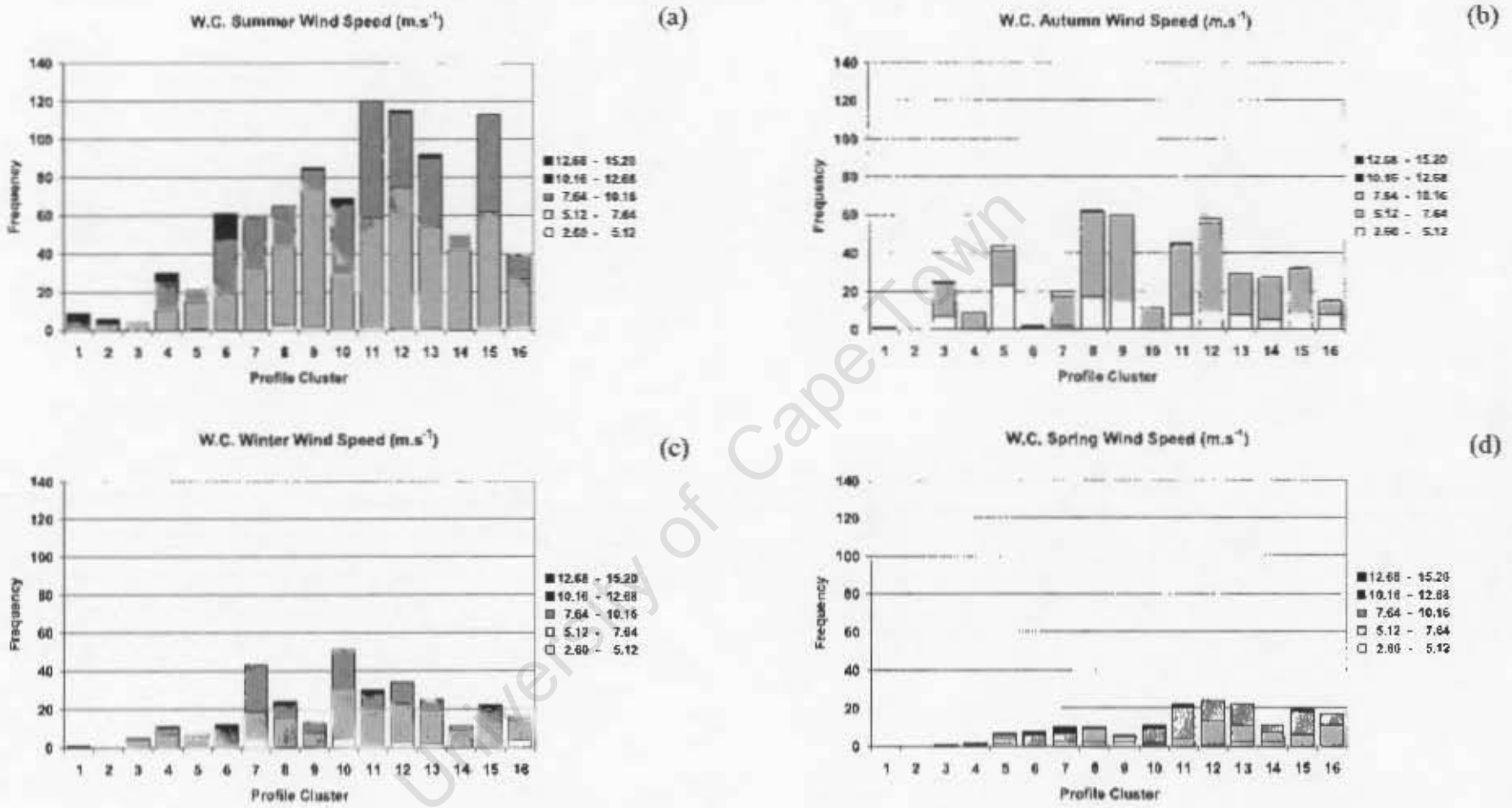

Fig. 3.7. Frequency distributions of profiles and their associated surface wind speed data intervals for each season; (a) summer (n = 939), (b) autumn (n = 440), (c) winter (n = 305) and (d) spring (n = 170).

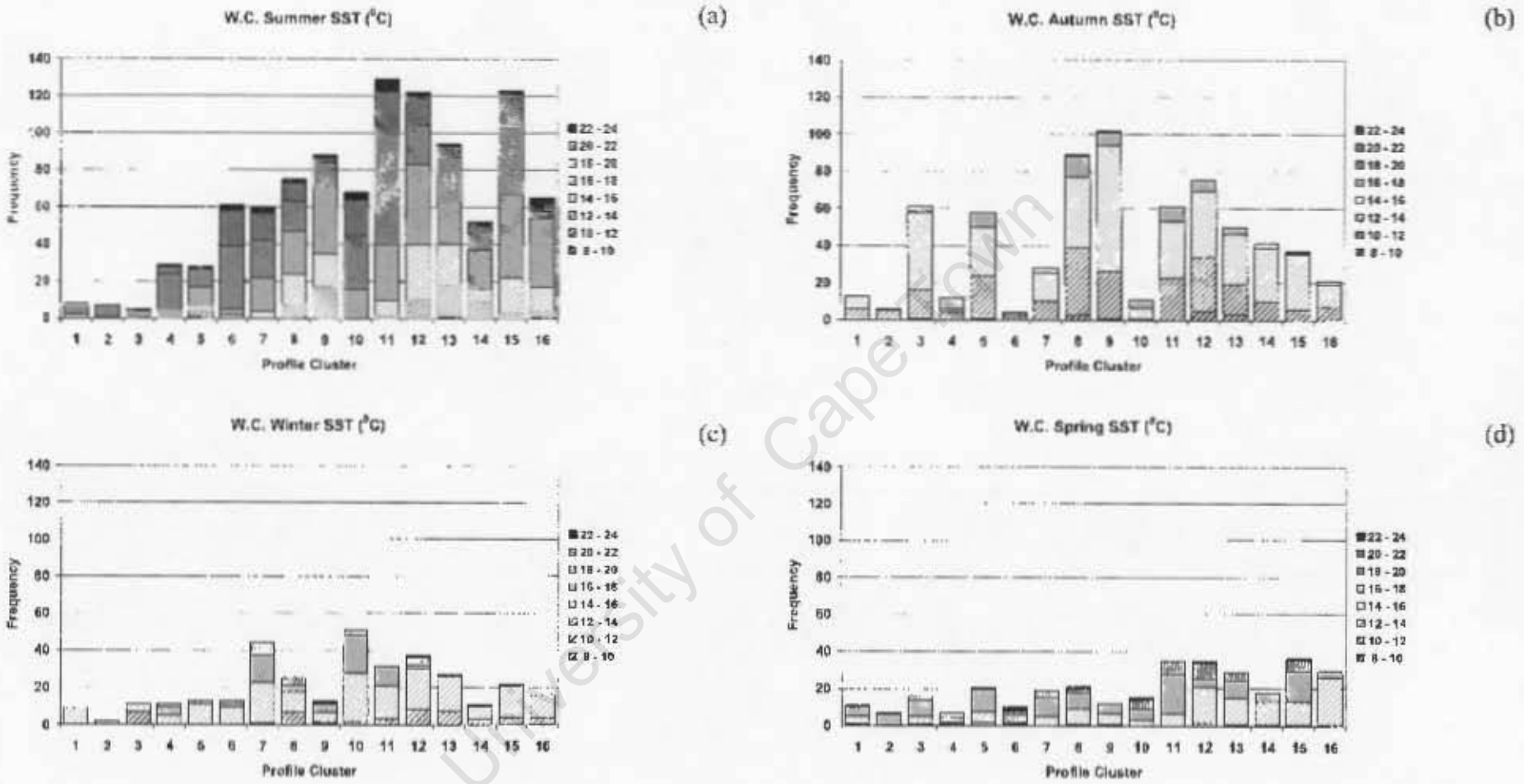

Fig. 3.8. Frequency distributions of profiles and their associated surface temperature data intervals for each season; (a) summer (n = 1015), (b) autumn (n = 669), (c) winter (n = 336) and (d) spring (n = 321).

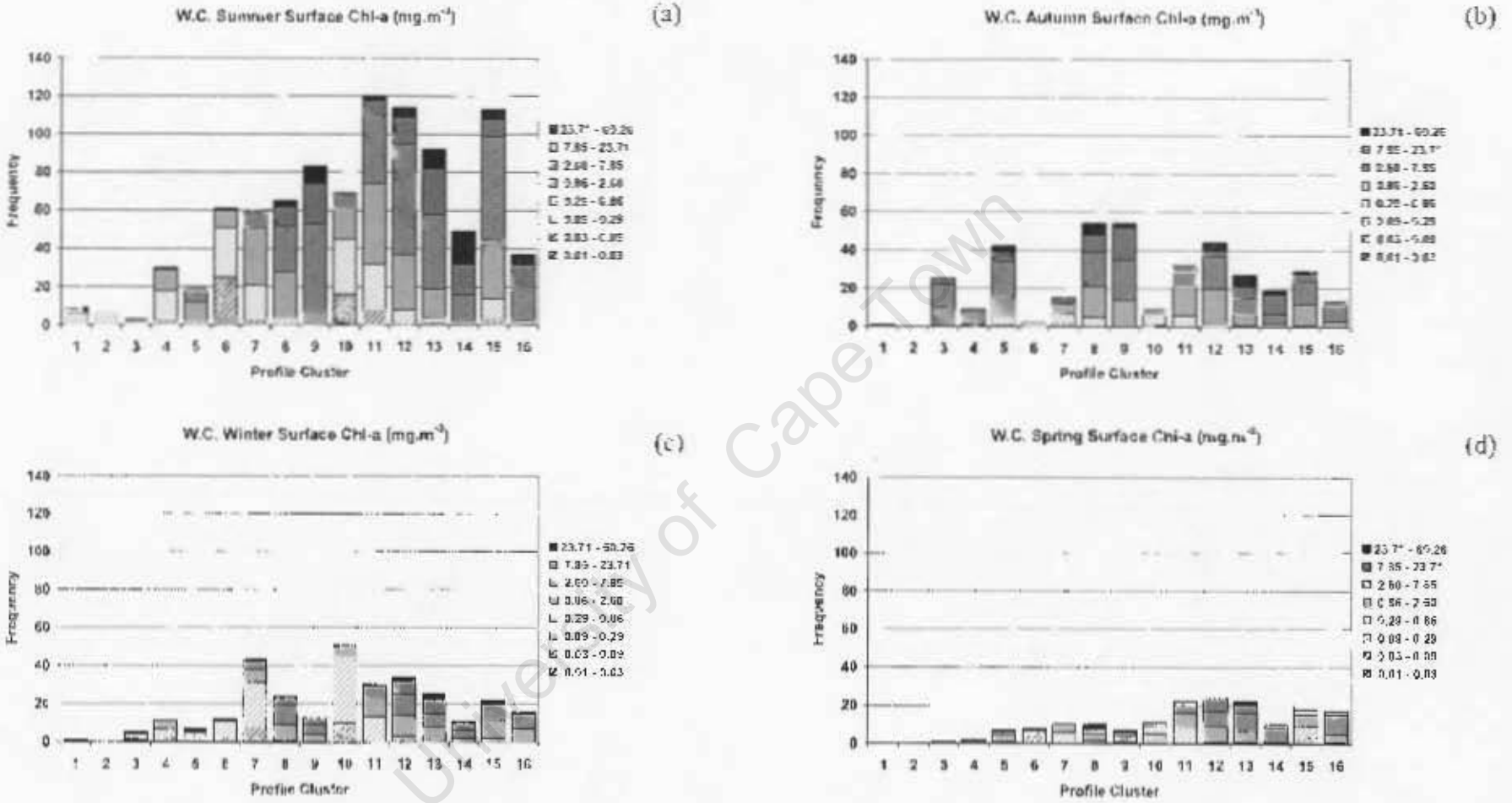

Fig. 3.9 Frequency distributions of profiles and their associated surface chlorophyll data intervals for each season; (a) summer (n = 932), (b) autumn (n = 397), (c) winter (n = 305) and (d) spring (n = 171).

clusters 9, 14 and 16 and to a lesser extent in clusters 8, 12, 13 and 15. Approximately one third of the profiles in cluster 14 have surface values >23.71 $mg.m^{-3}$. Clusters 6, 7 and 10 are almost entirely associated with surface values <2.60 $mg.m^{-3}$. During autumn the most common profiles belong to clusters 3, 5, 8, 9, and 11 to 14. Values >2.60 $mg.m^{-3}$ can be seen to dominate the more common profiles. Of the more frequent profiles in winter only clusters 8, 12 and 13 have a majority proportion >2.60 $mg.m^{-3}$. Clusters 7 and 10 are the most common winter profile clusters and are almost entirely comprised of surface values <2.60 $mg.m^{-3}$. Spring also has a high frequency of profiles with missing surface data. The profile distribution shows a similar trend to summer in that the profiles are concentrated in the higher integrated chlorophyll clusters. The surface chlorophyll distribution within each cluster also resembles summer.

