UNIVERSITY OF SOUTHAMPTON

Development of chlorophyll *a* prediction algorithms for hyperspectral CASI imagery using neural networks

Isabel M J Sargent

Submitted for the qualification of Doctor of Philosophy

Faculty of Engineering and Applied Science
School of Engineering Sciences

December 2000

To the memory of

Sheila Myfanwyn Hall

All colours will agree in the dark

Francis Bacon, 1625, Essays

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE

SCHOOL OF ENGINEERING SCIENCES

Doctor of Philosophy

Development of chlorophyll *a* prediction algorithms for hyperspectral CASI imagery using neural networks

by Isabel M J Sargent

This research addressed the problem of predicting chlorophyll a from remotely sensed imagery in waters which contain suspended sediments, dissolved organic pigments and over which lies an inhomogeneous atmosphere. Image and in situ data were obtained for the North Norfolk Coast region of the United Kingdom. These data were investigated for sources of error and noise and were found to be contaminated by several signals stronger than that of chlorophyll a. To determine whether a non-linear and more complex regression model was more applicable than conventional linear techniques to predicting chlorophyll a in Case 2 waters, the technique of neural network regression was compared to conventional linear regression techniques and was found to give more reliable results. Moreover, information was taken from the training and accuracy of linear and non-linear prediction models to infer details about the relationship between chlorophyll a and spectral data in this environment. Several techniques of feature selection were used to determine which regions of the spectrum were useful for predicting chlorophyll a, given the type of model used. The non-linear neural network feature selection technique consistently selected the noisy blue and near-infrared bands that were rejected by the linear feature selection techniques.

This research clearly demonstrates the applicability of the neural network technique to predicting chlorophyll a. Also, using the novel investigative techniques described in this thesis, new information has been gained about the regions of the spectrum that are useful to the development of robust chlorophyll a prediction models. Unlike previous studies using neural networks, this research used data with a high level of noise and strong effects from environmental factors - a realistic data set. The findings are therefore highly relevant to all remote sensing investigations of the chlorophyll a content of water.

Contents

List of Figures

List of Tables

Acknowledgments

I wish to thank the following people, without whom I would never have made it as far as writing this page:

My supervisor, Dr Adrian Tatnall, for his encouragement, advice and his unfailing confidence in my work.

Dr Martin Brown for showing me that matrix algebra is easy, really.

Dr Hugh Lewis to whom I am enormously in debt for his advice, encouragement, intelligent discussion, proof-reading and of course FLIERS_SFT. Will 10,000 KitKats suffice?

The Environment Agency's National Centre for Environmental Data and Surveillance for providing the data used in this research and particularly Dr Alison Matthews, Rebecca Allen and Kyle Brown who have all gone to great efforts to help me with this research.

The Engineering and Physical Sciences Research Council for funding the research.

Sharon 'the other girl' Benton for really knowing how it feels and much needed trips to Avebury.

My office chums, James Osborne, Franz Newland, Adrian 'Tigger' Tighe, Jon Zerihan, Ben Howell and Ismat Rudwan for a good laugh and a good moan and the *odd* game of footie.

The ISIS 96-ers, Phil 'Phluffy' Jackson, John Manslow, Tony 'Dodgy' Dodd, Alex Bailey, Alistair Channon, Oliver Feminella for being so welcoming on my first day!

Dr Ian 'Westie' West for his alternative insights into PhD-dom.

Dr Steve Beaumont from proof-reading from a sailor's perspective.

Julian Cope for entertaining me through the mid-night hours.

My 'ES family', with whom I have lived and drunk during my time in Southampton including, but by no means exclusively, Gary C, Sally, Airasa, Anneke, Ali A, Simon, Danni, 'Jinja' Simon, Dianne and Ali G-S.

...and finally, for their unfailing encouragement and understanding, my parents, without whom I would not have been here to start this work and Gary Llewellyn who must be mad to have stuck around until the end of it.

Glossary of terms and abbreviations used

Activation function ..	A function that scales values at each node (often of the form of a sigmoid, or a linear scaling or even a thresholding function)
Back-propagation	The process of adjusting the weights backwards through a neural network during training
Case 1 waters	Water whose dominant spectral influence is phytoplankton pigments
Case 2 waters	Water whose spectral signature is dominated by substances other than phytoplankton pigments
CASI	Compact Airborne Spectrographic Imager (an aircraft-borne remote sensing instrument)
CDOM	Coloured Dissolved Organic Matter. Organic breakdown products which are highly absorbing at blue wavelengths and are therefore also known as gelbstoffe or yellow substance
CZCS	Coastal Zone Colour Scanner (a satellite-borne remote sensing instrument)
Endmember	The 'pure' spectral characteristic of a scene component (e.g. vegetation,soil)

Euphotic zone.....	The upper layer of the water within which the downwelling photosynthetically available radiation is at least 1 % of its level just below the surface (Kirk 1994)
Eutrophication.....	The process by which water, over-rich in nutrients, becomes depleted in oxygen as a consequence of enhanced plant and algae growth
Feed-forward.....	The process of passing the inputs to a neural network through the internal weights and activation functions to derive an output
FLH.....	Fluorescence Line Height
Hidden layer.....	Layer of nodes in the network where the sum of the weighted outputs from the previous layer is calculated and then passed through an activation function before being passed on to the next layer of nodes
HPFS.....	Hand-picked feature selection
MLP.....	Multilayer Perceptron
MLFS.....	Multiple linear regression feature selection
NIR.....	Near Infrared
NNFS.....	Neural network feature selection
PCA.....	Principal component analysis
Pixel.....	Picture element - the smallest discrete sample from which a raster image is built
SeaBAM.....	SeaWiFS Bio-optical Algorithm Mini-workshop (a project set up to develop bio-optical algorithms for SeaWiFS data)

SeaWiFS..... Sea-viewing Wide Field-of-view Sensor (a satellite-borne remote sensing instrument)

SLP..... Single layer perceptron

Chapter 1

Introduction

1.1 Background

This thesis describes research into measuring the chlorophyll content of coastal waters, using remotely sensed spectral data and neural networks. Although much previous work has developed reliable methods for the open ocean, the conditions in the coastal margin are far more complex and have therefore attracted a great deal of investigation over recent years.

At the interface between the continents and the oceans, the coastal zones of the world are of vital importance to both terrestrial and marine, biological and chemical cycles (Mackenzie *et al.* 1991; Wollast 1993). As awareness of the impact of human activity on our environment from local to global scales has developed, the importance of the coastal regions has been recognised. Although coastal waters account for only 0.5 % of the global ocean's volume, they account for about 14 % of the global ocean productivity due to nutrient inputs from terrestrial systems via rivers and from bottom waters via currents and upwelling. Also, with about 60 % of the world's population living within a band between 200 m below sea-level to 200 m above sea-level, these waters are the most threatened by contamination and pollution by human activities. In turn, the response

of the coastal ecosystems to pollutants and environmental change can threaten these human activities (Cracknell 1999; Doerffer *et al.* 1999; Downing *et al.* 1993).

A global environmental issue that has received a great deal of attention over recent years is climate change. A major concern is that the emission of carbon dioxide (CO_2) from human activities may be enhancing the natural greenhouse effect that heats the Earth. As the basis of the ocean food chain, phytoplankton use solar energy to fix carbon. The rôle of phytoplankton in the removal of CO_2 from the atmosphere is debated (Miller *et al.* 1995; Downing *et al.* 1993; Wollast 1991) but attempts to model the change in climate resulting from changes in CO_2 have produced wildly differing results (Mitchell *et al.* 1987). It is therefore important that accurate estimates of the quantity and distribution of phytoplankton be available to climate modellers.

There has also been a great deal of concern about the effects of human activity on the health of the ocean. Many coastal regions have elevated quantities of nutrients as a result of runoff from terrestrial and marine agriculture, the disturbance of soils in river basins, industrial processes, tourism, marine agriculture and the suppression of riverine nutrient uptake caused by pollution (Smetacek *et al.* 1991). The resulting increase in algal growth, called eutrophication, threatens fisheries and tourism, and may result in a permanent loss of vital ecosystems. The ability to monitor for the presence of phytoplankton blooms can therefore both aid in the detection of pollution and provide warning of the bloom hazard itself (Environment Agency 1997).

Satellite and airborne remote sensing is the only method by which broad and contiguous regions can be sampled efficiently. In a region as dynamic as the coastal zone the use of remote sensing for monitoring change is crucial. Remote sensing is a useful tool where a study site is too vast or too remote to be studied adequately 'on the ground'. Since few measurements other than the spectral information may be made under most circumstances, algorithms using remote sensing data are usually based on inferences using the imagery alone.

The phytoplankton concentration in the surface of the oceans is usually estimated from

the spectral signature of the chlorophyll pigments it contains. These pigments have a distinctive pattern of absorption throughout the spectrum. However, this signature becomes hard to detect in regions where non-organic sediments and other pigments occur in the water. In coastal regions, the spectral signal can be dominated by light backscattered by inorganic particles. Pigments formed by the breakdown of organic matter are also prevalent in coastal waters and the spectral signature of these can be very similar to that of chlorophyll. Coastal regions are often near to regions of human settlement and sources of wind-borne particles. Therefore, in addition to the influences of other water constituents, the spectral signal detected in coastal regions can be strongly affected by particles in the atmosphere. Owing to the mixture of influences on the atmosphere at the coast, the distribution of these aerosols is often not uniform, making it difficult to correct for atmospheric effects.

The most common method used to determine the quantity of chlorophyll in water from spectral data is the colour ratio technique. This is an empirical method whereby the ratio of a chlorophyll-sensitive waveband to a less chlorophyll-sensitive waveband is regressed against chlorophyll concentration. Another empirical method that has shown great potential over recent years uses chlorophyll fluorescence at 685 nm to predict chlorophyll concentration. A third method that has popularly been used over recent years, models the radiative transfer within water containing known constituents have been developed. Inversion of these models provides a method of prediction of chlorophyll content from measured spectral information (Rast *et al.* 1999).

The research presented here aimed to investigate more deeply the detection of chlorophyll in remote sensing imagery where the dominant spectral signal was from environmental factors such as the atmosphere and suspended sediments in the water. The empirical approach was used to derive chlorophyll prediction algorithms. The data that were used were collected over a coastal region impacted by inorganic sediments and dissolved pigments and also covered by an inhomogeneous atmosphere. A thorough investigation of the information content of the data found noise from many sources to be present. Several feature selection techniques were used to determine useful wavelengths for predicting chlorophyll concentration. Linear and non-linear models of varying com-

plexity were developed for predicting chlorophyll from spectral information using both conventional regression and neural network approaches. Finally, the findings of the different techniques were drawn together to derive a more complete understanding of the linearity and complexity of chlorophyll-spectra relationship across the visible and near-infrared (NIR) spectrum in Case 2 waters.

1.2 Outline of thesis

The remainder of this chapter details how other research has estimated the chlorophyll content of water using remote sensing. Much previous work has concentrated on finding the best single feature to predict chlorophyll concentration. However, the complex way in which the optical elements of the scene interact indicates that several spectral features are necessary to characterise the nature of the chlorophyll-spectra relationship. The model used to relate spectral information to chlorophyll is also addressed. The non-linearity and complexity of this relationship suggests that current linear regression techniques are inadequate. Therefore, the research presented here develops non-linear models for predicting chlorophyll. Neural networks have produced many valid results in remote sensing studies, which are relevant to the current problem. Short reviews of empirical modelling techniques and neural networks in remote sensing are provided. The aims of this research are set out at the end of this chapter.

Obtaining suitable sets of spectral and *in situ* data for detecting chlorophyll in water can be a difficult task because rarely are ship-board and remotely-sensed measurements taken at the same time and place. Few photons are reflected back to the sensor from the water and so the remotely sensed signal is subject to high levels of noise. It is important that the nature of this noise is understood before the data are used for predicting chlorophyll. Chapter 2 describes how the data used in this research were obtained and calibrated. Various techniques were applied to investigate the suitability of these data for predicting chlorophyll and to highlight any potential problems with the data. Previous studies using neural networks for predicting chlorophyll used data

sets, which were not typical of those used for water quality monitoring in the UK. This is the first study to demonstrate the applicability of the neural network technique to a data set, which has a typical level of contamination by oceanic, atmospheric and sensor effects.

Before the use of neural networks is introduced in this research, chapter 3 describes their operation and discusses issues relevant to their implementation. This begins from an understanding of the familiar technique of multiple linear least-squares regression and demonstrates that the neural networks used in this research are an extension of this simple method. Thus, the two techniques may be directly compared. Both techniques allow many models to be developed such that a thorough comparison of model inputs, structure and accuracy may be easily achieved. However, few studies have used these regression algorithms to their fullest potential.

This research was undertaken in two parts. Firstly a primary study was performed to assess the validity of the techniques that have been proposed. The methods used to assemble the data and regress the spectral inputs against chlorophyll using both a linear least-squares method and a neural network approach are described in chapter 4. The results of the primary study are also presented in this chapter and carefully assessed taking into account not only the accuracy with which chlorophyll predictions were made using the models but also aspects of model design and inferences that may be made about the relationship between chlorophyll and spectra.

The second part of the research is then described in chapters 5 and 6. Here, the method of the primary study was refined to extract more information about the relationship between chlorophyll and spectra by using several different methods of selecting spectral features. Although feature selection is a well-researched technique, it is unusual for ocean colour research to apply automated techniques for finding useful spectral bands. Yet these techniques have the potential to highlight extremely useful combinations of bands that may otherwise have been ignored. Using several different sets of inputs, multiple regression and neural network regression are then applied to the data to make a more rigorous test of the validity of the neural network model that had been proposed.

Chapter 7 then discusses the findings of both the primary and secondary studies with reference to the knowledge about the data that had been gained in chapter 2. A summary of the research and the conclusions relating to the project aims are then presented in chapter 8 with recommendations for future work in this field.

The approach taken to analysing the results of this research was a practical and informative way of extracting new understanding about the relationship between remotely sensed data and chlorophyll in Case 2, coastal waters.

1.3 Detecting chlorophyll in remotely sensed imagery

When light from the sun enters a medium, it interacts with the molecules and particles therein and its characteristics are altered. The light measured above a water body has been attenuated by absorption and scattering within the atmosphere and as it crosses the water surface. The constituents of the water and reflection from a shallow sea-bed further attenuate the light before it passes back through the surface and the atmosphere to the sensor. Scattering and absorption have varying effects throughout the spectrum depending on the material involved. Remote sensing aims to determine the nature of these materials in the scene from the spectral information gathered from it.

Three decades ago, Clarke *et al.* (1970) first explored the use of remote sensing for detecting chlorophyll in natural waters. They compared airborne spectrometer data with *in situ* chlorophyll data and found a decline in backscattered blue light relative to green light as the chlorophyll increased.

Phytoplankton contain a combination of pigments that are principally used to absorb light for photosynthesis. Of the three chlorophylls, *a*, *b* and *c*, chlorophyll *a* is the pigment common to all species of phytoplankton (Fischer and Kronfeld 1990). It has

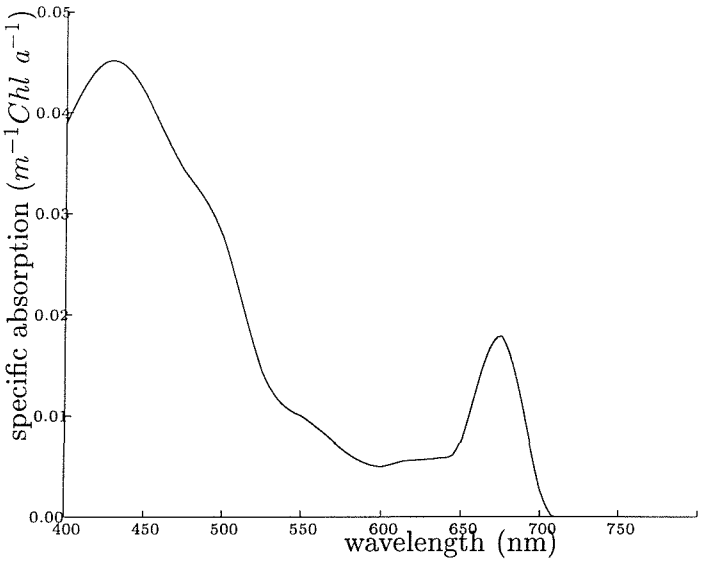


Figure 1.1: Absorption spectrum of chlorophyll *a* in water averaged from measured spectra of algae grown in culture (from Morel 1988)

a distinctive absorption signature in water (figure 1.1) showing absorption in both the blue and red wavelengths. There is a maximum reflectance in the green and NIR wavelengths, resulting in the green hue experienced by the human eye. A chlorophyll breakdown product, phaeophytin *a*, cannot be distinguished spectrally from chlorophyll *a* and so any spectral measurement of chlorophyll *a* in water is actually the sum of the two pigments (Gordon *et al.* 1980).

Absorption of light by water increases with wavelength. At 690 nm, the absorption begins to increase rapidly such that very little signal is detected beyond 720 nm (Malthus 1997).

Above the water's surface, a sensor measures the water-leaving radiance. Here, the reflectance from the water constituents has been altered by passing through the air-water interface. Under calm conditions and at a viewing angle of less than 48.8° to the vertical, the signal that is detected above the water surface has been affected little by passing through the water surface. However, as the roughness of the surface or the viewing angle increases, more of the signal is reflected downwards at the interface (Kirk

1994).

As the height of the sensor increases, the influence of the atmosphere becomes greater. Much of the contribution by the atmosphere is the scatter of light into the path of the sensor. Scatter by molecules, Rayleigh scatter, is greatest at shorter wavelengths. The contribution from this can be approximated if the illumination and observation angles are known. The contribution by scattering from aerosols is more complicated however, because knowledge of the aerosol type and optical thickness is required (Antoine and Morel 1999). The signal is also subject to absorption by molecules within the atmosphere. Principal absorption bands are caused by water at 644-660 nm and around 710 nm, and by oxygen absorption around 687 nm and 760 nm (Fischer and Schlüssel 1990; Gower *et al.* 1984). Some absorption by aerosols also occurs. Again this is difficult to predict because the type of aerosol must be known. Aerosols resulting from industrial processes may be quite common in some coastal regions and can be strongly absorbing (Gordon *et al.* 1997).

The sensor also receives radiance resulting from specular reflectance at the sea surface, known as sun-glint. This changes with the viewing and illumination geometry as with the reflection of upwelling radiance at the underside of the air-sea interface. Therefore, roughening of the sea surface causes an increase in sun-glint (Gordon 1997; Fraser *et al.* 1997).

Since the work of Clarke *et al.* (1970), a common method of determining the chlorophyll *a* content of water has been to use the ratio of the detected radiance in the blue region of the spectrum with that detected in the green. The ratio is a comparison of the region of absorption in the blue with the region of increased reflection in the green. As the chlorophyll *a* in the water increases, the radiance in the absorption band decreases and so a decrease in the blue-green ratio indicates an increase in the amount of chlorophyll *a* (Kirk 1994).

Principally, two blue-green ratios have been used to determine chlorophyll *a*. The first uses measured radiance at 443 nm over that at 550 nm and when the chlorophyll *a*

concentration gets so high that the absorption at 443 nm is near-absolute, radiance at 520 nm over that at 550 nm. As chlorophyll *a* concentration increases, the band of absorption around 443 nm becomes broader and hence absorption can be detected at 520 nm (Gordon and Morel 1983; Kirk 1994).

The remote sensing instrument, the Coastal Zone Color Scanner (CZCS), was launched in 1978 on the Nimbus-7 satellite. Throughout the seven years of its operation, both blue-green ratios were used to determine the concentration of chlorophyll in the global ocean. Studies, such as Gordon *et al.* (1980) found that a linear regression of this ratio against chlorophyll *a* concentration could be derived from ship-board measurements and applied to CZCS imagery very successfully. However in their seminal work, Clarke *et al.* (1970) also noted anomalies in the relationship between ocean colour and chlorophyll *a*, which they thought likely to be due to other constituents in the water, the water surface or the atmosphere. Although the blue-green ratio, rather than a measure at a single wavelength, does compensate for the effects of other materials in the water and the atmosphere (Gordon *et al.* 1980), many studies have found that inorganic sediments and coloured dissolved organic material (CDOM) in the water have had a strong influence on the light detected above certain waters (Morel and Gordon 1980; Mittenzwey *et al.* 1992; Han 1997).

Inorganic sediments are carried into the coastal zone by rivers or are resuspended by wave and current action at the coast (Williams *et al.* 1998). Where present, inorganic particles are usually the principal influence on water-leaving radiance because they scatter light within the water column and back to the sensor (Kirk 1994). The spectral profile of this scatter depends on the size and mineralogy of particles (Novo *et al.* 1989; Han *et al.* 1994). Generally speaking however, a brightening at all optical wavelengths is usually detected (Moore 1978). Some absorption of light may occur which is also very dependent on the mineralogy of the sediment (Han and Rundquist 1994).

CDOM consists of the pigments formed by the breakdown of organic matter. Some of this pigment is formed in the ocean by the breakdown of plankton but much is carried to the coastal zone by rivers. The major influence of this pigment on the received

spectral signal from the water surface is in absorbing light in the shorter wavelengths (Ferrari *et al.* 1996). The absorption of light decreases in an approximately exponential manner with increasing wavelength (Bolgrien *et al.* 1995). This results in a bias to the backscattered light towards the green and red wavelengths and, since the water absorbs more strongly in the red, the result can appear yellow to the human eye, hence the name often given to this pigment - gelbstoffe or yellow substance (Taylor and Smith 1967; Dekker 1993). Dissolved organic pigments do not always co-vary with chlorophyll *a* and, since they also absorb strongly at 440 nm they can result in an under- or over-estimation of the chlorophyll concentration (Carder *et al.* 1989).

To account for the complicating effects of water constituents other than chlorophyll *a*, radiative transfer models have been developed to model the effect of absorbing and scattering materials on the light entering the water (e.g. Fischer *et al.* 1986; Carder *et al.* 1991; Aiken *et al.* 1995b; Kondratyev *et al.* 1998). If the specific absorptions, the concentrations and the scattering characteristics are known for each substance in the water, then many of these studies have shown that the models reproduce measured spectral radiances accurately (Morel and Prieur 1977; Prieur and Sathyendranath 1981; Sathyendranath *et al.* 1989).

These models can then be used to produce spectral data for a wide-enough range of water-quality conditions, so that ocean colour algorithms may be developed and tested (e.g. Sathyendranath *et al.* 1989). Alternatively, the models can be inverted to develop ocean colour algorithms directly. To invert such a model the specific optical properties of all the water constituents in the model must be known (Morel 1980; Morel and Gordon 1980; Doerffer and Fischer 1994) for all wavelengths (Sathyendranath *et al.* 1989). Therefore, the spectral dependence of these optical properties is often assumed to obey a simple law (Morel and Prieur 1977).

By comparing measurements of turbidity and pigment content and in-water spectral measurements with modelled spectral values, Morel and Prieur (1977) were able to investigate the origin of the observed spectra. They showed that the scattering of light was at least as influential in determining the blue-green ratio as pigment absorption.

Their research separated green waters into Case 1 and Case 2 waters. In Case 1 waters, they noted a clear minimum in reflectance at 440 nm that developed as the chlorophyll in the water increased. For these waters, the water-leaving radiance is a product only of phytoplankton pigments. However, they found that the minimum at 440 nm was not visible in turbid Case 2 waters. Here the spectral signature is strongly affected by CDOM, by suspended sediment, or by both. The waters around the coast of the United Kingdom have been classed as Case 2 waters.

Mitchelson *et al.* (1986) found a significant difference between the blue-green ratio to chlorophyll slopes for Case 1 and Case 2 waters and proposed a general blue-green ratio algorithm for determining chlorophyll *a* in Case 2 waters. However, different studies have found that such algorithms are very location- and season-specific (Environment Agency 1997) and that no general algorithms can be produced using this ratio for use in Case 2 waters (Carder *et al.* 1989; Tanis and Pozdnyakov 1995).

In addition to the influence from water constituents, these coastal areas are often subject to non-uniform atmospheric effects, particularly when near industrial or urban zones, which increase the aerosol load of the atmosphere (Tassan and Ribera d'Alcalá 1993). Most of the work that has developed empirical relationships between remotely sensed reflectance and chlorophyll *a* has emphasised the need to correct for atmospheric influences (e.g. Gordon *et al.* 1980; Morel and Prieur 1977). This is because, due to the high absorption of the water itself, the radiances measured over water are very low. Over very clear, deep water, the scatter of light within the atmosphere can account for up to 95 % of the measured signal (Moore 1978). However, over more reflective waters, the contribution from the atmosphere is a little reduced.

Although a popular choice for chlorophyll *a* detection, the blue wavelengths have been found to be noisy as a result of sensor effects (Dekker *et al.* 1992a; Dekker *et al.* 1992b) and atmospheric effects (Aiken *et al.* 1995b). Lathrop and Lillesand (1986) also considered that the poor correlation between the blue Thematic Mapper band and water constituents was possibly because this band represented the integrated signal from deeper within the water body than the reference measurements of the water

constituents. Recently, Kondratyev *et al.* (1998) demonstrated how six different blue-green and green-green ocean/marine algorithms became unusable as soon as the water column became enriched with sediments and/CDOM for lake waters. Malthus *et al.* (1996) noted that at shorter wavelengths the signal is subject to a mixture of absorption and scattering effects from photosynthetic pigments, suspended sediments and CDOM. The response in the longer wavelength bands was however, dominated by phytoplankton pigments. Workers have therefore tried to find other regions of the spectrum from which to obtain information about the chlorophyll *a* content of the water.

Other efforts to detect chlorophyll used radiance in the NIR band, where there is also a band of absorption by chlorophyll *a*. Bukata *et al.* (1974) found that there was a relationship between reflectance in the NIR and the logarithm of chlorophyll concentration. Because CDOM absorption is minimized at longer wavelengths and the effect of suspended sediment on water-leaving radiance is approximately the same for wavelengths longer than 600 nm (Quibell 1991), some studies have suggested the use of band difference and band ratio algorithms which utilise reflectance in the green, red and NIR wavelengths (Rundquist *et al.* 1996; Hoogenboom *et al.* 1998). Mittenzwey *et al.* (1992) found that red wavelengths are most sensitive to changes in phytoplankton. They suggest a NIR to red ratio which seems to work well for waters with a medium to high chlorophyll *a* concentration. When looking at the validity of using Airborne Thematic Mapper data for assessing inland water quality, Malthus *et al.* (1996) also found that a ratio of bands in the red and NIR had the best correlation with chlorophyll *a*. However, Han *et al.* (1994) found that the rate of increase in reflectance with an increase in the level of sediment was not the same for all wavelengths, even in the red and NIR.

The early remote sensing instruments required broad spectral bands to ensure that enough radiation was collected for the signal to be detected. As sensor technology has advanced so instruments have been developed that can collect data for much narrower wavebands. Hyperspectral instruments have allowed near-complete spectra to be collected, so that narrow features in the spectrum, such as absorption bands, may be identified and interpreted. Airborne remote sensing instruments are now available

with approximately the 10 nm spectral resolution recommended by Morel and Gordon (1980), Dekker *et al.* (1992a) and Dekker *et al.* (1992b). With many more bands of information, features in the spectrum other than band ratios and differences may now be computed.

One feature extracted from hyperspectral data is the derivative of the spectrum, usually near a region of particular interest. This technique determines the rate of change of the measured spectrum with wavelength and so can be used to highlight peaks and troughs in the spectrum, or indeed points of inflection. Where reflectance was not highly correlated with chlorophyll *a* and especially for data that are affected by the atmosphere, Malthus and Dekker (1995) found that the first derivatives of the measured reflectance had a high correlation with chlorophyll *a* concentration, particularly in the red wavelengths. Building on this work, Farrington *et al.* (1994) found that, the first derivative spectra of the measurements from an airborne and a surface platform were very similar whereas the subsurface reflectance calculated from both these platforms showed shifts in magnitude. This indicated that more robust algorithms could be derived from the derivative spectrum. Rundquist *et al.* (1996) found that the first derivatives near 500, 700 and 800 nm had a good correlation with chlorophyll *a* concentration. In particular, they found that the derivative at 690.7 nm compared very favorably with a NIR to red ratio. Fraser (1998a) found that peaks in the first derivatives near 429 and 695 nm were likely to be particularly useful for chlorophyll *a* prediction. Goodin *et al.* (1993) found that a difference between the peak in the second derivative at 695 nm and its trough at 660 nm was very stable to changes in the sediment content of water.

Neville and Gower (1977) considered that the effects of the atmosphere and of particulates in the water cannot be simply removed by ratios such as the blue-green. This is because the wavelengths being used are situated far apart in the spectrum and so it is likely that these effects are likely to be different at each wavelength. Therefore, they introduced the idea of measuring chlorophyll *a* using the peak in spectral response at 685 nm. This peak is thought to be due to chlorophyll fluorescence (Gitelson 1992), which results from the re-emission of energy that has been absorbed at about 675 nm (Rundquist *et al.* 1996; Gitelson 1992). Only about 1 % of light absorbed by photo-

synthesising plankton is lost by fluorescence (Kirk 1994) but an increase in fluorescence may be observed to accompany an increase in chlorophyll *a* (Lorenzen 1966). However, there is still some discussion about whether the detected peak is due to fluorescence or is the result of a combination of high-scattering of light and absorption of chlorophyll at 670 nm (Gower *et al.* 1999).

The height and width of the fluorescence feature is expected to increase with an increase in chlorophyll *a*. Therefore, by measuring these dimensions an estimate of the amount of chlorophyll *a* may be made. The height of the observed peak has been found to have an approximately linear relationship with chlorophyll *a* concentration in the range from 1 to 20 $\mu g\, l^{-1}$ (Gower *et al.* 1999).

By simulating the typical responses of coastal waters, Fischer *et al.* (1986) compared the blue-green ratio to three fluorescence features. They concluded that the blue-green ratio was 'useless' for predicting chlorophyll *a* in these conditions. However, all the fluorescence features could be used. Their fluorescence features were: 1) the difference between reflectance at 685 nm and 670 nm - the wavelengths of chlorophyll *a* fluorescence and absorption, respectively, 2) the height of the reflectance above a straight line joining reflectance at 645 and 725 nm, and 3) the height above a straight line joining reflectance at 645 and 670 nm - again this accounts for chlorophyll *a* absorption at 670 nm. Several fluorescence line height (FLH) algorithms have been derived that use features similar to the latter two fluorescence measures of Fischer *et al.* (1986) (e.g. Fischer and Kronfeld 1990; Gitelson 1992; Gitelson *et al.* 1994). The usual method for determining FLH is demonstrated in figure 1.2.

Not only does the peak of measured response at 685 nm increase with an elevation of chlorophyll *a* concentration, but several studies have also found that the peak appears to move to longer wavelengths as chlorophyll *a* increases. This is due to a broadening of the 675 nm chlorophyll *a* absorption band (Gitelson 1992; Matthews 1994), to sediments in the water, to phytoplankton maturity or to stress (Matthews and Boxall 1994).

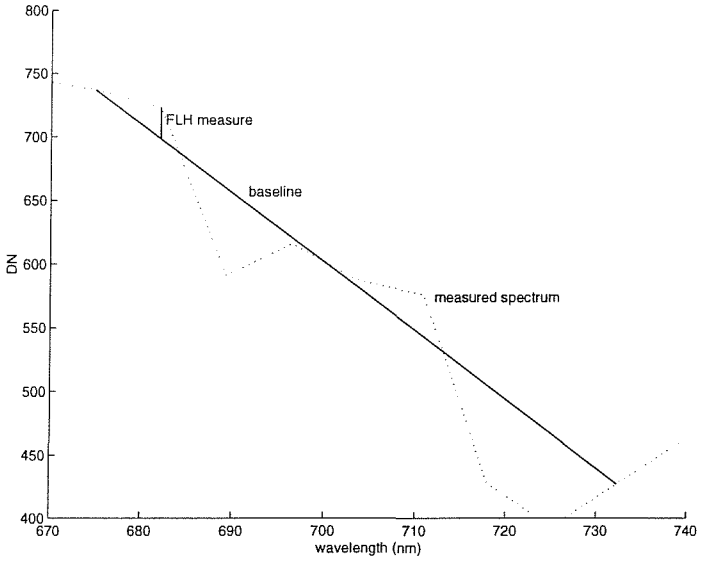


Figure 1.2: Diagram of FLH measurement. A straight baseline is drawn between the response at two wavebands (here, 675 and 732 nm as used by Gitelson 1992) and the height of the spectral response above this baseline at the fluorescence waveband (685 nm) is measured.

The fluorescence peak is in a wavelength region not greatly affected by Rayleigh scatter within the atmosphere (Gower *et al.* 1984). Unfortunately attenuation of light by water is much stronger at red wavelengths, which can result in a very weak fluorescence signal (Malthus 1997). As a result, the detected fluorescence signal originates only in the top 2 m of the water (compared to the top 5 m for shorter wavelengths) and hence is not a good indicator of chlorophyll *a* in the majority of the euphotic zone (Fischer and Kronfeld 1990). Fischer and Kronfeld (1990) also found that the FLH feature is affected by very turbid water. Surface reflectance has also been found to strongly affect this signal (Neville and Gower 1977).

Fischer and Schlüssel (1990) created a radiative transfer model to assess the impact of the atmosphere on the detection of the fluorescence signal through it and concluded that the signal could be detected if the sensor had a spectral resolution of 5 nm. However, few remote sensing instruments have such fine spectral resolution and, for

ocean colour applications, it is possible that the signal received within such a narrow wave band will be too low to keep the signal to noise ratio above their recommended 1000:1.

Fluorescence is a measure of the physiological state of the phytoplankton (Sathyendranath *et al.* 1989). This results in a high variability of the fluorescence signal in response to changes in environmental conditions such as nutrient levels and light availability (Bricaud *et al.* 1995). Ocean colour, on the other hand, is affected by the physical characteristics of the phytoplankton population such as cell size and shape (Bukata *et al.* 1991).

Studies trying to quantify chlorophyll *a* using its spectral signature have found that blue wavelengths are noisy and strongly affected by CDOM. Red wavelengths (including fluorescence) are absorbed by water and/or affected by scatter by both sediments in the water and aerosols in the atmosphere. Fluorescence itself has a very low signal which can have a variable response to chlorophyll *a* concentration. The single feature algorithms have been found to hold too little information for accurate chlorophyll *a* detection.

Morel and Prieur (1977) recommended that where other pigments and suspended sediments are present many more spectral features are required to determine how much chlorophyll *a* there is in the water. For example, Carder *et al.* (1991) found that the CDOM to chlorophyll ratio had a profound impact on upwelled radiance at 443 nm. Degradation products absorb more strongly at 412 nm whereas chlorophyll absorbs more strongly at 443 nm. Therefore, they suggested separating CDOM with a band at 412 nm and using this in ratio with a green band (they used the SeaWiFS (Sea-viewing Wide Field-of-view Sensor) band at 565 nm) with another ratio of 443 to 565 nm.