3.3.2 West Agulhas Bank

Wind Direction (Figure 3.10a-d)

On the WAB summer profiles in cluster 2 are almost exclusively associated with ESE winds. Profiles in clusters 9, 11 and 13 rarely occur during westerly winds whereas profiles in clusters 4, 7 and 8 do relatively often. Profiles in cluster 3 to 5 and 10 are mostly associated with the northwesterly winds in autumn although they do occur during southeasterlies. Profiles in clusters 7, 8 and 11 occur during both these predominant wind directions. In winter profile clusters 7, 10 and 11 are closely coupled with westerly to northwesterly wind. Only profiles in cluster 7 and 10 occur during northeasterly winds. Of the dominant spring clusters only profiles in cluster 4 were observed in association with northwesterly wind. Profiles in clusters 10 and 11 very seldom occurred during northeasterlies. Southeasterly winds favoured profiles in cluster 6 whereas southwesterlies favoured clusters 10 and 11.

Wind Speed (Figure 3.11a-d)

In summer clusters 5 to 8 and 10 have similar proportions of wind speed categories that are predominantly of medium speed. Cluster 11 has more variety with higher proportions of high and low wind speed categories. Cluster 1 almost consists entirely of high wind speeds. Autumn indicates a good association between wind speed and profiles where clusters 3 to 5 have higher wind speed and 10 and 11 have lower. Cluster 7 and 8 seem to indicate a transition. There is also a trend evident between predominant clusters in winter. Cluster 7 has an intermediate wind speed average between the higher and lower average of

cluster 10 and 11 respectively. In spring lower wind speeds are associated with profiles that fall into clusters 4 and 6 while wind speeds in clusters 2, 7 and 11 are higher. Clusters 8, 10 and 15 have very similar proportions.

Sea Surface Temperature (Figure 3.12a-d)

The vast majority of summer profiles have surface temperatures >18°C. Only profiles in clusters 1 and 11 have temperatures >22°C. Cooler water (<14°C) is evident in clusters 5 and 8. Autumn temperatures similarly have profile surface temperatures consistently in the range 14°C to 18°C. During winter there is a bit more variance among the clusters. Temperatures >16°C occur substantially less often particularly in cluster 7 and to a lesser extend in clusters 10 and 11. Again in spring the profiles are largely confined to two intervals, *viz.* 16°C to 18°C and 18°C to 20°C but more profiles fall into the former intervals with the exception of. cluster 6 which has almost equal frequencies in each of the two intervals.

Surface Chlorophyll (Figure 3.13a-d)

The most frequent profiles on the WAB during summer are categorised into clusters 6 and 7. Cluster 6 is dominated by profiles with surface chlorophyll values ranging from 0.09 to 0.29 $mg.m^{-3}$ and cluster 7 by the interval above. Only a few profiles have surface values >2.60 $mg.m^{-3}$. Clusters 3, 9 and 13 to 16 are rarely represented. Autumn has the majority of profiles in clusters 3 to 5, 7, 8, 10 and 11. Clusters 5 and 7 have almost all their profiles associated with surface temperatures ranging between 0.09 and 2.60 $mg.m^{-3}$ whereas clusters 8, 10 and 11 have few if any <0.29 $mg.m^{-3}$. Over the winter season only clusters 7, 10 and 11 have 30 or more profiles while the remainder have less than 20. These three clusters have increasing proportions of interval 0.86 to 2.60 $mg.m^{-3}$ and decreasing proportions of interval 0.09 to 0.29 $mg.m^{-3}$ respectively. The spring season shows a distinct dominance of 6 clusters, *viz.* 2, 4, 6, 7 10 and 11. Clusters 2 and 4 have 91% and 65% of their profiles with no data but showed the majority of profiles with intervals 0.29 to 0.86 $mg.m^{-3}$ as does cluster 10. Profiles in cluster 6 are split between this interval and the lower one. Cluster 11 has a higher proportion of interval 0.86 to 2.60 $mg.m^{-3}$ than 0.29 to 0.86 $mg.m^{-3}$.

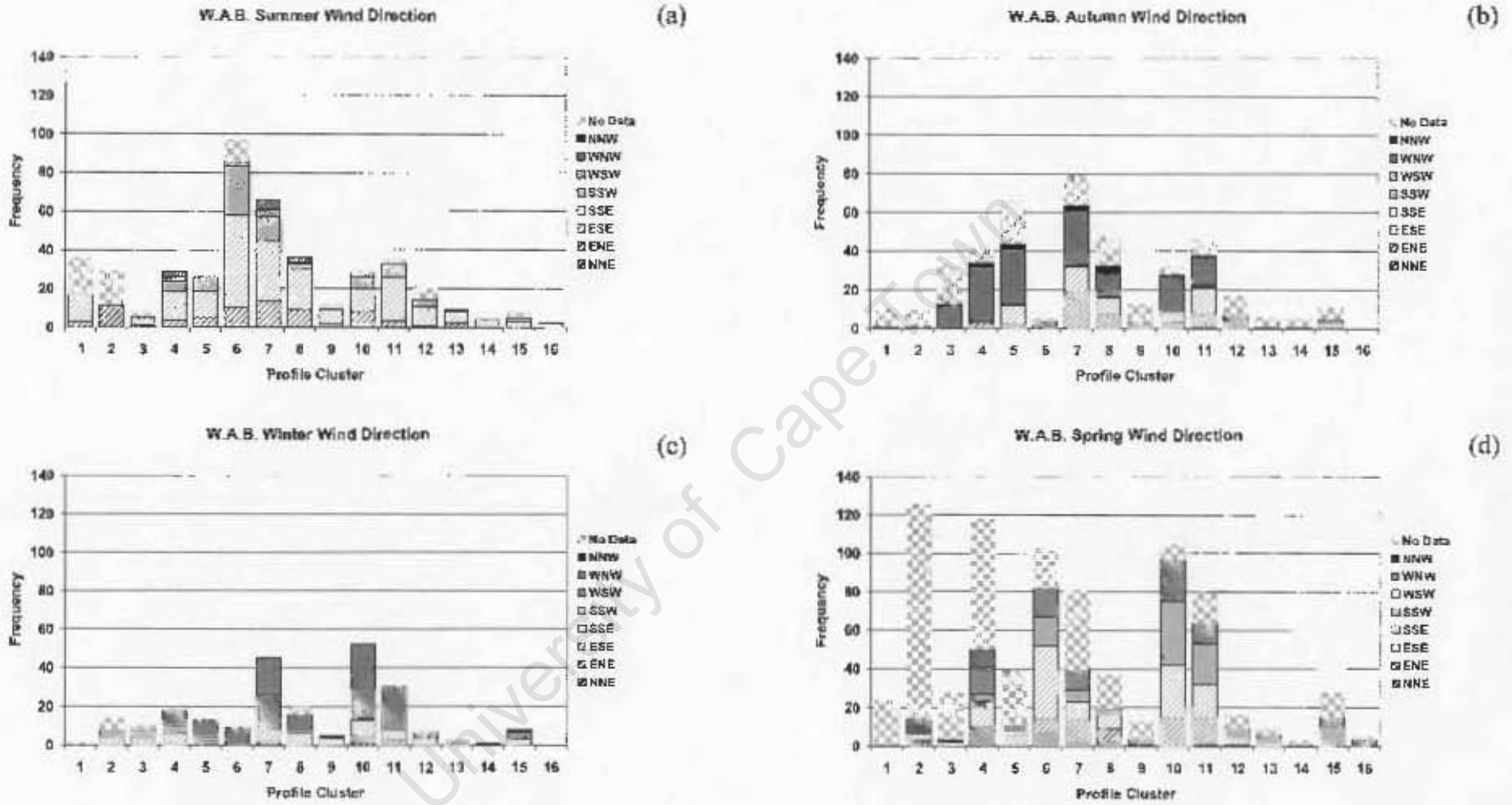

Fig. 3.10. Frequency distributions of profiles and their associated surface wind direction data intervals for each season; (a) summer (n = 385), (b) autumn (n = 278), (c) winter (n = 223) and (d) spring (n = 424). Including profiles with no surface data gives the total number of profiles in each cluster for the season.

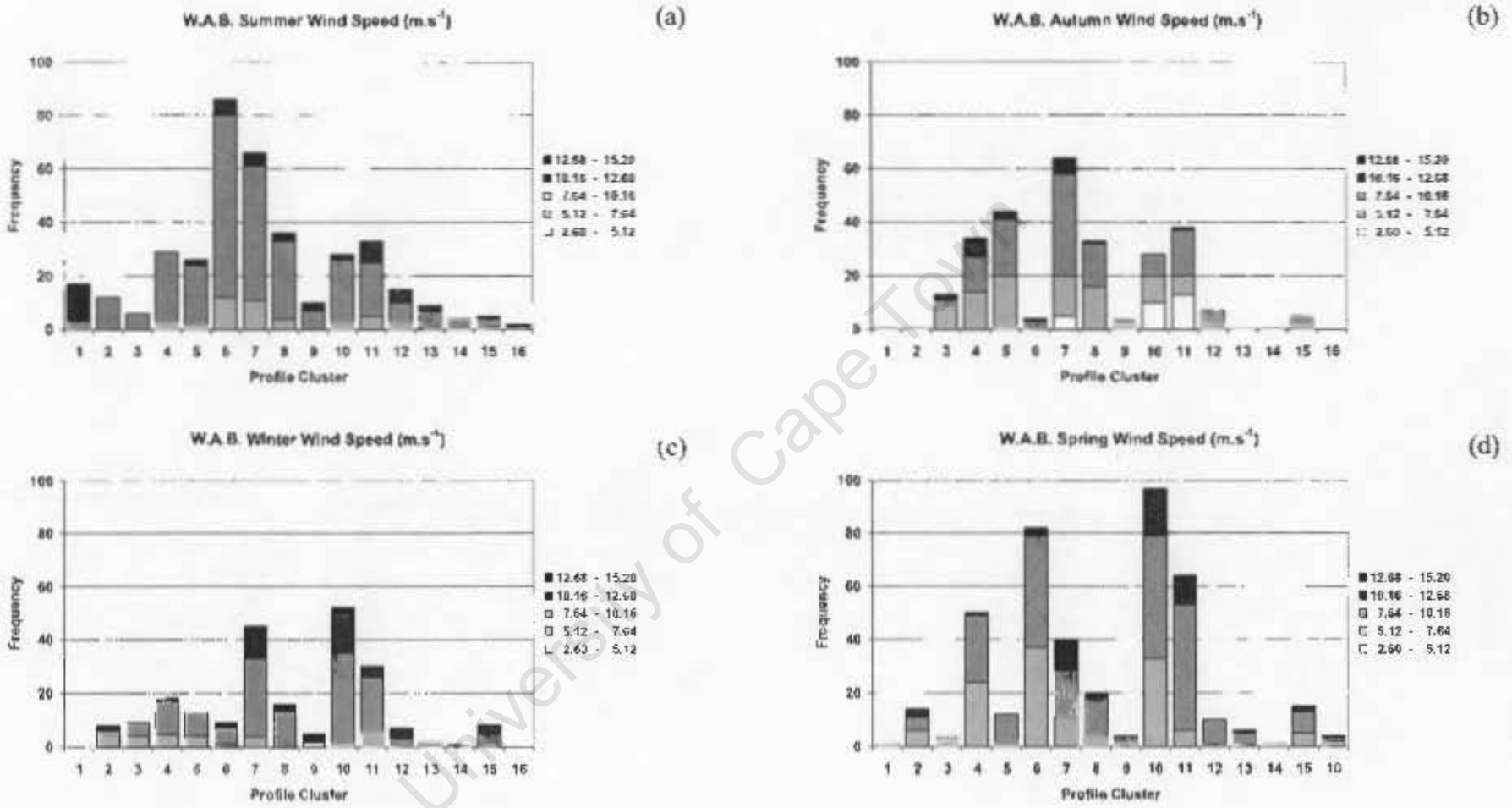

Fig. 3.11. Frequency distributions of profiles and their associated surface wind speed data intervals for each season; (a) summer (n = 385), (b) autumn (n = 278), (c) winter (n = 223) and (d) spring (n = 424).