Gower *et al.* (1984) and Fischer *et al.* (1986) have suggested that the entire visible and some of the NIR spectrum would be required to determine chlorophyll *a* in Case 2 waters. However, owing to the volume of data to be collected, this is usually impracticable. Also, since many wavelengths are highly correlated, much of the information

is redundant. Studies have begun to determine those wave bands which together hold as much information as the entire spectrum. Sathyendranath *et al.* (1989) used a model of ocean colour to find a group of five wave bands, centred at 400, 445, 520, 565 and 640 nm, which held almost as much information as the whole spectrum. This group appeared to be as successful at predicting chlorophyll *a*, suspended sediment and CDOM as the whole spectrum. Dekker *et al.* (1992a) and Dekker *et al.* (1992b) defined 9 bands in the 500-800 nm part of the spectrum with a minimum band width of 10 nm which provided almost all the information of the whole spectrum. Using multiple regression analysis, Wernand *et al.* (1997) found a group of 5 wavelengths from which the whole spectrum could be reconstructed.

When many fine bands of spectral data have been measured, it is often found that these bands are highly correlated. A transformation of such data has been used in several studies to reduce this correlation between features and to reduce the number of features used. Principal component analysis is a popular method by which this is achieved. This is described in more detail in section 2.7.4.

Using principal component analysis, Doerffer (1981) found two characteristic vectors of data acquired over the northern North Sea. The first of these had negative correlations with blue wavelength and positive correlations with the region around 685 nm and so was interpreted to represent chlorophyll *a*. The second was interpreted to represent sediment. Similarly, Gower *et al.* (1984) used eigenvector analysis to determine which parts of the spectrum contained the most useful information for deriving chlorophyll *a* concentration. They found that it is the absorption and scattering of phytoplankton that correlate with the response in the blue-green region of the spectrum and that the information in the red region of the spectrum was mostly found in the fluorescence feature.

1.4 Regression analysis

In modelling problems there are three elements for consideration, the inputs, the outputs and the model which relates them. In the case of this research, measures of the inputs (the spectral measures) and the outputs (the concentration of chlorophyll *a*) will be used. Therefore, the research aims to develop the optimal model to relate them (Brown *et al.* 1994).

Morel and Gordon (1980) described three different approaches to development of models which predict chlorophyll *a* from spectral information: empirical, semi-analytical and analytical. The empirical method statistically relates a small set of measured chlorophyll *a* values to measured spectral values and then applies the algorithm to further spectral values. The semi-analytical method determines the optical effects of materials that co-vary with chlorophyll and those that do not. These are then statistically related to the measured water constituents. The analytical method develops a radiative transfer model by computing the optical consequences of water constituents. The model must then be inverted to determine the concentrations of constituents that produce a measured spectrum. Whilst developing algorithms for use with SeaWiFS data, several semi-analytical and empirical algorithms were tested. The empirical algorithms were found to produce the more reliable predictions (Maritorena *et al.* 1997).

Much knowledge of the water body in question needs to be obtained for the two analytical methods. One of the most common uses for chlorophyll *a* prediction in the UK is for identifying water quality, particularly algal blooms, to enable rapid response by fisheries, environment agencies, etc. (Environment Agency 1997). Therefore the relatively straightforward empirical method is commonly favoured. This method usually uses the least squares regression approach to find the straight-line fit between the feature and chlorophyll *a*.

Most empirical algorithms developed to predict chlorophyll *a* concentration from water-leaving spectra assume a linear relationship. However, this assumption has often been found to be untenable over all but the shortest ranges of chlorophyll *a* concentration

(Sathyendranath *et al.* 1989). Also, the optical signature has been found to be very complex (Bukata *et al.* 1991; Sathyendranath *et al.* 1989). This is due to the interaction between the optical signatures of the water constituents and the optical properties of the water itself (Sathyendranath *et al.* 1989; Fischer and Kronfeld 1990).

The features used as an input to a model are recognised as one of the most important aspects of the model. No matter how complicated a regression model is, it will not extract information that is not there in the features. Thus much of the work aimed at developing algorithms to predict chlorophyll *a* has concentrated on finding the ideal features to regress against chlorophyll *a* and little work has concentrated on finding a more applicable model for the data. However, the linear model often used to relate the spectral information to chlorophyll *a* is inappropriate for Case 2 waters (Bukata *et al.* 1991). A non-linear and rather more complex model would be better (Keiner and Yan 1998) but this may require input features that have not been considered for linear models. Neural networks present a method by which a non-linear function may be optimally modelled without any advance knowledge of the data structure.

Spectral data that are collected over a coastal region are the product of a number of processes within the atmosphere and water and are likely to be affected by instrumental error. Since chlorophyll *a* is the only measure from the scene, the user experiences these effects as variability in the spectral response that is not attributable to chlorophyll *a*. Neural networks have been shown to be more robust to such noise in data (Paola and Schowengerdt 1995a).

Clearly, neural networks have the potential to overcome the difficulties experienced when prescribing the right model for the data set. Many remote sensing applications have used neural networks to solve scene modelling problems. The following section gives a quick overview of the use of neural networks with remotely sensed data.

1.5 Neural networks in Remote Sensing

Neural networks are a simple, iterative method of developing algorithms that relate two sets of variables. These algorithms may be linear or non-linear, and of varying complexity. The technique is therefore ideal for investigations in which linear and non-linear, and simple and complex algorithms are to be compared. The operation of neural networks is described in detail in chapter 3.

Algorithms used to derive information from remotely sensed imagery are of two main types. Regression algorithms, such as those described in the previous section, predict the values for a continuous variable for a given set of observations (inputs). Other algorithms assign observations to discrete classes, rather than regarding them as a continuous variable. These algorithms are known as classifiers and have been the most popular application of neural networks in remote sensing (Paola and Schowengerdt 1995b).

Early classifications using neural networks divided the image into 'pure' classes, usually by assigning each pixel to one single surface type (for example Hepner *et al.* 1990; Benediktsson *et al.* 1990; Heermann and Khazenie 1992). More recently, image classifications have been performed which distinguish the spectral classes within so-called 'mixed pixels'. These so-called 'fuzzy' classifications produce fraction images which show, for each class, the fraction of each pixel which is covered by that class (Adams *et al.* 1986; Atkinson *et al.* 1997; Foody *et al.* 1997; FLIERS 1999).

Some studies have found that the mixtures of class spectral information may be non-linear and have recommended that neural networks be used for spectral unmixing rather than alternative linear methods (Foody *et al.* 1997). Initially, neural networks used for unmixing pixels were trained using pure examples (endmembers) of the class spectra (Civco 1993; Warner and Shank 1997a) but better results have been achieved using examples of mixed spectra to train the network (Foody 1996; Foschi and Smith 1997; Clark and Cañas 1995). The function defined by the neural network to determine class proportions using spectral information is a regression function (Lewis *et al.* 1998).

The pixels in marine imagery are intrinsically a mixture of several different signals. It is therefore logical to extend the technique of spectral unmixing to the spectra measured over water. This concept was first tested on estimating the suspended sediment concentration of water using endmembers derived in the laboratory (Mertes *et al.* 1993) with successful results. If the mixing of signals is not linear, it is logical to 'unmix' this signal using a non-linear technique, such as neural networks.

Although many radiative transfer models assume simple spectral dependence laws these are unlikely to exist. Hence, Morel and Prieur (1977) suggested the use of some iterative fitting algorithm. Neural networks were recommended by Peters *et al.* (1998) to invert the increasingly complex radiative transfer models demanded by the use of imaging spectrometry. A practical application of this recommendation has been carried out recently by Schiller and Doerffer (1999). They used a model to reproduce the top-of-atmosphere reflectance over Case 1 and Case 2 waters and from this derived a set of data for a comprehensive range of phytoplankton, suspended sediments and CDOM concentrations. These data were then used to train a neural network, the trained network thus emulating the inverted model. In a similar study, Buckton *et al.* (1999) trained a neural network on data derived from a model incorporating oceanic, atmospheric and instrument effects. They found that even when the data contained noise from a number of sources, the retrieval of water constituents was within what they considered an acceptable accuracy for Case 2 waters.

An alternative method is to use the neural network as a specialist predictor of one water constituent, rather than of all the constituents. This method is equivalent to producing empirical models for predicting chlorophyll *a* from an ocean colour or fluorescence feature. Four studies have investigated the potential for using neural networks for chlorophyll *a* detection. The first, Keiner and Yan (1998), compared neural network regression with linear and log-linear regression of the three visible Thematic Mapper bands on chlorophyll *a*. A great improvement was noted with using the neural network method. The second study, Keiner and Brown (1999) used the SeaWiFS Bio-optical Algorithm Mini-workshop (SeaBAM) dataset. The neural network outperformed all the empirical and semi-empirical algorithms from SeaBAM. Recognising the value of

previous works using both model-derived and simulated space-borne imagery, Gross *et al.* (1999) and Gross *et al.* (2000) have also used a neural network to develop models for predicting chlorophyll *a* concentration. This work used data derived from the model of Morel (1988) and simulated SeaWiFS data with realistic noise added. In this work it was found that neural networks that have been trained with noise were highly robust to noise and outperformed polynomial models based on blue-green ratios.

Although highly relevant, none of these studies adequately addressed the problem of chlorophyll *a* prediction in Case 2 waters. Thematic Mapper data are expensive to obtain in the UK and have too broad a spectral resolution to show any of the important chlorophyll *a* features (Dekker *et al.* 1992a). Because they are largely based on extrapolated information, the SeaBAM data do not have realistic noise levels. They are also largely of Case 1 regions (Maritorena *et al.* 1997). The data derived from the model of Morel (1988) and the simulated SeaWiFS data used by Gross *et al.* (1999) were all for Case 1 waters only. It is evident that these studies using neural networks need to be extended and applied to hyperspectral data containing noise and contamination of the chlorophyll *a* signal with the atmosphere and other water constituents.

1.6 Aims

The initial premise of this research is that when the remotely-sensed spectrum is contaminated by the signal from inorganic sediment and non-chlorophyllous pigments, the relationship between chlorophyll *a* and spectral values may be noisy and become non-linear and complex. Although the fluorescence signal has been found to have a more linear relationship to chlorophyll *a*, this feature can be noisy, difficult to find and does not give information about more than the first few metres of water.

A wealth of features may be extracted from remote sensing data for prediction of scene characteristics and it is likely that several features will be needed to predict chlorophyll *a* in Case 2 waters. Recent studies into the use of neural networks for

calculating chlorophyll *a* concentration have shown promising results but there is a need to extend this work to the type of data that would be used for water quality studies in the UK. The present research will use a data set that was obtained for the purpose of determining chlorophyll *a* in a region of mixed coastal waters. Previous work using such data has demonstrated the difficulties with using conventional algorithms for determining chlorophyll *a* concentration (Environment Agency 1997; Hill 1998) and has indicated the desirability of using non-linear multiple regression techniques.

The present research investigates the potential for improving empirical methods of predicting chlorophyll *a* by using neural networks. The research will also use realistic data similar to those regularly used by organisations such as the UK Environment Agency, so that the results will be relevant to all bodies concerned with monitoring chlorophyll *a* in coastal waters. Usually, hyperspectral airborne data are used for such studies. These data have a high spatial resolution as well as a high spectral and radiometric resolution. Often many of the spectral measurements are discarded and only a few of the narrow bands used to determine chlorophyll *a*. Feature selection and model selection will be performed and the response of the neural network to new configurations examined. This can provide a great deal of information about the data and can help describe the nature of the relationship between chlorophyll *a* and spectral information. The approach shall be highly analytical, applying mathematical and image processing techniques to the complex problem of detecting chlorophyll *a* in Case 2 waters. By interpreting the results at each stage, this research aims to determine:

1. The separability of the chlorophyll *a* spectral signal in the presence of other environmental factors and sensor noise
2. The severity of the contaminating signals and noise
3. The applicability of neural networks to predicting chlorophyll *a* concentration from spectra contaminated by other environmental factors and sensor noise
4. A comparison of conventional linear regression and neural network regression techniques

5. The nature of the relationship between chlorophyll *a* and different regions of the spectrum

6. The identification of the most appropriate spectral regions for the prediction of chlorophyll *a* concentration

Chapter 2

Data preparation

2.1 Introduction

The first part of this chapter describes how the data were prepared for use in the research. The second part describes a number of preliminary analyses that were performed with the aim of understanding the data before their use.

A key aspect to this research was that it should use data that were typical of those that have been and that shall be used for environmental investigations of the coastal zone. To the analyst with no immediate access to remote sensing instruments and survey vessels, obtaining a suitable data set can pose difficulties. Section 2.2 describes potential sources from which data were sought for this research. The data obtained for this research are then discussed in section 2.3. The preprocessing of the data is described in sections 2.5 and 2.6.

It is important that the quality of the data is fully understood before undertaking an analysis. In so doing, the potential of the data to provide enough information for the task as well as the restrictions imposed by the data, may be determined. A typical data set is far from perfect however, and so a thorough investigation of its sources of error was undertaken. This examined sensor noise (section 2.7.5), atmospheric effects

(section 2.7.6) and errors due to co-registering the image and *in situ* data (section 2.7.7).

2.2 Acquisition of data

Supervised methods of developing regression algorithms require sets of input and output patterns. That is, paired sets of observations of the phenomenon which is to be predicted and the information from which is it to be predicted. In this case, chlorophyll *a* was to be predicted from remotely sensed imagery.

Remote sensing is particularly useful for oceanographic purposes because it is often difficult to take an adequate amount of samples directly from the ocean for a valid study of a subject. However, this inaccessibility is notorious for causing problems with obtaining viable data sets for investigations such as that presented here. Many cruises take *in situ* water measurements and many satellites and aircraft carry sensors over regions of the ocean but few of these activities coincide. A significant portion of this research entailed tracking down a suitable set of *in situ* and image data for the research. Since the findings of this search may be of relevance to other works, a brief summary is given here.

For a decade since the demise of the Coastal Zone Colour Scanner (CZCS) in 1986 there were no space-borne sensors which were optimised for ocean applications. Sensors which were recently launched were investigated. The Modular Optoelectronic Scanner (MOS) was launched in March 1996 and is still operating and data is available from DLR, the German centre for air- and space-research. The Ocean Colour and Temperature Scanner (OCTS) data is available from NASDA (the Japanese space agency) for August 1996 to June 1997. The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) was launched in August 1997 and is still operational with data available for download from the satellite receiving station at Dundee. Imagery from all these sensors was downloaded from the internet. Unfortunately, no *in situ* data were located for any of the imagery during the data acquisition period of this research.

In situ data were sought from the British Oceanographic Data Centre (BODC). Their advice was to find suitable imagery and contact them to find out if cruise data were available. The Reports Of Oceanographic Cruises and Oceanographic Programmes (ROSCOP) database was also searched for suitable data. Unfortunately, due to the time taken to process cruise reports, not until late into this research did the database hold information about cruises sailing after the launch of any of the above sensors.

To secure a valid *in situ*/imagery data set it is necessary to time cruises for satellite overpasses or send airborne sensors out during scheduled cruises. Because there are a number of factors which can prevent *in situ* data from being collected for specific time periods of cruises, the latter scenario is usually the most reliable. The data finally secured for this research were collected in this way - three flights were made over the path of a cruise which regularly sampled the coastal waters of the UK.

2.3 Description of the data

The Environment Agency's National Centre for Environmental Data and Surveillance have an archive of imagery and *in situ* samples. These data were originally obtained for their *Case Study 1* (Environment Agency 1997). From this, one location was chosen for this research which will be referred to as *Norfolk 30/05/96*.

These data comprised three 72 band Compact Airborne Spectrographic Imager (CASI) images (see figure 2.1) of approximately 20 m spatial resolution in *enhanced spectral mode* (the details of the bands are shown in appendix A). These had been flown at about 3000 m over the north Norfolk coast on 30th May 1996 and each image contained a small portion of land for the purpose of geometric correction (section 2.7.5). Concurrent with these over-flights, a cruise followed the coastline taking continuous fluorescence measurements by pumping sea-water into a fluorometer. This flow-through fluorometer emits light at 440 nm and detects light at the 685 nm chlorophyll *a* fluorescence band. At about 606000 m east, the cruise also diverted to take measurements

on a transect that was perpendicular to the main cruise direction. Several samples of water were also taken along the cruise track. The water samples were processed in the laboratory to determine the chlorophyll *a* concentration of the water in $\mu g\, l^{-1}$. These data conform to the quality standards set out by the Environment Agency, which are detailed in Environment Agency (1997) and summarised in table 2.1.

Table 2.1: Environment Agency data quality requirements

Data	Quality criteria
Laboratory	• Date, time and location of sample must be known • Extreme values ($> 20\mu g\, l^{-1}$) must be checked with the laboratory • At least 7 samples must be spatio-temporally associated with the fluorometer measurements and • at least 6 of these must have values greater than the laboratory limit of detection
Fluorometer	• Date, time and location of measurement must be known • Data must be free of noise • At least 10 fluorometer readings must be recorded within 250 metres of each laboratory sample site and • within 30 minutes of the laboratory sample being taken • At least 15 km of continuous track fluorometric readings must lie within the geographical region of the CASI imagery
CASI image	• Date and time of image must be known and images must be geometrically corrected • Edge brightening and glint must be minimal • Images should be free of cloud
General	• The chlorophyll *a* concentrations in the region should vary by at least $4\mu g\, l^{-1}$ over a 10 km stretch of water

The *Norfolk 30/05/96* data contained the laboratory measured chlorophyll *a* for 17 water samples. This information was used to calibrate the 3142 fluorometer measure-

ments. The calibrated fluorometer data were used as a measure of chlorophyll *a*. This calibration is described in section 2.5. No ground measurements of downwelling irradiance were made *in situ* and so radiance measurements could not be converted to reflectance. Thus image spectral measurements shall be referred to as digital numbers or DN.

2.3.1 The location of the research

The north Norfolk coast is located in the eastern region of the UK. The region of the research faces north and north-east into the North Sea (see figure 2.2). The north Norfolk coast is designated a Heritage Coast and Area of Outstanding Natural Beauty because of its exceptional scenic qualities. The region also contains a Ramsar site (wetland of international importance) and Special Protection Area, both of which commit the UK to conserving the habitats in the region (Barne *et al.* 1995). The region is affected by a plume of sediment during parts of the year although around the time the data were collected (May) the plume becomes discontinuous (Dyer and Moffat 1998). Also, the North Sea has been found to have high levels of coloured dissolved organic matter (CDOM) (Morel and Gordon 1980).

The regime of the north-facing coast tends to be that of sediment deposition (Dyer and Moffat 1998) as is demonstrated by the spit features and marshes from Brancaster Bay to Blakeney Point. From Sheringham eastwards the north-east coast has a more erosional regime as demonstrated by the cliffs at Cromer. North-easterly and southerly winds occur in equal amounts at the time of year when the data were collected. East of 600000 m, the water is around 10 m deep within the region of the images. West of 600000 m, the water is shallower (Barne *et al.* 1995). April to May is the time of a phytoplankton bloom although it is likely that this bloom will have receded somewhat by the time of sampling (the end of May). The recession of a phytoplankton bloom can result in increased levels of CDOM as the organic matter breaks down (Carder *et al.* 1989).

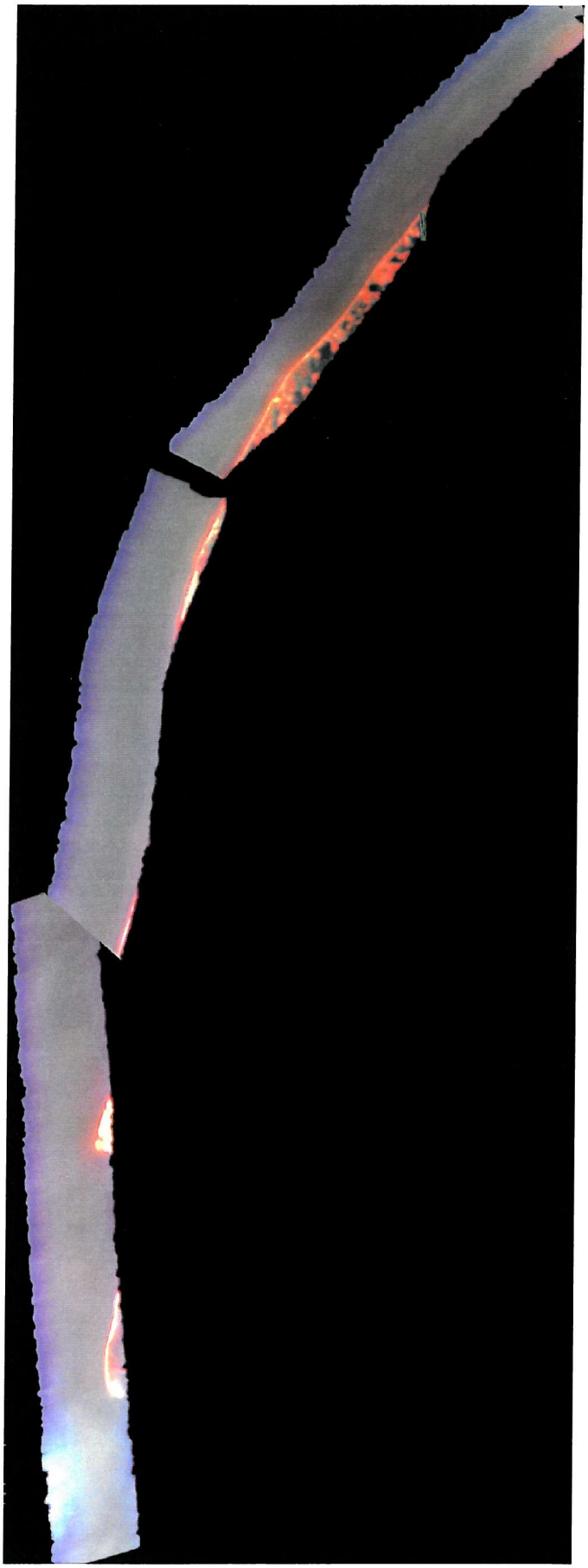
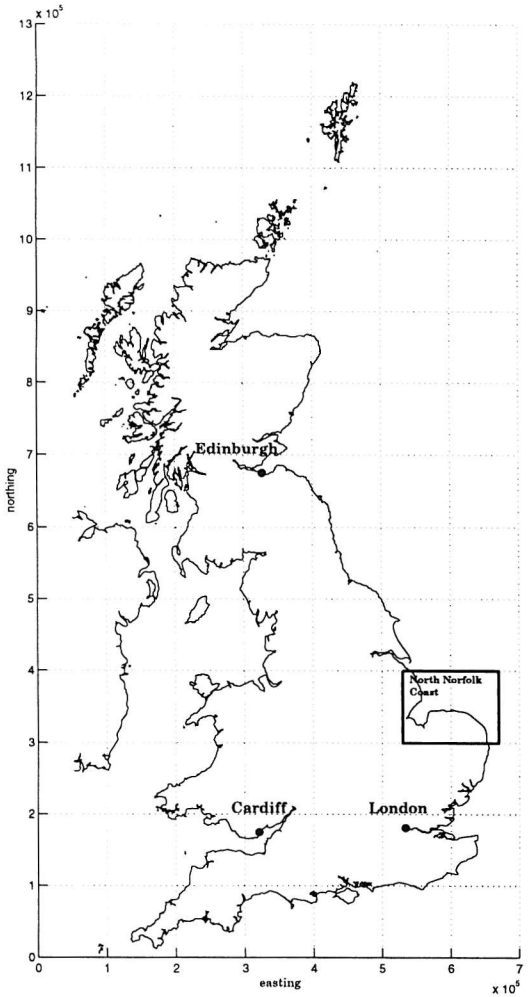
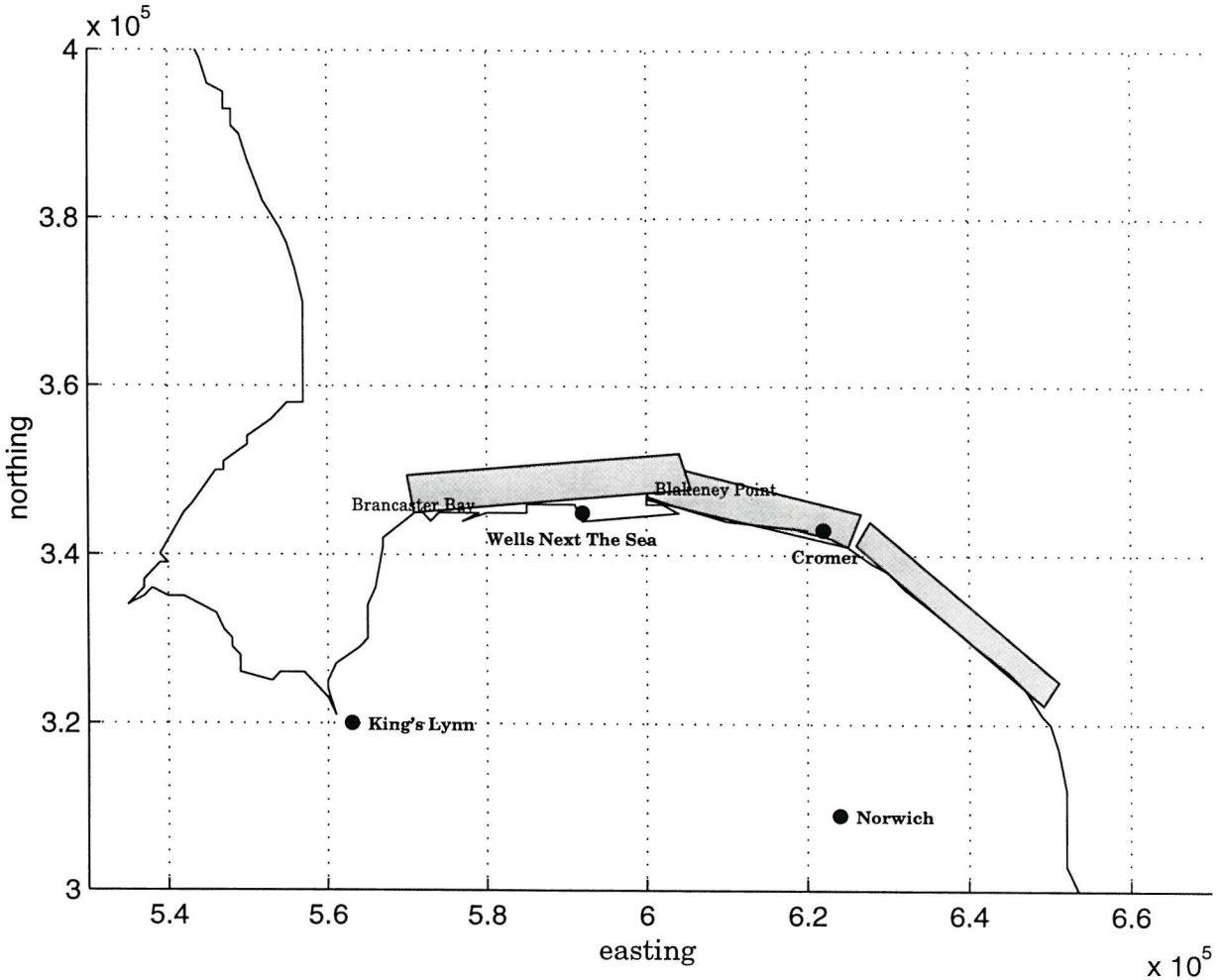


Figure 2.1: CASI data used for this research with bands 6 (441 nm), 20 (540 nm) and 41 (689 nm) set to blue, green and red, respectively. From left to right the images are 1877, 1876 and 1875.



(a) The United Kingdom



(b) The North Norfolk Coast

Figure 2.2: The location of the images off the North Norfolk Coast

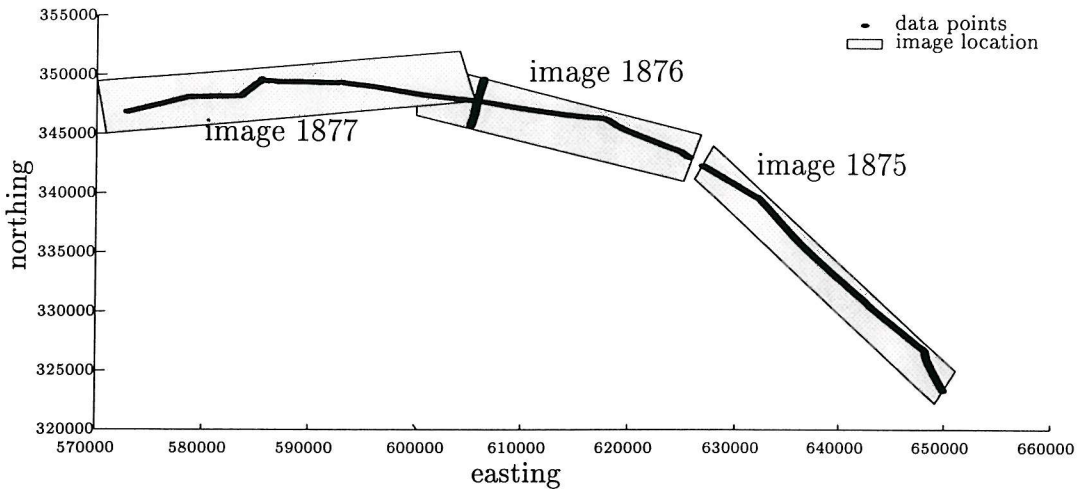


Figure 2.3: The locations of the three images are shown here with the location of the *in situ* cruise samples superimposed.

2.3.2 Description of CASI imagery

CASI is a aircraft-mounted pushbroom spectrographic imager. Light entering the detectors is collimated and then dispersed into wavelengths onto a charged coupled device (CCD) array for which the rows represent the image rows and the columns represent the spectral bands (Babey and Anger 1989; Keller and Fischer 1998). The CCD array creates the digital output, which is then recorded onto magnetic tape ready for processing.

CASI was designed to be convenient and inexpensive to operate whilst providing high sensitivity and resolution (Babey and Anger 1989). It has the ability to change the spectral and spatial configuration to suit the intended use. Three modes are available: *spatial mode*, *spectral mode* and *enhanced spectral mode*. *Spatial mode* has full spatial coverage but only about 15 spectral bands in the range 400-920 nm. *Spectral mode* has full spectral coverage of 288 spectral bands but the sampled pixels do not give contiguous spatial coverage. *Enhanced spectral mode* allows contiguous spatial coverage with the possibility of a large number of spectral bands produced by combining bands in the full set of 288.

The spatial resolution of the imagery is determined by a number of factors. The width of the pixel is determined by the field of view and the altitude of the sensor. Typically, at a height of about 3000 m, the pixels have a width of about 4 m. The length of the pixel is determined by the speed of the aircraft and the data integration time. In *spatial mode* the integration time can be short enough to ensure square pixels. As the number of desired bands increases, the integration time needs to increase to ensure that enough photons are collected. Thus, the pixels of the *enhanced spectral* and *spectral modes* are longer in the flight direction. For example, the 72 band set used in this research would have had an original pixel size of about 4 x 15 m. During geometric correction the data are resampled such that the pixels may be considered square although some loss of resolution is incurred (Andrew Wilson, NERC Centre for Ecology and Hydrology, personal communication; Kyle Brown, Environment Agency, personal communication). These data were resampled to 20 x 20 m.

2.4 Software used for this research

Geometric correction of the imagery was performed by staff at the Environment Agency's Centre for Environmental Data and Surveillance within the PCIWorks toolkit for which CASI data are formatted. Images were then imported into ERDAS Imagine and ENVI for image processing and extraction of spectral values (section 2.6). Calibration of *in situ* samples, least-squares regression and analysis of point-data and results were performed within Matlab. The FLIERS SFT2.0 software (Lewis 1997) was used for neural network algorithm training and testing.

2.5 Calibration of the fluorometer data

Since Lorenzen (1966) many cruises have measured *in vivo* fluorescence to determine the *in situ* chlorophyll *a* content of water (for example, Neville and Gower 1977; Steele and Henderson 1979; Gordon *et al.* 1982; Strutton *et al.* 1996; Strutton *et al.* 1997).

These data are quick to obtain and thus enable many measurements to be made within a reasonable time-window of the aircraft over-pass. Alternative methods use only laboratory processed samples of sea water, which are time consuming to collect and so often result in too few samples per image to be viable. A recent study by Nieke *et al.* (1997) found that *in vivo* fluorometer detection of chlorophyll *a* compared well to standard laboratory as well as laser detection methods.

Chlorophyll *a* fluorescence is a measure of the reaction of the pigment to the light source. This is a product of its physiological condition. It is therefore a particularly suitable measure of chlorophyll *a* for studies interested in primary production. If quenching (where the turbidity of the water is high and much light is absorbed) does not occur and there is no scattering of light by particles in the water, the relationship between fluorescence and chlorophyll *a* is linear (Lorenzen 1966; Yentsch and Menzel 1963). Variation in the measured fluorescence can be caused by many factors including the configuration of the instrument, the physiological state of the phytoplankton and the phytoplankton species and even the time of day (Estrada *et al.* 1996). It is not possible to determine all these factors *in vivo* so some error in determining the chlorophyll *a* concentration may be expected (Cunningham 1996; Hanelt 1996; Carder *et al.* 1989). Additionally, some error is expected in the determination of chlorophyll *a* in the samples taken for the calibration of the fluorometer measurements because it is not possible to remove all accessory pigments or ensure that identical techniques are employed in all laboratories (Gieskes and Kraay 1983; SCOR-Unesco 1966; Aiken *et al.* 1995b).

The fluorometer measurements were calibrated to chlorophyll *a* in $\mu g\, l^{-1}$ using the 17 water samples which had been processed for chlorophyll *a*. It was assumed that a linear relationship existed between the fluorometer readings and the chlorophyll *a* concentration of the water and that the relationship remained constant throughout the cruise (Yentsch and Menzel 1963; Holm-Hansen *et al.* 1965; Lorenzen 1966).

A fluorometer reading was found for each laboratory datum using the time fixes provided with both sets of measurements. Four of the laboratory data were discarded

because no fluorometer reading could be found within a short time interval, or because no fix had been provided.

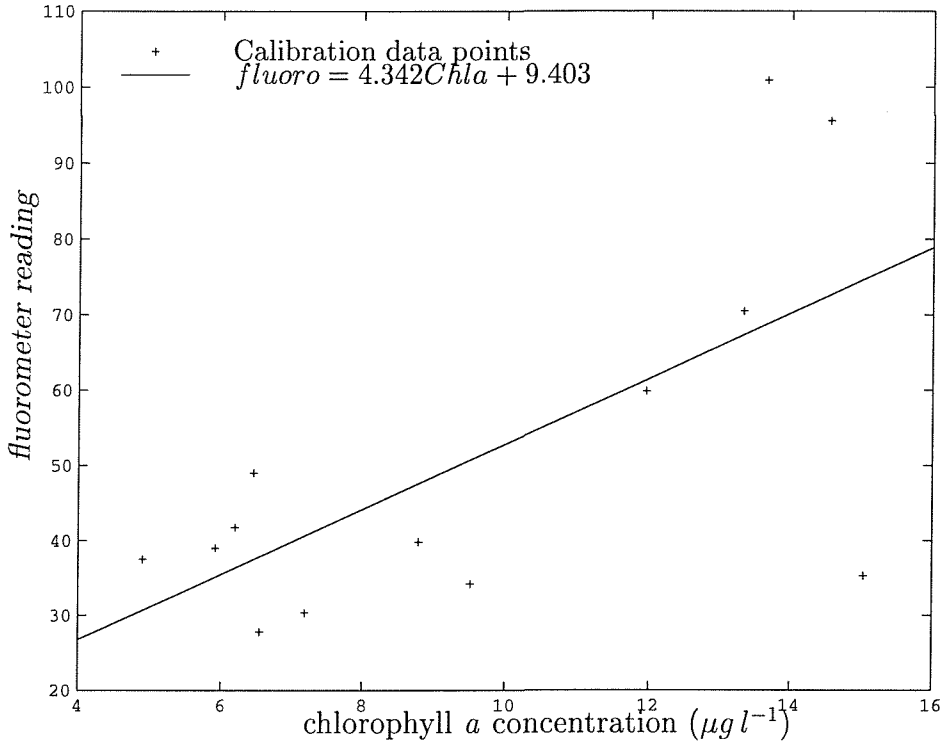


Figure 2.4: Calibration of the fluorometer data using the laboratory water samples.

A first-order polynomial was fitted to the remaining 13 data points using the least squares method. The relationship

$$fluoro = 4.342Chla + 9.403 \tag{2.1}$$

was found where *fluoro* indicates the fluorometer reading and *Chla* indicates the chlorophyll a concentration in $\mu g\, l^{-1}$. Inverted, this produced the calibration

$$Chla(\mu g\, l^{-1}) = 0.230 fluoro - 2.163 \tag{2.2}$$

as illustrated in figure 2.4. This relationship was used to calibrate all 3142 data points to chlorophyll a in $\mu g\, l^{-1}$.

The correlation of this calibration is 0.664 which is significant to within the 10 % level. A better correlation would be desirable if absolute chlorophyll a predictions were needed. However, it was intended that this research make a comparison of algorithms

for predicting fluorometer-measured chlorophyll *a*. This measure will be referred to as chlorophyll *a* from this point forward.

2.6 Extraction of spectral information

It is common to use some form of averaging window when extracting information from images of water scenes (for example Doerffer *et al.* 1989; Malthus and Dekker 1995). This is intended to smooth out noise and/or account for current and tidal effects and uncertainty in the registration of images to sample points. However, the research should account for the underlying scale of variation of chlorophyll *a* in the water which is often of the order of a few tens of metres in coastal regions (Steele and Henderson 1979).

A moving window was chosen that found for each pixel the weighted mean for a window around that pixel. The pixels within this window were weighted using a two-dimensional Gaussian function such that the mean was biased towards values in the middle of the window. Such an averaging window smoothes out the noise in the data whilst giving priority to the measurements made nearest to the location of the pixel. Initially, this was assessed for several window sizes. This assessment was performed only for image 1876 as it would have been too computer intensive to have used all three images. The original research using these data (Environment Agency 1997) performed this correlation test for a FLH calculated from the imagery and using a window size of up to 21 by 21 pixels. Window sizes above 21 by 21 pixels were considered too large to be valid. This window size represented an area of approximately 420 m^2 on the ground. The digital numbers in several image bands for several different moving window sizes were calculated and compared to the chlorophyll *a* measurements using the correlation coefficient:

$$\rho = \frac{\sum_{i=1}^{n}(Chl_i - \bar{Chl})(DN_i - \bar{DN})}{\sqrt{\sum_{i=1}^{n}(Chl_i - \bar{Chl})^2 \sum_{i=1}^{n}(DN_i - \bar{DN})^2}} \tag{2.3}$$

where DN_i and Chl_i are the ith digital number and chlorophyll *a* measurement respectively, $\bar{DN}$ and $\bar{Chl}$ are the means of these observations and n is the number of observations. This measure indicates the amount of association between the chloro-

phyll *a* concentration and the spectral measurements.

Table 2.2 gives the values for these coefficients for various window sizes.

Table 2.2: Correlation coefficients for different moving window sizes around fluorometer data points

	Weighted mean for CASI band:					
$\Sigma^{\dagger}$	15	20	30	40	60	$\mu^{\ddagger}$
1	0.5080	0.5323	0.6863	0.6852	0.4318	0.5687
7	0.6502	0.6305	0.7114	0.7267	0.5410	0.6520
11	0.6483	0.6339	0.7103	0.7265	0.5468	0.6531
15	0.6477	0.6352	0.7096	0.7270	0.5515	0.6542
21	0.6463	0.6364	0.7090	0.7273	0.5540	0.6546

† Size of sides of window in pixels (one pixel has an edge of approximately 20 m).

‡ Mean correlation coefficient for each window size.

Band 40 (682 nm) is the band closest to the fluorescence peak and hence the correlations with the fluorometrically-derived chlorophyll *a* concentration are the highest. Generally, as the window size increased, there was an increase in the correlation of chlorophyll *a* measurements with spectral data obtained from the window. This increase was consistent with the findings of the Environment Agency in their study using this data (Environment Agency 1997) and, although the increase in the mean values was not statistically significant, a moving window of 21 by 21 pixels was selected for comparability with their work.

The spectral data used in the rest of this research was then extracted from the imagery using the weighted mean method and a window of 21 by 21 pixels. The three sets of spectral values, corresponding to the three images, were then concatenated. Because the images 1876 and 1877 overlapped, some data points were removed by giving priority

to the data which had a shorter time interval between the over-flight and the *in situ* sampling. This resulted in pixels from image 1877 being chosen where they coincided with pixels in image 1876.

Several more data points were removed because they fell outside of the image range and others were removed because the window around the data point lay partly outside the image. This resulted in 2300 chlorophyll *a* measurements (calibrated fluorometer readings) with corresponding spectral readings in each CASI band.

2.7 Pre-study analysis of data

Some preliminary analyses were performed to enable a full knowledge of the data. This involved determining some basic statistics of the data and using cluster analysis and principal component analysis to determine the information content of the spectral data.

It is widely acknowledged that spectral data collected over water surfaces is subject to high noise levels. This is because the signal received from the water body itself is so low that any other source of interference can have a significant impact. These impacts can be dependent on the wavelength and region of the image. Further analyses assessed the potential sources of error in the data.

2.7.1 Statistical analysis of the data

The chlorophyll *a* data ranged in value from 3.12 $\mu g\, l^{-1}$ to 22.61 $\mu g\, l^{-1}$ with a mean value of 8.96 $\mu g\, l^{-1}$. A histogram of chlorophyll *a* values is shown in figure 2.5. Much of the data had chlorophyll *a* concentrations of between 5 and 7 $\mu g\, l^{-1}$ with over 1000 points having values in this range. There were also two lesser peaks in chlorophyll *a* values at about 13 $\mu g\, l^{-1}$ and 21 $\mu g\, l^{-1}$.

The mean spectrum and the standard deviation of the spectral values were determined

for each image (figure 2.6). The mean spectra displayed a typical peak in the blue and green wavelengths (refer to appendix A for a description of wavelengths corresponding to each band) and a sharp decline as water-absorption increases towards longer wavelengths. A narrow trough was visible at approximately 440 nm (band 6), which corresponded to the chlorophyll *a* absorption band here (Wrigley *et al.* 1992). Several peaks and troughs were visible in the blue and green region, which corresponded well between images. Troughs at bands 45 (718 nm) and 51 (761 nm) were well defined and probably resulted from atmospheric water (Fischer and Schlüssel 1990) and oxygen (Wrigley *et al.* 1992) absorption, respectively. The standard deviation of these spectra ranged from around 500 DN in the blue to about 100 in the NIR. Although the overall shape of the spectra were the same, there was some difference between the spectral values of the three images with 1875 having the highest values and 1877 having the lowest values. This may be a consequence of some environmental factor such as increased scattering by the atmosphere or suspended sediment, which caused overall brightening in some images more than others, a decrease in the sensitivity of the sensor throughout the course of the flight or even differences in the calibration of each image. No adjustment was made to the image values to reduce these differences between the images.

2.7.2 Correlation analysis of the data

Figure 2.7 illustrates the correlation (equation 2.3) found between each CASI band and chlorophyll *a*. With 2300 data points all correlations above ~ 0.1 may be considered to be significant. The highest correlations were in the red wavelengths, particularly around 689 nm (band 41). The lowest correlations were for bands 1 to 4 (401 to 427 nm). Band 72 (914 nm) had the only negative correlation with chlorophyll *a* at -0.73. This was found to be because quite a few values were very low or zero and these tended to correspond to the higher chlorophyll *a* concentrations. This was not observed in any other band and was due to some error with the calibration of the imagery. Without these values, band 72 had a correlation of 0.35 to chlorophyll *a*

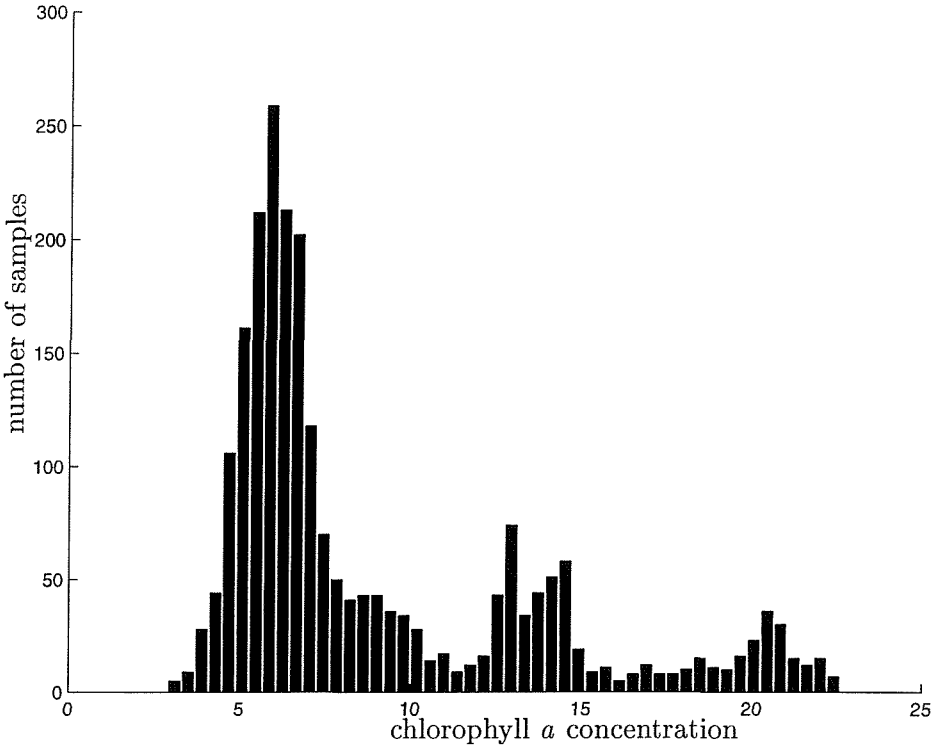


Figure 2.5: Histogram of chlorophyll *a* concentration. The mode of the values was between 5 $\mu g\, l^{-1}$ and 7 $\mu g\, l^{-1}$.

concentration. All the data points, including those with zero values in band 72 were kept in the data set.

Many chlorophyll *a* detection algorithms use a ratio of two bands. The correlation of each 2-band ratio with chlorophyll *a* is plotted in figure 2.8. The highest correlations were for the ratio of bands 41 (689 nm) to 39 (675 nm) and for bands 29 (604 nm) to 27 (589 nm) and for ratios of longer red wavelengths with green wavelengths.

The correlation between the bands was also assessed. Figure 2.9 illustrates this. Most wavebands had a high correlation with each other. The lowest correlations were for blue wavelengths and band 72 with the rest of the spectrum. A distinct boundary between the visible and the NIR was apparent as comparatively low correlations between bands of the two spectral regions.

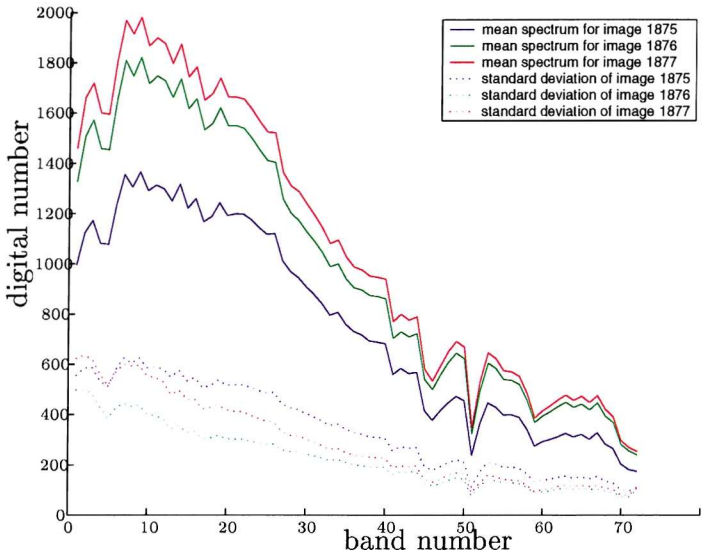


Figure 2.6: Spectral mean and standard deviation of images. The mean digital number and standard deviation of this value is plotted for each band in each of the three images. These values have been calculated for the pixels representing water only.

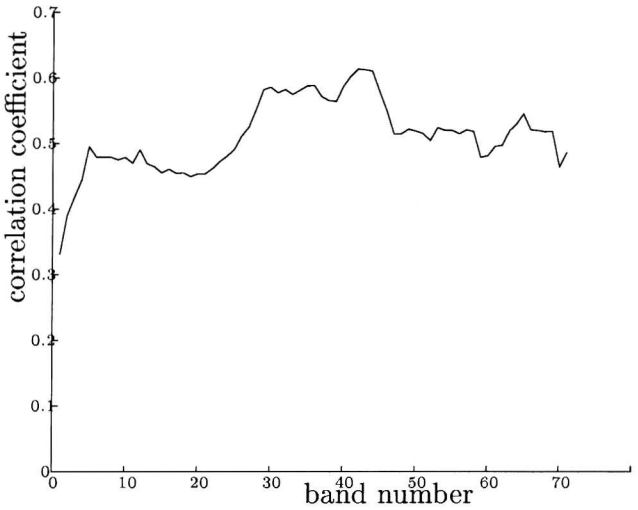


Figure 2.7: The correlation of CASI bands 1 to 71 with chlorophyll *a*. Band 72 had a correlation coefficient of -0.73 (see text).

2.7.3 Cluster analysis of data

To assess whether there was enough information in the spectral bands for prediction of chlorophyll *a*, an initial visual technique was used. This examined the clustering in the

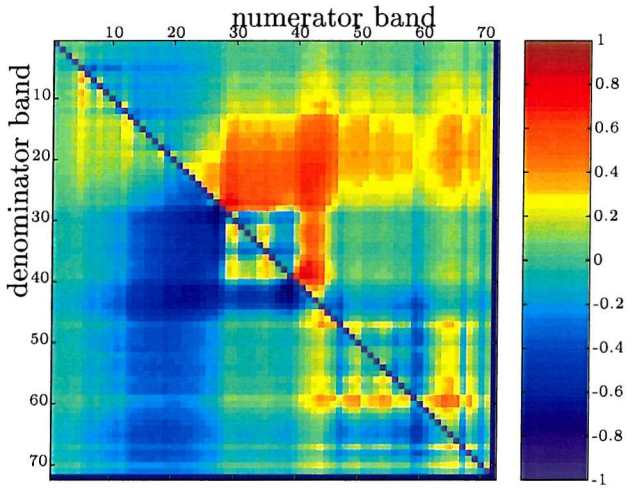


Figure 2.8: Correlation of the 2-band ratios of the 72 CASI wave bands with chlorophyll *a*.

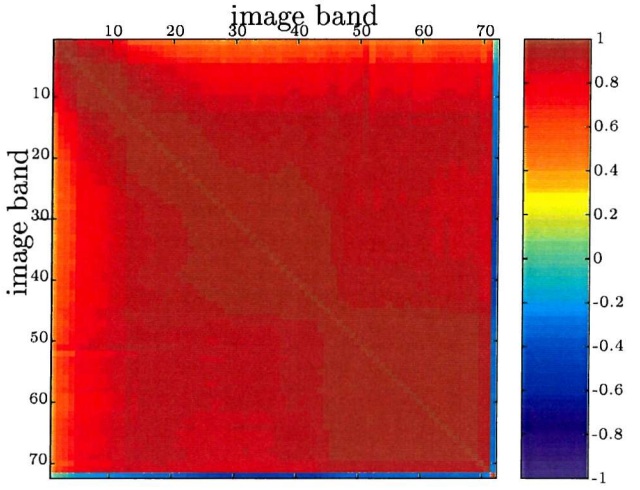
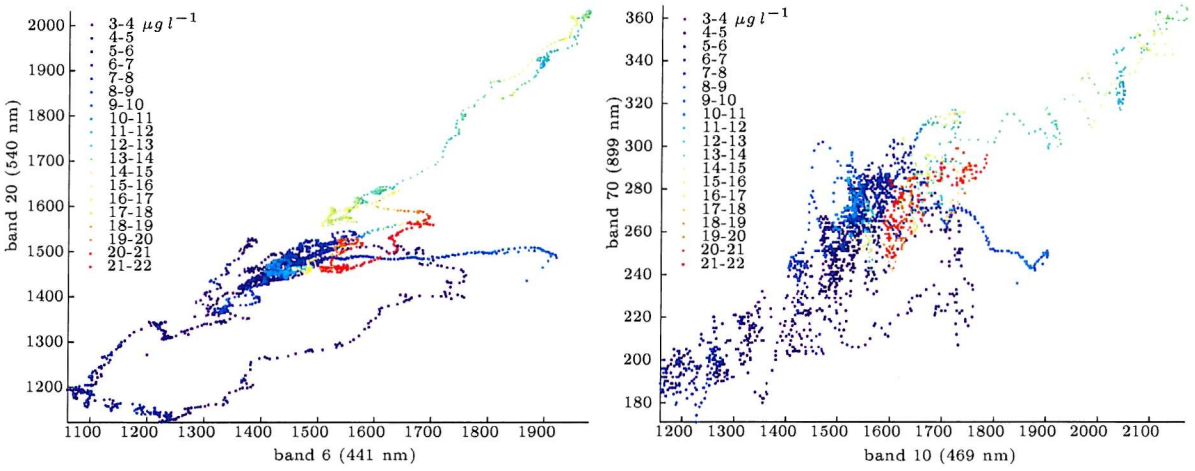


Figure 2.9: Correlation of the 72 CASI wave bands (normalized to zero-mean, unit variance) with each other.

feature (band) domain of data representing similar chlorophyll *a* concentrations. This entailed extracting spectral measurements for data points with short ranges of chlorophyll *a* concentration and plotting them in two-dimensional feature space. Figure 2.10 demonstrates this for two pairs of bands. In these figures, each short spectral range is represented by a different colour such that the progression of data points in feature

space can be followed as the chlorophyll *a* concentration increases.

Figure 2.10a shows a typical example for much of the *Norfolk 30/05/96* data. Curvilinear features are visible which may have been due to changes in chlorophyll *a* along the cruise track and which correlated with gradual changes in spectral values due to atmospheric, oceanic or sensor effects. Such features were not so evident when plotting short ranges of chlorophyll *a* against bands 10 (469 nm) and 70 (899 nm) in figure 2.10b. From these figures it was evident that clustering of similar chlorophyll *a* values does occur in feature space but with a change in chlorophyll *a* (as represented by a change in colour) the change in location of the cluster in feature space was not uniform. This indicates the complexity and non-linearity of the relationship between chlorophyll *a* and spectral values.



(a) band 6 against band 20

(b) band 10 against band 70

Figure 2.10: Cluster analysis of data. These figures demonstrate how well data with similar chlorophyll *a* concentrations cluster together in feature space.

The covariance matrices for each range of chlorophyll *a* concentration were also calculated. The trace of these was taken as a measure of the overall variance within each chlorophyll *a* range and those with a high variance were investigated further by finding two bands in the covariance matrix with a high covariance and plotting the data

points against these bands. This technique highlighted a number of anomalies in the data. The data points in the range 6-7 $\mu g\, l^{-1}$ displayed two distinct clusters in feature space for which one cluster had higher spectral values for both bands. The data points from these two clusters were plotted according to their geographical positions and the clusters were found to correspond to images 1877 and 1875. The values in image 1877 had the higher spectral values, as was found in section 2.7.1.

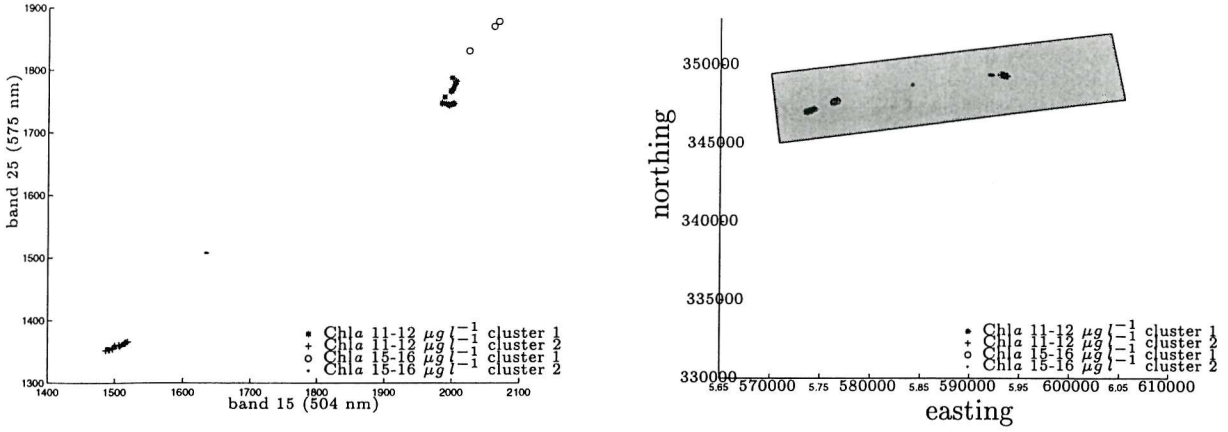
Other ranges of concentrations with high variances were 11-12 $\mu g\, l^{-1}$ and 15-16 $\mu g\, l^{-1}$. Distinct clusters were found in these ranges in spectral space, which corresponded to distinct clusters in geographical space. These highlighted patches of brighter values to the west of image 1877 at about 577000 m east (see figure 2.11) and also in image 1875 at 630000 m east. These bright patches may have been due to haze, or a highly reflective suspensoid, reflection from the bottom or sun-glint on the water surface. Cluster analysis highlighted these bright regions because, by choosing small ranges of chlorophyll *a* concentration, very localised areas in the data were being analysed.

2.7.4 Principal component analysis of the data

Principal component analysis (PCA) transposes the data, X onto a new set of axes such that the maximum variance in the data is projected onto the minimum number of axes:

$$Y = \Phi X \tag{2.4}$$

where Y is the transposed data and Φ the matrix of transform vectors. Fischer (1985) used principal component analysis to determine how many independent optical components of the water could be detected given variation in scattering and absorption by sediments and the atmosphere. Fischer (1985) used a radiative transfer model to generate the data and so were able to control effects such as sensor error. With measurements at 6 wavelengths they found that at most two water substances caused a significant change in spectral information. Also, using model-generated data, Sathyendranath *et al.* (1989) found that enough orthogonality existed between suspended sediments, chlorophyll and yellow substance to permit their retrieval at least in some cases. They



(a) feature space

(b) geographic space

Figure 2.11: Identifying anomalies in the spectral data with cluster analysis. The left figure shows data points for image 1877 with chlorophyll *a* concentrations between 11 and 12 $\mu g\, l^{-1}$ and 15 and 16 $\mu g\, l^{-1}$ plotted in feature space defined by bands 15 (504 nm) and 25 (575 nm). The two clusters that are found for each range of chlorophyll *a* are displayed with different symbols. These same clusters are then plotted in geographic space in the right figure.

found that the first principal component was highly correlated with suspended sediments. Gower *et al.* (1984) found that the 2nd and 3rd principal components of their reflectance spectra, collected off the coast of Canada, related to chlorophyll. Fischer *et al.* (1986) found the 2nd principal component of their model-derived data correlated to chlorophyll if CDOM was not present. However if CDOM was present, the 2nd principal component correlated with CDOM and chlorophyll with the 4th principal component.

A similar technique, factor analysis (Comrey 1973; Otsu 1984; Cooley and Lohnes 1971), identifies correlations between factors (eigenvectors of the correlation matrix) and variables. Using this technique, Doerffer (1981) found two eigenvectors in 16-band airborne sensor data from the northern North Sea the first of which corresponded to chlorophyll and the second to sediment. However, when applying factor analysis to

the covariance matrix, Doerffer *et al.* (1989) found that neither chlorophyll *a* nor Gelbstoffe signals could be separated out from Thematic Mapper measurements made over the sediment-rich German Bight waters of the North Sea. Their principal factors corresponded to suspended sediments, atmospheric turbidity and temperature. This second study had fewer than half the number of spectral bands of the first and this may explain why the more complex chlorophyll *a* signal could not be separated.

An assessment of the principal components of the *Norfolk 30/05/96* data was undertaken to provide an insight into the linearly separable components of the data. Principal component analysis may be performed using a number of techniques. Usually the eigenvectors of the covariance matrix, Σ_X are found such that

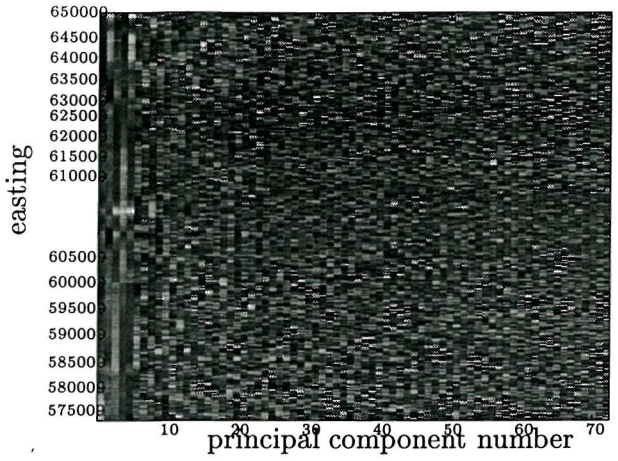
$$\Sigma_X = \Phi \Sigma_Y \Phi^t \ . \tag{2.5}$$

With the restriction that Y is orthogonal, Σ_Y is a diagonal matrix describing the variance of the data along each transpose axis. These are the eigenvalues corresponding to the vectors in Φ. Sometimes the eigenvectors are determined for the correlation matrix of the data. A less common method uses singular value decomposition (SVD). In this case the matrices U, S and V are found such that

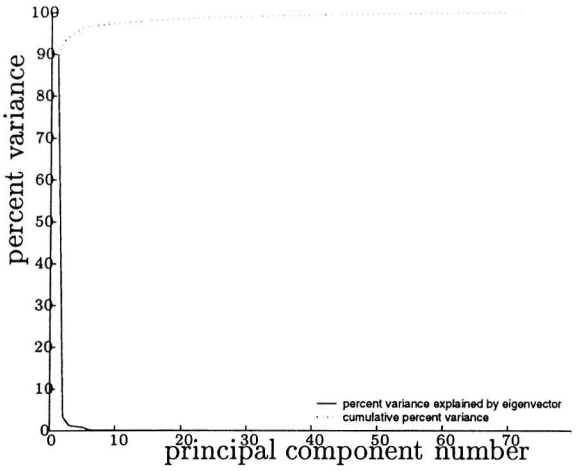
$$X = USV^t \ . \tag{2.6}$$

The columns of U contain left singular vectors of X, which are the unit eigenvectors of XX^t (which may be equated with the matrix of covariance between the data points) and the rows of V are the right singular vectors of X or the unit eigenvectors of X^tX (which may be equated with the matrix of covariance between the variables). S contains the square roots of the eigenvalues of X, known as the singular values of X. The principal components are found from $U * S$ $(= (V^t)^{-1}X)$. This method was used by Buckton *et al.* (1999) to produce features for input to a neural network for predicting oceanic constituents.

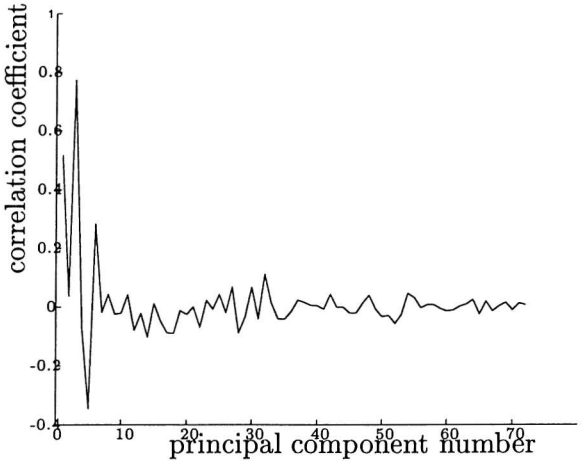
The standard method was found to produce some principal components that are less noisy than those with a higher eigenvalue. This is an effect that has been found with high dimensional data with highly correlated bands (Green *et al.* 1988; Lee *et al.*



(a) normalised principal components



(b) eigenvalues



(c) correlation

Figure 2.12: Principal component analysis of spectral data. (a) displays the principal components of the data normalised to zero mean, unit variance as grey values; (b) shows the eigenvalues corresponding to the eigenvectors, 95 % of the variance is accounted for by the first 4 eigenvectors and 98 % by the first 16; (c) shows the correlation of each of the principal components with chlorophyll *a*

1990). The SVD method produced some more useful principal components that shall be presented here.

The principal components of the *Norfolk 30/05/96* data were normalised and their values plotted according to a grey scale (figure 2.12a). A visual inspection of these showed

that the first five principal components varied steadily over space whereas the higher order principal components were very noisy. Because the bands were highly correlated with each other, it was expected that much of the variance in the data should be accounted for by very few principal components. This was supported by the magnitude of the first few eigenvalues compared to the majority of the eigenvalues (figure 2.12b), which indicated that the first 5 principal components accounted for 96 % of the variance. These first five components may have represented uncorrelated elements in the scene which change the spectral characteristics of the data. The correlation coefficients between all the principal components and chlorophyll *a* were calculated (figure 2.12c) and the third principal component was found to have a very high correlation coefficient of 0.77.

The spectral dependence of the principal components was found by inspecting the eigenvectors (figure 2.13). The eigenvector corresponding to the largest eigenvalue echoed the shaped of the mean spectra for the images (figure 2.6). If the first principal component was to be correlated with some element of the scene it would have been one that had an equal effect throughout the spectrum such as suspended sediment, bottom reflectance or sun-glint. Inspection of the first principal component (see also figure 2.22a) found peak values that correspond to brighter regions in the imagery, particularly at 577000 m east and 630000 m east. These were the same bright regions that were highlighted by cluster analysis of chlorophyll *a* concentration ranges 11-12 $\mu g\, l^{-1}$ and 15-16 $\mu g\, l^{-1}$ in section 2.7.3. Lesser peaks throughout this principal component corresponded to brighter regions in the images, the origin of which was difficult to determine. Again, it was possible that this effect was due to atmospheric haze or to particulates in the water that were increasing the scatter into the sensor. Alternatively, it was possible that changes in the attitude of the sensor resulted in increased reflectance from the surface of the water or transmission through the atmosphere.

The second eigenvector had large negative values in the blue region of the spectrum indicating that the second principal component was dominated by these wavelengths. The manner in which the values in the eigenvector approached zero with increasing wavelength was similar to the shape of the absorption curve of CDOM and so the

second principal component may have corresponded to this water constituent. This characteristic was very similar to the second eigenvector found by Sathyendranath *et al.* (1989). Very low values were observed in the second principal component (see also figure 2.22b) at 586000, 632000 and 648000 m east.

The third eigenvector, corresponding to the principal component that correlated most with chlorophyll *a*, and the fifth and sixth eigenvectors had a shape similar to the chlorophyll *a* absorption spectrum with peaks in the blue and red and a deep trough in the blue-green and green. However, the shape of the fourth eigenvector was similar to the chlorophyll *a* reflectance spectrum and it may have been that these four principal components represented changes in pigment concentrations in the water.

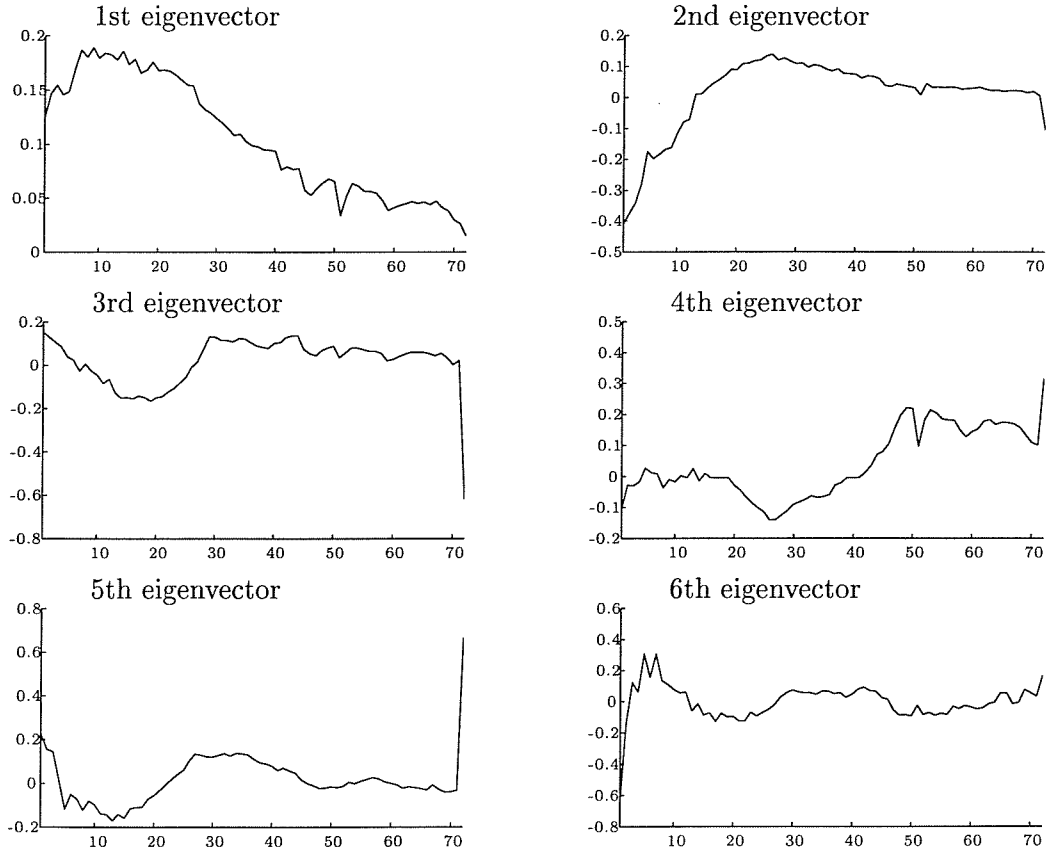


Figure 2.13: The eigenvectors corresponding to the six highest eigenvalues. The values of the vectors (the coefficients of the principal component transform) are plotted against the band number.