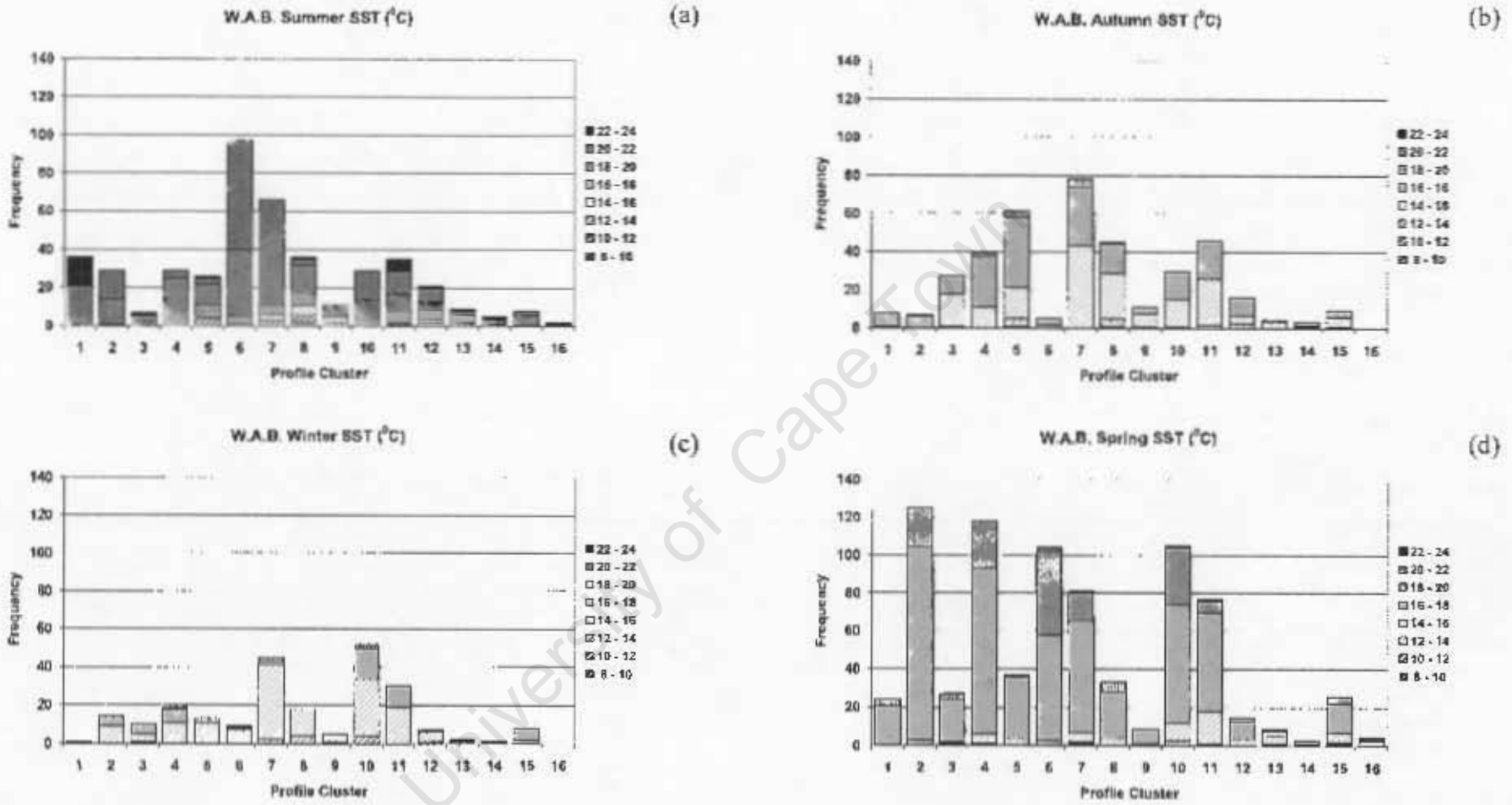

Fig. 3.12. Frequency distributions of profiles and their associated surface temperature data intervals for each season: (a) summer (n = 446), (b) autumn (n = 392), (c) winter (n = 234) and (d) spring (n = 798).

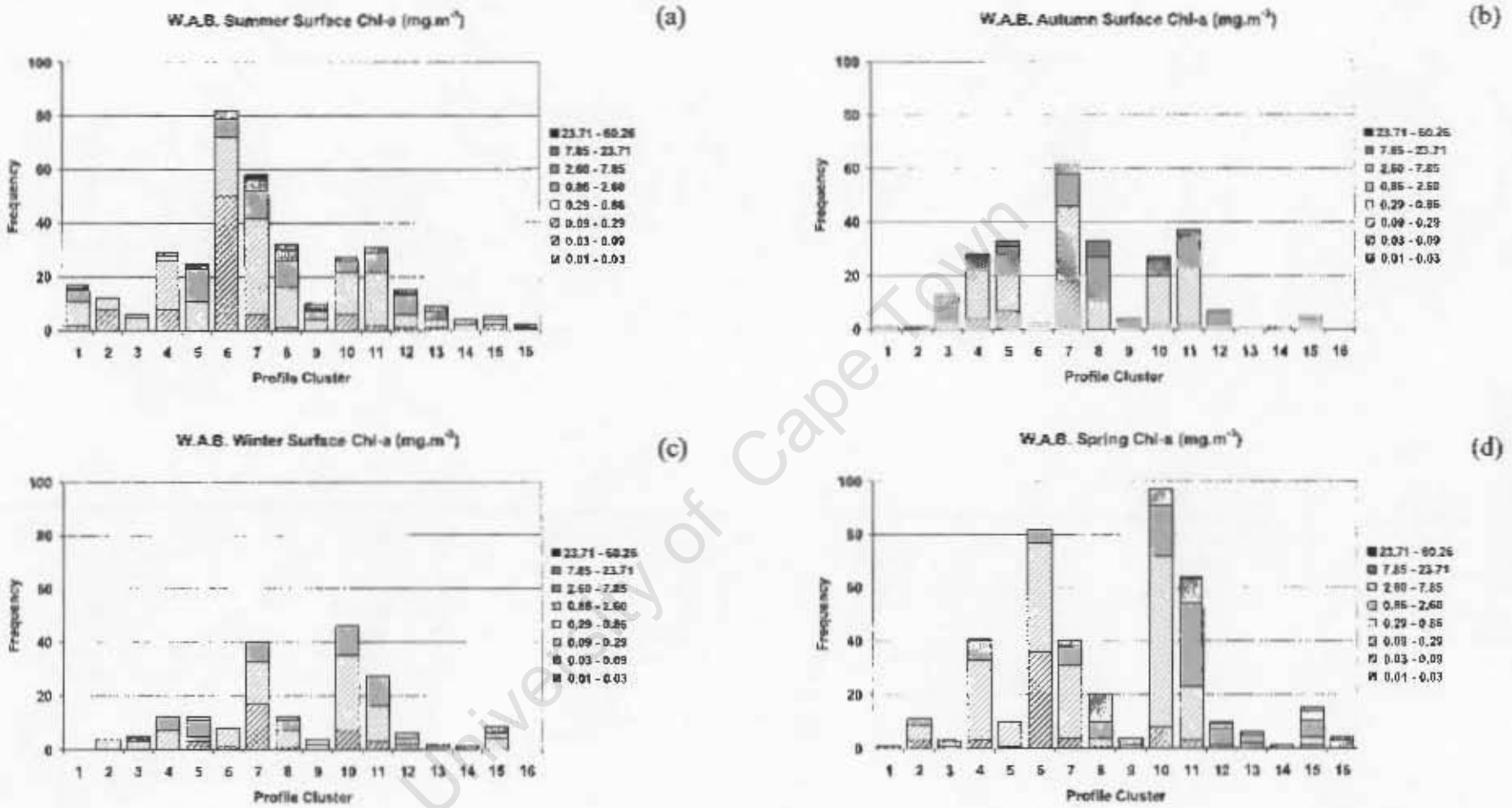

Fig. 3.13. Frequency distributions of profiles and their associated surface chlorophyll data intervals for each season: (a) summer (n = 364), (b) autumn (n = 256), (c) winter (n = 187) and (d) spring (n = 409).

3.3.3 *East Agulhas Bank*

Wind Direction (Figure 3.14a-d)

With the exception of clusters 5, 7, 14 and 16 summer profile clusters are dominated by southeasterly winds. Profiles in cluster 14 and 16 only occur during either easterly or westerly winds. During northeasterly winds only clusters 9 and 16 are represented. In winter there is a clear change seen in the dominant profile clusters of associated wind direction. Across clusters 4 and 5, 7 and 8, 10 to 12 and 15 there is a decrease occurrence of easterly winds and an increase in westerly to northwesterly winds. A very similar pattern is seen in spring from cluster 1 to 15 (but excluding 2 and 3 which only occur during westerly winds) although there is more of a southeasterly component and northwesterly winds are only evident in cluster 4.

Wind Speed (Figure 3.15a-d)

The majority of summer profiles in most clusters were observed during periods of low weekly averaged wind speed. The exceptions are cluster 1 which has a majority of medium wind speed and clusters 4 and 5 which are roughly equally divided between low and medium. Cluster 10 had only low wind speed and only in clusters 8, 11, 13 and 15 are there records of high wind speed. In autumn clusters 3, 4 and 5 had the largest occurrence of high wind speed. Cluster 11 was the only cluster with very low wind speed. Very high wind speed were recorded in clusters 8, 10 and 11 during winter but were similar to the remaining clusters, with the exception of cluster 5, in terms of proportional distribution of low, medium and high wind speed. Cluster 5 has no record of high wind speed. In spring profiles in clusters 1 and 2 have the lowest. Clusters 6 to 8 have more occurrences of stronger winds and which increase further in clusters 10 to 12 and 15.

Surface Temperature (Figure 3.16a-d)

The satellite-derived SST that is associated with each profile is strongly seasonal and consistent for all clusters evident in the season. Summer SSTs range primarily between 18°C and 22°C with a few profiles with higher and lower intervals in cluster 1, 5, 7 and 11. Autumn includes the range 16°C to 18°C and winter profiles are dominated by the interval 16°C to 18°C although there are also relatively high numbers in interval 14°C to 16°C. Spring indicates a warmer season in terms of SST with its temperatures falling in the range of 16°C to 20°C with slightly more in the interval 18°C to 20°C.

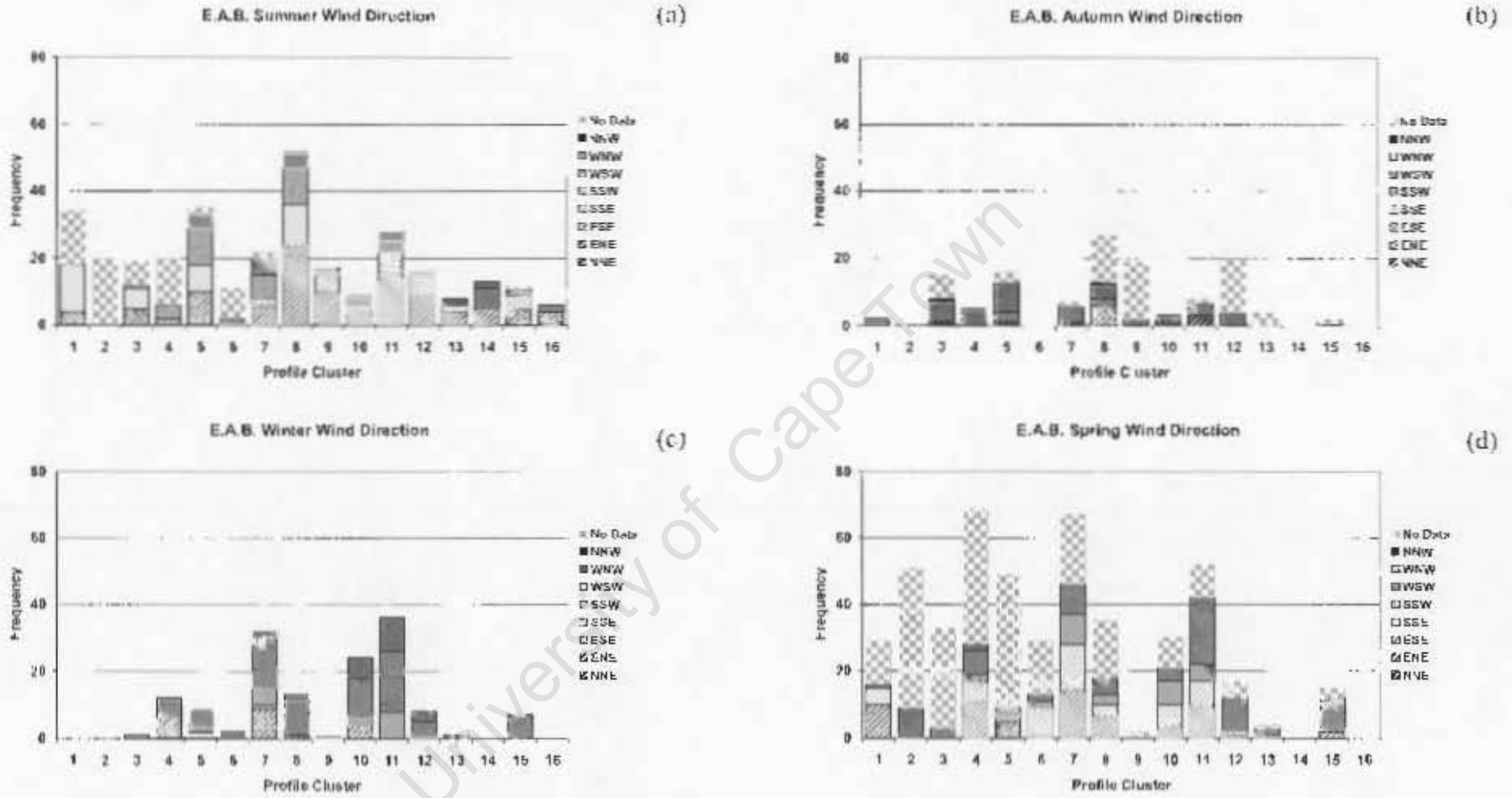

Fig. 3.14. Frequency distributions of profiles and their associated surface wind direction data intervals for each season: (a) summer (n = 251), (b) autumn (n = 64), (c) winter (n = 146) and (d) spring (n = 232). Including profiles with no surface data gives the total number of profiles in each cluster for the season.