All the principal components showed a large range in values at 606000 m east, the region where the cruise took a transect perpendicular to the main cruise direction (section 2.3). The 3rd, 4th and 5th principal components are directly compared to chlorophyll *a* concentration in figure 2.14. Although highly correlated with chlorophyll *a*, these principal components clearly do not follow the same pattern as chlorophyll *a* over the track of the cruise.

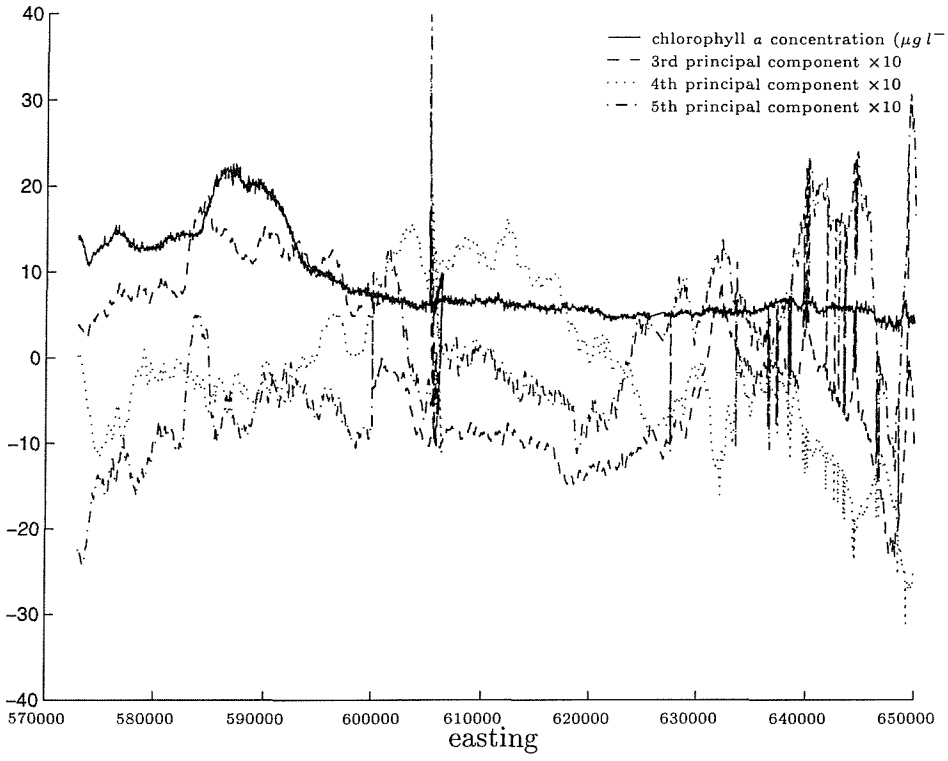


Figure 2.14: Chlorophyll *a* concentration and the values of the third, fourth and fifth principal components multiplied by 10.

Owing to the high values of the first two principal components, these were identified as indicators of factors that would result in error in the prediction of chlorophyll *a*. The PCA technique is linear and thus non-linear elements in the data may not be separated using this technique. Therefore, the presence of chlorophyll *a* may have been intrinsic in more than one of the lower order principal components.

2.7.5 Sensor effects

The pitch, roll and yaw of the aircraft can have a significant effect on pushbroom scanners and this effect can be different for each line of the image. CASI data are provided with data from the onboard vertical gyroscope (which accounts for pitch and roll) and Global Positioning System (GPS) data, which account for changes in position. Cosandier *et al.* (1992) found that standard CASI geometric correction using these two sources of information worked well where the surface was flat, as is a water surface. The data were obtained already geometrically corrected by the Environment Agency. This entailed using the gyroscope and GPS data to correct the image and then performing a secondary, manual correction using information from the scene. This resulted in an error of correction of around 10 m towards the coast, which increases away from the coast (Rebecca Allen, Environment Agency, personal communication).

The most noticeable sensor effect in this imagery was a brightening and darkening of whole rows or columns of the image data prior to the application of the 21-pixel window (section 2.6) (figure 2.15). Striping occurs in all imagery collected from surfaces of low reflectance (Bagheri *et al.* 1998). It was thought that the brightening of certain rows was caused by a subtly changing row integration time as the aircraft passed over the scene or perhaps by fluctuations in power supply (Bizzi *et al.* 1996). However, the variation was not consistent across the wavebands. This indicated that the changes were possibly occurring whilst the light signal was being separated by the reflection diffraction grating into wavebands before being recorded. There was a similar, apparently random variation in the values of whole columns, which was likely to be due to particular Charged Coupled Devices within the array that registered slightly erroneous light levels.

The contribution by the row and column variation to the overall variation in the data was quantified from ungeometrically corrected images as follows (see figure 2.16). For the image (*I* in figure 2.16), the mean value for each row (or column) was calculated for each band to produce the 'row-mean', *A*, ('column-mean', *B*). The difference between this value and the mean of the surrounding row-means (column-means) was then de-

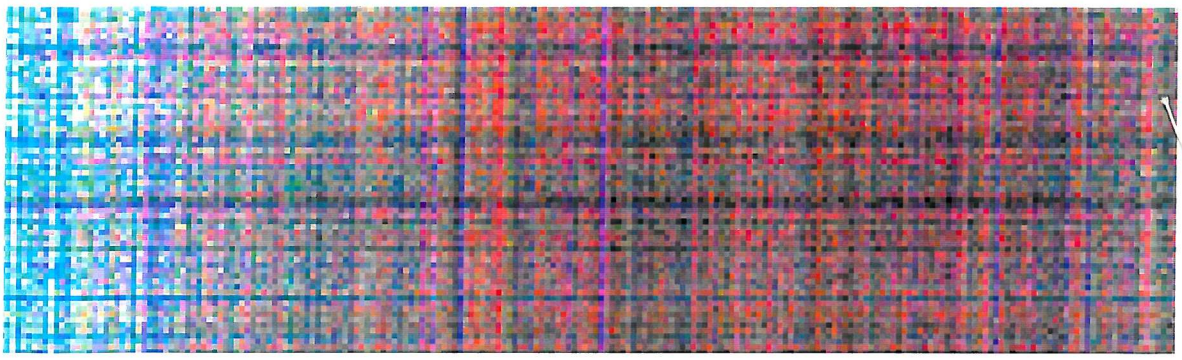


Figure 2.15: Variation in digital numbers (DN) for whole rows or columns shown in bands 70, 6 and 1 = red, green and blue. Here the grey levels for each band have been stretched from 0 to 255 for this section of the image to enhance the effect.

termined ('row-difference', C/'column-difference', D). It was found that the local mean of row-means was very similar for 5 to 21 rows (columns) and so the row-difference (column-difference) was calculated within the locality of 21 rows (columns). The result was a matrix that is 72 (the number of bands) by the number of rows (columns) of the image.

The row-mean matrix was analysed per-row, F (along the flight direction) and per-band, E. No structure was found along the flight direction but a clear variation in this noise effect was found by averaging the row difference for each band. The dotted line in figure 2.17 shows how variation that occurred from one row to the next varied with the image band in the original images. The maximum mean difference between a row and its surrounding rows for the original image was about 45 digital numbers for the shortwave blue bands. The minimum mean difference was about 6 digital numbers. With a standard deviation of between 400 and 600 DN for the blue wavelengths (figure 2.6) this accounted for as much as 6 percent of the standard deviation at these wavelengths for the original images. As the signal declined at longer wavelengths the contribution from row-to-row variation to the standard deviation approached 25 percent. By passing the window over the image, (section 2.6) the effects of this noise were greatly reduced as shown by the solid lines in figure 2.17. The variation from row to

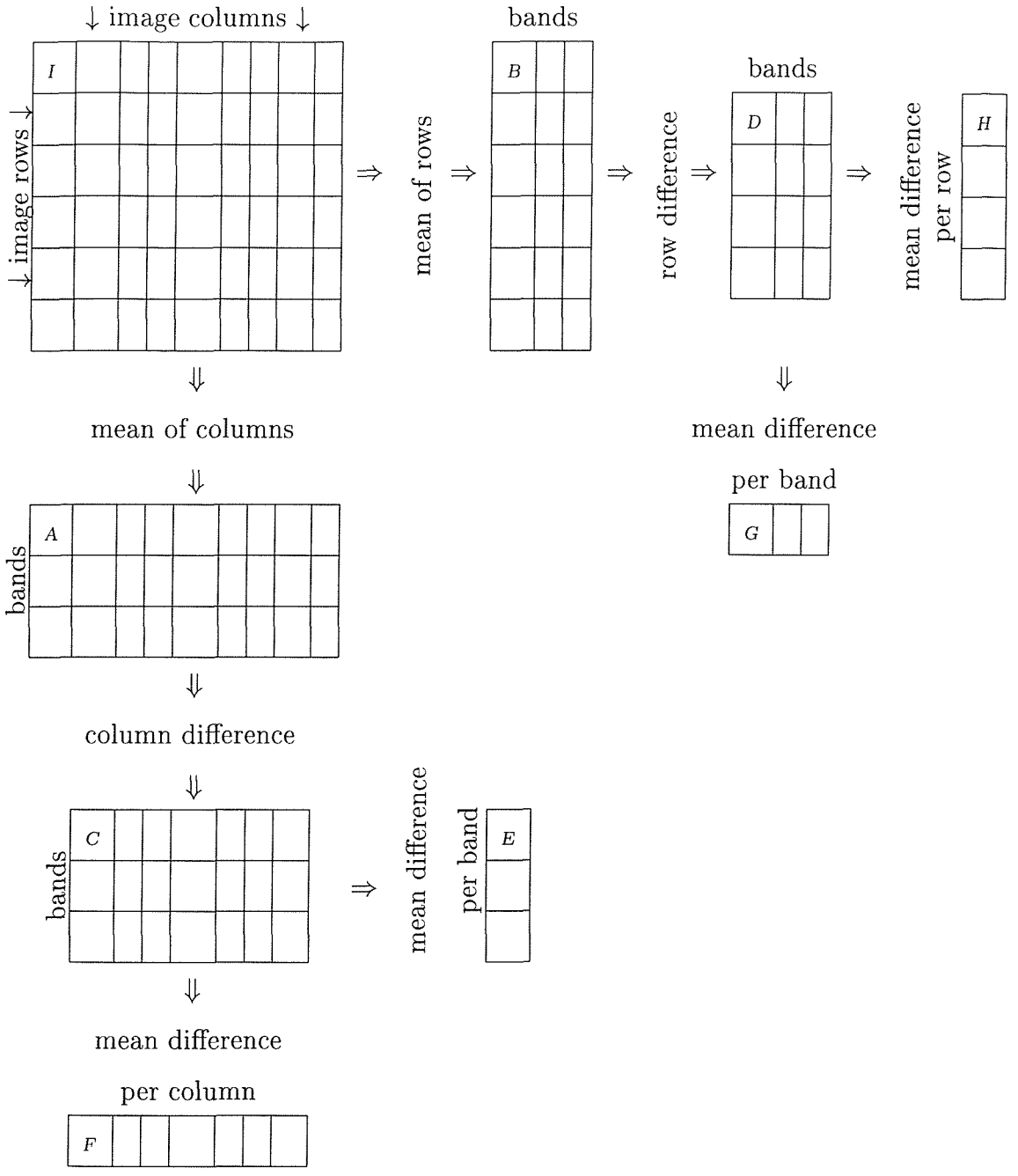
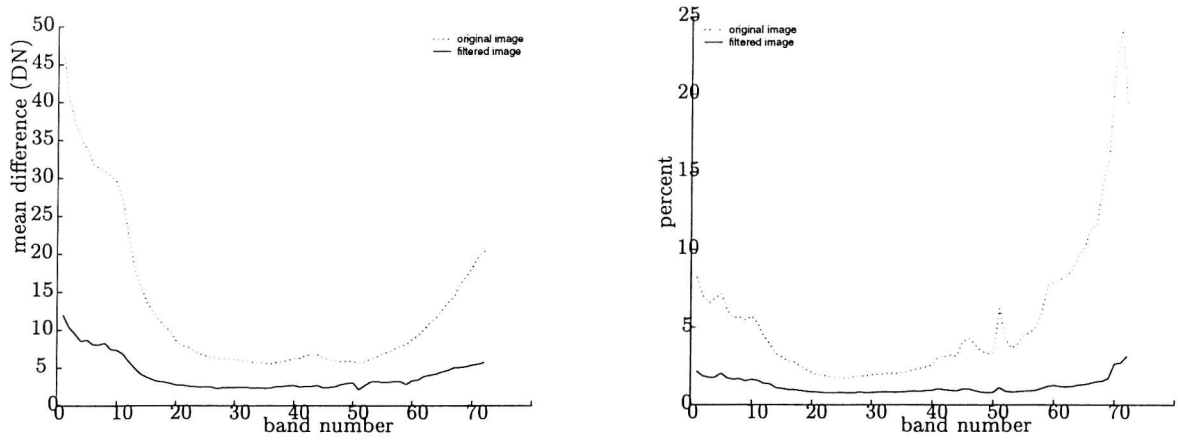


Figure 2.16: This figure illustrates how the row and column variation in the images was assessed. This is explained fully in the text on page 51.

row in the image was now found to account for a maximum of 3 % of the standard deviation.

A similar analysis was performed on the column-mean matrix. One effect was apparent



(a) Mean difference

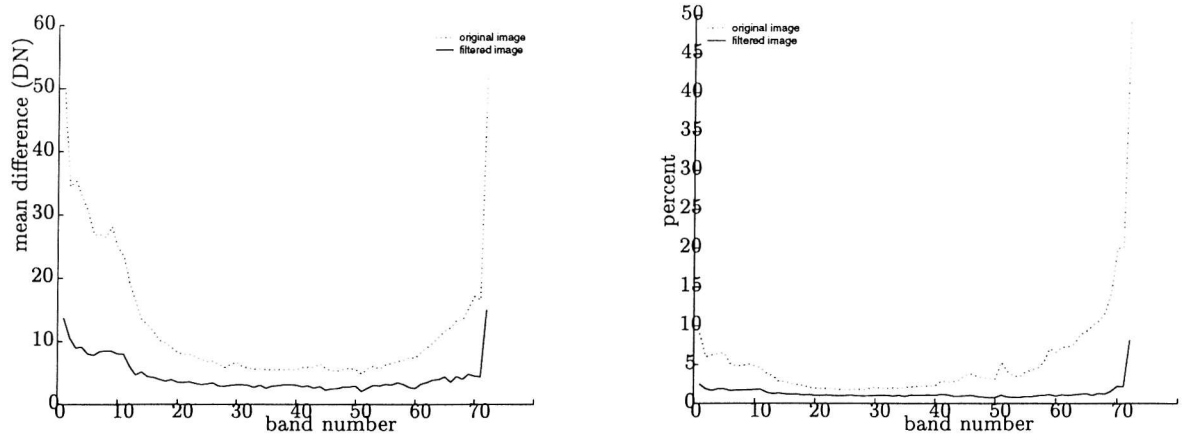
(b) Percent difference

Figure 2.17: The mean difference between the mean value for a row and its surrounding 20 rows is shown in the left plot (a) against band number. The dotted line shows this noise before the 21-pixel moving window was applied and the solid line shows the effect of the noise after the moving window was applied.

across the flight direction (H in figure 2.16) but this is more relevant to section 2.7.6 and so will be discussed there. Per-band (G in figure 2.16), this matrix showed a similar effect to the row-mean matrix (figure 2.18). However, more noise was evident in the near-infrared. This noise accounted for approximately three to seven percent of the standard deviation at blue wavelengths, dropping to around three percent for green and red wavelengths and then rapidly increasing in the near-infrared from five percent to almost 50 %. Again, applying the moving window to the image data reduced this noise effect to only a few percent.

2.7.6 Atmospheric, surface, bottom and adjacency effects

The greatest contribution to radiance received over water is usually from the atmosphere. Rayleigh scatter from molecules and Mie scatter from aerosols adds radiance throughout the spectrum, although it is wavelength-dependent. By flying an aircraft



(a) mean difference

(b) percent difference

Figure 2.18: The mean difference between the mean value for a column and its surrounding 20 columns is shown in the left plot (a) against band number. The dotted line shows this noise before the 21-pixel moving window was applied and the solid line shows the effect of the noise after the moving window was applied.

at different heights over the same stretch of water, Gower and Borstad (1981) found that the atmosphere contributes 0.4 % to the radiance at 440 nm for every 100 m and 0.15 % at 560 nm (also Singh *et al.* 1997). The variations in radiance caused by the atmosphere can be greater than the variations caused by water constituents (Quenzel and Kaestner 1981).

Many authors insist that atmospheric effects must be corrected for if the chlorophyll *a* signature is to be extracted (e.g. Hoogenboom *et al.* 1998; Gordon 1997). The most popular method by which this is achieved is dark-pixel subtraction whereby the water-leaving radiance is assumed to be zero in the NIR (e.g. Antoine and Morel 1999; Moore 1978). The measured value at these wavelengths is then solely due to the atmosphere and this radiance can be subtracted from all bands. However, in turbid waters it cannot be assumed that the radiance is zero because NIR light can be backscattered by sediments in the water (Gordon 1981; Peters *et al.* 1998). Other methods for correcting for atmospheric effects are based on models (for example Land and Haigh

1996; Gordon *et al.* 1997). However, without ground truth measurements these are hard to specify or verify (Lazar *et al.* 1998).

Atmospherically corrected formats of images 1875-1877 were available for this research. These data were processed using the atmospheric correction algorithms developed by Plymouth Marine Laboratory as part of the COAST project (Aiken *et al.* 1995a). This required subsampling the 72 channel imagery to 15 channels of up to 25 nm width before applying the algorithm. With these data, it was found, in some pixels, that the fluorescence peak around 685 nm became more prominent as a result of this correction. However, the data were extremely noisy with many pixels having zero values in one or more bands. A previous study using these data had found a reduced accuracy for chlorophyll *a* algorithms developed from it (Environment Agency 1997). It was decided that only the original imagery would be used for the research, since this would allow the exploration of algorithms that did not require the extra stage of atmospheric correction.

The most dominant atmospheric effect in this data was brightening towards the edges of the image (figure 2.19). This was due to the increased path length of the radiance received as a result of the increased viewing angle over the scene at the edges of the scanner. This tended to be slightly greater along the north edge of the images than the south, possibly due to the interaction of viewing and solar zenith angles causing greater scatter into the sensor from the atmosphere on one side. In addition, the CASI sensor has an asymmetric field of view such that nadir is found some columns away from the centre column of the image. This results in a greater viewing angle at one edge of the image than the other (Andrew Wilson, NERC Centre for Ecology and Hydrology, personal communication).

Using the column-mean method described in section 2.7.5, it was found that this effect is most prominent with the shortest wavelengths and was negligible beyond ~560 nm. Figure 2.20 illustrates how the average spectral value per image column varies across the image and with wavelength. For the blue and green wavebands the average digital numbers were highest to the north of the scanline, declined steadily towards the centre

and began to increase to the south of the scanline for all three images. At the south of the scanline the mean values for all images tended to drop off because the values of the land pixels had been set to zero. In the image with the least land (1877) the mean digital numbers increased towards the south following a smooth curve, although the values did not approach the high values that were found to the north. At the very edge of the images the increase in spectral values in the blue wavelengths was as much as 1000 digital numbers. This was a very significant effect on the imagery because it was greater than the standard deviation of the imagery. However, correction of this brightening would be relatively simple as the brightening was a function of column number and appeared to fit a quadratic-like function. The ENVI image processing package (only available very late on in the research) provides a simple tool for correcting such brightening. This increased radiance had little effect upon the spectral data at the 2300 data points because these were largely located along the centre columns of the images. Only the transect of measurements at 606000 m passed into this region of increased DNs.

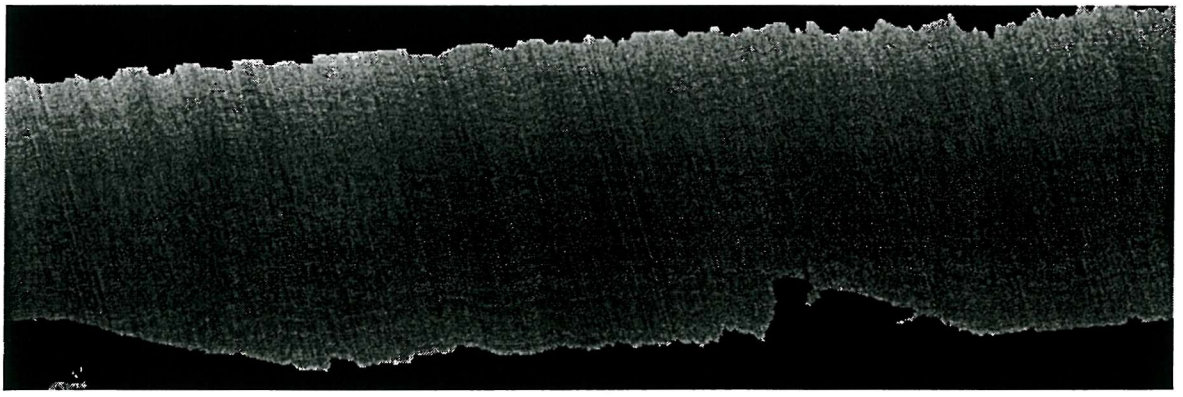
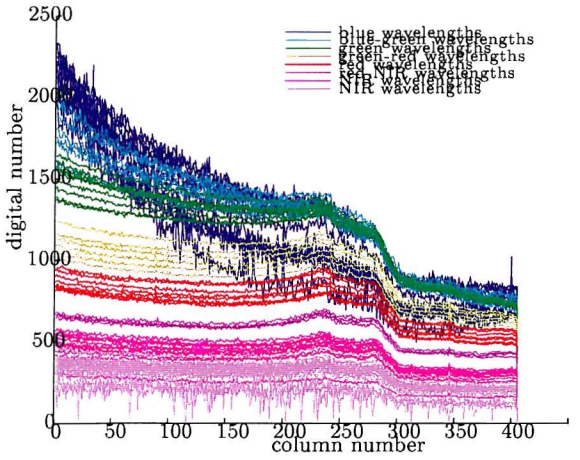
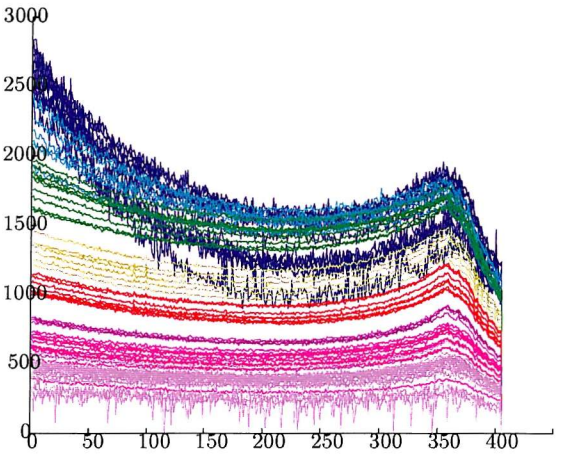


Figure 2.19: The blue bands, in this case band 4 (427 nm), show clearly the changes in spectral values across the scanline, here shown for part of the geometrically corrected image 1877.

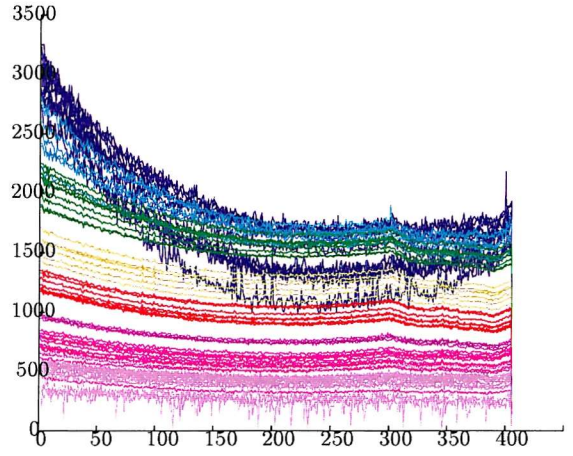
Another effect common in images over water is sun- and sky-glint. This is specular reflection from the water surface and so depends on the interaction between illumination and viewing geometry as well as the clarity of the atmosphere (which affects



(a) Image 1875



(b) Image 1876



(c) Image 1877

Figure 2.20: These three figures demonstrate how the mean spectral value changes across the scanline of the image. Column-mean values for all wavebands have been plotted and the plot lines have been coloured according to the region of the spectrum that they represent.

the brightness of the sky). Because the viewing angle is important, the effects of glint may change with the attitude of the sensor platform during image acquisition (Wrigley *et al.* 1992). Glint is particularly common where the surface is roughened by winds because the area of surface that reflects light straight to the sensor is increased. The overall effect of glint on spectral measurements is to brighten values across the spectrum. Tassan (1981) found that sun-glint was a high source of error in chlorophyll *a* prediction although sky-glint is usually not significant enough to necessitate its correction (Moore 1978). Sturm (1981) and Fraser *et al.* (1997) provide methods for correcting for glint but, as with atmospheric correction, these require knowledge of the

atmospheric, meteorologic and oceanographic conditions at the time of sampling.

Reflection from the bottom of the sea has been identified as a source of error in shallow waters. Gould and Arnone (1997) found that bottom reflection was significant in their data and Fraser (1998a) and (1998b) found that error in their turbidity and chlorophyll *a* predictions was partly due to bottom reflection and this error varied with the bottom sediment type. Lee *et al.* (1994) derived an expression for the remote sensing reflectance received from the bottom as a function of the bottom albedo, the water depth and its absorption coefficient and Pérez-Ruzafa *et al.* (1996) proposed a model from which the chlorophyll *a* content of water may be determined even when the water is shallow and the bottom is covered in meadows of macroalgae. East of 600000 m in the *Norfolk 30/05/96* site, the water depth quickly descends to greater than 10 m. To the west the water is shallower and some possibility of bottom reflectance was considered.

The point spread function of the sensor and scattering within the atmosphere determine from where in the scene and with what weighting photons are collected for each pixel. This usually results in a bias towards the central region of the pixel which decreases towards the edge of the pixel, with a few photons being gathered from the areas represented by adjacent pixels (Justice *et al.* 1989; Fisher 1997). Brightening of spectral values close to land is often attributed to reflection from the bottom sediments, however, it can also result from light being gathered from nearby regions of land (Reinerman and Carder 1995).

Haze in the atmosphere, or surface or bottom effects may have caused the bright regions at 577000 m and 630000 m that were highlighted by cluster and principal component analysis. Without *in situ* measurements, it was difficult to determine which were the strongest influences on the data. Yet, the potential for error in chlorophyll *a* predictions due to these factors was recognised.

2.7.7 Errors due to co-location of data

A great deal of effort was made to ensure that the image and *in situ* data sets were well registered to each other. However, error may have been introduced at any stage in the registration. For example, the instruments giving readings for the locations of *in situ* samples and the GPS and gyroscopic instruments giving information about the location of the aircraft during image acquisition may have some slight error. The Environment Agency CASI used a 'differentially corrected' GPS receiver using a commercial signal. There can be some time delay in receiving this signal and so a certain amount of error in positioning was expected (Kyle Brown, Environment Agency, personal communication). More error may occur because water samples used for calibrating the fluorometer and the fluorometer itself were not read from the same parcel of water, or represent samples of different sizes in the parcel of water. Image geometric correction is not perfect and some pixels may not have represented the correct region of the water body, especially away from the coast where the geometric correction is less certain. Also, the sample represented by a pixel is not the same volume as the parcel of water represented by the fluorometer measure at the same location. Errors such as these were treated as noise in the data because very little could be done to mitigate for them. One source of co-locational error that was investigated further was that due to the movement of water between sample from ship and aircraft.

2.7.7.1 Tidal and current effects

Unlike terrestrial remote sensing, the scene of marine remote sensing is constantly moving as a result of currents within the water. This results in error in the co-location of image and *in situ* measurements where these have not been sampled at the same time. Although much effort was made to ensure that the samples were taken concurrently the greater speed of the aircraft meant that the images were sampled within a few minutes whereas the same region in the water took up to four hours to sample. Table 2.3 demonstrates the differences in the times of the samples.

Table 2.3: Table showing the difference in the times that the images and the *in situ* measurements were sampled.

Image number	Time of flight (GMT)	Time of *in situ* samples for region of images	Greatest time difference (hrs:mins)	Least time difference (hrs:mins)
1875 start	14:41	10:46	3:56	2:11
finish	14:49	12:41		
1876 start	14:55	12:44	2:10	0
finish	15:02	16:07		
1877 start	15:16	15:42	2:57	0:50
finish	15:26	18:22		

(*in situ* sample points did not always correspond to very edges of images)

The images covered a region only 8 km out to sea. Currents are driven by waves nearshore and by tides offshore (Kieran Millard, HR Wallingford, personal communication). Thus, ideally current measurements would have been gathered simultaneously to the data so that the motion of the water between *in situ* and aircraft sampling could be calculated and possibly corrected for. Several sources were contacted regarding such measurements including the Environment Agency, the companies involved with Bacton Gas Terminal (BP-AMOCO, Shell, Phillips, British Gas and Interconnector), Anglian Water, Great Yarmouth Coast Guard, EMC and GeoTeam and who had worked on a pipeline to Bacton during 1996 but data that were available were not suitable for correcting the *in situ* data.

Using the Admiralty tidal stream atlas and the Admiralty tide tables (Great Britain Ministry of Defence Navy Hydrographic Department 1976; Great Britain Hydrographic Department 1996) the tidal currents were estimated (table 2.4). At the beginning of the *in situ* sampling of the region of image 1875 the tidal currents were flowing from the east-south-east at a speed of 1.5 knots. The currents were in the same direction,

Table 2.4: The tidal currents at the time of sampling

time	speed (knots)	speed (ms^s)	direction (clockwise from north)
10:00	1.5	0.77	285°
11:00	1.4	0.72	285°
12:00	1.1	0.57	270°
13:00	~0.6	~0.31	270°
14:00	0	0	270°
15:00	0.9	0.46	105°
16:00	1.6	0.82	105°
17:00	1.9	0.98	105°
18:00	1.6	0.82	105°

slowing to 1.4 knots during the ship sampling of the middle of the region of image 1875. As the ship moved into the region of image 1876 the currents were slowing to about 1.1 knots from the east and the water became slack during the time that the middle of this region was being sampled. As the ship passed into the region of image 1877 the currents were from a west-north-west direction at a speed of 0.9 knots. The currents remained in the same direction increasing to a speed of 1.9 knots and then declining, as the ship headed west. During the overflights of images 1875 and 1876 (whilst the ship was in the region of the centre of image 1876) the currents were slack increasing to 0.9 knots from a west-north-west direction during the acquisition of image 1877.

Using the Admiralty tide information, the relative velocity between corresponding parcels of water sampled *in situ* and from the aircraft was calculated (figure 2.21). The Matlab code for this is given in appendix B. The resulting distance between the sampled parcels of water was quite large - up to 8 km to the west of the region and 14 km to the east. However, because the images represent only up to 8 km from shore, the speed of the tidal currents will have been reduced by friction at the coast. They are also significantly lower than the maximum tidal current speeds calculated by Sager and Sammler (1968). Therefore, although tidal currents were considered to be a source

of error in these data, they were not thought to be as influential as the Admiralty tide calculations suggest.

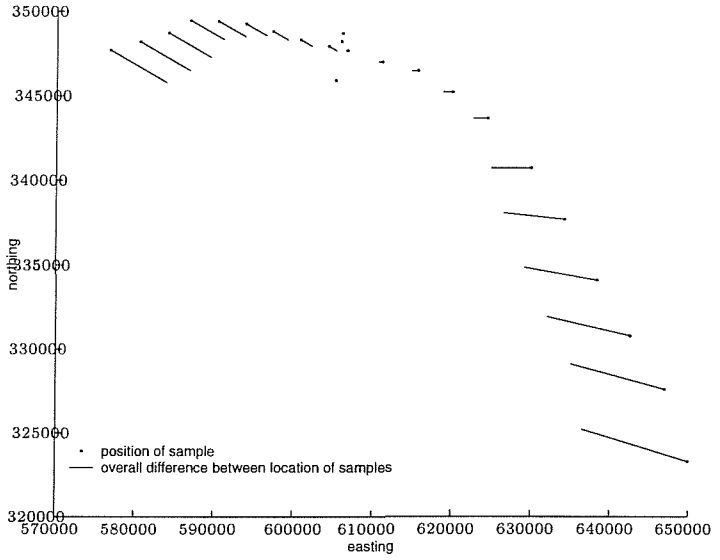


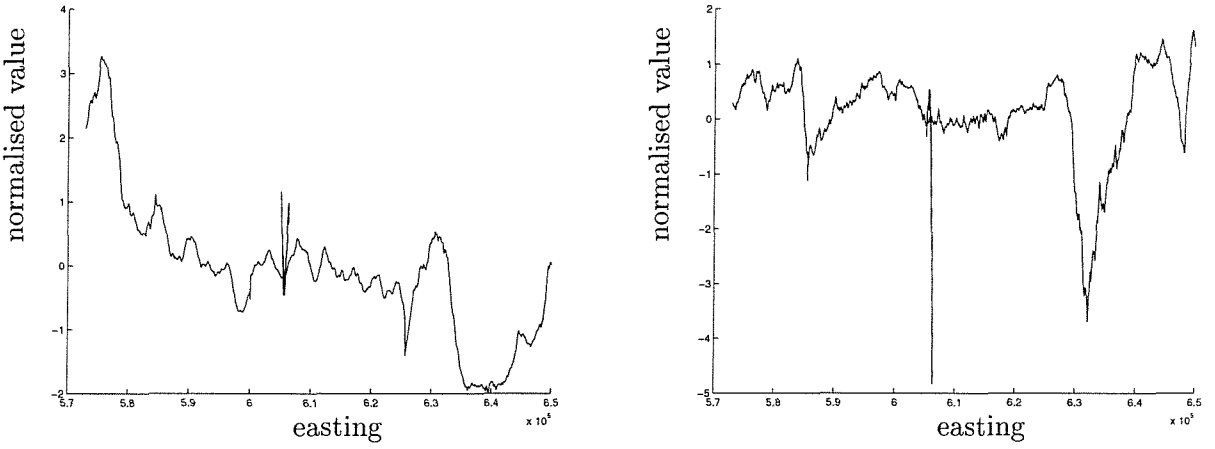
Figure 2.21: The maximum relative distance between corresponding parcels of water sampled *in situ* and from the aircraft, for every hundredth sample point.

2.7.8 Total error in data

The previous sections have detailed a study that was made of a number of factors which may have influenced the relationship between chlorophyll *a* and spectral values. A number of effects were identified and these are summarised in table 2.5. A moving window was applied to the data to remove the effects of sensor noise and possible problems with co-registering data. However, it was not possible to correct for other effects in a simple manner because such correction required reliable information about environmental factors such as water constituent concentration, water depth, atmospheric and meteorological conditions and the current velocities. It was considered important to bear these effects in mind when developing the chlorophyll *a* prediction models.

In the subsequent analysis the presence of these various causes of error were consid-

ered. However, regions of the spectrum and geographical regions that were strongly affected by these factors were retained in the data to determine the overall effect on the prediction of chlorophyll *a*. Figure 2.22 shows the first two principal components plotted against the easting of the corresponding data point for easy reference during the assessment of error.



(a) 1st principal component (normalised)

(b) 2nd principal component (normalised)

Figure 2.22: The spatial dependence of sources of error. If the first two principal components are considered to indicate the presence of factors in the scene that may cause error in chlorophyll *a* prediction, (a) and (b) show how these factors change over the course of the cruise transect.

Summary of chapter 2

There are many sources of image and *in situ* oceanographic data. However, it is often difficult to obtain data that coincide. The data used in this research required some preprocessing to determine the chlorophyll *a* concentration and spectral measures corresponding to each cruise data point and the methods used to process these data have been presented in this chapter.

The data were assessed for their relevance to predicting chlorophyll *a* from spectral information using cluster analysis and principal component analysis. It was found that there was a change in spectral value with chlorophyll *a* although it was clear that other factors dominated the spectral signal that was detected.

The data were assessed for factors such as image sensor noise, atmospheric effects and the effects of co-registering the *in situ* and image data. In particular, the data displayed some significant sensor and atmospheric effects as is typical for oceanographic data. In previous studies extensive steps have been taken to correct for this but correction can introduce its own error if it is undertaken without accurate information about the sources of error. The movement of water between the *in situ* and spectral sampling was estimated using Admiralty tide information. However, it was considered that this would be the maximum effect of currents as they would be reduced by friction so close to the shore.

A moving window was applied to the spectral data to mitigate sensor noise and current effects. No atmospheric correction was undertaken, however, rather it was intended that algorithm development should attempt to produce algorithms that were robust to the errors that had been identified.

The detailed analysis of error in the data that has been described here is fundamental to the research that is detailed in the following chapters. It also represents one of the novel aspects of this research because the techniques used to investigate these data have not been used in previous ocean colour research.

Table 2.5: Effects of environmental factors that were identified in the data.

cause	how identified	dependence	severity
bright patches	cluster analysis	spatial/a little spectral	significant increase in spectral values
spectral variance not caused by chlorophyll *a*	principal component analysis	spatial	a dominant effect on spectral values
edge brightening	means of image columns/PCA	spatial and spectral	little effect on data points
row/column noise	means of rows or columns	spectral	minimised by using moving window
tidal current effects	-	spatial	possibly very severe at the far west and east of the data but the proximity of the coast is likely to have reduced the velocity of the currents

Chapter 3

Multiple regression techniques

3.1 Introduction

Regression analysis requires two important design decisions - the choice of the best features for input and a balance in the model between complexity and generalisation (Brown *et al.* 1994). This chapter describes the two types types of algorithms that were used to develop models in this research - multiple linear least-squares regression and non-linear neural networks. These two methods were compared in the research presented here to determine if the neural network technique would produce models that were more accurate than the more conventional multiple linear regression technique. The two techniques are related and this chapter begins by describing the linear method and then expanding on this to explain neural networks. Finally, some specific points about using neural networks are detailed.

3.2 From multiple linear regression to neural networks

To perform a linear least squares regression, the vector of coefficients, β, of the input vector $\mathbf{x}$ is identified which allows the prediction of the output y. For a regression with n input features, the expected value of y given the inputs $\mathbf{x}$ is calculated as

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \ . \tag{3.1}$$

In matrix notation, if $\mathbf{X}$ is a $p \times (n+1)$ matrix of input features with all elements in the first column having the value of 1 (to determine the offset, β_0) and the following columns having the values of all other input features $\mathbf{x}$ for p observations and $\mathbf{y}$ is the $p \times 1$ vector of measured output values for each of the p observations, then the model may be expressed as

$$\mathbf{y} = \mathbf{X}\beta + \mathcal{E} \ . \tag{3.2}$$

$\mathcal{E}$ is a vector of the error of the model

$$\begin{aligned} \mathcal{E} &= \frac{1}{2}\sum_{q=1}^{p}(\hat{y}_q - y_q)^2 \\ &= \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) \end{aligned} \tag{3.3}$$

where y_q is the measured output for observation q and $\hat{y}_q$ is the output predicted for the qth set of inputs. This error changes with β and so the minimum may be found using differentiation by satisfying

$$\begin{aligned} \tfrac{d}{d\beta}[(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)] &= 0 \\ = -2\mathbf{X}^T\mathbf{y} + 2(\mathbf{X}^T\mathbf{X})\beta & \end{aligned} \tag{3.4}$$

The best approximation of β, assuming that $\mathbf{X}^T\mathbf{X}$ is nonsingular, is therefore (Myers

1986)

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \ . \tag{3.5}$$

which is the linear, least squares solution and thus minimises the sum of squares error.

A feed-forward neural network regression algorithm may be visualised as rows of nodes interconnected with layers of weights. In a single layer of the network, each node (i) in the first row is connected to each node (j) in the second row by a weight. The value arriving at node j is calculated by weighting all the outputs from the previous layer and adding them together

$$o_j = \sum_{i=1}^{n} w_{ij}x_i \ + \ b \tag{3.6}$$

where w_{ij} is the weight connecting node x_i in the first row with o_j in the second row and b is the bias (offset). The value is then passed over an activation function such as a sigmoid, hyperbolic tan (tanh) or linear function. The simplest neural network of this type has only one layer of weights connecting two rows of nodes and is usually referred to as a single layer perceptron (SLP) (figure 3.1). If this has linear activation functions at the output nodes then the result is a linear model. Such a network is optimized to a set of observed inputs and outputs during the training process that minimises an error function. During this, the error (equation 3.3) of the predicted output is iteratively minimised. The expansion of equation 3.6 is

$$o_j = b \ + \ w_{1j}x_1 \ + \ \cdots \ + \ w_{nj}x_n \ . \tag{3.7}$$

Because o_j is the expected value of y given the inputs, x, equation 3.7 is equivalent to equation 3.1. Since during training, the sum of squares error is being minimised, an optimized linear neural network produces the same algorithm as the linear least squares method (equation 3.1).

A more complex network is produced by interconnecting more than two rows of nodes with layers of weights to produce a multilayer perceptron (MLP). A two layer MLP (figure 3.2) has one hidden set of nodes at which the weighted sums of the input layer are passed over an activation function. The resulting outputs from the hidden nodes are then weighted in the second layer and summed at the output nodes before being

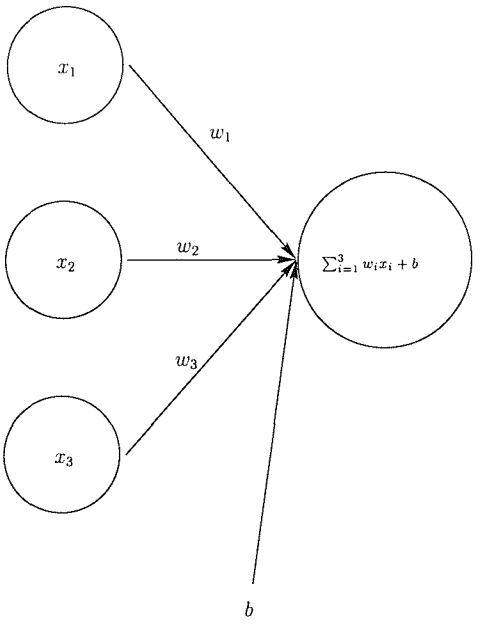


Figure 3.1: A single layer perceptron with three inputs, one output and a bias term.

passed over another activation function. At each output o_k of a two-layer network the resulting value is therefore calculated as

$$o_k = \mathcal{G}\left(\sum_{j=1}^{m} w_{jk} h_j + b_k\right) \tag{3.8}$$

$$where \quad h_j = \mathcal{F}\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right)$$

Here, x_i is the value at the input node i of n inputs, h_j is the value output from the hidden node j of m hidden nodes, w_{ij} is the weight connecting nodes i and j and b_l is the bias for layer l. $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$ are the activation functions for the hidden and output sets of nodes, respectively. In this case it is assumed that the same activation function is applied at all the nodes in the same row.

In effect, the input information is being transformed from n-dimensional input space to m-dimensional space in the hidden layer and then again to o-dimensional output space. This allows non-linear functions to be modelled. An increase in complexity is gained by adding hidden nodes or adding layers to the network. However, it is worth emphasising that the neural network is still converging towards a range of mean values for the data set, rather than modelling the transfer within the scene. The method by which the transformation is achieved by back-propagation of the error is described in

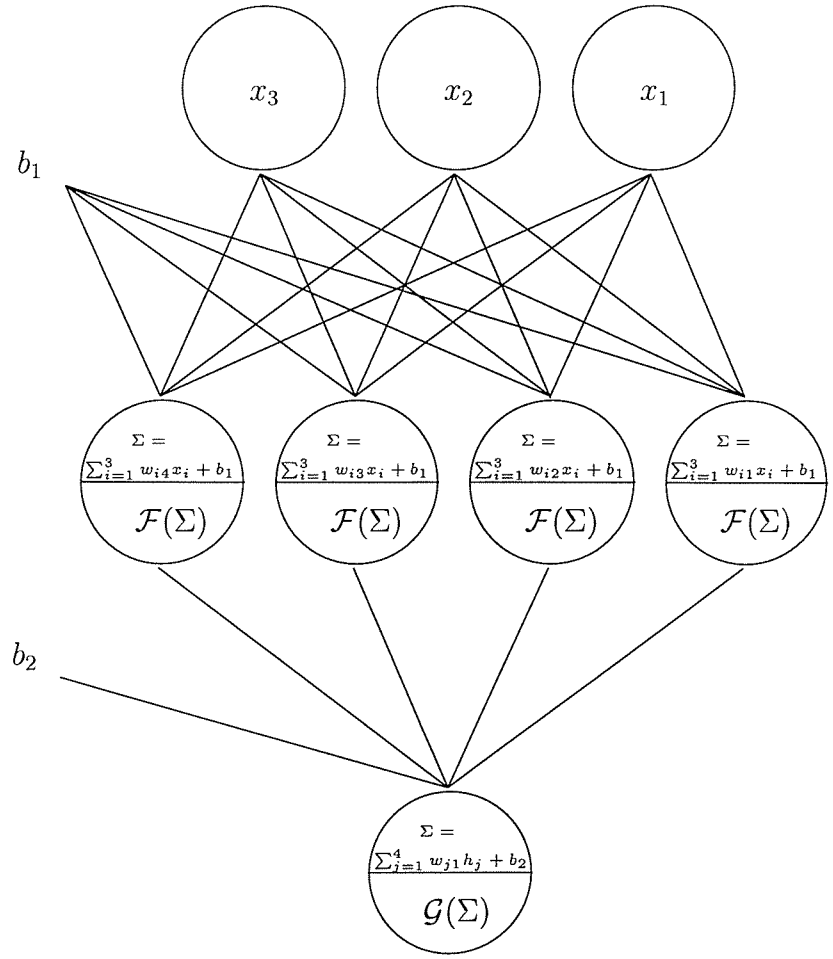


Figure 3.2: A multilayer perceptron with three input nodes, four hidden nodes and one output. Notation is the same as in equation 3.8

the following section.

3.2.1 The back-propagation rule

The network is trained with patterns of known inputs (x) and outputs (y). The inputs are fed forward through the network (equation 3.8) and the error calculated (equation 3.3) (Rumelhart *et al.* 1986). The network is optimized to find the best function to fit the training data set by iteratively updating the weights according to the error of the prediction of y. Each weight w_{ij} is changed by an amount Δw_{ij} which is proportional

to the local error gradient in weight-space $\frac{\partial \mathcal{E}}{\partial w_{ij}}$

$$\Delta w_{ij} = -\eta \frac{\partial \mathcal{E}}{\partial w_{ij}} \tag{3.9}$$

where η is the learning rate which controls the magnitude of each weight change. If z_i is the output from node i, a_j is z_i weighted by w_{ij} and $\mathcal{H}_i(\cdot)$ is the activation function applied to node i such that

$$z_i = \mathcal{H}_i(a_i) \tag{3.10}$$

$$a_j = \sum_{i=1}^{n} w_{ij} z_i + b, \tag{3.11}$$

we can use the chain rule to calculate the value of $\frac{\partial \mathcal{E}}{\partial w_{ij}}$:

$$\frac{\partial \mathcal{E}}{\partial w_{ij}} = \delta_j z_i \tag{3.12}$$

$$where \quad \delta_j = \frac{\partial \mathcal{E}}{\partial a_j} \tag{3.13}$$

$$and \quad z_i = \frac{\partial a_j}{\partial w_{ij}} \tag{3.14}$$

δ_j can be viewed as the error at node j and it is this that is fundamental to the calculation of the weight adjustment (from equations 3.9 and 3.12):

$$\Delta w_{ij} = -\eta \delta_j z_i \tag{3.15}$$

To adjust the weights in the final layer, δ_k is simply calculated using the error at the output of the network. Firstly, the chain rule is used to expand equation 3.13 to

$$\delta_k = \frac{\partial \mathcal{E}}{\partial z_k} \frac{\partial z_k}{\partial a_k}. \tag{3.16}$$

From equation 3.10 it can be seen that

$$\frac{\partial z_k}{\partial a_k} = \mathcal{H}'_k(a_k) \tag{3.17}$$

and thus:

$$\delta_k = \mathcal{H}'_k(a_k) \frac{\partial \mathcal{E}}{\partial z_k} . \tag{3.18}$$

Therefore, where a linear activation function is used for the output nodes, the change in weights is calculated using

$$\delta_k = \hat{y}_k - y_k . \tag{3.19}$$

δ_j is determined for any hidden node, j using the chain rule such that

$$\delta_j = \frac{\partial \mathcal{E}}{\partial a_j} = \sum_{k=1}^{n} \frac{\partial \mathcal{E}}{\partial a_k} \frac{\partial a_k}{\partial a_j} \ . \tag{3.20}$$

Since $a_k = \mathcal{H}_j(a_j)w_{jk}$ and using equation 3.13, equation 3.20 can be re-written to obtain the formula for calculating δ for any hidden node:

$$\delta_j = \mathcal{H}'_j(a_j) \sum_{k=1}^{n} w_{jk}\delta_k \ . \tag{3.21}$$

This formula defines how much of the error at the output is contributed to by the values calculated at each hidden node. If, as in the research presented here, a two-layer network were trained with a hyperbolic tan (tanh) function as the activation function in the hidden layer, this formula becomes

$$\delta_j = (1 - y_k^2) \sum_{k=1}^{n} w_{jk}\delta_k \ . \tag{3.22}$$

By adjusting the weights in this way, the network is performing gradient descent optimization - it is descending down the error surface in weight space. This feed-forward/back-propagation process is repeated, iteratively adjusting the weights, until either a predetermined number of iterations are reached or the reduction in error with each iteration is below a desired amount. The full derivation of this is given in Bishop (1995), Haykin (1994) and Rumelhart and McClelland (1986).

3.3 Specifics of neural network training

The function derived during training is an empirical estimation of the relationship between the inputs and outputs. A balance must be struck between finding a function that fits the training data well and producing a function that will fit all subsequent data well. Where p is the number of training patterns and the matrix of measured inputs is $\mathbf{X}$ the solution to the regression problem is therefore:

$$\mathbf{y} = s(\mathbf{X}, \mathbf{w}^*) + \mathcal{E} \tag{3.23}$$

$\mathbf{y}$ is the set of desired outputs, $\mathbf{w}^*$ is a vector of unknown optimal weights, $\mathcal{E}$ is the error and $s(\mathbf{X}, \mathbf{w})$ is the mean value of $\mathbf{y}$ in response to $\mathbf{X}$ (Steppe *et al.* 1996). This emphasises that the regression coefficients (weights) of the neural network, like other least-squares regression techniques, determine the expected value of the desired outputs for a given range of input values.

As with all regression techniques, deriving a good *generalisation* is dependent on the data used. The neural network is trained on a limited set of data which, as with any set of real data, will not encompass the full range of possible patterns existing in nature. The temptation with training is to strive to achieve the lowest training error. However, because overly complex networks may be built, there is a danger that the training will fit the network precisely to the training data but will predict new data poorly. This results in a loss of generality (and the network will not accurately predict the testing patterns). Over-fitting may be avoided by keeping the complexity to a minimum and using a validation data set during training to determine when the error for this set begins to deviate from that of the training set.

A good generalisation gives the neural networks a robustness to noise in the data that other methods do not have (for example, Clark and Cañas 1995). This is important with hyperspectral ocean colour data which is subject to strong noise effects as was demonstrated in chapter 2 particularly at the shorter wavelengths where a great deal of useful information is held (Moore 1978; Sathyendranath *et al.* 1989).

Neural networks are often said to be a non-parametric method of determining regression functions but in practice are actually semi-parametric. Many traditional methods of determining relationships, including least-squares analysis, rely on some assumption of a normal distribution. Where such a distribution is not present, the assumptions of these measures are not valid (Paola and Schowengerdt 1995a). Non-parametric models may be used when the distribution of the data is not normal or is unknown. However, neural networks usually require some parametric assumptions in training, such as using the assumption of a normally distributed output error, in order to specify the model by which the error is minimised. In this case the error is reduced using the least squares

method, that is by minimising the square of the error (equation 3.3).

The technique of neural network regression has been considered for the development of chlorophyll *a* prediction algorithms because it has a number of advantages over other methods. As has already been discussed, many studies have found that the chlorophyll *a*-spectra relationship is non-linear. Neural networks offer a simple method of finding the optimal non-linear model to fit the data. The neural network has the advantage over least-squares curve fitting in that it does not require prior understanding of the curvature or the complexity of the regression to be performed (Bishop 1995).

Neural networks are also useful when a number of different data types are to be used. Multisource information presents problems with more traditional techniques because the statistical properties and scales of the different data types can cause a bias in the resulting algorithm. A neural network however, will assign automatically reliability factors to each input during training in that it will find the combination of weights that will use optimally all the input information (Benediktsson *et al.* 1990; Civco and Waug 1994). However, when using data of various scales, the user should be aware that a certain amount of the initial training of the neural network is concerned with adjusting the weights to the scale of the data being used and so training may take longer. The current research uses only spectral information as an input to the algorithm but this can have different ranges. There is the potential for other information, such as measures of the atmospheric and sedimentary contribution to the image, to but used as inputs to a neural network .

Furthermore, using neural networks, a number of comparable algorithms of varying complexity may be developed alongside each other. This is very relevant to the current problem because the non-linearity and complexity of the relationship between chlorophyll *a* spectral data has not been determined. Using neural networks is thus a useful method to determine the characteristics of the problem and enables greater insight for future work. However, some key problems inherent in the neural network method need to be addressed.

It has been said that neural networks are particularly useful in studies for which not all the training data are available at one time because the network may be re-trained easily when new information becomes available (Schiller and Doerffer 1999). However, other techniques (such as the k-nearest neighbour technique) are better for this purpose because the neural network actually has to be completely retrained with all available data, to ensure that model is re-optimised.

Firstly, the error may be described as a surface in weight-space, such that, as we change the weights the output error increases or decreases. The gradient descent learning rule aims to follow the error surface towards a minimum value. This can result in a non-optimal solution however, if the network converges on a minima in the error surface which is not the global minimum (Paola and Schowengerdt 1995a). A solution to this would be to run each training several times with different initialisation weights, to ensure that the network always converges on the same solution, or to find the best solution.

Secondly, a problem often reported with pattern recognition problems is known as the 'curse of dimensionality'. This describes the poor response of a recognition system when presented with too many input features. For every possible combination of input and output values, enough sets of training input and output data are required by the regression function during training to model the relationship well. As the number of input features increases, there is an exponential increase in the number of example patterns required by the training algorithm to derive a prediction of the output. With a limited training data set, some input-output patterns may be under-represented and the training algorithm will not succeed in finding a reliable solution. This can be avoided by using only a limited number of inputs by employing feature extraction and selection techniques. It was clear that with 72 input bands some form of feature selection would be necessary.

Summary of chapter 3

The development from using the linear least-squares technique to determine the regression coefficients to that using a neural network technique has been presented. The two techniques are easily compared because they derive the regression model by minimising the mean square error. Both techniques can have one or more input variables, the neural network technique can produce linear (SLP) or non-linear (MLP) models, with the linear models being identical to those developed using the standard least squared method. Neural network-derived models can be of varying complexity. Their advantage is that any number of inputs with any scale may be used and many linear and non-linear models of varying complexity may be developed for comparison. The importance of considering the assumptions made in the process of model fitting, i.e. that the output error has a normal distribution and also the need for a balance between a model fitting the training data and the generality of the model, were highlighted. Finally, the volume of data was addressed. So that a sufficient number of training examples is provided to the function, it was considered to be impossible to regress all 72 input bands against chlorophyll *a*.

The multi-layer perceptron attempts to determine the regression coefficients iteratively and allows non-linear functions to be developed. This approach uses gradient descent to find the minimum in the error surface. Whilst this may not be a global minimum and a problem of over-fitting the model to the training data exists, methods for avoiding these drawbacks are simple to implement.

The research presented here explicitly chose the neural network technique of regression because the results would be directly comparable to the more common method of linear least squares regression. In so doing, the research would be able to investigate more thoroughly than previous studies using neural networks the nature of the relationship between chlorophyll *a* and spectral information by investigating the structure of the models used for the regression.

Chapter 4

Primary study

4.1 Introduction

A primary study was performed to assess the validity of the neural network regression technique and to identify any refinements to the techniques that may be required. This investigated whether straight-forward chlorophyll *a* prediction models could be used with the *Norfolk 30/05/96* data. A simple selection of spectral features was undertaken (section 4.2), the data were divided into training, validation and testing sets (section 4.3) and then a training (4.4) and testing (4.5) of conventional linear least squares- and neural network-derived models was performed. All aspects of the training of the neural networks and test results were used to draw conclusions about the nature of relationship between chlorophyll *a* and spectral information.

4.2 Extracting a subset of spectral features

With the CASI hyperspectral data, a large volume of information was available. This was reduced by picking out a number of spectral features with which to develop the models. Bands were chosen from each section of the spectrum - blue, green, red and

near infra-red (NIR) - according to their correlation with chlorophyll *a* (figure 2.7), their correlation as ratios with chlorophyll *a* (figure 2.8) in the training set, their correlation with each other (figure 2.9) and also with bands close to known features in the spectrum were considered. This resulted in a subset of eight bands being extracted from the training, validation and testing sets. These bands, and the basis by which they were chosen, are detailed in table 4.1.

Table 4.1: The subset of eight bands that were chosen for model development

band number	wavelength † (*nm*)	comments
6	441 (blue)	Close to chlorophyll *a* absorption peak and had one of the highest correlations with chlorophyll *a* in the blue.
20	540 (green)	Close to green reflectance peak and was within a minima in the correlation of green wavelengths with chlorophyll *a* . One of the better correlations with chlorophyll *a* as a ratio with band 6.
27	589 (green-red)	Ratio of band 27 and 29 gave the highest correlation with chlorophyll *a* of the whole training set.
29	604 (red-green)	Ratio of band 27 and 29 gave the highest correlation with chlorophyll *a* of the whole training set. Low response to chlorophyll *a*.
39	675 (red)	Chlorophyll *a* absorption peak. Also one of highest correlations of ratios (with band 41).
41	689 (red-NIR)	Good correlation with chlorophyll *a* as single band and in ratio with band 39.
44	711 (NIR-red)	Highest red band correlation with chlorophyll *a* . Close to peak in reflectance.
47	732 (NIR)	Low correlation with chlorophyll *a* in the NIR.

†see appendix A for full details of wavebands

The fluorescence feature near 685 nm was also extracted. This was the same feature as was measured by the flow-through fluorometer although in this case it was solar-induced. The peak was not very clear in the CASI spectra but a drop in reflectance was almost always evident from band 40 (682 nm) to band 41 (689 nm). This was probably due to absorption by oxygen at 687 nm (Fischer and Schlüssel 1990). The fluorescence line height (FLH) feature is usually measured above a baseline described by two bands on either side of the peak (see equation 4.1 and figure 1.2). Several baselines were tried and the correlation coefficient of the FLH measurement with chlorophyll *a* was used to determine which FLH measurement would be used in model development (table 4.2). The correlation of the FLH and chlorophyll *a* should be high and linear since they are both measuring the fluorescence efficiency of the chlorophyll.

$$F = L_{\lambda_F} - \frac{L_{\lambda_1}(\lambda_2 - \lambda_F) + L_{\lambda_2}(\lambda_F - \lambda_1)}{\lambda_2 - \lambda_1} \tag{4.1}$$

where F is the fluorescence band,
1 and 2 are the first and second baseline bands,
L_{λ_n} is the radiance at band n and
λ_n is the wavelength of band n.

The baseline described by bands 39 and 47 (Gitelson92) had the highest correlation of 0.7282 and so was chosen for the rest of this research. This feature was subsequently calculated for the validation and testing sets.

4.3 Dividing data into training and testing sets

Three sets of *in situ* and spectral measurements were produced by dividing the data randomly. Because there were over 2000 data points available, a large training set could be constructed from only 30 % of the data, and a validation set from 20 % of the data. This allowed a very large testing set (50 %) to be created thus ensuring that the accuracy tests were valid. The training set was to be used to create both types of regression model, the validation set was to be used in training the neural networks and

Table 4.2: Trials of different Fluorescence Line Height measurements. FLH was measured at band 40 (centred on 682 nm) above a straight baseline described by two bands in the spectrum.

name of FLH measure	reference	baseline wavelengths†		baseline bands		correlation coefficient
Gitelson92	(Gitelson 1992)	675	730	39	47	0.73
Gitelson94a	(Gitelson *et al.* 1994)	650	715	36	45	0.53
Gitelson94b	(Gitelson *et al.* 1994)	670	730	38	47	0.72
FischerK90a	(Fischer and Kronfeld 1990)	645	725	35	46	0.56
FischerK90b	(Fischer and Kronfeld 1990)	645	670	35	38	0.67

†see appendix A for full details

the testing set was for testing the accuracy of the models.

By dividing the data using a random method, the statistical properties of the training, testing and validation sets were very similar. A subset of the validation set was chosen arbitrarily for validation of the neural network during training (this is described in detail in section 4.4.2) so that, if the network was to reduce the training error by predicting the mean chlorophyll *a* for every pattern and so over-fitting to the training data, then an increase in error would be evident for the validation set.

These sets are summarised in table 4.3 and their distributions are shown in the histogram in figure 4.1.

4.4 Deriving the model coefficients

Both the simple linear regression and the neural network models were produced using the *Norfolk 30/05/96* training data. The testing data were kept aside to test the models (section 4.5). The neural network also required the validation set to determine the value

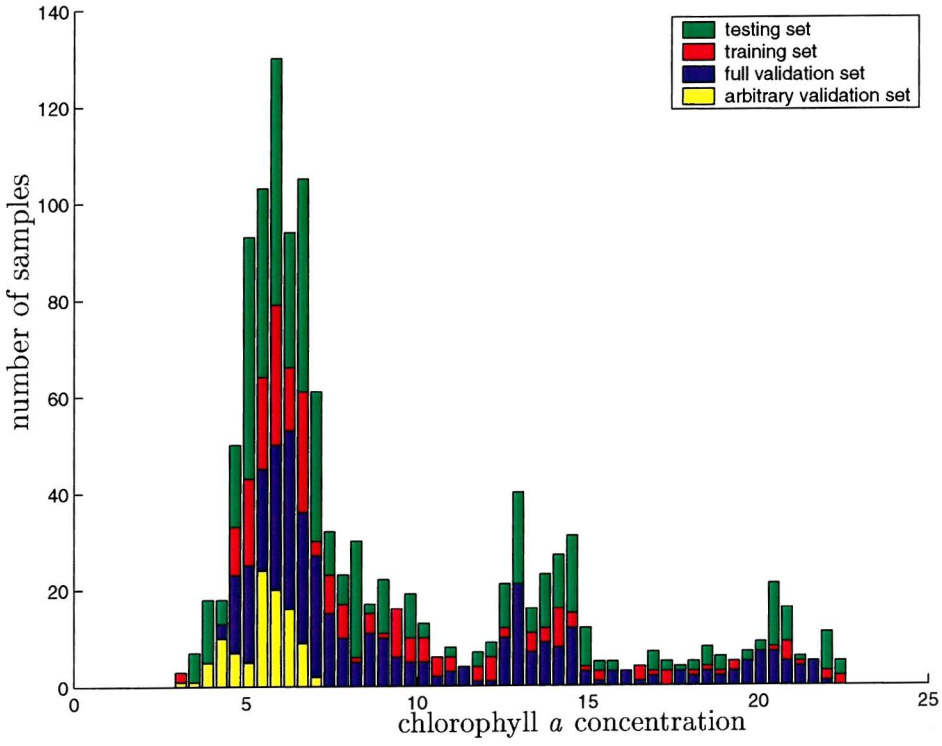


Figure 4.1: Histogram of the training, validation and testing sets for the primary study

Table 4.3: Statistics of the data sets which were used to train and assess the models in the primary study.

		chlorophyll a ($\mu g\, l^{-1}$)			
set	proportion of data	minimum	maximum	mean	standard deviation
training	30 %	3.15	22.62	8.86	4.63
validation	20 %	3.14	22.16	9.02	4.76
arbitrary validation set	~4 %	3.14	7.03	5.50	0.86
testing	50 %	2.91	22.60	8.98	4.73

of some of the parameters in its architecture. The following two sections describe, first, the determination of the coefficients for the simple linear models and second, the method by which the neural network model coefficients were derived. The technique of finding the model coefficients is commonly called 'training' in neural network literature. This term will be used for both methods of finding model parameters.

4.4.1 Training the simple linear models

It is common to use models with a single input in studies which determine the amount of chlorophyll a in water. Using the subset of bands defined in section 4.2, several sets of band ratios were produced. These were either based on ratioing a chlorophyll a absorption band with a region of the spectrum which had low correlation with chlorophyll a or on ratios that were found to have a high correlation with chlorophyll a (as described in table 4.1). Least squares regression was used to produce a linear approximation of the relationship $feature = f(Chl)$ for each feature using equation 3.5. This was then inverted to produce the model which would predict chlorophyll a concentration from the ratios ($Chl = Gradient * feature + Intercept$). The resulting linear models are summarized in table 4.4 and illustrated in figure 4.2.

Table 4.4: The inverted least-squares regression model of the chlorophyll a to spectra relationship

feature	description	correlation with chl a	gradient	intercept
blue-green ratio	$\frac{band\ 6}{band\ 20}$	0.13	489.13	-484.89
red-NIR ratio	$\frac{band\ 41}{band\ 47}$	0.16	584.26	-746.27
red-green ratio 1	$\frac{band\ 39}{band\ 20}$	0.50	554.99	-303.25
red-green ratio 2	$\frac{band\ 29}{band\ 27}$	0.77	480.75	-439.25
red-red ratio	$\frac{band\ 41}{band\ 39}$	0.74	795.49	-632.06
FLH	see section 4.2	0.73	0.99	-19.34

4.4.2 Training the neural networks

A single-layer network (figure 3.1) and a two-layer network (figure 3.2) were chosen for this research. A single-layer network can only model linear functions; this architecture was used to test whether a linear function was an adequate model. A network with two or more layers can model increasingly complex non-linear functions. Various spectral features were chosen as inputs. The output of the network was always chlorophyll a concentration in $\mu g\, l^{-1}$.

Training a multilayer perceptron can prove complicated because there are many adjustments that can be made to the network to improve the training. Several preliminary trainings were performed and tested against the validation set and, based on the outcomes, assumptions were made as to the best activation functions and learning rate to be used throughout the course of this research. For the two-layer network, a hyperbolic tan (tanh) activation function was used at the hidden nodes and a linear function used at the output nodes. A tanh function is a sigmoid that scales the data from -1 to 1. For both types of network, the learning rate remained 0.001 throughout this research.

A smaller set of spectral features than for the linear regression model was used to train

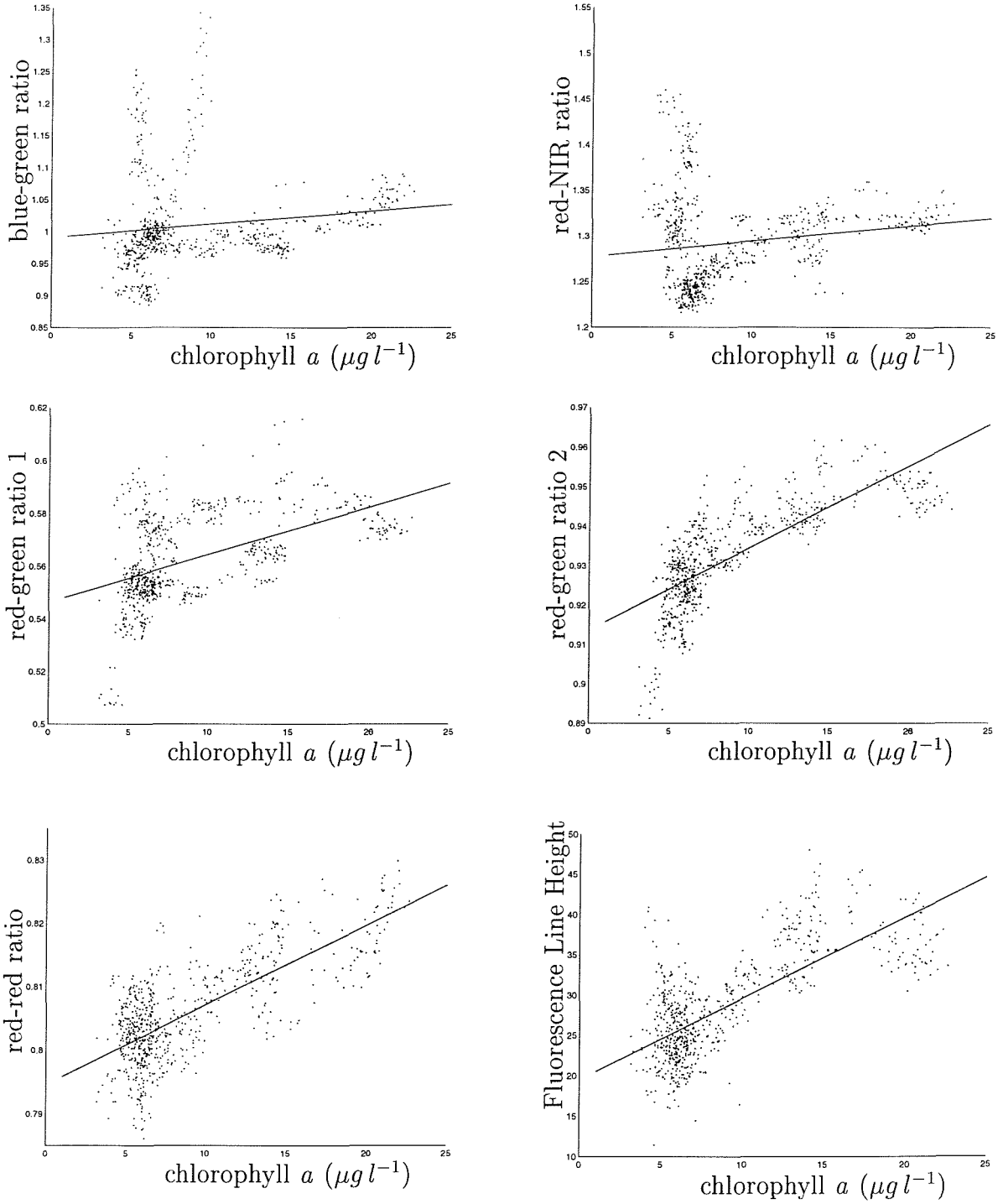


Figure 4.2: The linear model is illustrated by plotting a straight line over the scatter plot of data in the training set

the neural networks. These features were: bands 6 (441 nm), 20 (540 nm), 41 (689 nm) and the FLH measurement. The features were taken from the three main regions of the visible domain - blue (band 6) green (band 20) and red (band 41). This combination of four features allowed the exploration of the characteristics of these regions of the spectrum and of the FLH feature. With only these four features, 15 combinations of inputs were possible. The number of nodes in the hidden layer was varied for each possible combination of input. The possible numbers of nodes were: zero (the one-layer network), two, six or 10. By manipulating only these two parameters of the network, 60 networks were set up for training.