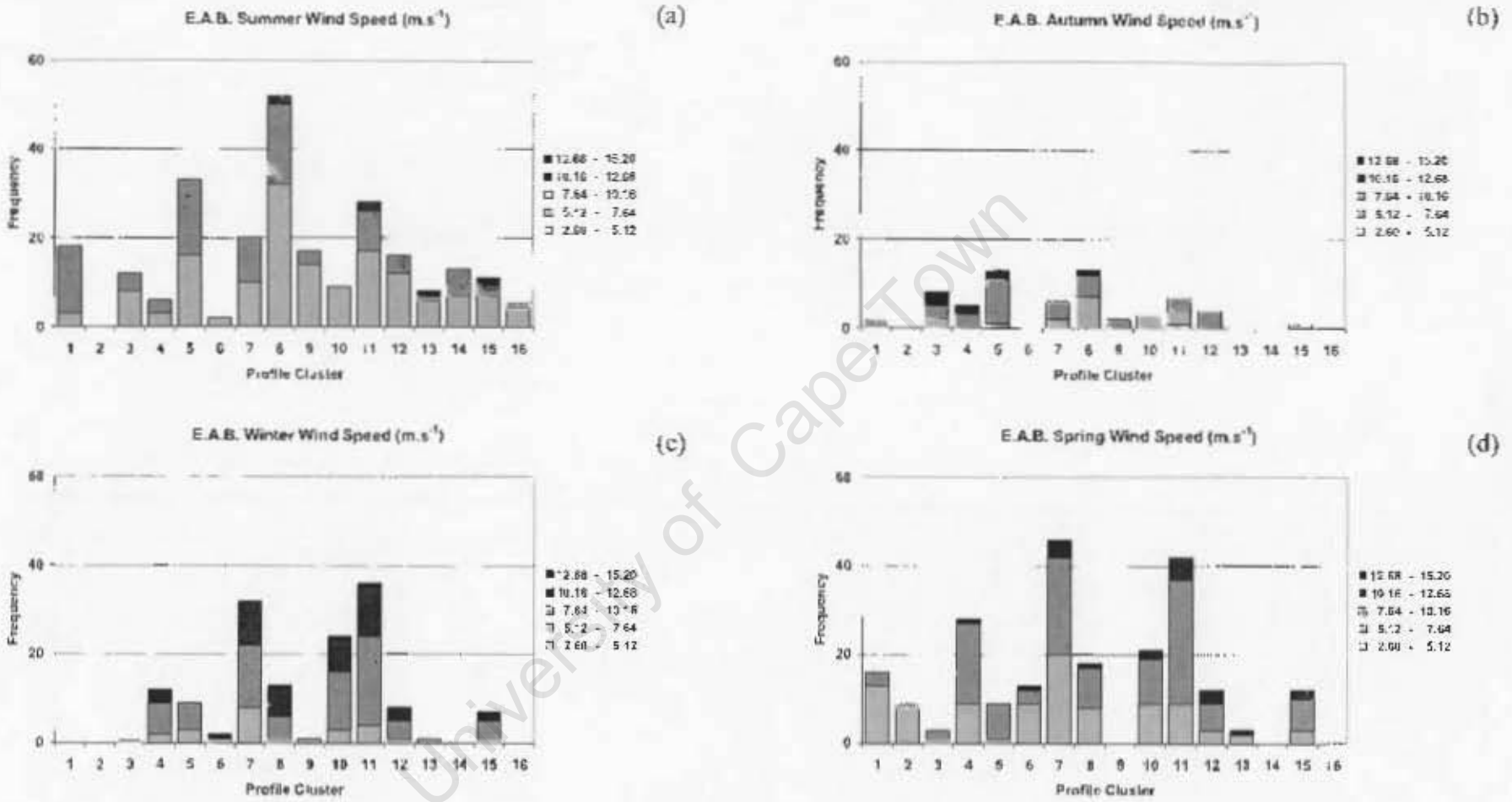

Fig. 3.15. Frequency distributions of profiles and their associated surface wind speed data intervals for each season; (a) summer (n = 251), (b) autumn (n 64), (c) winter (n = 146) and (d) spring (n = 232).

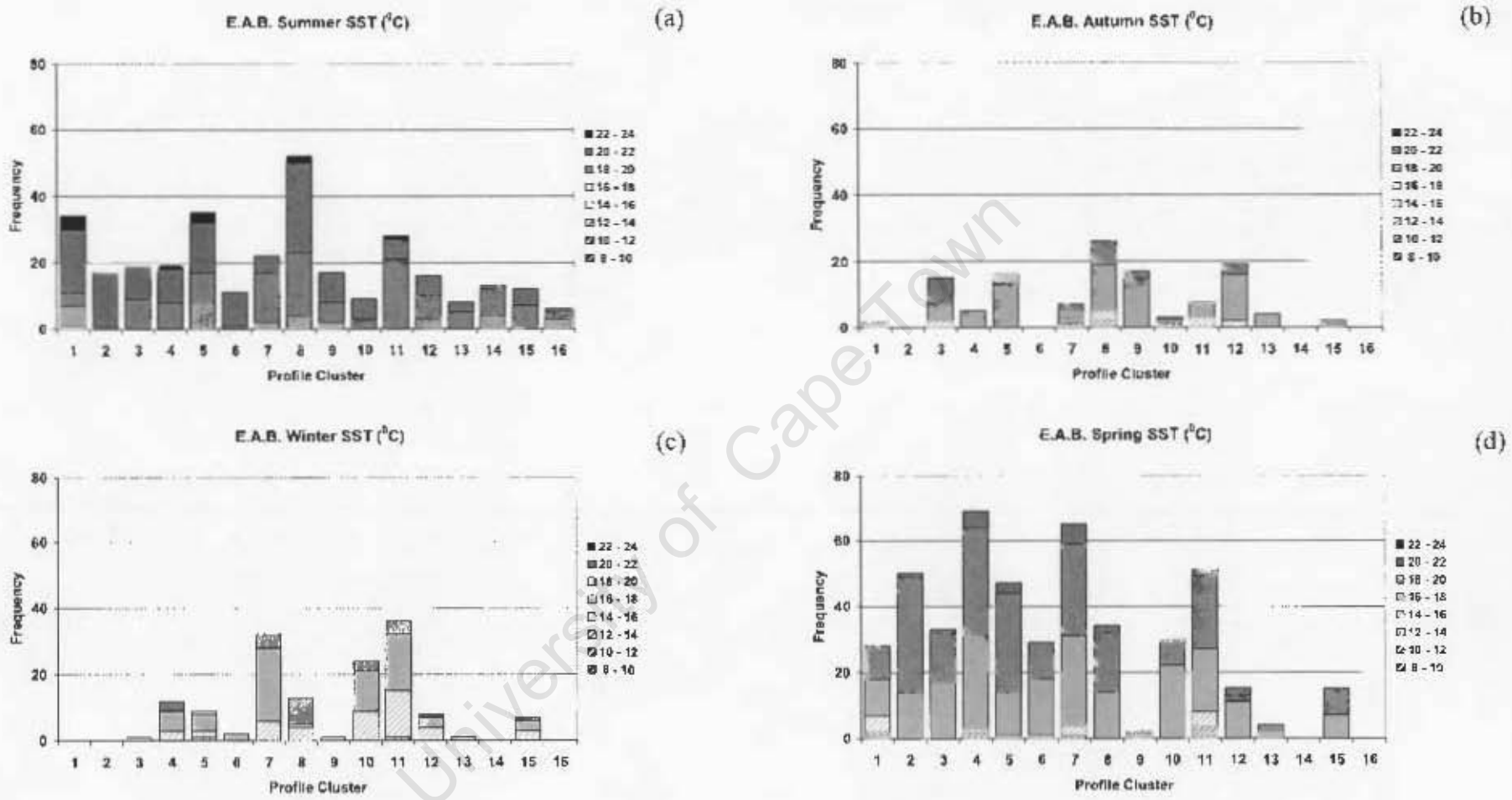

Fig. 3.16. Frequency distributions of profiles and their associated surface temperature data intervals for each season; (a) summer (n = 318), (b) autumn (n = 124), (c) winter (n = 146) and (d) spring (n = 472).

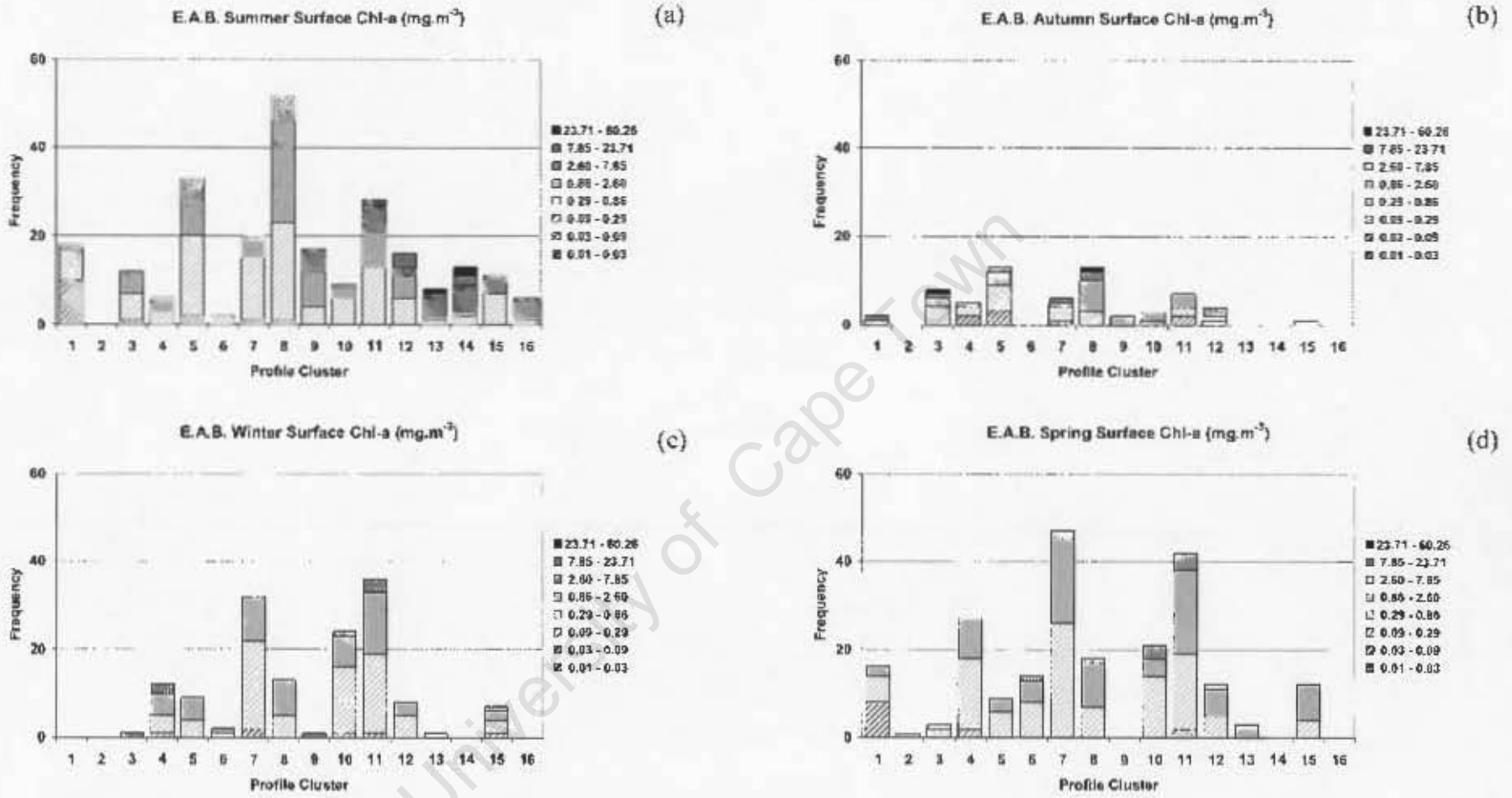

Fig. 3.17. Frequency distributions of profiles and their associated surface chlorophyll data intervals for each season; (a) summer (n = 251), (b) autumn (n = 64), (c) winter (n = 146) and (d) spring (n = 225).

Surface Chlorophyll (Figure 3.17a-d)

The first cluster consists mostly of profiles with surface chlorophyll in the range 0.09 to 0.29 $mg.m^{-3}$ but many are associated with the adjacent larger interval. Clusters 5, 8and 11 have similar distributions with the majority in the interval 0.29 to 2.60 $mg.m^{-3}$ and a varying but significant proportion in the interval 2.60 to 7.85 $mg.m^{-3}$. Very few profiles were collected during autumn on the EAB and many of them had no surface data. The profiles are mainly distributed among clusters 3, 5, 8, 9 and 12. Surface values vary in each of these profiles. Little more can be said based on such a small number of samples. More than 60% of winter profiles belong to three clusters, *viz.* 7, 10 and 11. Surface temperatures again range between 0.29 to 2.60 $mg.m^{-3}$ and a few warmer temperatures are observed in cluster 11. During spring the profiles are concentrated in the first eight clusters particularly 2, 4, 5 and 7 and also in cluster 11. Unfortunately more than half the profiles do not have surface values. The temperatures that were observed range across the same intervals as previous seasons have done. Cluster 1 however shows almost 50% in the lower concentration interval, *viz.* 0.09 to 0.29 $mg.m^{-3}$.

Chapter 4: Discussion

4.1 Profile Clustering

Three primary forms of upwelling have been described in the three regions discussed: wind-driven upwelling; temperate zone blooms; and shelf-edge upwelling. Wind-driven upwelling dominates the West Coast (Shannon, 1985) and the WAB (Probyn, 1994) whereas temperate zone blooms are confined to the WAB in spring (Mitchell-Innes *et al.*, 1999) and shelf-edge upwelling to the EAB and to a lesser extent the WAB (Probyn *et al.*, 1994). A sequence of clusters centroid profiles, *viz.* 3, 5, 8, 9 and 14 depict a typical pulsed upwelling scenario following strong southerly winds (Figure 4.1.a). The development begins as a well-mixed homogenous layer (to approximately 20m depth) of low phytoplankton biomass (~1.0 $mg.m^{-3}$) increases but remains mixed in the upper 20m due to surface wind (clusters 3, 5 and 8). Profiles in cluster 9 and cluster 14 show the development of a near surface peak (~13.0 $mg.m^{-3}$) that develops as the water column stabilizes. This scenario is in agreement with that discussed in Pitcher *et al.* (1996). The decline of a bloom can be depicted from profiles 15, 11, and 10 (Figure 4.1.b) as the surface nutrients are depleted and chlorophyll concentrations subsequently decrease. The peak deepens either as a result of phytoplankton sinking out of the surface layer (Pitcher *et al.*, 1998) or motile species actively seeking higher nutrient concentrations at the pycnocline (Probyn *et al.*, 1990; Mitchell-Innes and Walker, 1991). Offshore and typically over the continental slope dinoflagellates are able to maintain a subsurface peak usually between 30 and 50m depth under low nutrient conditions (Probyn *et al.*, 1994). The vertical chlorophyll distribution tends to be temporally and spatial stable and can be represented by profiles from clusters 2 and 6 (Figure 4.1.c). Beyond the thermal front the water column is characteristic of open ocean oligotrophic water and is exposed to high wind turbulence which maintains a low biomass homogenous water column as depicted by cluster 1 and cluster 4 (Figure 4.1.d). Due to the high degree of forcing variability over the three regions it is unreasonable to attempt to depict every scenario discussed in the literature. It is clear that the clusters do provide a range of profiles that can depict the transition from inshore near-surface blooms to offshore dinoflagellates dominated water and that the magnitude and depth of the profile are in agreement with *in situ* studies.

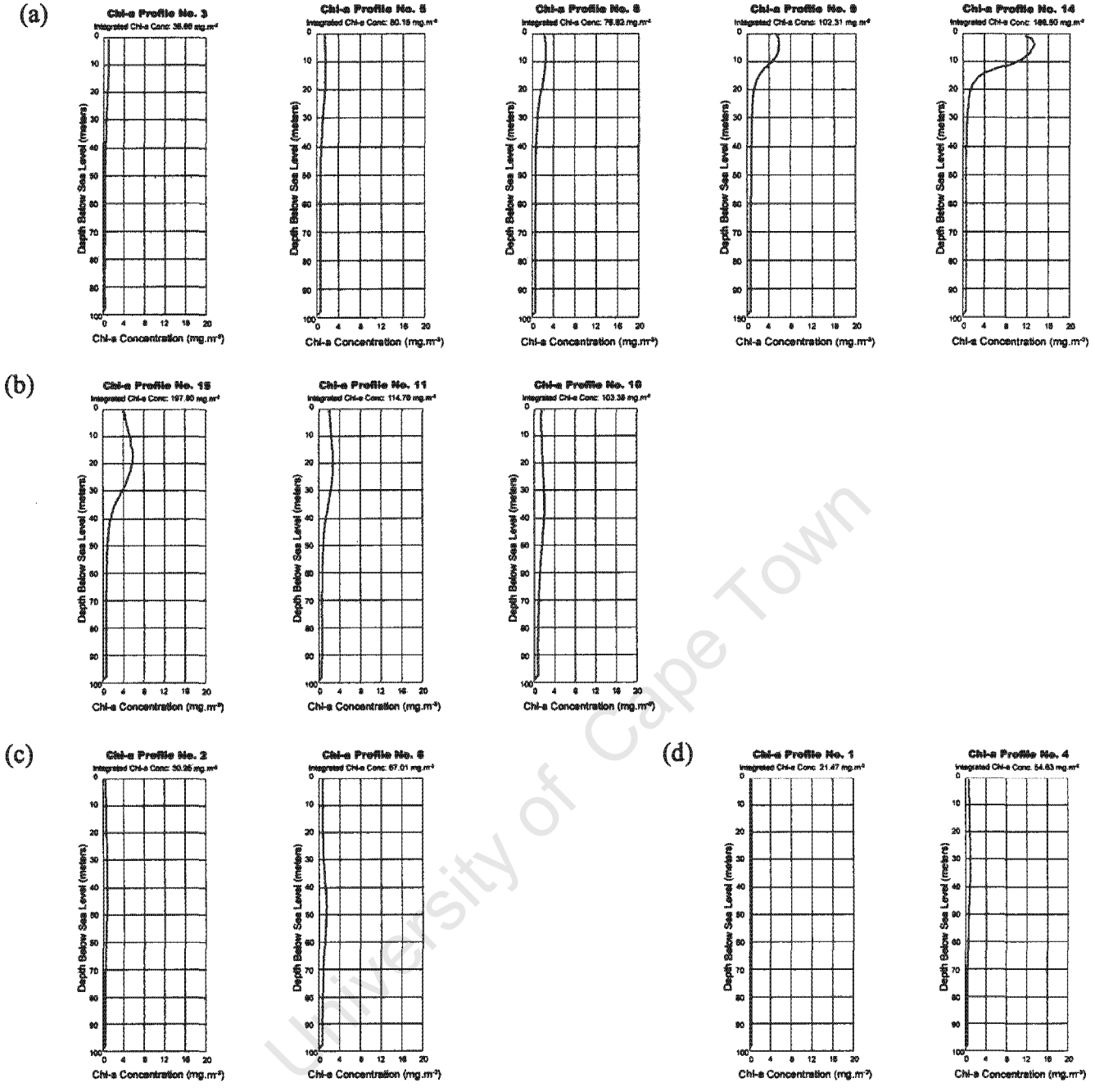

Fig. 4.1. A scenario of: a bloom event **(a)**, nutrients are introduced to the surface by upwelling winds (profile 2 and 5), a quiescent phase follows and the surface layer stratifies resulting in a near surface bloom (profile 14); a bloom decline **(b)** the biomass is grazed and less buoyant cells sink out the water column as it stratifies; offshore water **(c)** where motile dinoflagellates form a deep peak; and open ocean water **(d)** which is oligotrophic and well mixed by high speed winds.

4.2 Satellite-Derived Surface Data

The collection of weekly averaged satellite wind data that is related to the chlorophyll-a profiles, depicts the seasonal migration of the South Atlantic Anticyclone and the

subsequent changes to the mean wind direction well. An example of this is the seasonal migration of passing frontal systems and the subsequent increase in occurrence of northwesterly winds recorded over the region in winter (Shannon, 1985). The seasonal migration of the mid-latitude westerly wind belt and the resulting reversal of predominantly summer easterlies to winter westerlies over the Agulhas Bank (Hardman-Mountford *et al.*, 2003) is also clearly evident. Interestingly north-northwesterly winds only occur in the autumn data for all three regions. The satellite wind data indicate that there is a broad spectrum of wind directions experienced within each season of each region. The variability in the weekly averaged data emphasises the deficiency of seasonally averaged and possibly monthly averaged data. On the EAB during autumn when northwesterly winds dominate, there is also a clear seasonal northeasterly component. Similarly seasonal or monthly averaged data cannot include the highly variable wind direction observed over the Agulhas Bank in spring since frontal systems and associated northwesterly winds generally pass over in the range 24-48 hours. The winds are likely a reflection of the complexity of the smaller scale dynamical of the regions investigated here, which have been linked to localised upwelling cells. Thus even times scales of weeks could bias the results.

Seasonal wind speed data indicate few occurrences of very high and very low speed. Only in autumn and winter on the WAB and only in winter on the EAB are wind speeds above 12.5 $m.s^{-1}$ encountered coinciding with the mid-latitude westerlies (Probyn *et al.*, 1994). In a pulsed upwelling region such as the southern Benguela where strong winds blow for short periods of time, in the region of a few days and are followed by quiescent periods or wind reversals (Shannon, 1985), weekly averages may undermine the strength of their signal. Nevertheless their existence is evident in the data (for instance where relatively large proportions of high and low wind speeds do occur). The importance of mean wind speed and wind direction in studies of phytoplankton biomass dynamics has been discussed by Brown and Hutchings (1987) and Nelson (1992) for example.

Sea surface temperature data depict the enhanced temperature gradient associated with intense summer upwelling inshore and warmer mature upwelling and open ocean water offshore along the West Coast and WAB. The influence of the warm water of the Agulhas Current is depicted in the SST over the bank but more so over the EAB than WAB (Boyd and Shillington, 1994). The high proportion of low SST in autumn on the West Coast is

unexpected and suggests that upwelling may occur more frequently during this season. Studies by Shannon and Nelson (1996) and Hardman-Mountford *et al.* (2003) noted a wind forcing maximum in autumn over the northern extent of the Benguela which could explain the enhanced upwelled water. However the autumn wind data do not corroborate this. It is feasible that the low temperatures and low wind speeds are a bias of the autumn *in situ* sampling.

Surface chlorophyll data depicts less seasonal variation over the Agulhas Bank, particularly in winter and spring in comparison to the West Coast. This is in accordance with Mitchell-Innes *et al.* (1999) who describe even chlorophyll distribution due to deep mixing from strong winter westerlies and increased water column stability due to intrusions of cold water over the shelf in spring. The Agulhas Bank has upwelling favourable winds predominantly in summer and autumn but current-driven upwelling year round (Boyd and Shillington, 1994) and this may account for the only slight increase in chlorophyll seen in these seasons. The West Coast has very high surface chlorophyll concentrations throughout the year which is consistent with the persistent upwelling favourable winds (Longhurst, 1995). Autumn on the West Coast shows highest surface chlorophyll-a and coincides with the lowest SST and lowest wind speeds. One possible explanation for this is that the upwelling winds may be of longer duration although weaker (Longhurst, 1995) leading to more continuous upwelling.

Although the sampling strategy used in the collection of profiles is neither temporarily nor spatially random, the mesoscale physical dynamics of the regions investigated are well represented in the associated satellite data. More localised patterns can only be detected by an association with local upwelling dynamics represented by the profiles.