The naming convention for these networks will be *slp + number of inputs + code* (for single layer networks) or *mlp + number of inputs + number of hidden nodes + code* (for multi-layer networks) where the *code* signifies the combination of inputs used and is specific to the number of inputs. For example, *mlp36a* has three inputs of the *a* combination (bands 6, 20 and 41) and six hidden nodes. When discussing a number of neural networks with similar characteristics the wild card * shall be used. For example all 1-input, 2-hidden node networks will be referred to as *mlp12**. The names of all the networks and the number of iterations to which each of the 60 networks was trained to is given in table 4.5.

The neural networks were trained with the training data set. The 'arbitrary' validation data set was then used to estimate the ideal number of iterations for training the networks as follows. At the end of each iteration the arbitrary validation data points were passed through the network and the output error of this was calculated. Because the network weights were adjusted by the back-propagation of the training set only, it was expected that the validation error would decrease for a number of iterations, then increase when the network began to over-train. The number of iterations at which the validation error was minimised was then assumed to be the ideal number of iterations.

The 60 networks were all trained for 2000 iterations and then retrained to the number of iterations at which the validation set displayed the minimum error. These networks were then used to compare the models developed using a neural network model, to

those using linear regression.

key to table 4.5	
network name	see section 4.4.2, page 86
input features	the features used as inputs to the networks
number of hidden nodes	the number of hidden nodes. 0 indicates a 1-layer neural network
training MSE at 2000	the mean squared error of the training set at 2000 iterations
validation MSE at 2000	the mean squared error of the validation set at 2000 iterations
lowest validation MSE at: ..	the number of iterations at which the validation MSE was lowest
training MSE here	the MSE of the training set at which the validation MSE was lowest
validation MSE here	minimum validation MSE achieved for this training
correlation	correlation between the predicted and measured test set values. This is explained in section 4.5
RMSE	the root MSE between the predicted and measured test set values. This is explained in section 4.5

Table 4.5: Training and results of linear and non-linear neural networks in the primary study.

network name	input features	number of hidden nodes	training MSE at 2000	validation MSE at 2000	lowest validation MSE at:	training MSE here	validation MSE here	correlation	RMSE
slp1a	6	0	17.157	3.383	13	17.2314	3.1208	0.4919	4.1362
slp1b	20	0	17.4713	5.3474	10	18.063	4.3608	0.4635	4.2796
slp1c	41	0	14.3462	5.304	16	14.3744	5.235	0.611	3.7548
slp1d	FLH	0	10.0575	10.9864	7	13.2581	7.673	0.76	3.5307
mlp12a	6	2	0.6105	0.0664	172	0.6591	0.0366	0.6277	3.6828
mlp12b	20	2	0.7282	0.3039	110	0.7171	0.2227	0.5701	3.8896
mlp12c	41	2	0.4263	0.4388	10	0.6403	0.2274	0.6732	3.6924
mlp12d	FLH	2	0.3633	0.4804	7	0.5442	0.4456	0.7667	3.3970

continued on next page

table 4.5 continued from previous page

network name	input features	number of hidden nodes	training MSE at 2000	validation MSE at 2000	lowest validation MSE at:	training MSE here	validation MSE here	correlation	RMSE
mlp16a	6	6	0.6219	0.1056	115	0.663	0.0338	0.6259	3.6916
mlp16b	20	6	0.7402	0.3239	131	0.7164	0.2184	0.5709	3.8882
mlp16c	41	6	0.4257	0.4366	8	0.6448	0.2211	0.657	3.7030
mlp16d	FLH	6	0.3618	0.491	6	0.5097	0.4526	0.766	3.2639
mlp110a	6	10	0.6192	0.0983	115	0.6655	0.0332	0.6248	3.6961
mlp110b	20	10	0.7288	0.3031	99	0.7183	0.2217	0.5689	3.8956
mlp110c	41	10	0.4256	0.4351	6	0.6847	0.2269	0.6346	3.8236
mlp110d	FLH	10	0.3618	0.4893	4	0.5774	0.4591	0.7626	3.5049
slp2a	6,20	0	16.9167	3.6551	14	16.9654	3.4699	0.4998	4.1108
slp2b	6,41	0	14.3617	5.1752	14	14.4894	4.7835	0.6105	3.7746
slp2c	6,FLH	0	9.9126	8.8655	9	11.2301	6.9917	0.7675	3.2252
slp2d	20,41	0	10.5211	9.6166	12	13.4466	4.4229	0.6659	3.6343
slp2e	20,FLH	0	10.0518	10.8472	8	12.1053	7.8029	0.7611	3.3637
slp2f	41,FLH	0	9.8724	9.2292	9	11.2888	7.6961	0.7686	3.2352
mlp22a	6,20	2	0.6067	0.1025	463	0.633	0.0735	0.6455	3.6140
mlp22b	6,41	2	0.2488	0.0618	$\geq$2000	0.2488	0.0618	0.8789	2.2840
mlp22c	6,FLH	2	0.3192	0.2434	8	0.5094	0.2328	0.7701	3.2432
mlp22d	20,41	2	0.3764	0.3987	8	0.6726	0.2214	0.629	3.7893
mlp22e	20,FLH	2	0.343	0.488	8	0.5188	0.3244	0.7637	3.2889
mlp22f	41,FLH	2	0.291	0.4152	5	0.4857	0.2894	0.7915	3.1753
mlp26a	6,20	6	0.4769	0.0684	1780	0.4984	0.0634	0.7306	3.3269
mlp26b	6,41	6	0.231	0.0598	1351	0.2369	0.0587	0.8845	2.2447
mlp26c	6,FLH	6	0.2208	0.0634	$\geq$2000	0.2208	0.0634	0.8959	2.1290
mlp26d	20,41	6	0.3721	0.3799	5	0.7023	0.2223	0.6074	3.8786
mlp26e	20,FLH	6	0.3184	0.5135	5	0.5617	0.3434	0.7617	3.4378
mlp26f	41,FLH	6	0.2782	0.3973	4	0.4854	0.3006	0.7797	3.1578
mlp210a	6,20	10	0.4907	0.0881	$\geq$2000	0.4907	0.0881	0.7554	3.2873
mlp210b	6,41	10	0.2283	0.0586	$\geq$2000	0.2283	0.0586	0.8929	2.1876
mlp210c	6,FLH	10	0.2243	0.0727	$\geq$2000	0.2243	0.0727	0.8934	2.1556
mlp210d	20,41	10	0.3727	0.3644	5	0.7058	0.2158	0.5959	3.8876
mlp210e	20,FLH	10	0.3189	0.5102	4	0.5248	0.3261	0.7588	3.2988
mlp210f	41,FLH	10	0.282	0.4218	4	0.5509	0.3027	0.7656	3.3938
slp3a	6,20,41	0	9.4785	6.283	15	12.4302	3.3058	0.7009	3.4720
slp3b	6,20,FLH	0	9.8235	8.6167	9	11.1597	6.8509	0.7698	3.2130
slp3c	6,41,FLH	0	9.8632	8.7347	10	10.7658	7.4115	0.7693	3.1549
slp3d	20,41,FLH	0	8.7713	11.5411	9	10.9406	8.2164	0.7781	3.1821
mlp32a	6,20,41	2	0.1624	0.1262	$\geq$2000	0.1624	0.1262	0.9278	1.7694
mlp32b	6,20,FLH	2	0.2052	0.1264	$\geq$2000	0.2052	0.1264	0.9175	1.9356
mlp32c	6,41,FLH	2	0.1621	0.1017	$\geq$2000	0.1621	0.1017	0.9305	1.7569
mlp32d	20,41,FLH	2	0.2832	0.4344	7	0.5264	0.2718	0.7752	3.3213
mlp36a	6,20,41	6	0.1432	0.1136	$\geq$2000	0.1432	0.1136	0.9347	1.6871
mlp36b	6,20,FLH	6	0.1341	0.1478	518	0.1803	0.1035	0.923	1.8451
mlp36c	6,41,FLH	6	0.1164	0.0803	$\geq$2000	0.1164	0.0803	0.9526	1.4620

continued on next page

table 4.5 continued from previous page

network name	input features	number of hidden nodes	training MSE at 2000	validation MSE at 2000	lowest validation MSE at:	training MSE here	validation MSE here	correlation	RMSE
mlp36d	20,41,FLH	6	0.2548	0.3638	4	0.533	0.2613	0.7564	3.3377
mlp310a	6,20,41	10	0.1322	0.0876	$\geq$2000	0.1322	0.0876	0.9398	1.6180
mlp310b	6,20,FLH	10	0.134	0.1658	631	0.1908	0.1408	0.9246	1.8115
mlp310c	6,41,FLH	10	0.109	0.0583	$\geq$2000	0.109	0.0583	0.9538	1.4498
mlp310d	20,41,FLH	10	0.2577	0.3532	4	0.5253	0.2688	0.7528	3.2994
slp4a	6,20,41,FLH	0	8.2217	8.1801	10	10.2596	7.2131	0.785	3.0689
mlp42a	6,20,41,FLH	2	0.1279	0.0989	$\geq$2000	0.1279	0.0989	0.9464	1.5410
mlp46a	6,20,41,FLH	6	0.0813	0.0459	$\geq$2000	0.0813	0.0459	0.9669	1.2117
mlp410a	6,20,41,FLH	10	0.0694	0.0279	1640	0.0746	0.0269	0.9685	1.1838

4.5 Results

The models developed in the previous section were tested by comparing their predictions of test set chlorophyll *a* values with the measured values. Two measures were used for this comparison. The first is the correlation between the predicted output and the desired output (this was calculated as in equation 2.3 but is reiterated here)

$$\rho = \frac{\sum_{q=1}^{p}(y_q - \bar{y})(\hat{y}_q - \bar{\hat{y}})}{\sqrt{\sum_{q=1}^{p}(y_q - \bar{y})^2 \sum_{q=1}^{p}(\hat{y}_q - \bar{\hat{y}})^2}} \tag{4.2}$$

where y_q is the qth point in the testing set, $\hat{y}_q$ is the prediction for this point, $\bar{y}$ is the mean of chlorophyll *a* measurements in the testing set, $\bar{\hat{y}}$ is the mean of the predictions of these measurements and p is the number of data points in the testing set. This measurement is the normalized covariance of the predicted and desired values.

The second measure is the Root Mean Squared Error (RMSE) which was calculated as

$$\varepsilon = \sqrt{\frac{1}{p}\sum_{q=1}^{p}(y_q - \hat{y}_q)^2}\ . \tag{4.3}$$

This measured the accuracy of the results and increased with the error of the prediction. Because both linear least-squares regression and neural network regression train by minimising the mean squared error, these models always try to predict the mean output value given the inputs values. When presented with a set of data for which they are

poorly specified, the models will tend to predict values near the mean for the whole training set. Where the training and testing data have similar statistical characteristics, a model that predicts only the mean value of the data can be identified by predictions with a RMSE that is approximately equal to the standard deviation of the desired values.

4.5.1 Testing the simple linear models

As can be seen in figure 4.2, the relationship between band ratios and chlorophyll *a*, as measured by the fluorometer, was quite complicated. The blue-green and red-NIR ratios were particularly non-linear, displaying filaments of data points. For these two relationships the geographical position of the data points in the filaments were plotted (figure 4.3).

From these plots, there appeared to be a number of regions in which the relationship between spectra and chlorophyll *a* concentration was different. These may have been caused by environmental factors such as sediment or CDOM in the water, wind roughening of the surface causing glint or foam or atmospheric effects, such as haze. The tiny region characterised by the green filament in the plot of blue-green ratio against chlorophyll *a* concentration corresponded to the region in all the principal components for which a great range in values was noted in section 2.7.4. The red region in this plot corresponded to the region in which a peak in values was found in the first principal component. Because the two ratios did not appear to display the same division of regions, it was thought that each was affected by a different environmental factor.

The other three ratios were better distributed around the line of the model. The red-green ratio 1 showed a slight separation of data points into regions in the plot (figure 4.2), perhaps again as a result of environmental effects. The red-red ratio appeared to have the most linear ratio-chlorophyll *a* relationship.

The FLH measure was a special case since it was considered to be similar to the

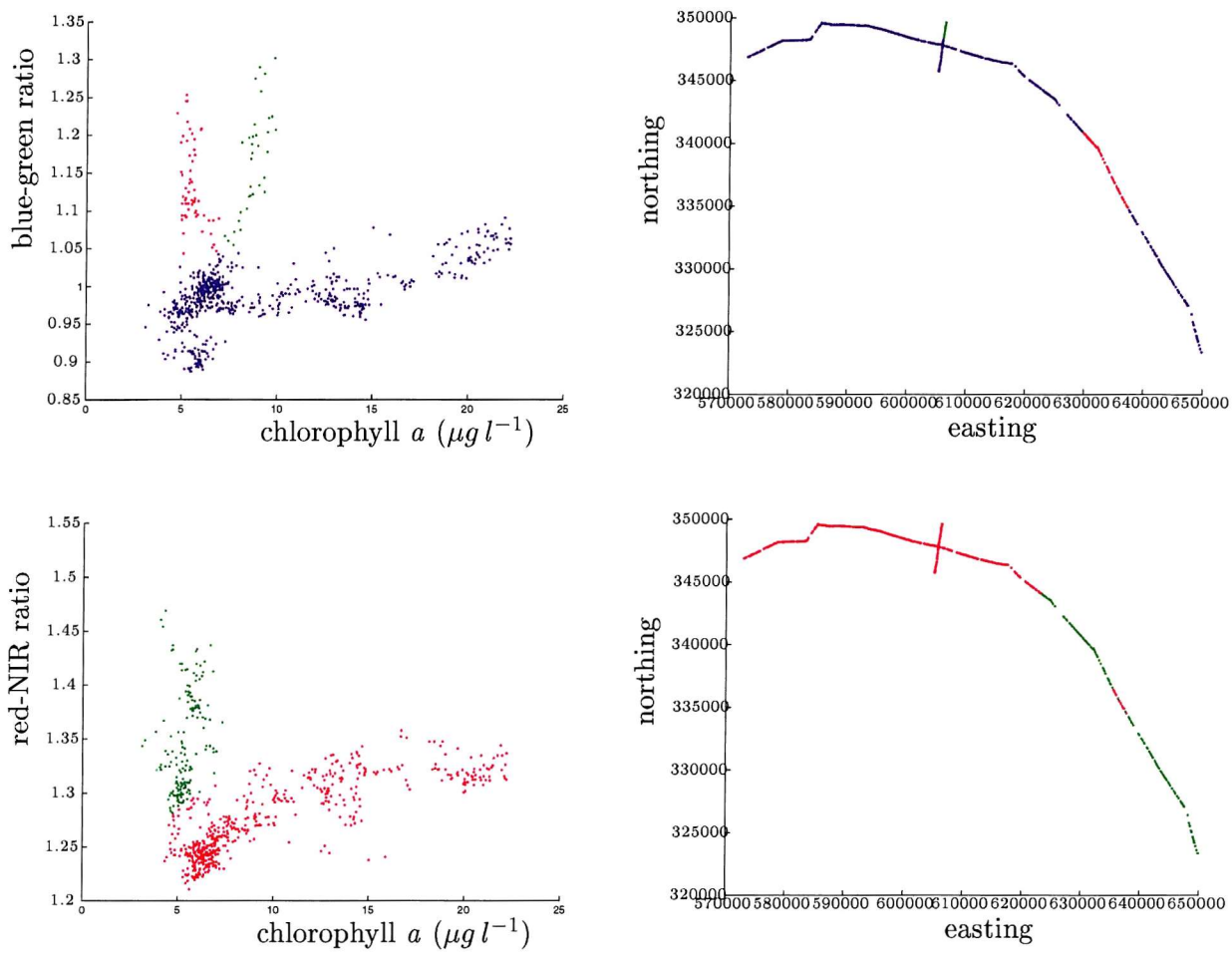


Figure 4.3: Different water bodies within the Norfolk region. The coloured filaments in the ratio-chlorophyll a plots are plotted geographically along the cruise transect.

chlorophyll a measure (as this was measured using the fluorometer). The gradient of this relationship was indeed nearly 1:1, however there was quite a large offset of about 40 % of the FLH measure and the relationship appeared slightly non-linear. These deviations from the 1:1 relationship illustrated the interference in the system caused by other water constituents, the water surface and the atmosphere.

Table 4.6 gives the correlation and RMSEs of the chlorophyll a values predicted by the linear models. There was quite a distinction between the performances of the first three models and second three models. The blue-green, red-NIR and red-green 1 ratios

did not perform well, the first two had error values which indicated that predicted chlorophyll *a* was well outside the range of measured values. The second red-green ratio, the red-red ratio and the FLH measure gave the better results, with errors below the standard deviation of the chlorophyll *a* for this set (4.73 $\mu g\, l^{-1}$) and good correlations. These models were based on features which were close to each other in the spectrum. As pointed out by Neville and Gower (1977) the effects of the atmosphere, CDOM and suspended sediment can be found to be approximately uniform over a short spectral range. Therefore the better performance of the second red-green ratio, the red-red ratio and the FLH measure may have been because they were measured over a short spectral range.

Table 4.6: The correlation and RMSE for the inversion of linear least-squares models

Feature	ρ	ε
blue-green ratio	0.16	34.78
red-NIR ratio	0.16	27.90
red-green ratio 1	0.50	8.40
red-green ratio 2	0.76	4.09
red-red ratio	0.76	4.28
FLH	0.76	4.27

$\rho \equiv$ correlation, $\varepsilon \equiv$ RMSE

4.5.2 Comparison of both types of linear models

The results in the previous section were compared to the linear neural network results (tables 4.6 and 4.7). Two of these models were directly comparable - the FLH least-squares models with *slp1d* and the blue-green ratio models with *slp2a*. The correlations for the FLH linear least-squares model and the *slp1d* model were the same but the neural network produced a lower error. The blue-green ratio did not perform at all well compared to *slp2a*. The error of the *slp2a* model prediction was just less than the standard deviation of the testing set. Although none of the other linear neural

networks were directly comparable, it was clear that those with more than one input performed better than the least-squares models. This indicates that the problem has a greater level of complexity than one input can resolve.

Table 4.7: The correlation and RMSEs for the linear neural network

network	features	ρ	ε
slp1a	blue	0.49	4.14
slp1b	green	0.46	4.28
slp1c	red	0.61	3.75
slp1d	FLH	0.76	3.53
slp2a	blue, green	0.50	4.11
slp2b	blue, red	0.61	3.77
slp2c	blue, FLH	0.77	3.23
slp2d	green, red	0.67	3.63
slp2e	green, FLH	0.76	3.36
slp2f	red, FLH	0.77	3.24
slp3a	blue, green, red	0.70	3.47
slp3b	blue, green, FLH	0.77	3.21
slp3c	blue, red, FLH	0.77	3.15
slp3d	green, red, FLH	0.78	3.18
slp4a	blue, green, red, FLH	0.79	3.07

$\rho \equiv$ correlation, $\varepsilon \equiv$ RMSE

4.5.3 Comparison of linear and non-linear neural network models

Here, the neural network models were used to compare the effect that altering different neural network parameters had on the final prediction. From this it was possible to draw conclusions about the nature of the data. Table 4.5 gives the correlation and RMSE for all 60 networks. Table 4.8 summarizes these data by finding the mean

correlation and RMSE for networks with a given number of inputs, number of hidden nodes, input feature or combination of features.

Table 4.8: These tables summarize the goodness of the neural network predictions for *Norfolk 30/05/96*.

number of inputs	$\sum$	$\mu\rho$	$\mu\varepsilon$	$\mu\mathcal{I}$
1	16	0.64	3.74	51.81
2	24	0.74	3.21	571.50
3	16	0.86	2.39	950.44
4	4	0.92	1.75	1412.50

number of hidden nodes	$\sum$	$\mu\rho$	$\mu\varepsilon$	$\mu\mathcal{I}$
0	15	0.68	3.54	11.00
2	15	0.77	2.96	719.87
6	15	0.79	2.82	795.13
10	15	0.79	2.83	834.13

input includes band	$\sum$	$\mu\rho$	$\mu\varepsilon$	$\mu\mathcal{I}$
blue	32	0.81	2.59	1090.22
green	32	0.75	3.01	609.66
red	32	0.79	2.82	723.63
FLH	32	0.83	2.74	591.56

input includes bands	$\sum$	$\mu\rho$	$\mu\varepsilon$	$\mu\mathcal{I}$
blue, green	16	0.83	2.42	1192.50
blue, red	16	0.88	2.12	1440.00
blue, FLH	16	0.88	2.15	1177.19
green, red	16	0.80	2.74	732.44
green, FLH	16	0.83	2.65	553.56
red, FLH	16	0.84	2.56	731.63

$\sum \equiv$ The number of networks' results that the mean correlation and RMSE were calculated from (see table 4.5), $\mu\rho \equiv$ Mean correlation coefficient, $\mu\varepsilon \equiv$ Mean RMSE $\mu\mathcal{I} \equiv$ mean number of iterations at which the validation set had reached minimum

The top left table in table 4.8 shows how the average correlation, RMSE and number of iterations required during training changed with the number of input nodes. There was a very clear increase in correlation and decrease in error with more input features, indicating that each input added more useful information. With the increase in inputs,

there was an increase in the number of weights in the network. This was reflected by the requirement for more iterations for the network to find minimum error.

The top right table compares the number of hidden nodes. Where no nodes were used (a one-layer network), the predictions were not as good. A marked improvement is apparent for the two-layer networks and those networks with 6 hidden nodes resulted in slightly better prediction than those with 2 hidden nodes. However, no more improvement in prediction resulted from having 10 hidden nodes. Again the increase in the number of hidden nodes caused the networks to require, on average, more iterations to train.

The two bottom tables indicate the effect that particular bands have on the outcome of the model and whether certain bands complement each other in the model. The greatest difference here was that band 20 (green) did not appear to contribute as much to the model as the three other bands. Band 6 (blue) in combination with FLH or with band 41 appeared to produce much better models than any other combinations of features. The number of iterations required was particularly high when the blue band was included in the inputs. This indicated that this input had a more complex relationship with chlorophyll *a*.

The best correlations and error values were always achieved for the networks with all four inputs, indicating that it was the inputs that had the greatest effect on the resulting model. The best network overall was the most complex one using all four inputs and 10 hidden nodes.

Figure 4.4 shows the regression functions derived by the neural network when only one input was used for the four different configurations of hidden nodes. The shape of the non-linear models closely resembled the tanh activation function used in the hidden nodes. Greater non-linearity was apparent for the blue and green bands and the FLH was most linear.

Figure 4.5 illustrates the distribution of the error of prediction for all the networks. In most cases the error was roughly symmetrical about zero. The most skew error

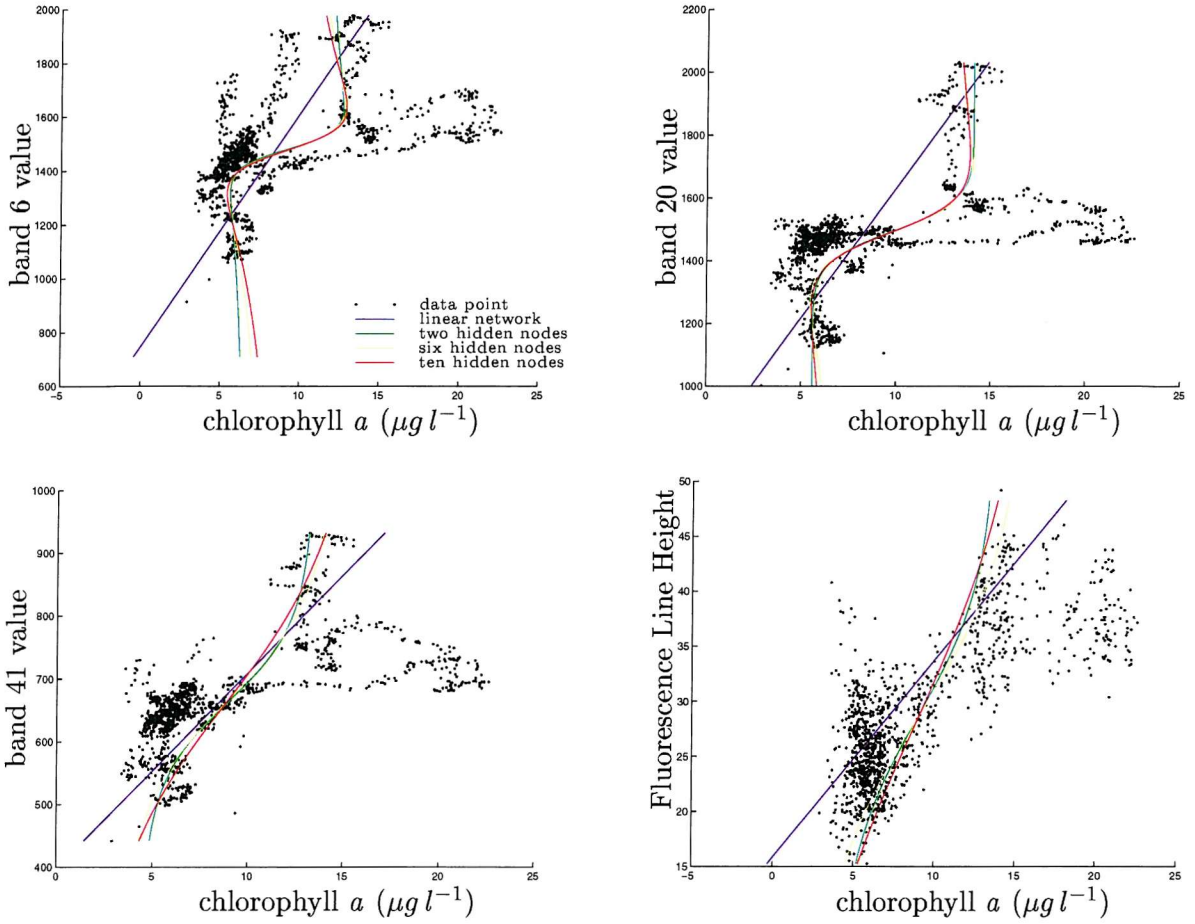


Figure 4.4: The regression functions derived by the neural network for one-input networks

distributions (*slp1a* - *c*, *slp2a* - *b*, *slp2d*, *slp3a* and the *mlp2*d*s) did not have FLH as an input. This was particularly apparent with the linear networks. Many of the non-linear networks had a leptokurtic error distribution (with a high peak compared to width) which was symmetrical about zero. This was particularly evident for the 3- and 4-input networks and also the 2-input networks with the *b* combination of inputs (bands 6 and 41), reflecting the finding on table 4.8 that it was the combination of inputs that contributed most to prediction accuracy. The non-linear 3-input networks with the *d*-combination of inputs performed poorly compared to other non-linear 3-input networks. This input combination did not contain the blue band and so was another indicator of the value of this band.

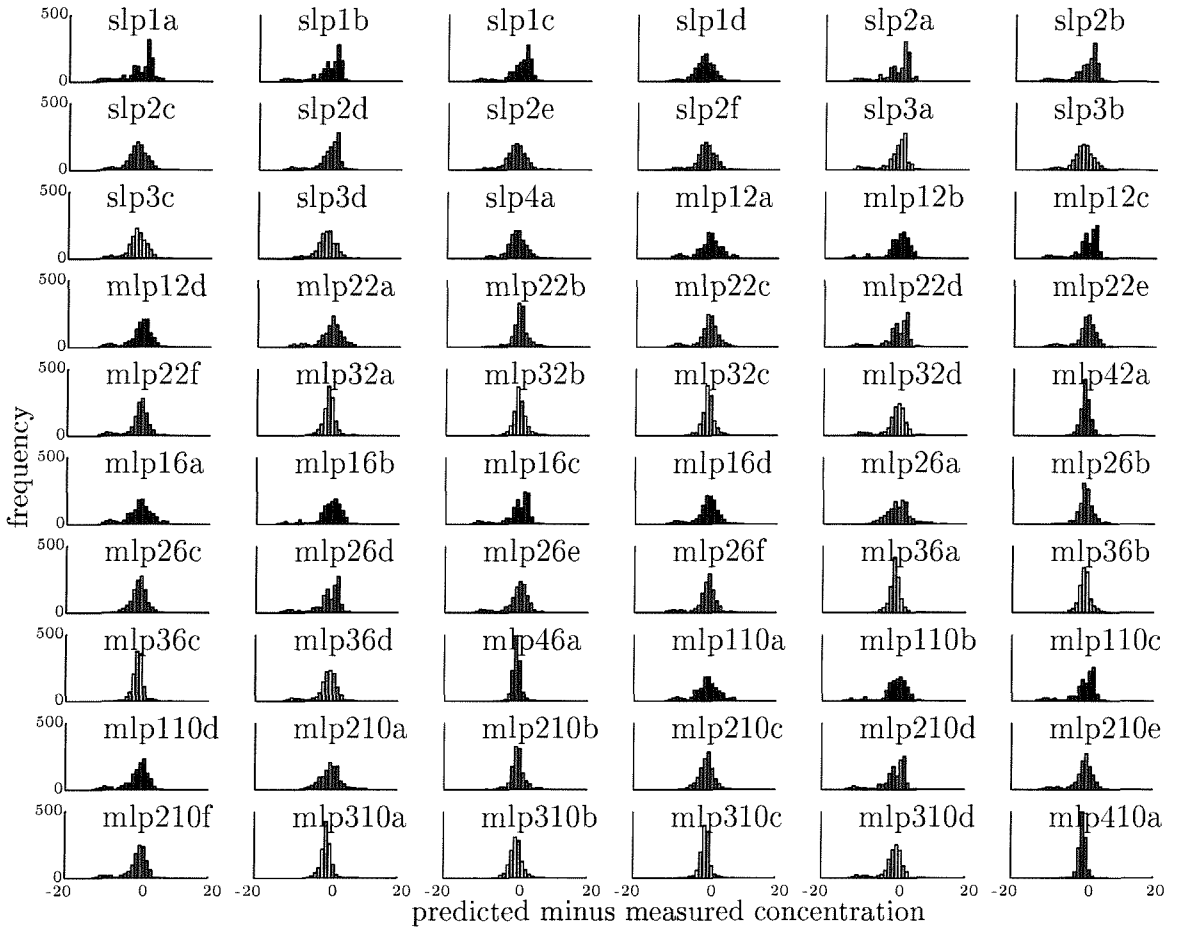


Figure 4.5: Distribution of prediction error. The histograms show the distribution of the error of chlorophyll *a* as predicted by each of the neural network models.

Figure 4.6 shows the values predicted by the neural networks plotted against the measured values. Many of the networks, particularly the one-input networks, *mlp2*c-f* and *mlp3*d* did not show a good one to one relationship. These networks tended to predict a value of around the mean for the training set (8.86 $\mu g\, l^{-1}$). Therefore, even those predictions with an error that was less than the standard deviation of the data were not very reliable. For some of the other networks especially those with three or four inputs, the values are reasonably well distributed around the one to one line.

Another indicator of the importance of the combination of inputs was that, if networks with the same architecture (number of inputs and hidden nodes) were ranked according

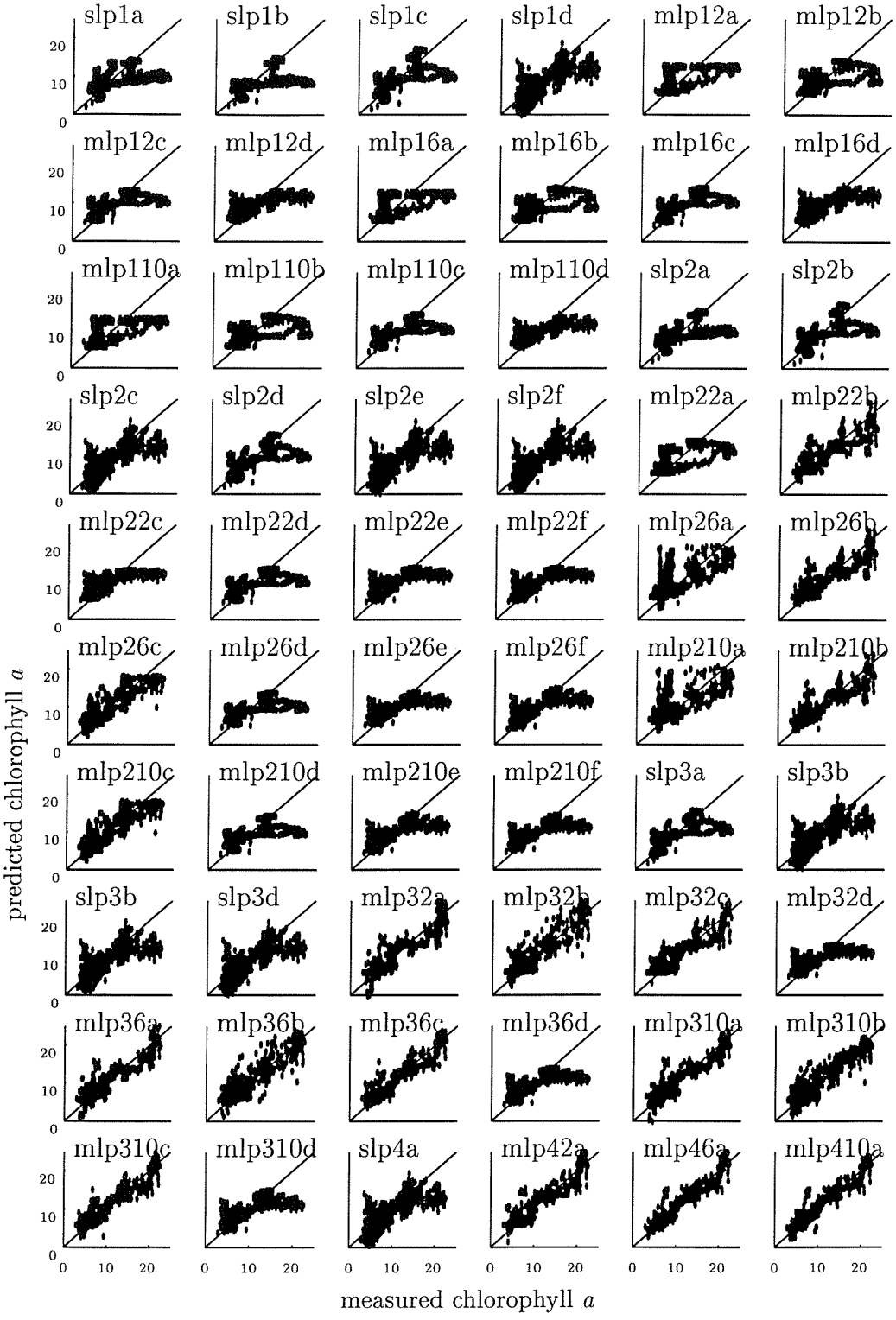


Figure 4.6: Measured versus predicted chlorophyll *a* values in the primary study

to their output error, then it was generally found that networks with the same input features had the same rank. The main exceptions to this were the linear networks for

which the best ranking networks were always those with FLH or FLH and band 6 as inputs.