4.3 Satellite Surface Data and Profiles

The predominant strong southeasterlies along the West Coast in summer produce a variety of subsurface chlorophyll-a profiles which can be attributed to the complexity of the coast and the resulting localised upwelling maxima and minima (Shannon, 1985). Changes in meteorological forcing in the regions discussed produce rapid and often dramatic changes to the vertical structure of the water column and subsequently the biological dynamics (Pitcher *et al.* 1996; Mitchell-Innes *et al.*, 1999). Seasonally averaged profiles have been shown to be inadequate for estimations of phytoplankton distribution (Richardson *et al.*,

2002). A crucial objective of the Phytoplankton Prediction System is to predict profile shape on a daily basis from satellite data. In order to accomplish this individual profiles need to be associated with concurrent surface data. The new data set will be used to train the Dynamic Bayesian Network. The DBN improves predictability over the Bayesian Network by learning not only the causal probabilities relating concurrent surface data to each profile by also the temporal probabilities; surface data from a previous time step or time step sequence associated with the profile. This section discusses the relationship between the shape of the profile and concurrent satellite-derived surface variables. Including prior temporal data in the analysis was beyond the scope of this paper.

The relationship between the profile and each individual surface variable can be easily evaluated. For example, on the West Coast profiles in clusters 9, 14 and 16 have a more pronounced westerly wind component which is indicative of weaker summer winds (Figure 4.2.). The three profiles indicate the development of a typical near surface bloom that is facilitated by the coupled onshore advection of warmer surface water and stratification of the surface layer (Andrews and Hutchings, 1980). Sea surface temperature can be related to profile shape but similar temperatures can predict different profiles. On the West Coast the coolest SSTs are frequently linked to clusters 8, 9, 12 and 13. These clusters have profiles that suggest recently upwelled water. On the Agulhas Bank the coolest surface temperature can reflect deep subsurface chlorophyll peaks such as in clusters 7, 10 and 11 where upwelling is mainly subsurface (Mitchell-Innes *et al.*, 1999). Surface chlorophyll concentration also has a strong link to the profiles but is a more effective predictor of profile where the profile peak is near the surface. The simplest form of a temporal probability model uses only one observable variable such as surface chlorophyll or SST. The intended DBN structure requires all the satellite-derived surface data to be used in conjunction to predict a profile. Murphy (2002) indicates that increasing the number of parent variables in a BN is one method of improving prediction accuracy. This is the idea behind using Bayesian Networks. However, a Dynamic Bayesian Network, in addition to a causal network structure, uses time series information or temporal sequences. A number of examples where concurrent surface data can explain the profile shape are discussed below. The examples have been grouped into upwelling sequences in order to illustrate the temporal link between sets of surface data. The first sequence, depicted in Figure 4.2., is applicable to a summer upwelling bloom and decline on the West Coast and WAB.

Following upwelling, weaker winds from the west advect warmer surface water inshore and the upper mixed layer stratifies. An increase in phytoplankton biomass ensues. The surface waters warm with age and the wind speed begins to increase signalling the end of the quiescent period. As the stronger southeasterly winds mix the surface layer, the stratified water is broken down and the phytoplankton sink out of the water column. In the second example on the WAB (Figure 4.3.) the upwelling favourable winds are stronger than on the West Coast. Following upwelling the high wind speed prevents a near-surface bloom as it mixes the upper layer. In both these examples the surface data confirms the profile. These sequences may change according to the environmental variables. For instance, if strong winds begin before a bloom sequence can be completed, a different profile and associated surface properties will be evident. This might lead to a profile in cluster 13 following the cluster 9, cluster 14 sequence (Figure 4.4.. In this case the southeasterly wind mixes the plankton to deeper water and SST drops again.

The two examples used above indicate good relationships between averaged surface data and profile shape and illustrate that the clusters show individual surface expression. The suggested temporal relationship is also good. For instance, in the first example it is clear how a two time-step sequence of increasing SST, high surface chlorophyll and near constant low westerly wind can predict a cluster 16 profile. A problem may lie in the nature of the southeast Atlantic upwelling zone, *viz.* pulsed wind-driven upwelling. Phytoplankton typically bloom and decay over periods of 6 to 8 days (Brown and Hutchings, 1987) and these events may have little relationship to events in previous weeks.

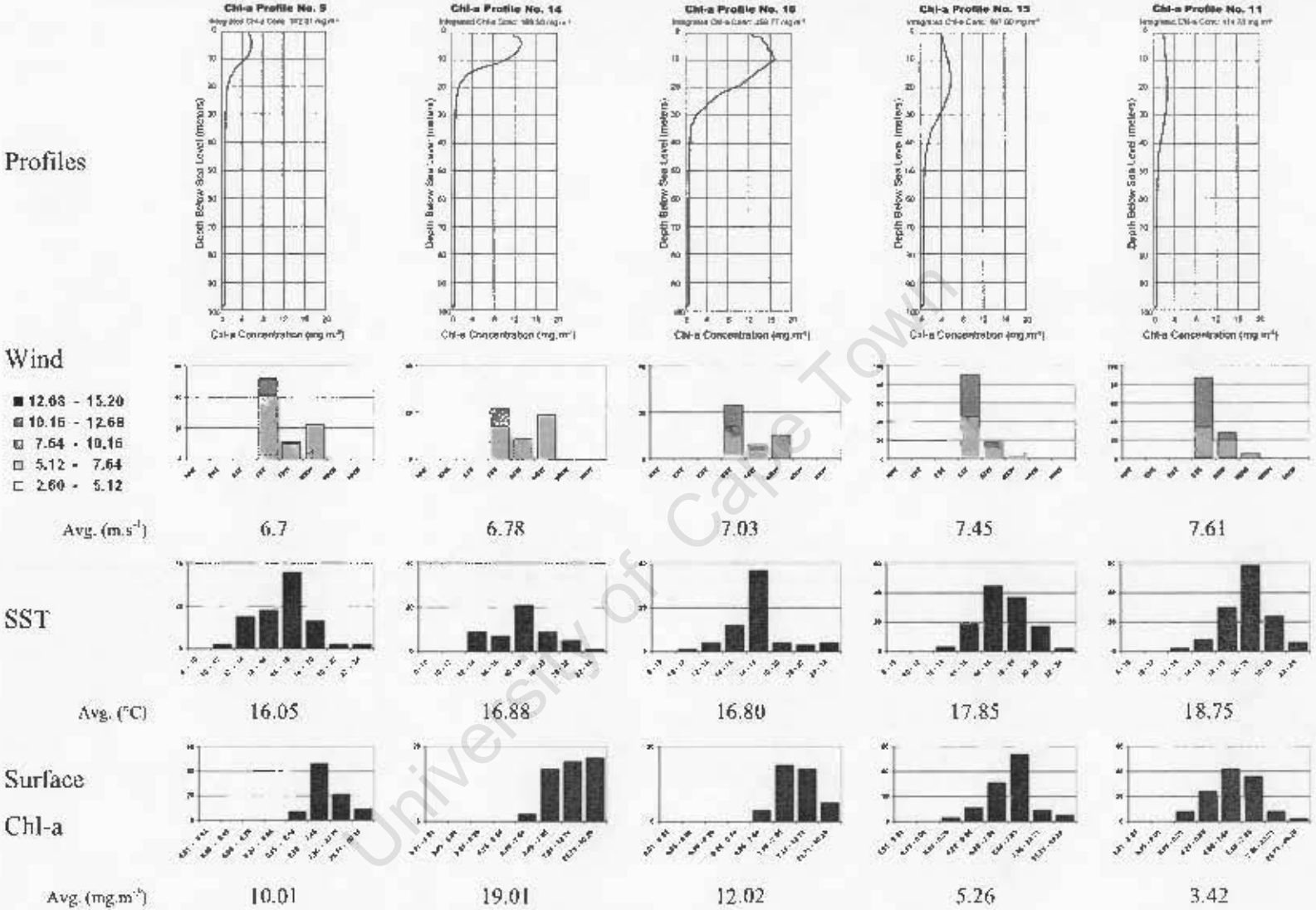

Fig. 4.2. An upwelling scenario from the West Agulhas Bank depicting new nutrients introduced to the surface layer. Weaker westerly winds enable a near surface bloom as warmer advected water stratifies the surface layer. The phytoplankton biomass sinks down the water column in the late stages marked by in stronger southeasterly winds.

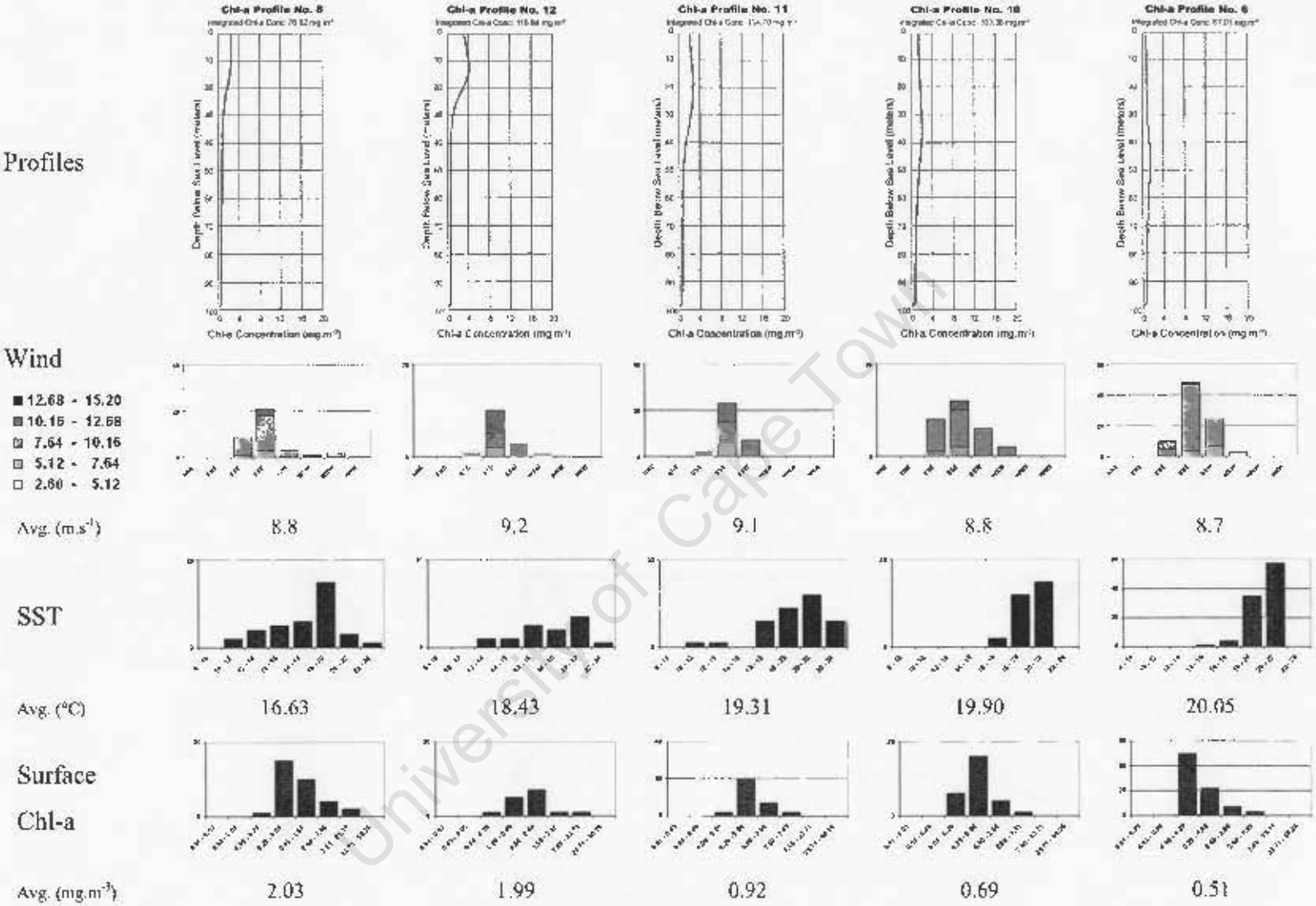

Fig. 4.3. Profiles from the WAB during summer indicate high, predominantly southeasterly winds favourable for upwelling. High wind speeds keep the upper layer mixed and prevent a surface bloom.

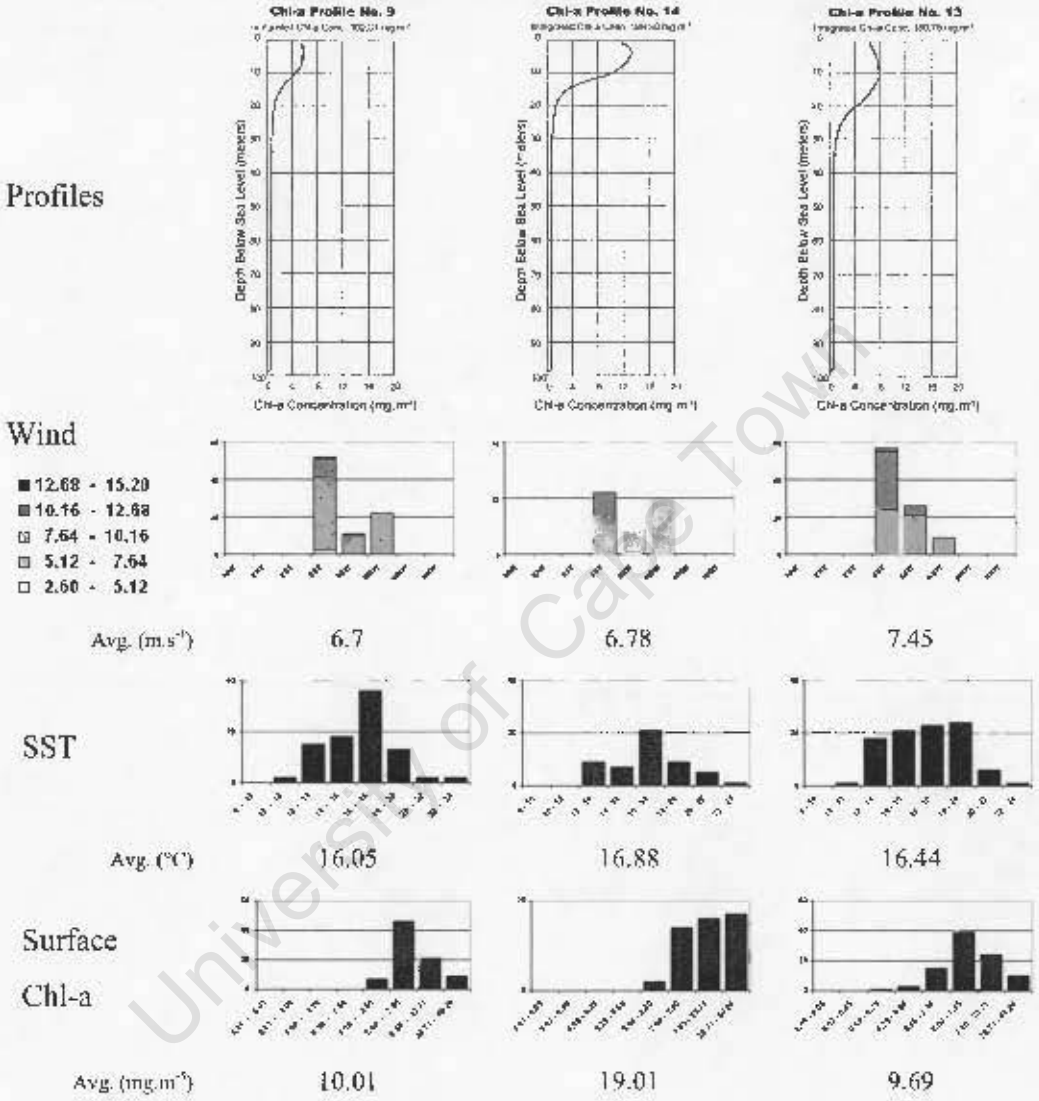

Fig. 4.4. An upwelling scenario from the West Coast that depicting new nutrients introduced to the surface layer. A short quiescent period Strong limits the near-surface bloom as the upper layer is mixed, indicated by cooler temperatures.

4.4 Conclusion

Methods for accurate predictions of phytoplankton production in dynamic regions such as coastal upwelling regions require a better temporal and spatial resolution of vertical phytoplankton distribution. Currently biomass distribution in the vertical dimension and on the required scales can only be inferred from satellite-derived surface data. The subject of this paper has been the development of a more accurate method of inference. Previous research has had some success with clustered profiles and surface data but few if any have suggested incorporating wind data or sequential data.

The proposed system utilizes a Dynamic Bayesian Network program that has proven to be adept in dealing with temporal data (Murphy, 2002). The DBN will be trained with a data-base of *in situ* profiles and associated satellite-derived surface data. Of the 5813 in situ profiles collected along the west and south coast of southern Africa between 1988 and 2006, 5457 were within the regions analysed in this paper. Each of these profiles were associated with weekly averaged wind data and 5-day averaged SST and surface chlorophyll data from satellite imagery. The relationships between the profiles and surface data were analysed. There is generally a good consensus between the profile centroid surface expression and the satellite-derived surface chlorophyll concentration. Similarly there is a rational argument for the association between SST and profile shape in terms of upwelling and offshore water. Wind data is more difficult to analyse with phytoplankton data for two reason: firstly the data are averaged over a week and typical upwelling events are characterised by short pulsed wind speeds and directions; and secondly there is a time lag between wind dynamics and upwelling so that wind on a particular day may only be manifested in the profile a day or two later. This point is also valid when training the DBN with weekly averaged surface data. It seems reasonable to expect a poor relationship between data more than a week apart in an upwelling region characterised by episodic events on timescales of days. In the present study it was not feasible to attempt creating a training database using daily averages as the required computing power and processing time was not available. Nevertheless this study has shown that there is a close association between the suite of weekly averaged variables and profile shape. The logical sequences produced by the suite of data are encouraging and provide the justification for including wind data in the prediction of chlorophyll profiles. Furthermore the causal relationships of data over time are well suited for the proposed Dynamic Bayesian Network. The training

and testing of the DBN using the current data will indicate the success of the temporal aspect.

References

Afanasyev, Y. D., Nezlin, N. P. and Kostianoy, A. G. (2001) Patterns of seasonal dynamics of remotely sensed chlorophyll and physical environment in the Newfoundland region. *Remote Sensing of Environment,* **76**, 268-282.

Akima, H. (1970) A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the Association for Computing Machinery*. **17** (4), 589-602.

Andrews, W. R. H. and Hutchings, L. (1980). Upwelling in the southern Benguela Current, South Africa. *Prog. Oceanogr.*, **9**, 1-81.

Atkins, W. R. G. (1928) Seasonal variations in the phosphate and silicate content of sea-water during 1926 and 1927 in relation to the phytoplankton crop. *J. Mar. Biol. Ass. U.K.*, **15**, 191-205.

Bakun, A. (1993) The California Current, Benguela Current and Southwestern Atlantic shelf ecosystems: A comparitive approach to identifying factors regulating biomass yields. In Sherman, K., Alexander, L. M. and Gold, B. (eds), *Large marine ecosystems-stress mitigation and sustainability*. American Association for the Advancement of Science, Washington DC, pp. 199-221.

Barber, R. T. and Hilting, A. K. (2002) History of the study of plankton productivity. In Williams, P. J. B., Thomas, D. N. and Reynolds, C. S., le, B. (eds), *Phytoplankton Productivity*, Blackwell Science, pp. 16-43.

Barlow, R. G., Aiken, J., Sessions, H. E., Lavender, S. and Mantel, J. (2001) Phytoplankton pigment, absorption and ocean colour characteristics in the southern Benguela ecosystem. *S. Afr. J. Sci.* **97**, 230–238.

Barlow, R. G., Cummings, D. G. and Gibb, S. W. (1997) Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. *Mar. Ecol. Prog. Ser.*, **161**, 303-307.

Behrenfeld, M. J. and Falkowski, P. G. (1997) Photosynthetic rates derived from satellite-based chlorophyll concentrations. *Limnol. Oceanogr.*, **42**, 1-20.

Bigelow, H. B. (1927) Physical oceanography of the Gulf of Maine. *Bull. U.S. Bureau Fisheries,* **40**, 511-1027.

Borchers, P. and Hutchings, L. (1986) Starvation tolerance, development time and egg production of *Calanoides carinatus* in the southern Benguela Current. *J. Plankton. Res.*, **8**(5), 855-874.

Boyd, A. J. and Shillington, F. A. (1994) Physical forcing and circulation patterns on the Agulhas Bank. *S. Afr. J. Sci.*, **90**, 114-122.

Brandt, K. (1899) Ueber den Stoffwechsels im Meere. *Wiss Meeresunters., Abt. Keil,* **4**, 213–230.

Brown, P. C. and Hutchings, L. (1987) The development and decline of phytoplankton blooms in the southern Benguela upwelling system. Drogue movements, hydrography and bloom development. In Payne, A. I. L., Gulland, J. A. and Brink, K. H. (eds.), *The Benguela and Comparable Ecosystems. S. Afr. J. Mar. Sci.*, **5**, 357–391.

Brown, P. C., Painting, S. J. and Cochrane, K. L. (1991). Estimates of phytoplankton and bacterial biomass and production in the northern and southern Benguela ecosystems. *S. Afr. J. Mar. Sci.*, **11**, 537–564.

Chambers, C. O. (1912) *The relation of algae to dissolved oxygen and carbon-dioxide. With special reference to carbonates. Missouri Botanical Garden 23rd Annual Report,*

Board of Trustees, St Louis, Missouri.

Chapman, P. and Shannon, L. V. (1985) The Benguela ecosystem. 2. Chemistry and related processes. *Oceanography and Marine Biology Annual Review* **23**, 183–251.

Cullen, J. J. (1982) The deep chlorophyll maximum: Comparing vertical profiles of chlorophyll a. *Can. J. Fish. Aquat. Sci.*, **39**, 791-803.

Cushing, D. H. (1971) Upwelling and the production of fish. *Adv. Mar. Biol.*, **9**, 255-334.

Demarq, H., Richardson, A. J. and Field, J. G. (2007) Generalised model of primary production in the southern Benguela upwelling system. *Mar. Ecol. Prog. Ser.* (in press).

Dickey, T., Zedler, S., Yu, X., Doney, S., Frye, D., Jannasch, H., Manov, D., Sigurtson, D., McNeil, J., Dobeck, L., Gilboy, T., Bravo, C., Siegel, D. and Nelson, N. (2001) Physical and biogeochemical variability from hours to years at the Bermuda testbed mooring site: June 1994–March 1998. *Deep-Sea Res.*, **48**, 2105–2140.

Dugdale, R. C., Davis, C. O. and Wilkerson, F. P. (1997) Assessment of new production at the up welling centerat Point Conception, California, using nitrate estimated from remotely sensed sea-surface temperature. *J. Geophys. Res.*, **102**, 8573–8586.

Dutkiewicz, S., Follows, M., Marshall, J., and Gregg, W. W. (2001) Interannual variability of phytoplankton abundances in the North Atlantic. *Deep-Sea Res., Part II*, **48**, 2323-2344.

Epply, R. W., Stewart, E., Abbott, M. R. and Heyman, U. (1985) Estimation ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California Bight. *J. Plankton Res.*, **7**, 57-70.

Falkowski, P. G. (1981) Light-shade adaptation and assimilation numbers. *J. Plankton Res.*, **3**, 203-216.

Gill, A. (1992) *Atmosphere–Ocean Dynamics.* Academic Press, San Diego.

Gaarder, T. and Gran, H. H. (1927) Investigation of the production of plankton in the Oslo Fjord. *Rapp. Proc-Verb. Réun. Cons. Int. Explor. Mer,* **42**, 1-48.

Gran, H. H. (1912) Pelagic plant life. In Murry, J. and Hjort, J. (eds), *The Depths of the Ocean.* MacMillan and Co., London, pp. 307-387.

Gran, H. H. (1931) On the conditions for the production of plankton in the sea. *Rapp. Proc-Verb. Réun. Cons. Int. Explor. Mer*, **7**, 343-358.

Hardman-Mountford, N. J., Richardson, A. J., Agenbag, J. J., Hagen, E., Nykjaer, L., Shillington, F. A. and Villacastin, C. (2003) Ocean climate of the South East Atlantic observed from satellite data and wind models. *Prog. Oceanogr.* **59**, 181-221.

Harrison, W. G. and Platt, T. (1986) Photosynthesis-irradiance relationships in polar and temperate phytoplankton populations. *Polar Biol.*, **5**, 153-164.

Hensen, V. (1887) Ueber die Bestimmung des Planktons oder des im Meere treibende Materials an Pflanzen und Thieren. *Ber. Komm. Wiss. Unters. Dt. Meere,* **5**, 1-109.

Hill, A. E., Hickey, B. M., Shillington F. A., Strub, P. T., Brink, K. H., Barton, E. D. and Thomas, A. C. (1998) Eastern ocean boundaries: A pan-regional review. In Robertson, A. R. and Brink, K. H. (eds), *The Sea. Volume II. The global coastal ocean. Regional studies and synthesis.* John Wiley and Sons Ltd, New York, pp. 29-68.

Holm-Hansen, O., Lorenzen, C. J., Holmes, R. W. and Strickland, J. D. H. (1965) Fluorometric determination of chlorophyll. *J. Cons. Perm. Int. Explor. Mer*, **30**, 3-15.