One of the linear regressions and three of the trained networks were used to predict the chlorophyll *a* for the whole of the Norfolk 30/05/96 image data (figure 4.7, page 103) For much of the center of the images, the linear regression with the blue-green ratio (figure 4.7(a)) predicted chlorophyll *a* concentrations of zero or below. This image clearly showed that the brightening at the edges of the CASI images has a very strong effect on this linear regression model as very high chlorophyll *a* concentrations were predicted for the edges of the images. The neural networks were *mlp22b* with bands 6 and 41 as inputs and 2 hidden nodes, *mlp36c* with bands 6 and 41 and FLH as inputs and 6 hidden nodes and *mlp410a* with all four inputs and 10 hidden nodes. The range of chlorophyll *a* predictions was not extreme for most of these images. Overestimates may have occurred in the north of the region, possibly where edge brightening had affected the images. Very few of the training or testing data points were affected by edge brightening because the cruise was located along the centre columns of the images. Therefore, this effect was not explicitly corrected for by the models during training. The two more complex networks appeared a little more sensitive to noise in the images whereas *mlp22b* (two inputs, 2 hidden nodes) produced a much smoother chlorophyll *a* prediction.

4.6 Discussion of results of primary study

From this primary analysis of the *Norfolk 30/05/96* dataset it was clear that conventional techniques of linear regression of band ratios were not adequate for the prediction of chlorophyll *a* from these data. The inverted least-squares analysis was a very simple method of producing models which predicted the chlorophyll *a* in the water from light detected above the water. The neural networks were rather more complicated to train but were found to be a simple way of combining several features and of developing non-linear regression models. Also, the results of the different network architectures

allowed some insights into the nature of the data being modelled.

The simpler networks reached a validation error minimum much sooner than the more complex networks. A training neural network will always produce a linear model in the initial iterations. Those networks which required the fewest iterations tended to have FLH as an input. Those networks which required the most training iterations tended to have blue band 6 as an input. This indicated that FLH had a near-linear relationship with chlorophyll *a* and that blue had the most non-linear. From the improved performance with the two-layer networks, it appeared that a non-linear relationship did exist between all ocean colour features and chlorophyll *a*, which explained the poor performance of the linear regression functions.

For the two-layer networks, the number of hidden nodes did not have a great impact on the prediction, even when there were fewer hidden nodes than inputs, (for example *mlp32** and *mlp42**). This indicated that, although a non-linear model was necessary a particularly complex one was not. However, it was found that different signatures did occur in different localities (figure 4.3) when observing the blue-green and red-NIR ratios. This fitted with the findings in chapter 2 that a number of influences on the spectral data, other than chlorophyll *a*, varied geographically. The complexity may have occurred in only certain parts of the spectrum. This was reflected by the increase in the number of iterations required by the blue input band to reduce the prediction error.

Perhaps the strongest agreement found from other studies was that it was the choice and combinations of inputs to the models that had the most significant effect on their outcome. With the non-linear networks, particular combinations of inputs consistently performed better than others.

As has been found in several studies, the FLH measure gave good predictions for the linear regressions. As was expected, the relationship between this and the chlorophyll *a* measure (which was based on fluorescence measurements) was the most 'linear' because the single-layer networks which had FLH as an input always gave better results than

other single-layer networks. This idea of linearity was supported by the short training time required by networks with FLH as an input.

The performance of the blue band 6, was more of a surprise. Recent studies have avoided this part of the spectrum because of noise or the interference of other water constituents and the atmosphere (e.g. Dekker *et al.* 1992a). This study confirmed that in the original image data the blue band was noisy and the complex relationship between shorter wavelengths and chlorophyll *a* was highlighted by the plots of blue-green ratios against chlorophyll *a* (figure 4.2) and blue against chlorophyll *a* (figure 4.4). However, band 6 provided good predictions as an input to non-linear and complex regressions.

It was noted that neither the blue-green ratio model (figure 4.2) nor the blue-input neural network (figure 4.4) found a negative slope between the spectral feature and chlorophyll *a*. This indicated that the absorption of blue wavelengths by chlorophyll *a* was not being detected. Other studies have found a similar effect in the coastal zone (e.g. Fischer *et al.* 1986).

Although the green band, band 20, did not perform well, the four-input networks always performed better than any others. This indicated that this band did contribute some information to the predictions. With the increase in accuracy of prediction, which occurred with more inputs, there was no tail off in the increase in accuracy indicating that each band contributed information to the prediction. The FLH feature consistently gave good predictions and was the most suitable of all the features when only linear modelling was possible. The blue wave band appeared to have the most non-linear relationship to chlorophyll *a*, this relationship was strong and so the blue wave band enabled some good predictions of chlorophyll *a*.

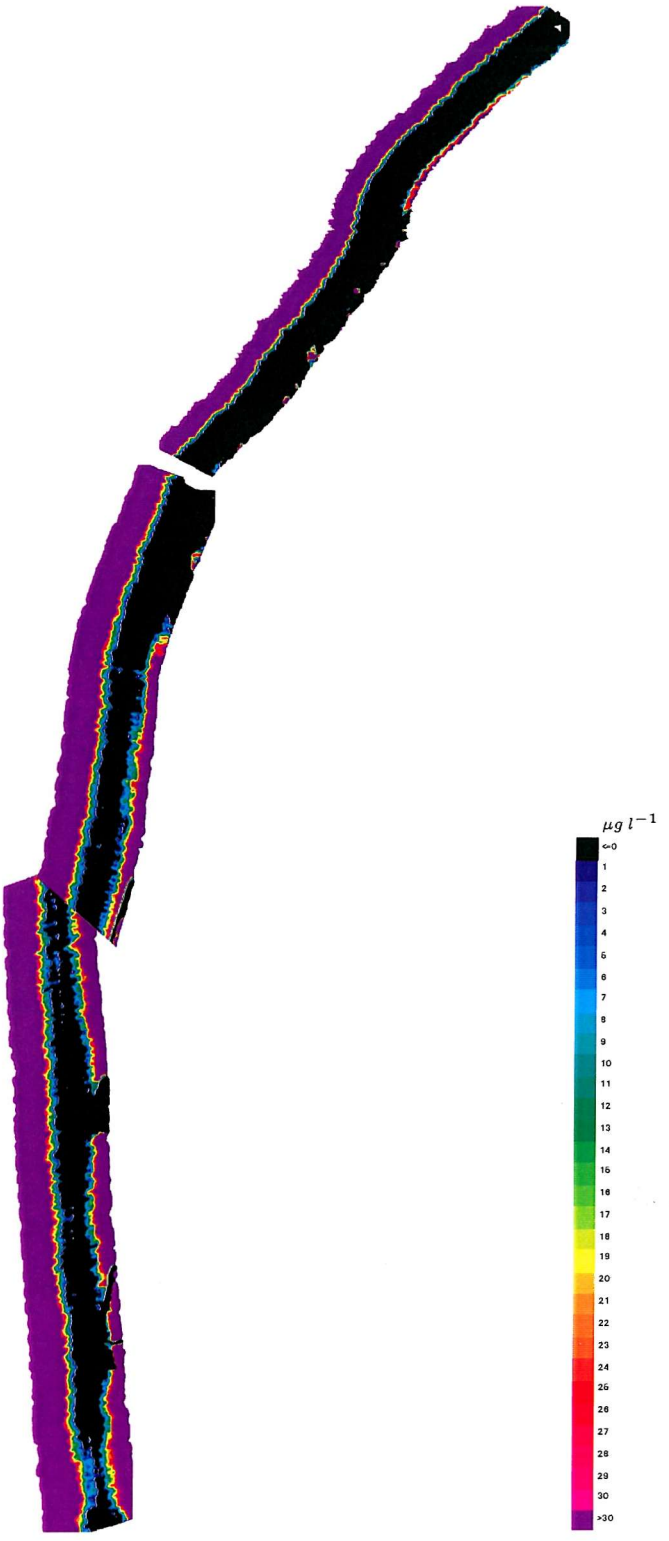
A factor that was not tested in the neural network models was whether the closeness of features played a rôle in improving these regressions. Of the other linear regressions, those ratios produced from spectrally-close wave bands gave better performances. This is an issue worth exploring in future studies.

The results showed that, even in an environment contaminated by atmospheric effects, suspended sediment and CDOM, it was still possible to detect chlorophyll *a* in the *Norfolk 30/05/96* imagery. The effects of this contamination was notable by the need for several sources of information (wavebands) and non-linear models. The potential of the neural network technique had been verified such that the research could be expanded to investigate further the issues found to be most important to the prediction of chlorophyll *a*.

Summary of chapter 4

The primary study showed that there is a need for better models than the single-ratio linear models that are often used for Case 2 waters. Increasing the number of inputs gained increased accuracy in prediction and certain inputs were found to contribute more to the accuracy than others. The features that performed well depended on the model being used. The FLH performed well with the linear models but the blue band performed best with the non-linear models. The non-linear models performed better than the linear models. The symmetrical error distribution demonstrated by most of the neural networks indicated that these models were well-specified and that neural networks were a valid method of defining non-linear chlorophyll *a* prediction models.

The technique of comparing the performances of different neural network architectures that has been described in this chapter is a novel use of neural networks and has produced some new insights into problem of predicting chlorophyll *a* from spectral data that has been collected over Case 2 waters. In particular, it has been shown that non-linear but not very complex models predict much better than simple linear models. Also, the number and type of input has been found to be the most important aspect to creating good chlorophyll *a* prediction models. This second point was investigated further, as described in the following chapter.



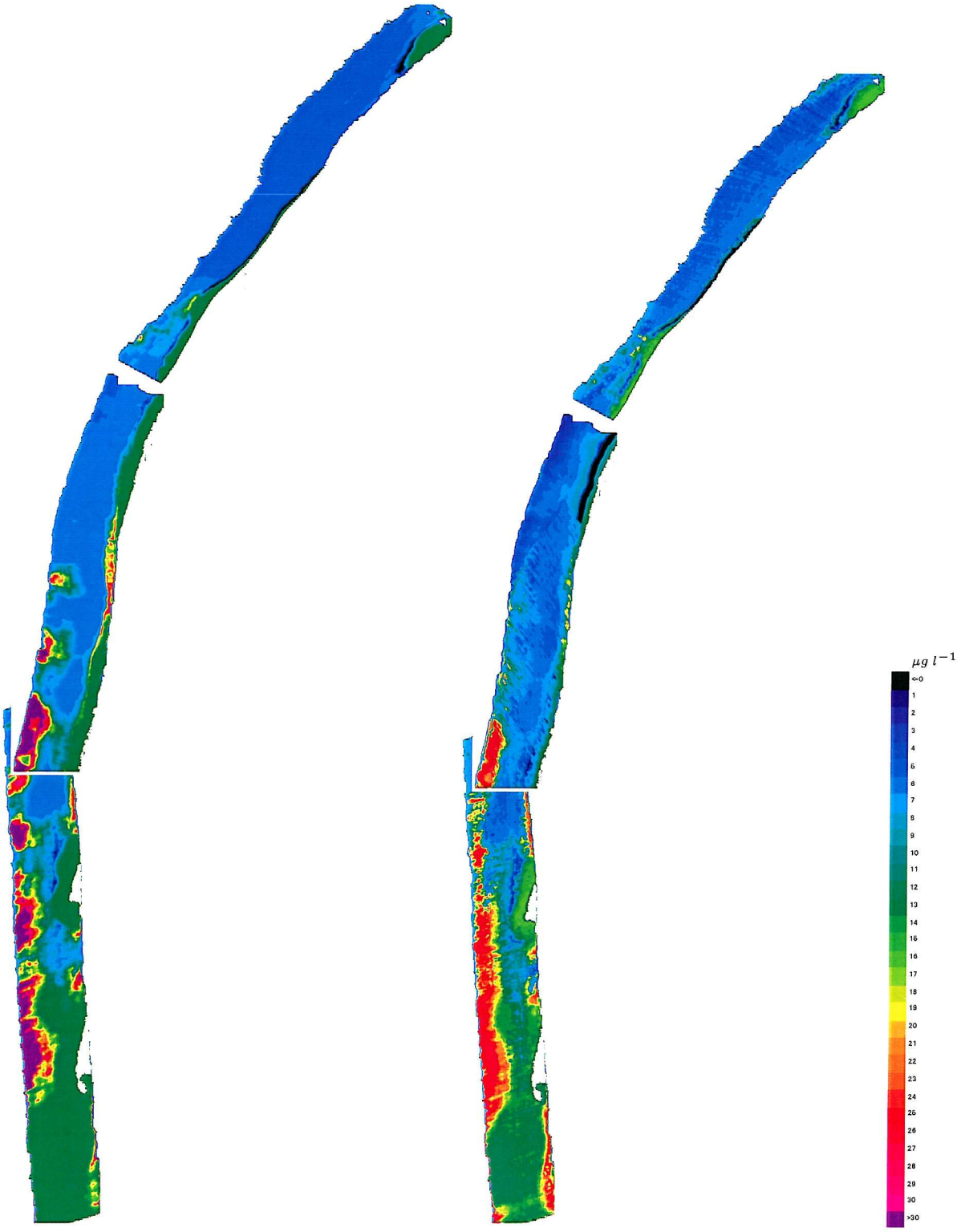
(a) Linear regression of the blue-green ratio

Figure 4.7: One of the linear regressions and three of the trained neural networks were used to produce the following three images. These neural networks increase in complexity:

mlp22b with two inputs (bands 6 and 41) and two hidden layers

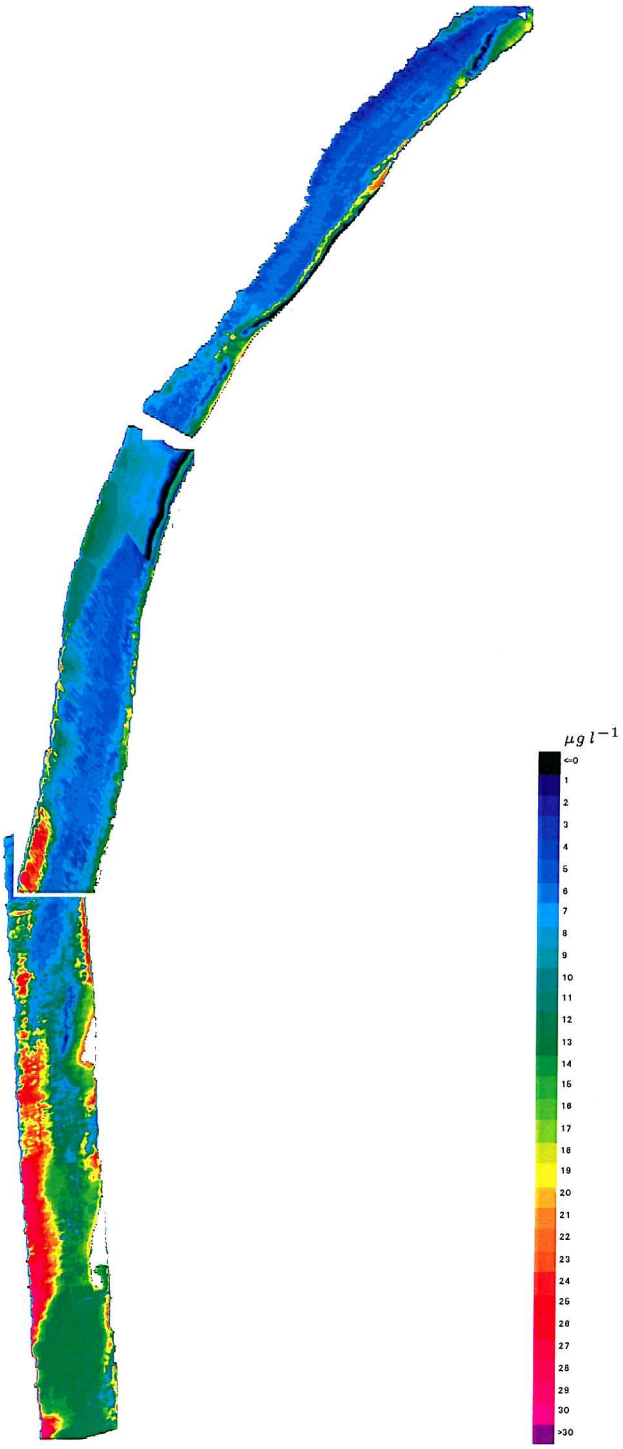
mlp36c with three inputs (bands 6 and 41 and FLH) and six hidden nodes

mlp410a with four inputs (bands 6, 20 and 41 and FLH) and 10 hidden nodes.



(b) *mlp22b* at 2000 iterations

(c) *mlp36c* at 2000 iterations



(d) *mlp410a* at 1640 iterations

Chapter 5

Development of method

5.1 Introduction

In the primary study it was demonstrated that increasing the number of inputs to the model produced more accurate predictions of chlorophyll *a*. It was also found that certain inputs provided more useful information to the model than others. This chapter describes three methods of selecting features for the prediction of chlorophyll *a* and then shows how models were developed with the chosen sets of features.

5.2 Creating a training, validation and testing set

A different approach was used to select the data in this secondary study. It was considered that having a large number of training data points with very similar values (figure 2.5) may bias the models during training. Therefore the new training set was selected so that the chlorophyll *a* concentrations within this set were evenly distributed throughout the range of values. This was achieved by first dividing the data randomly into training, validation and testing sets containing respectively 673, 469 and 1158 data points (table 5.1).

Table 5.1: Statistics of the data sets which were used to train and assess the models in the secondary study.

		chlorophyll a $\mu g\, l^{-1}$			
set	proportion of data	minimum	maximum	mean	standard deviation
training	30 %	3.14	22.22	8.96	4.75
selected training set	~4 %	3.14	22.10	12.69	5.57
validation	20 %	3.17	22.59	9.09	4.91
testing	50 %	3.40	22.61	8.91	4.62

One hundred data points were then selected from the training set by finding the data points with nearest chlorophyll a concentration to values spaced evenly through the range of the training data set. The resulting set had a similar range to the training, validation and testing sets but the mean value was much higher and the standard deviation was increased. This new set will be referred to as the training set from this point forwards.

5.3 Feature selection

As was demonstrated in the primary study, the selection of the input bands was the most important factor in improving the chlorophyll a prediction. Particular bands gave better results than others, although combining all four inputs gave the best results. It was therefore important that the features for model development be carefully chosen.

Hyperspectral data provides many bands of fine spectral resolution. However, it is not possible computationally to use all these bands as inputs to the model (Dekker *et al.* 1991). Nor would we want to, since each band multiplies the curse of dimensionality (section 3.3) and including irrelevant and redundant bands can often result in a poorer

model (Myers 1986; White 1992). Therefore it may be considered that there is a lower and an upper limit to the number of features which would produce a good regression.

The practice of feature selection is well documented and, over recent years, several new techniques of selecting features have been developed. Much of the work into feature selection for remotely sensed images has concentrated on choosing features for classification (e.g. Richards 1995; Mather 1996; Piramuthu 1996). Feature selection methods that are more generally applicable to regression problems are available however. There are two parts to feature selection algorithms, the first is the method of assembling sets of features for testing and the second is the metric that is used to test the sets of features (Bishop 1995).

Selected bands may be 'narrow' or 'broad'. Narrow band selection picks individual bands from the original image whereas broad band selection combines original bands. Narrow bands are sensitive to narrow features in the spectrum such as absorption features. However, these bands are subject to noise and so, often, broad bands may be favoured.

Transforms of the data such as principal component analysis (section 2.7) (Otsu 1984; Benediktsson and Sveinsson 1997; Malki and Moghaddomjoo 1991)) and similar techniques (Singh and Harrison 1985; Roger 1996; Karhunen *et al.* 1997) are often used to reduce the dimensionality of the data. These methods order the features according to the variability in the input data that they explain. Such transforms are optimal in the sense that, in selecting the lowest order eigenvectors, the mean squared error between the original data and the transformed and reduced data is minimised. However, on their own these new features do not provide much information about the spectral dependence of the model (Fukunaga and Koontz 1970).

Research such as that by Townshend (1984) have selected those original wavebands that contribute most to the lowest order eigenvectors or have selected broadband regions corresponding to the lowest order eigenvectors (Price 1990). Conversely Csillag *et al.* (1993) have removed those wavebands which contribute most to the highest

order eigenvectors. The transform resulting from principal component analysis is very dependent on the data set used (Eklundh and Singh 1993) and the inclusion of features such as land, clouds or even sensor effects can result in axes orientated very differently to the direction expected.

Because transforms that are based on variance are not always relevant to high dimensional data (page 46), Green *et al.* (1988) and Lee *et al.* (1990) developed transforms that are based on the signal to noise ratios of each band. Since noise is considered to have little or no spatial autocorrelation, other feature selection methods are based on the spatial autocorrelation of image bands (Warner and Shank 1997b; Warner *et al.* 1999). Although popular, unsupervised feature selection methods do not provide a measure of the performance of the subsequent model because they do not account for the relationship between the spectral data and the variables that are to be predicted from them.

In the circumstances of studies such as this, where model outputs are available, features should be selected according to their suitability to predicting the output. Two new metrics are now available as a basis on which to select subsets of features - correlation of the model outputs with the sets of features and the ability of the features to predict a set of outputs from a new set of inputs (respectively, these metrics are equivalent to measures of class separability and classification accuracy commonly used for feature selection in image classification studies).

As has been demonstrated by Elashoff *et al.* (1967) and Cover (1974) the best subset of features is not necessarily made up of the best individual features. Therefore the feature selection methods should select features which 'work together' well. There are $2^d - 1$ possible subsets of features in a data set of d features. It clear is not feasible to test all these possible subsets with hyperspectral data. The number of possible subsets is reduced to $\frac{d!}{(d-\tilde{d})!\tilde{d}!}$ if the desired number of features in the subset, $\tilde{d}$ is known (Bishop 1995) but this may still be too many subsets to compare.

If the number of desired features is not known, it is necessary to find a sequential

method of selecting feature subsets. Two main approaches to this are *forward selection* and *backward elimination.* Forward selection begins by choosing the best single feature (according to the chosen metric) and then determining which feature works best with the first feature. Features are added sequentially until some stopping criterion is met, for example at the iteration when the improvement in the metric is not statistically significant. Backward elimination begins with the entire feature set and removes features one by one until some stopping criterion is met.

This research looked into the many possible features which may be extracted from a spectral image data set. A remote sensing image may be considered as a three-dimensional block of information with measurements in one spectral and two spatial directions. Depending on from which orientation this image is viewed, a number of different types of features may be extracted. These include broad bands, band ratios and differences, spectral derivatives and transforms of the data onto new axes. Spatial features have also been considered as it has been shown that the presence of phytoplankton results in a spatial structure which is different from that of tracer features such as sediment and temperature (see section 8.2).

In line with the aims of this research the choice of input was kept simple. The original spectral bands in the images were investigated for input which made the resulting models more generally applicable (Paola and Schowengerdt 1995b). These bands were narrow and better for the retrieval of information about fine features in the spectrum (Malthus *et al.* 1996). The FLH feature was shown to have a linear relationship with chlorophyll *a*, however this feature is more suited to high concentrations of chlorophyll *a* (Sathyendranath *et al.* 1989) when the water is not too turbid (Fischer and Kronfeld 1990). This feature was not used for the rest of the research although its value for predicting chlorophyll *a* was recognised.

Since the models were to be developed using neural networks as well as a more conventional technique, many methods of feature selection were not so relevant. For example, if the selection procedure were to be based on a measure of a feature's linear correlation to chlorophyll *a* then the purpose of using a non-linear neural network would

be negated. Several methods of feature selection have been developed for use with neural networks but they are often highly computer intensive. Linear feature selection methods can be simpler to use.

The three methods chosen for this research used information about the output from the model to select features. The first method (section 5.3.1) used the wealth of knowledge about chlorophyll *a* in water and about the environmental conditions of Case 2 locations to select spectral features. This is the method commonly used in the design of new sensors. The second (section 5.3.2) and third (section 5.3.3) methods were a linear and a non-linear method which tested the accuracy of the models' predictions to determine the features to remove. Stepwise methods are useful for highlighting sets of features that work well together that may not have been identified by a study of the spectral characteristics of chlorophyll *a* alone. Such methods are not commonly used in studies of ocean colour. Also, these methods are more relevant to feature selection from hyperspectral data.

5.3.1 Hand-picking features based on other chlorophyll *a* studies (HPFS)

Much work has already been undertaken into the most useful wavelengths for predicting chlorophyll *a*. Section 1.3 outlined much of this work. Before undertaking an intensive feature selection process, it was worth utilising these findings to construct a set of features for inputs to the models. The research was broadly separated into the following three areas:

1. Research in specific features for chlorophyll *a* prediction. Many of these features were discussed in section 1.3 (summarised in table 5.2).

2. Research into those features which are affected by other environmental factors such as the atmosphere or CDOM. These features may be used to correct for the effect of this factor or should be avoided (summarised in table 5.3).

3. Research into reconstructing the measured spectrum from as few measured wavebands as possible. Some workers have determined which wavelengths are the most useful for reconstructing a spectrum detected over a water body containing chlorophyll *a*. These features are the ones that are typical of the responses of the spectrum to changes in the constituents in the water (summarised in table 5.4).

This method of feature selection shall be known as hand-picking feature selection or HPFS.

Table 5.2: Features used for detecting chlorophyll in water

Citation	Type of feature	Wavelengths (nm)	Situation	Comment
Gordon *et al.* 1980	ratio	443 / 550	Gulf of Mexico	CZCS algorithm for low chlorophyll *a* concentration
Gordon *et al.* 1980	ratio	520 / 550	Gulf of Mexico	CZCS algorithm for high chlorophyll *a* concentration
Tassan 1981	difference	440 - 550	model	Difference algorithm performed better than ratio of same bands and overcame some problems of sun glitter, foam and atmospheric effects
Giannini 1981	difference	585 - 662	coastal waters	Preliminary study, possibly low sediment
Mitchelson *et al.* 1986	ratio	440 / 550	Irish Sea Case 2 waters	All algorithms empirically derived for Case 2 waters were statistically similar
Mittenzwey *et al.* 1992	ratio	705 / 670	inland waters	Least sensitive to other substances in the water. Based on a comparison between laboratory and *in situ* measurements
Goodin *et al.* 1993	raw band	720	experimental tanks	As sediment varies reflectance at this wavelength remains constant
Talcott 1995	ratio	485 / 570	coastal waters	Landsat TM data
Aiken *et al.* 1995b	ratio	490 / 555	ocean *in situ* spectral	Ratio responds best to range of chlorophyll *a* levels (where

continued on next page

continued from previous page				
Citation	Type of feature	Wavelengths (nm)	Situation	Comment
			measurements and modelled spectra	carotenoid pigments co-vary with chlorophyll *a*
Pérez-Ruzafa *et al.* 1996	ratio	443 / 560	oligotrophic lake water	Use natural log of attenuation coefficients
Rundquist *et al.* 1996	ratio	maximum NIR/ minimum red	experimental tanks, no sediment	When chlorophyll *a* concentration is low
Rundquist *et al.* 1996	1^{st} derivative	690.7	experimental tanks, no sediment	When chlorophyll *a* concentration is relatively high
Hoogenboom *et al.* 1998	ratio	~530/ 600	model of North Sea water	Sensitivity of this combination not changed by addition of tripton
Fraser 1998a	1^{st} derivative	429 and 695	inland waters	Peak in derivative at these wavelengths corresponds to steep reflectance slopes associated with chlorophyll *a*

Table 5.3: Features used for determining the influence of other factors

Citation	Substance	Wavelengths (nm)	Situation	Use to correct for or avoid?
Abbott *et al.* 1994 Wrigley *et al.* 1992 Carder *et al.* 1991	CDOM	412	oceanic and coastal waters	correct for
Morel and Gordon 1980	CDOM	400	Case 2 waters	correct for
Morel and Gordon 1980	atmospheric aerosol	880 + at least 2 other bands in the NIR	Case 2 waters	correct for
Morel and Gordon 1980	turbidity	610	Case 2 waters	correct for
Aiken *et al.* 1995b	atmosphere	412	ocean *in situ* spectral measurements and modelled spectra	avoid
Dekker *et al.* 1992a	noise	< 500 nm	inland waters	avoid
Sathyendranath *et al.* 1989	CDOM	400-430	model with sediment and CDOM	correct for
Singh *et al.* 1997	atmospheric aerosol	760	inland waters	correct for
Gordon *et al.* 1980	atmosphere	670	Gulf of Mexico	correct for
Mittenzwey *et al.* 1992	bottom reflectance	< 550	inland waters	avoid
continued on next page				

continued from previous page				
Citation	Substance	Wavelengths (nm)	Situation	Use to correct for or avoid?
	and CDOM			
Pérez-Ruzafa *et al.* 1996	bottom reflectance	443 / 560 ratio	shallow coastal waters	correct for
Quibell 1991	sediment	~665 minus ~710	experimental tanks	correct for
WrigleySKFCM92 citing NASA 1982	sediments	620	coastal waters	correct for

Table 5.4: Bands used for reconstructing spectrum

Citation	Method used	Situation	Wavebands (nm)
Wernand *et al.* 1997	Multiple regression analysis	coastal water	412, 492, 556, 620 and 672
Sathyendranath *et al.* 1989	Eigenvector analysis	model of Case 2 water	400, 445, 520, 565 and 640
Dekker *et al.* 1992a Dekker *et al.* 1992b	Knowledge-based method	inland water	510-530, 555-575, 590-610, 620-640, 645-655, 660-670, 670-685, 695-715 and 770-800

5.3.2 Multiple linear regression feature selection (MLFS)

See table 5.5 on page 115 for a glossary of the notation used in this section.

The linear multiple regression model was used to select features for regression. This was performed using a backward elimination technique. At each iteration, this method removes one feature and compares the error of the resulting prediction with that of the original set to determine if there is any decrease in accuracy. This is achieved as follows:

1. Using equation 3.5, the regression coefficients (β) were determined from the training set for the full set of I_+ inputs
2. The β were then used to make predictions of chlorophyll a from the validation set with the full set of inputs
3. The mean squared error $\mathcal{E}_+$ was determined for these predictions with the full set of inputs

Table 5.5: Notation used in description of multiple regression feature selection method

Symbol	Description
I_+	Number of inputs in the full feature set
I_-	Number of inputs in the reduced feature set
$\mathcal{E}_+$	Minimum error achieved for full configuration
$\mathcal{E}_-$	Minimum error achieved for reduced configuration
$Ptrain$	Number of patterns in the training data set
$Pvalid$	Number of patterns in the validation data set
P	$Ptrain + Pvalid$
df_+	Degrees of freedom for the full configuration
df_-	Degrees of freedom for the reduced configuration
x_i	Measurements in band i
$x_{\mathcal{E}_-}$	Band which, when removed from the I_+ inputs results in the lowest error for all the I_+ sets of I_- inputs

4. Stages 1 to 3 were then repeated I_+ times, each time with a reduced set of I_- inputs where $I_- = I_+ - 1$, produced by removing band x_i

5. The lowest error achieved at stage 4, $\mathcal{E}_-$ by removing band $x_{\mathcal{E}_-}$ was then compared to $\mathcal{E}_+$ to determine if $\mathcal{E}_- - \mathcal{E}_+ \leq 0$. If this was the case, $x_{\mathcal{E}_-}$ was removed from the data, I_+ was set to I_-, i set to $i - 1$ and $\mathcal{E}_+$ set to the value of $\mathcal{E}_-$

6. Stages 4 and 5 were repeated until the removal of any more bands resulted in $\mathcal{E}_- - \mathcal{E}_+ > 0$

The criterion used to determine whether $\mathcal{E}_- - \mathcal{E}_+ \leq 0$ was a statistical one. Clearly, with noise in the system, slight variations in the error may not have been statistically significant. The statistical function allowed natural variation in the error due to noise to be distinguished from that caused by real alterations in the model. This was achieved in this function by assuming that the error was normally distributed and hence its natural

variation could be determined. This assumption was reasonable, as was demonstrated in the primary study.

This was tested by comparing the test statistic L where

$$L = \frac{\frac{\mathcal{E}_- - \mathcal{E}_+}{df_- - df_+}}{\frac{\mathcal{E}_+}{df_+}} \tag{5.1}$$

to the corresponding values in the f-distribution for the numerator degrees of freedom $df_- - df_+$ and denominator degrees of freedom df_+ for the significance level of 95 %. The hypothesis being tested was

$$H_0 : R\beta \neq 0 \tag{5.2}$$

and the alternative hypothesis was

$$H_A : R\beta = 0 \tag{5.3}$$

where β was the vector of coefficients of the model and R was a vector of zeros and a one at the position corresponding to the coefficient being tested. At each iteration the coefficient corresponding to the band $x_{\mathcal{E}_-}$ was being tested and when $\mathcal{E}_- - \mathcal{E}_+ \leq 0$ (the prediction with one fewer inputs was at least as good as with all the inputs) the null hypothesis was rejected.