Hooker, J. D. (1874) On the marine algae of St. Thomas and the Bermudas, and on *Halophila baillonis* Asch. Contributions to the botany of the expedition of H.M.S. Challenger. *J. Linn. Soc. Bot.*, **14**, 311–317.

Jury, M. R. and Brundrit, G. B. (1992) Temporal organization of upwelling in the southern Benguela ecosystem by resonant coastal trapped waves in the ocean and atmosphere, *S. Afr. J. of Mar. Sci.*, **12**, 219–224.

Lewis, M. R., Cullen, J. J. and Platt, T. (1983) Phyotplankton and thermal structure in the upper ocean: Consequences of nonuniformity in chlorophyll profiles. *J. Geophys. Res.*, **88**, 2565-2570.

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

Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C. M. (1995) An estimation of global primary production in the ocean from satellite radiometer data. *J. Plankton Res.*, **17**, 1245-1271.

Lorenzen, C. J. (1966) A method for the continuous measurement of in vivo chlorophyll concentration. *Deep-Sea Res.*, **13**, 223-227.

Lorenzen, C. J. (1970) Surface chlorophyll as an index of the depth, chlorophyll content and primary production of the euphotic layer. *Limnol. Oceanogr.* **15**, 479-481.

Lutjeharms, J. R. E., Cooper, J. and Roberts, M. (2000) Upwelling at the inshore edge of the Agulhas Current. *Cont. Shelf Res.*, **20**, 737-761.

Mann, K. H. (2000) Ecology of Coastal Waters, with Implications for Management, 2nd edn. Blackwell Science, Boston.

Mann, K. H. and Lazier, J. R. N. (1996) *Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Ocean*. Blackwell Science Inc, Oxford.

Marshall, S. M. and Orr, A. P. (1927) The relation of the plankton to some chemical and physical factors in the Clyde Sea area. *J. Mar. Biol. Ass. U. K.*, **14**, 837–868.

Marshall, S. M. and Orr, A. P. (1928) The photosynthesis of diatom cultures in the sea. *J. Mar. Biol. Ass. U. K.*, **15**, 321–360.

Macedo, M. F., Duarte, P., Mendes, P. and Ferreira, J. G. (2001) Annual variation of environmental variables, phytoplankton species composition and photosynthetic parameters in a coastal lagoon. *J. Plankton Res.*, **23**, 719-732.

Millán-Núñez, R., Alvarez-Borrego, S. and Trees, C. C. (1997) Modeling the vertical distribution of chlorophyll in the California Current System. *J. Geophys. Res.*, **102**, 8587–8595.

Mills, E. L. (1989) *Biological Oceanography: An Early History, 1870-1969*. Cornell University Press, Ithaca.

Mitchell-Innes, B. A., Richardson, A. J. and Painting, S. J. (1999) Seasonal changes in phytoplankton biomass on the western Agulhas Bank, South Africa. *S. Afr. J. Mar. Sci.*, **21**, 217-233.

Mitchell-Innes, B. A., Silulwane, N. F. and Lucas, M. I. (2001) Variability of chlorophyll profiles on the west coast of southern Africa in June/July 1999. *S. Afr. J. Sci.* **97**, 246–250.

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.

Morel, A. (1978) Available, usable, and stored radiant energy in relation to marine photosynthesis. *Deep-Sea Res.*, **25**, 673-688.

Morel, A. (1991) Light and marine photosynthesis: A spectral model with geochemical and climatological implications. *Prog. Oceanogr.*, **26**, 263-306.

Morel, A. and Berthon, J-F. (1989) Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated on view of remote-sensing applications. *Limnol. Oceanogr.*, **34**, 1545-1562.

Moroshkin, K. V., Bunov, V. A. and Bulatov, R. P. (1970) Water circulation in the eastern South Atlantic Ocean. *Oceanology*, **10**, 27-34.

Murphy, K. (2002) *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, U. C. Berkeley.

Nathansohn, A. (1906) Über die Bedeutung vertikaler Wasserbewegungen für die Produktion des Planktons im Meere. *Königl. Sächs. Gesellsch. d. Wissensch., Liepzig, Abhandl. d. Math.-Phys. Klasse*, Bd. 29, no. 5.

Nelson G. (1992) Equatorward wind and atmospheric pressure spectra as metrics for primary productivity in the Benguela system, *S. Afr. J. Mar. Sci.*, **12**, 19-28.

Nelson, G. and Hutchings, L. (1983) The Benguela upwelling area. *Prog. Oceanogr.*, **12**, 333-356.

Neumann, A. R., Doerffer, H., Krawczyk, M. D., Dowell, R., Arnone, C. O., Davis, M., Kishino, A., Tanaka, C., Hu, R. P., Bukata, H. R., Gordon, J., Campbell, A. R. and S. Sathyendranath, S. (2000) Algorithms for case 2 waters. In Sathyendranath, S. (ed.), *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters.*

Reports of the International Ocean-Colour Coordinating Group, **3**, IOCCG, Dartmouth, Canada, pp. 47-76.

Ørsted A.S., (1844) in Wolff, T. & Peterson, M.E. (1991) A brief biography of A.S. Ørsted, with notes on his travels in the West Indies and Central America and illustrations of collected polychaetes. Ophelia Supplement 5, 669-685.

Parsons, T. R., Maita, Y. and Lalli, C. M. (1984) *A Manual of Chemical and Biological Methods for Seawater Analysis*. Pergamon Press, Oxford.

Pingree, R. D., Holligan, P. M., Mardell, G. T. and Head, R. N. (1976) The influence of physical stability on spring, summer and autumn phytoplankton blooms in the Celtic Sea. *J. Mar. Biol. Ass. U.K.*, **56**, 845-873.

Pitcher, G. C., Boyd, A. J., Horstman, D. A. and Mitchell-Innes, B. A. (1998) subsurface dinoflagellates population, frontal blooms and the formation of red tide in the southern Benguela upwelling system. *Mar. Ecol. Prog. Ser.*, **172**, 253-264.

Pitcher, G. C., Brown, P. C. and Mitchell-Innes, B. A. (1992) Spatio-temporal variability of phytoplankton in the southern Benguela upwelling system. In Payne, A. I. L., Brink, K. H., Mann, K. H. and R. Hilborn (eds.), *Benguela Trophic Functioning. S. Afr. J. Mar. Sci.* **12**, 439-456.

Pitcher, G. C., Richardson, A. J. and Korrubel, J. L. (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.

Platt, T. (1986) Primary production in the ocean water column as a function of surface light intensity. *Deep-Sea Res.*, **33**, 149-163.

Platt, T., Caverhill, C. M. and Sathyendranath, S. (1991) Basin-scale estimates of oceanic primary production by remote sensing: The North Atlantic. *J. Geophys. Res.*, **96** (15), 147159.

Platt, T. and Sathyendranath, S. (1988) Oceanic primary production: estimation by remote sensing at local and regional scales. *Science*, **241**, 1613-1620.

Platt, T. and Sathyendranath, S. (1993) Estimators of primary production for interpretation of remotely sensed data on ocean color. *J. Geophys. Res.* **98** (14), 561-576.

Platt, T. and Sathyendranath, S. (1999) Spatial structure of pelagic ecosystem processes in the global ocean. *Ecosystems*, **2**, 384-394.

Platt, T., Sathyendranath, S., Caverhill, C. M. and Lewis, M. R. (1988) Ocean primary production and available light: Further algorithms for remote sensing. *Deep-Sea Res.*, **35**, 855-879.

Platt, T., Sathyendranath, S. and Longhurst, A. (1995) Remote sensing of primary production in the ocean: Promise and fulfilment. *Philos. Trans. R. Soc. London Ser. B,* **348,** 191-200.

Probyn T. A., Mitchell-Innes, B. A., Brown, P. C., Hutchings, L. and Carter, R. A. (1994) A review of primary production and related processes on the Agulhas Bank. *S. Afr. J. Sci.*, **90**, 166-174.

Probyn, T. A., Waldron, H. N. and James, A. G. (1990) Size-fractionated measurements of nitrogen uptake in aged upwelled waters: Implications for pelagic food webs. *Limnol. Oceanogr.*, **35**, 202-210.

Richardson, A. J., Silulwane, N. F., Mitchell-Innes, B. A. and Shillington, F. A. (2003) A dynamic quantitative approach for predicting the shape of phytoplankton profiles in the

ocean, *Prog .Oceanogr.*, **59**,

Ryther, J. H. (1963) Geographic variations in productivity. In M. N. Hill, M. N. (ed.), *The Sea.* John Wiley & Sons, New York, pp. 347–380.

Ryther, J. H. (1969) Photosynthesis and fish production in the sea. *Science,* **166**, 72-76.

Sathyendranath, S. (2000) General Introduction. In Sathyendranath, S. (ed.), *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. Reports of the International Ocean-Colour Coordinating Group, No. 3.* IOCCG, Canada, pp.5-21.

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

Shannon, L.V. (1985) The Benguela ecosystem volume 1. Evolution of the Benguela, physical features and processes. *Oceanogr. Mar. Biol. Ann. Rev. 23*: 105-182.

Shannon, L. V., Agenbag, J. J. and Buys, M. E. L. (1987) Large-and mesoscale features of the Angola-Benguela front. *S. Afr. J. Mar. Sci.,* **5**, 11-34.

Shannon, L. V. and Nelson, G. (1996) The Benguela: Large scale features and processes and system variability. In Wafer, G., Berger, W. H., Siedler, G. and Web, D. J. (eds), *The South Atlantic: Present and past circulation.* Springer-Verlag, Berlin, pp. 163-210.

Shannon, L. V., Nelson, G. and Jury, M. R. (1981) Hydrological and meteorological aspects of upwelling in the southern Benguela Current. In Richards, F. A. (ed.), *Coastal and Estuarine Sciences (1). Coastal Upwelling.* American Geophysical Union, Washington DC, pp. 146-159.

Silulwane, N. F. (2001) *A novel quantitative, sub-provincial approach to characterizing the shape of chlorophyll profiles*. M.Sc. dissertation, Uni. Cape Town.

Simpson, J. H. (1981) The shelf-sea fronts: Implications of their existence and behaviour. *Phil. Trans. R. Soc. Lond. A.*, **302**, 531-546.

Smith, R. C. (1981) Remote sensing and depth distribution of ocean chlorophyll. *Mar. Ecol. Prog. Ser.*, **5**, 359-361.

Smith, R. L. (1982) A comparison of the structure and variability in the flow of field of three coastal upwelling regions: Oregon, Northwest Africa and Peru. In Richards, F. (ed.), *Coastal Upwelling*. American Geophysical Union, Washington DC, pp. 107-118.
Steeman-Nielsen, E. (1952) The use of radioactive carbon (C14) for measuring organic production in the sea. *J. Cons. Perm. Int. Explor. Mer*, **18**, 117-140.

Smith, R. C. and Baker, K. S. (1978) The bio-optical state of ocean waters and remote sensing. *Limnol. Oceangr*, **23**, 247-259.

Strub, P.T., Mesías, J. M., Montecino, V., Rutllant, J. and Salinas, S. (1998) Coastal ocean circulation off western South America. In Robinson, A. R. and Brink, K. H. (eds.), *The Sea. Volume II. The global coastal ocean. Regional studies and synthesis.* John Wiley and Sons Ltd, New York, pp 273-313.

Sverdrup, H. U. (1953) On the condition for vernal blooming of the phytoplankton. *J. Cons. Perm. Int. Explor. Mer*, **18**, 287-295.

Taguchi, S., Kasai, H. and Saito, H. (1994) Estimation of vertical distribution of chlorophyll *a* off east Hokkaido by Gaussian curve fitting. *Proc. NIPR Symp. Polar Biol.*, **7**, 17–31.

Uz, B. M. and Yoder, J. A. (2004) High frequency and mesoscale variability in SeaWifs chlorophyll imagery and its relation to other remotely sensed oceanographic variables. *Deep-Sea Res.*, **51**, 1001-1017.

Waldron, H. N. and Probyn, T. A. (1992) Nitrate supply and potential production in the Benguela upwelling system. In Payne, A. I. L., Brink, K. H., Mann, K. H. and R. Hilborn (eds), *Benguela Trophic Functioning. S. Afr. J. Mar. Sci.*, 12. pp. 865-871.

Whipple, G. C. (1899) *The Microscopy of Drinking Water*. John Wiley and Sons, New York.

Wolff, T. and Peterson, M. E. (1991) A brief biography of A. S. Ørsted, with notes on his travels in the West Indies and Central America and illustrations of collected polychaetes. *Ophelia Supplement*, **5**, 669-685.

Yentsch, C. S. and Menzel, D. W. (1963) A method for the determination of phyotplankton chlorophyll and phaeophytin by fluorescence. *Deep-Sea Res.*, **10**, 221-231.

Appendix

The original clusters produced by the clustering program that have been arranged in order of increasing integrated chlorophyll-a concentration.