The degrees of freedom were the number of patterns used to train the algorithm minus the number of constrictions placed on the data. In this work two sets of data were used to train the networks, *Ptrain* and *Pvalid*, and so the number of patterns $P = Ptrain + Pvalid$. The number of constrictions placed on the data were the β regression coefficients (including the intercept), therefore $df_- = P - I_- + 1$, $df_+ = P - I_+ + 1$ and since at every iteration, $I_+ = I_- + 1$, $df_- - df_+ = 1$.

This method of feature selection shall be known as multiple regression feature selection or MLFS.

5.3.3 Neural network feature selection (NNFS)

See table 5.6 for a glossary of the notation used in this section.

Table 5.6: Notation used in description of neural network feature selection method

Symbol	Description
s	Number of networks used throughout the function
J_+	Number of hidden nodes in the full model configuration
J_-	Number of hidden nodes in the reduced model configuration
I_+	Number of inputs in the full feature set
I_-	Number of inputs in the reduced feature set
$\mathcal{E}_+$	Minimum error achieved for full configuration
$\mathcal{E}_-$	Minimum error achieved for reduced configuration
$Ptrain$	Number of patterns in the training data set
$Pvalid$	Number of patterns in the validation data set
P	$Ptrain + Pvalid$
u	Number of weights
df_+	Degrees of freedom for the full configuration
df_-	Degrees of freedom for the reduced configuration
x_i	Measurements in band i

The technique chosen here was originally described by Steppe *et al.* (1996) and was an extension of the method described in section 5.3.2. This method was chosen because it was specifically designed to select features for neural network models. Moreover, this technique, which shall be referred to as the neural network feature selection method also determined the number of weights required for the selected input features. The method was a backward elimination technique which begins with all the input features and carefully selects features for removal as with the multiple regression feature selection. This was achieved using the following procedure:

1. Several (s) networks with all the I_+ inputs to be assessed and maximum number

of hidden nodes J_+ were initiated with randomised weights. This was to allow for the possibility that certain initialisations of weights would result in the network converging on a non-global minimum. It was assumed that s was enough networks to ensure that at least one would converge on the global minimum.

2. The s networks were trained until each was deemed to have converged (using, for example, the validation set error as in the primary study or a maximum number of iterations stopping criterion). The training sum squared error was then assessed for each of the s networks and the minimum was taken to be the overall training error $\mathcal{E}_+$.

3. The s networks were then re-initiated and trained with one fewer hidden node (J_- hidden nodes). The minimum error of these networks $\mathcal{E}_-$ was sought and this was compared to $\mathcal{E}_+$ to determine if $\mathcal{E}_- - \mathcal{E}_+ \leq 0$. If the current training error $\mathcal{E}_-$ was equal to or less than the previous training error, the new model configuration was accepted and $\mathcal{E}_+$ was set to $\mathcal{E}_-$ and $J_+ = J_-$. If $\mathcal{E}_- > \mathcal{E}_+$, the new model was rejected. The assumption was that with this configuration of inputs, the $J+$ hidden nodes were required to adequately specify the model.

4. The s networks were then re-initiated and trained but this time one input was removed to leave I_- inputs. This step was repeated I_+ times with each input being withheld in turn. This resulted in a $s \times I_+$ matrix of errors in which the lowest error, $\mathcal{E}_-$, corresponding to the removal of feature $x_{\mathcal{E}_-}$, was located. Again $\mathcal{E}_-$ was compared to $\mathcal{E}_+$ to determine if $\mathcal{E}_- - \mathcal{E}_+ \leq 0$. If this was true, the accepted network configuration was that with feature $x_{\mathcal{E}_-}$ rejected leaving $I_+ = I_-$. Otherwise all I_+ features were retained within the model.

5. Stages 3 and 4 were repeated until no change was made to I_+ or J_+. The resulting configuration was considered to be that with the optimal inputs and number of hidden nodes for solving the problem.

Again, $\mathcal{E}_- - \mathcal{E}_+ \leq 0$ was tested by comparing L (equation 5.1) to the f-value. This tested the hypothesis:

$$H_0 : Rw^* \neq 0 \tag{5.4}$$

and the alternative hypothesis was

$$H_A : Rw^* = 0 \tag{5.5}$$

where w^* was a vector of optimal weights derived for the model with I_+ inputs and J_+ hidden nodes. R is a vector of zeros and ones where the ones are in the locations that correspond to the weights in w^* which were connected to the hidden node or input node that was being tested. Therefore, the test determined whether the weights connected to the input or hidden node of interest were equal to zero.

In the case of the neural network, the restrictions on the residuals were the weights (including biases) as these were equivalent to the coefficients of regression in the multiple regression problem. Thus, if the number of weights, u,= $J(I + 1) + (J + 1)$ (for a one-output network) then $df_+ = P - u$. When a hidden node was removed $df_- = P - (u - (I_+ + 2))$ and when an input was removed $df_- = P - (u - J_+)$.

This method of feature selection will be known as neural network feature selection or NNFS.

5.3.4 Initialisation conditions

So that the degrees of freedom were a positive value it was necessary that the number of weights in the network were fewer than the number of patterns training the network. Therefore, this restricted the number of inputs and hidden nodes in the network. All 72 wavebands could not be applied to this algorithm. Instead, the bands were separated into nine subsets, labelled A to I, with the wavebands spaced evenly throughout the spectrum (table 5.7). It was recognised that this removed the opportunity to test whether including adjacent bands in the models enabled a better correction for the effects of environmental factors on the spectral values. To enable a direct comparison with the neural network feature selection sets and the models derived from these sets, the multiple regression feature selection technique was initiated with the same subsets, A to I. Because the number of model coefficients was fewer for the MLFS method, it was

also possible to initiate this feature selection technique with all 72 image bands. The neural network feature selection method was initiated with each of these nine subsets of features and 18 hidden nodes. For each iteration of the function, 10 networks were trained, that is $s = 10$, and each network was trained for 4000 iterations.

Table 5.7: CASI bands in each of the 9 subsets used for the neural network feature selection method

Subset	Band numbers
A	1, 10, 19, 28, 37, 46, 55, 64
B	2, 11, 20, 29, 38, 47, 56, 65
C	3, 12, 21, 30, 39, 48, 57, 66
D	4, 13, 22, 31, 40, 49, 58, 67
E	5, 14, 23, 32, 41, 50, 59, 68
F	6, 15, 24, 33, 42, 51, 60, 69
G	7, 16, 25, 34, 43, 52, 61, 70
H	8, 17, 26, 35, 44, 53, 62, 71
I	9, 18, 27, 36, 45, 54, 63, 72

5.4 Model development

The selected features were to be related to chlorophyll *a* using a model. In the primary study it was seen that even linear neural networks perform better than simple linear least-squares regression if more than one input is used. Multiple linear least-squares regression is equivalent to linear neural networks (section 3.2) and is a popular method for relating inputs and outputs (e.g. Aiken *et al.* 1995b) which can be performed simply. It is more difficult to determine the optimal structure for non-linear models and so all non-linear models were developed using the neural network. The following two sections describe the two chosen methods of model development - multiple linear

least-squares regression and non-linear neural networks. These models were used to relate all the selected sets of features (sections 5.3.1 to 5.3.3) to chlorophyll *a*.

5.4.1 Multiple linear least-squares regression

Using equation 3.5, the coefficients for each selected set of features were calculated using the training set of patterns. The spectral values from the test set were then applied to these coefficients to predict chlorophyll *a*.

5.4.2 Neural networks for developing models

The neural networks were trained (using the feed-forward/back-propagation method described in section 3.2) with the training data set. The neural network feature selection determined how many hidden nodes were required for the A to I sets of inputs. The number of hidden nodes for the hand-picked and multiple linear regression feature selection data sets was chosen with reference to those for the neural network feature sets.

The networks were trained for 5000 iterations. Other parameters such as the learning rate and the activation functions were the same as those used in the primary study. It was intended that the results were the minimum achievable using this method.

Summary of chapter 5

The focus in this section has been on the choice of features for the models. Three methods of selecting features that were used in this research have been described. The first method was one that has been commonly used and was based on prior knowledge of the problem by using the information that was contained in the literature about the best parts of the spectrum to use for predicting chlorophyll *a*. The second and third

methods were iterative and selected sets of features according to their performance with linear (MLFS) and non-linear (NNFS) regression functions. Because of the number of coefficients being fitted in the neural network, only a limited number of bands could be tested at any one time in the NNFS technique. The 72-band data set was therefore divided into 9 subsets for the NNFS and, for comparative purposes, the MLFS. This resulted in the selection of 20 different sets of features. Table 5.8 summarises the feature selection techniques used and the subsets of band that were applied to them.

Table 5.8: The feature selection methods applied to the band subsets

band subset	HPFS	MLFS	NNFS
all 72	✓	✓	
A		✓	✓
B		✓	✓
C		✓	✓
D		✓	✓
E		✓	✓
F		✓	✓
G		✓	✓
H		✓	✓
I		✓	✓

Using multiple linear regression and neural network regression, chlorophyll *a* prediction models were then developed. These took as inputs the sets of bands that had been selected by all the feature selection techniques with the selected sets of bands as inputs. This resulted in 40 models - 20 each of the multiple linear regression and the neural network regression.

The feature selection described in this chapter has not been previously applied to ocean colour data. These methods provided an automatic method of determining from which regions of the spectrum bands should be taken to build robust chlorophyll *a* prediction models.

Chapter 6

Results

6.1 Introduction

This chapter details the results of the feature selection and the development of chlorophyll *a* prediction models. The first section (6.2) discusses the features selected by the three methods of 'hand-picking', multiple regression feature selection and neural network feature selection and then compares the different methods. Section 6.3 then describes the predictions of chlorophyll *a* as made using these selected sets of bands and the multiple regression and neural network regression methods.

6.2 Feature selection

The following three sections describe the results of the three feature selection methods that were used. These results are then compared in section 6.2.4.

Table 6.1: These features were chosen with recommendations from literature about the prediction of chlorophyll *a*.

Wavelength (nm)	wave band	Why chosen
412	2	correction for CDOM
443	6	detection of chlorophyll *a*
560	23	detection of chlorophyll *a*
670	38	correction for sediment
710	44	correction for sediment
880	67	correction for atmosphere

6.2.1 Hand-picking features based on other chlorophyll *a* studies

Using the literature a set of six wavebands was chosen (table 6.1). Several studies recommended the waveband around 412 nm for correction for CDOM as well as reconstruction of the spectrum (Abbott *et al.* 1994; Carder *et al.* 1991; Wrigley *et al.* 1992). Although Aiken *et al.* (1995b) recommended avoiding this wavelength because of atmospheric effects, this band was selected. Bands near 443 and 560 nm have been the most commonly used for chlorophyll *a* detection and so these were included in this set. Quibell (1991) recommended using bands at 665 and 710 nm in combination to correct for sediment. Hence, bands at 670 and 710 nm were chosen. These bands could also be used in combination with 880 nm for correction for atmosphere such that a long wavelength red band was substituted for one of the 2 NIR bands suggested by Morel and Gordon (1980) in combination with 880 nm.

Table 6.2: Output of multiple regression on the subsets A-I.

Band subset	Bands selected	Central wavelengths (nm)	(Bands rejected)
A	1, 10, 19, 28, 37, 46, 55, 64	406, 469, 533, 596, 661, 725, 790, 855	
B	11, 20, 29, 38, 56, 65	476, 540, 604, 668, 797, 863	(2, 47)
C	12, 21, 30, 39, 57, 66	483, 547, 611, 675, 804, 870	(3, 48)
D	13, 22, 31, 49, 67	490, 554, 618, 747, 877	(4, 40, 58)
E	5, 14, 41, 50	434, 497, 689, 754	(23, 32, 59, 68)
F	6, 33, 42, 51	441, 632, 696, 761	(15, 24, 60, 69)
G	7, 34, 43, 52, 70	448, 639, 704, 768, 899	(16, 25, 61)
H	8, 17, 26, 35, 44, 53	455, 518, 582, 646, 711, 776	(62, 71)
I	9, 18, 45, 54, 72	462, 526, 718, 783, 914	(27, 36, 63)

6.2.2 Multiple regression feature selection (MLFS)

Table 6.2 summarises the results of the multiple regression feature selection when the nine groups of bands (as described in table 5.7) were presented to the algorithm. No bands were rejected from subset A, two bands were rejected from subsets B, C and H, three from D, G and I and four from E and F. Appendix C gives details of how the multiple regression feature selection arrived at the final selection of bands.

To give the selection more meaning, the selected bands have been tabulated according to the regions of the spectrum from which they are taken (table 6.3). As was expected, the selected subsets contained bands from all regions of the CASI spectrum. Almost all the red wavebands were selected and green wavelengths were most poorly represented.

Figure 6.1 summarises the bands selected by the multiple regression feature selection when all the CASI wavebands were presented to the MLFS algorithm. Thirty-four bands were selected from the full 72-band set. This set comprised mostly of longer NIR wavebands and the green wavebands and fewer blue and red wavebands.

Table 6.3: Comparison of combinations of bands selected by the multiple regression feature selection. The rows represent the groups of bands that were selected from and the columns represent the regions in the spectrum. The bracketed values are the centre wavelength of the selected band.

Sub-set	Spectral region (band central wavelength in nm)							
	blue	blue-green	green	red	red-NIR	NIR	NIR	NIR
A	•(406)	•(469)	•(533)	•(596)	•(661)	•(725)	•(790)	•(855)
B		•(476)	•(540)	•(604)	•(668)		•(797)	•(863)
C		•(483)	•(547)	•(611)	•(675)		•(804)	•(870)
D		•(490)	•(554)	•(618)		•(747)		•(877)
E	•(434)	•(497)			•(689)	•(754)		
F	•(441)			•(632)	•(696)	•(761)		
G	•(448)			•(639)	•(704)	•(768)		•(899)
H	•(455)	•(518)	•(582)	•(646)	•(711)	•(776)		
I	•(462)	•(526)			•(718)	•(783)		•(914)

6.2.3 Neural network feature selection (NNFS)

Table 6.4 summarises the bands selected and the number of hidden nodes required for each of the nine subsets of CASI features applied to the neural network feature selection. Appendix D gives the details of how the function chose these configurations.

While the NNFS function was being developed, it was applied to the A to I subsets several times. Table 6.4 shows the results of the final run. Between three and five bands were selected during each run of the neural network feature selection from the eight original bands in the subsets, requiring between 6 and 14 hidden nodes. The same combinations of bands tended to be selected, particularly for subsets A, F and I. The number of times that each band was selected for this run and the previous

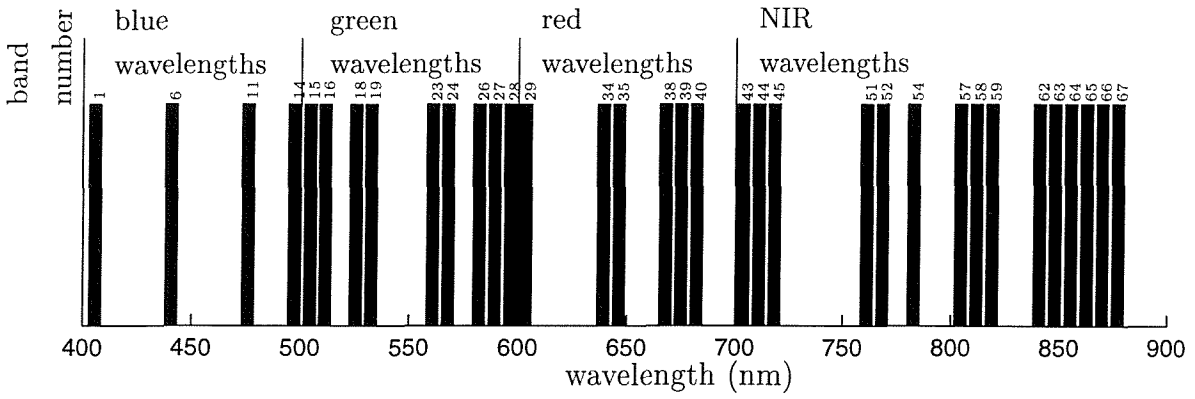


Figure 6.1: The bands selected by the multiple regression feature selection from the entire band set

Table 6.4: Summary of output of the neural network feature selection

Band subset	Bands selected	Central wavelengths (nm)	Number of hidden nodes	(Bands rejected)
A	1, 10, 19, 28, 55	406, 469, 533, 596, 790	14	(37, 46, 64)
B	2, 20, 29	413, 540, 604	10	(11, 38, 47, 56, 65)
C	3, 21, 30	420, 547, 611	10	(12, 39, 48, 57, 66)
D	4, 40, 49	427, 682, 747	13	(13, 22, 31, 58, 67)
E	5, 23, 41, 59	434, 561, 689, 819	6	(14, 32, 50, 68)
F	6, 33, 42, 60	441, 632, 696, 826	13	(15, 24, 51, 69)
G	7, 25, 43, 61	448, 575, 704, 833	7	(16, 34, 52, 70)
H	8, 35, 44, 53	455, 464, 711, 776	13	(17, 26, 62, 71)
I	9, 27, 54, 72	462, 589, 783, 914	8	(18, 36, 45, 63)

three is illustrated by the histogram in figure 6.2. The number of hidden nodes was a little more variable. This indicated that the final error for each training was influenced by initial weight conditions such that occasionally, particularly low final errors were achieved with only a few hidden nodes.

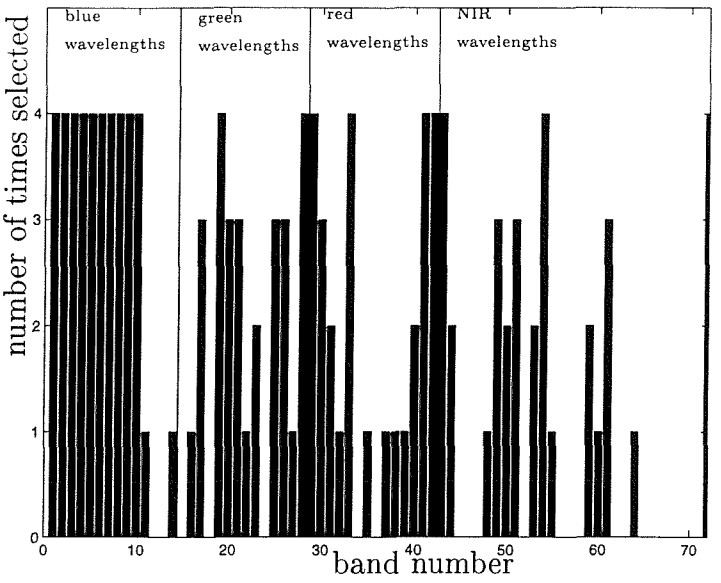


Figure 6.2: The consistency of the neural network feature selection in selecting CASI bands

Particular regions of the spectrum were consistently selected by the function during the A to I runs. The most notable was that the blue bands 1 to 10 (402 - 473 nm) were always selected. This was true for all runs of the function. Conversely, blue-green bands 11 to 16 (472 - 509 nm) were nearly always rejected. Band 19 (528 - 537 nm), bands 28 to 29 (592 -608 nm), band 33 (628 - 636 nm), bands 41 to 43 (685 - 708 nm) near the fluorescence feature and band 54 (779 - 787 nm) were all selected every time the function was run. Near infrared bands 62 to 71 (837 - 910 nm) were rejected consistently but band 72 (909 - 918 nm) was always selected.

Again, the bands that were selected are tabulated with respect to the regions of the CASI spectrum (table 6.5). In general some particular combinations of selected bands can be described. For subsets A - C the neural network feature selection derived a combination of a short wavelength blue band, a central green band and a short wavelength red band. Subsets E - H also had similar combinations to each other. The short wavelength blue band was combined with a band on the red-NIR transition and either a long wavelength green or a mid-wavelength red and one other NIR band.

Table 6.5: Comparison of combinations of bands selected by the neural network feature selection.

Sub-set	Spectral region (band central wavelength in nm)							
	blue	blue-green	green	red	red-NIR	NIR	NIR	NIR
A	•(406)	•(469)	•(533)	•(596)			•(790)	
B	•(413)		•(540)	•(604)				
C	•(420)		•(547)	•(611)				
D	•(427)				•(682)	•(747)		
E	•(434)		•(561)		•(689)		•(819)	
F	•(441)			•(632)	•(696)		•(826)	
G	•(448)		•(575)		•(704)		•(833)	
H	•(455)			•(646)	•(711)	•(776)		
I	•(462)		•(589)			•(783)		•(914)

Subset D was similar but no green or mid-red band was selected. Subset I also had a similar combination with the exception that a band on the red-NIR transition was replaced by a band in the middle of the NIR (band 72).

On average (mean and mode) about 10 hidden nodes were required in the final configuration of the network. Except for subsets E and G at least twice as many hidden nodes to inputs were required. With regard to the previous runs of the neural network feature selection, all except for subsets E and G arrived at a number of hidden nodes close to the average for the subset. For these subsets, the value for this final run was lower than any previously obtained.

6.2.4 Comparison of the results of the different feature selection methods

Six bands were chosen using the 'hand-picking' method. This was about the average number of bands selected from the nine subsets with the multiple linear regression feature selection. However, the neural network method, resulted in a same size or smaller set of features. When selecting from the entire data set, the MLFS selected 34 bands, many more than had been permitted by splitting the data into subsets. This indicated that a great deal of information about chlorophyll *a* was held in many of the image bands and that by separating the data into band sets A to I, the feature selections were constrained.

There were several similarities between the different sets of selected features. All but band 2 (413 nm)of the hand-picked features were also present in the MLFS sets selected from the whole data set. Four of the hand-picked bands were included in the feature sets selected from the nine subsets of the features using either the MLFS or NNFS. The MLFS and NNFS both selected bands 1, 6 19, 28, 29, 43 and 54 (406, 441, 533, 596, 604, 704 and 783 nm, respectively). The selection of band 1 was surprising as this is a very noisy band. Band 6 is in a region of the spectrum popularly chosen for predicting chlorophyll *a* at 440 nm. A recent study found that a band around 530 nm in combination with a band at 600 nm was particularly useful for detecting chlorophyll *a* in the North Sea (Hoogenboom *et al.* 1998) and this corresponded to bands 19, 28 and 29. Band 43 is close to the fluorescence peak and so may have been selected for this reason. However other bands that are closer to the peak were not always selected.

The MLFS did not select band 40 (682 nm) when working with the subset D. This is the band that is closest to the peak in fluorescence. However, the NNFS did select this band. In the primary study the fluorescence feature was found to have a generally linear relationship to chlorophyll *a* and so it was surprising that the linear feature selection did not select this band. It was possible that other bands in the D subset contained more valuable information.

Bands 41 to 43, (685-708 nm) which span the region of the fluorescence peak at higher chlorophyll *a* concentrations, were selected by both the MLFS and NNFS from the subsets E to G. However, only band 43 (704 nm) was selected from this region when all the bands were available indicating that much the same information was contained in all these bands. Similarly, band 72, an extremely noisy band with many zero values, was selected by both the MLFS and NNFS as part of subset I. However, it was not selected when the entire feature set was available. This is a particularly surprising outcome and may indicate that this band contained important information for the prediction of chlorophyll *a* that was not contained in other bands in subset I but that was available in another, much less noisy band in the whole band set.

On the whole the linear, MLFS selected more bands than the non-linear, NNFS. Bands on the red and NIR were favoured marginally over other regions of the spectrum by MLFS when selecting from the subsets A to I. The NNFS method always selected short- to mid-wavelength blue bands. When presented with the whole feature set, the MLFS selected bands throughout the spectrum although noisy blue and NIR bands were avoided in general. The number of bands that were selected from the full feature set indicates that many bands were required to characterise a linear relationship between the spectral data and chlorophyll *a*.

6.3 Model development using the selected feature sets

The results of the development of models for chlorophyll *a* prediction are presented in this section for multiple regression and neural network regression. The goodness of the prediction was evaluated using quantitative and qualitative measures. Firstly, statistical measures were used that were based on the comparison between predicted and measured chlorophyll *a* concentrations in the testing set. Secondly, the results were compared using visual measures which indicated whether errors in the predictions occurred for particular chlorophyll *a* concentrations or locations along the cruise.

6.3.1 Statistical assessment of models

Table 6.6 gives the results of the prediction of chlorophyll *a* concentration. These were compared to the measured values in the testing set using the root mean squared error of prediction (equation 4.3) and the correlation between measured and predicted values (equation 4.2).

6.3.1.1 Multiple linear regression

The hand-picked feature set performed well in comparison to the other feature sets with this model because the error of 2.77 $\mu g\, l^{-1}$ and correlation of 0.8 for this feature set were equal to the average for all the feature sets. In general, the multiple regression models performed marginally better with the MLFS sets A-I with an average error of 2.72 $\mu g\, l^{-1}$ and correlation of 0.81 in comparison to with the NNFS sets with an average error of 2.94 $\mu g\, l^{-1}$ and correlation of 0.77. The error and correlation values were also a little less variable for the MLFS sets than the NNFS sets. The lowest errors and highest correlations for the MLFS and NNFS sets A to I we achieved using set I, however the lowest error (1.75 $\mu g\, l^{-1}$) and highest correlation (0.93) overall was achieved with the 34-band set selected using MLFS applied to all 72 bands of the CASI imagery. The worst performing model was that using the NNFS set D, with an error of 3.64 $\mu g\, l^{-1}$ and a correlation of 0.62. There was little evidence that the number of features in each subset was in itself important for deriving a good model as some models with few features, such as the 4-band NNFS set I performed better than models using more input features, such as the 8-band MLFS set A.

6.3.1.2 Neural networks

The hand-picked set of features resulted in a neural network model error of 2.51 $\mu g\, l^{-1}$ and correlation of 0.87. This was a little worse than the average error and correlation for all the neural network models of, respectively, 2.23 $\mu g\, l^{-1}$ and 0.89. The neural network

models with the MLFS sets A to I with an error and correlation of 2.05 $\mu g\, l^{-1}$ and 0.91, performed better on average than those with the NNFS sets, which had an error and correlation of 2.44 $\mu g\, l^{-1}$ and 0.87. The model using the 34-inputs selected using MLFS from all the bands performed well, with an error and correlation of 1.56 $\mu g\, l^{-1}$ and 0.95. However, the lowest error and highest correlation (1.28 $\mu g\, l^{-1}$ and 0.96) were achieved by the neural network model using the MLFS set I. Again the performance of the models was not strongly related to the number of features used as inputs.

6.3.1.3 Comparison of models developed

The RMSE obtained using these models was compared to that which would result from always predicting the mean of the training set. In this case, this 'benchmark' error was 5.97 $\mu g\, l^{-1}$. For all the models the RMSE was below this mean value indicating that the models were not simply predicting the mean value for the training set. The correlations are all positive and statistically significant.

The 7th and 8th columns of figure 6.6 show the difference between the error and correlations of the different models using the same sets of input features. In all but one model (that using NNFS set C) the neural network models performed better than the multiple linear regression using the same set of inputs. This was indicated by the positive difference in RMSE and negative difference in correlation. However, the difference in the error was only between 0.10 $\mu g\, l^{-1}$ (NNFS set C) and 1.0 $\mu g\, l^{-1}$ (NNFS set I) and the difference in correlation ≈ 0.1.

Comparing the averages for the feature selections it can be seen that both types of model had lower errors and higher correlations, on average, when using the features selected using MLFS. The mean RMSE for the multiple linear regression models with the MLFS and NNFS sets A to I are, respectively, 2.72 and 2.94 $\mu g\, l^{-1}$. For the neural network regression models these errors are 2.05 and 2.44 $\mu g\, l^{-1}$. The improvement was therefore particularly marked with the neural network models. Both types of model performed moderately well with the hand-picked set of features and extremely well

with the 34-band set of features. It was possible that the models using the NNFS-derived subsets and the neural network models using 34-inputs were not trained for long enough to arrive at an optimal weight configuration. The NNFS subsets were selected from the one neural network in 10 which had the most optimal initial weight conditions (section 5.3.4), therefore it was likely that, in the final training, the initial weights of the one network trained for each set of inputs also resulted in non-optimal weights after only 5000 iterations. These subsets of features were also considered to have a complex relationship with chlorophyll *a* concentration and so would need longer to derive this during training (page 100).

6.3.2 Visual assessment of models

In the following pages the visual assessment of the predictions is described. This was aimed at determining the sources of good and poor chlorophyll *a* prediction.

For each data point the predicted chlorophyll *a* concentration was plotted against the measured chlorophyll *a* concentration and the grouping of points around the one-to-one line was assessed. This allowed the identification of particular ranges in chlorophyll *a* concentration that could be predicted well or predicted poorly. The diagonal line indicates where the predicted value was equal to the measured value. Points above this indicated that the model predicted a chlorophyll *a* concentration that was higher than the measured value. Points below this line indicated that the model predicted a chlorophyll *a* concentration that was lower than the measured value. The vertical distance of the point from the line represented the magnitude of this difference.

The difference between the predicted chlorophyll *a* concentration and the measured chlorophyll *a* concentration was also plotted against the easting co-ordinate of the data point and regions of over- or under-prediction in the data were located. This enabled a direct comparison of prediction error with the known sources of error as derived in section 2.7. If there was some locational reason why predictions at some data points were good or poor this could be identified as 'structure' in these plots. The

horizontal line indicated no difference between measured and predicted values. Points above this line indicated that the model predicted a chlorophyll *a* concentration that was higher than the measured value. Points below this line indicated that the model predicted a chlorophyll *a* concentration that was lower than the measured value. The vertical distance of the point from the line represented the magnitude of this difference.

6.3.2.1 Multiple linear regression

In figure 6.3 the data points generally formed a curve about the one-to-one line with predicted values lower than measured values at the lowest and highest measured chlorophyll *a* concentrations and higher predicted than measured values for the medium chlorophyll *a* concentrations. However, this was not evident with the 34-input model where the cluster was evenly distributed about the one-to-one line throughout the range of measured chlorophyll *a* concentration. The alignment to the one-to-one line was particularly poor for models with NNFS sets B, C, D and I. Here lower chlorophyll *a* concentrations are reasonably-well predicted but overall the models tended to predict a concentration of around 11-12 $\mu g\, l^{-1}$. This was close to the mean for the training set and so the good RMSE found was a little misleading.

The difference between measured values and those predicted using the multiple linear regression model showed a similar variation with geographical location for all models (figure 6.4). The models with MLFS sets A, B, F, G and H displayed a more even distribution of error over space and the model using the MLFS from all bands displayed the least structure over space. Other models however, particularly those with NNFS sets B, C, D, and E, displayed much more structure over space. The main feature in this variation of error over space, which is evident in all the plots to some degree, is a large under-prediction, sometimes of more than 10 $\mu g\, l^{-1}$, at about 585000 m east. This was located approximately in the middle of image 1877 and appeared to correspond to a large trough in the second principal component (figure 2.22). A peak in values at 630000 m east for many of the predictions corresponded to a peak in the first principal component and a large trough in the second principal component. This

peak, sometimes followed by a trough (especially in the case of NNFS set D), was also evident in many of the plots.

Another feature that was also common in most of these plots was an narrow region of over-prediction of up to 10 $\mu g\, l^{-1}$ at about 606000 m east, or conversely, under-prediction in the models with NNFS sets D and I. This corresponded to the region in which the principal components had a high variance. This geographical location and also the filament depicted by the colour green where the blue-green ratio was plotted against chlorophyll *a* in diagram figure 4.3 and was probably due to the brightening of spectral values due to increased path radiance at the edge of the image.

Also, all the plots displayed another region of over-prediction at 625000-635000 m east. This was attributed to the brightening of pixel values in the corresponding location of image 1875 as can be seen in figure 2.1. A region in which the difference between the measured and predicted chlorophyll *a* concentration varied a great deal for many models was the very far east of the data. Here, the difference between the times of the over-flights and the *in situ* sampling was greater than $2\frac{3}{4}$ hours and so the error may not have been in the prediction but rather the co-location of the *in situ* measurement with particular spectral values.

6.3.2.2 Neural networks

In figure 6.5 the data points from all the models tended to cluster tightly about the one-to-one line with little evidence of bias. One feature of note was the over-prediction of values where measured concentrations were about 5-7 $\mu g\, l^{-1}$. This may have resulted from the method chosen to select the training data set which picked data points evenly along the range of possible values. In so doing, the number of training samples was reduced and it was possible that not enough examples from the chlorophyll *a* concentrations in the range 5-7 $\mu g\, l^{-1}$ were provided during training for the network to learn the range of spectral values that corresponded to these chlorophyll *a* concentrations.

Only a few structures were evident in the plot of the difference between the measured

and predicted chlorophyll *a* concentrations against location (figure 6.6) and these are most evident for the models using the NNFS sets A to H as inputs. A large over-prediction was visible at about 630000 m east. This was greatest for the models which displayed a filament at 5-7 $\mu g\, l^{-1}$ in figure 6.5. These data points for which the over-prediction occurred were located in the region of brightening in image 1875 as noted in section 6.3.2.1. There were also many data points at an easting of 605000 m east for which the predictions had been extremely high or low compared to the measured chlorophyll *a* concentration.

6.3.2.3 Comparison of models developed

In the visual assessment of the prediction made by the multiple linear regression and the neural network models the multiple linear regression models displayed more bias in the predictions both over the range of measured chlorophyll *a* concentrations and due to data from particular regions of the study site. The latter effects were often attributable to environmental factors such as sediment in the water, atmospheric effects or to effects caused by changes in viewing angle. The neural network predictions fitted most closely the measured values over the whole range of chlorophyll *a* concentrations and also seemed less affected by the environmental factors. In particular, the neural network models appeared to be unaffected by the factor that caused large errors in the multiple linear regression model at about 585000 m east. Since a trough was found in the 2nd principal component at this location, it was thought that CDOM may have been high in this region.

Where the neural network models were affected by local factors these were often different to the ones that affected the multiple regression models. The neural networks were, however, affected by the choice of training data. The magnitude of the error evident in these plots was variable with the set of inputs used for the models. Generally, the models with the MLFS sets showed the least structure with measured chlorophyll *a* concentration and over the course of the cruise.

Summary of chapter 6

In this chapter the features that were selected by three methods of feature selection have been presented. The multiple regression feature selection tended to select more features given the same number of features in the initial set. Given the whole set of 72 bands, this method selected 34 bands that provided adequate information for the regression. Features from throughout the spectrum were selected together, including noisy bands in the blue and NIR. However, the neural network feature selection tended to favour the short wavelength blue bands. Models for predicting chlorophyll *a* concentration were developed and an assessment of their predictions has also been presented here. The multiple regression models did not perform as well as the neural network models. Also, both models performed better with the MLFS sets than the NNFS set. The best performances were achieved using the 34-band feature set with the multiple linear regression models and the MLFS set I with the neural network models. Most of the error in prediction could be attributed to environmental factors that were evident in the data.

The results of the feature selection provide a new understanding of the nature of the relationship between chlorophyll *a* and spectral information given the region of the spectrum. It has been shown that the spectral features that have traditionally been used for ocean colour research are not the most appropriate when non-linear models are available. The technique of removing noisy data has also been shown to be unnecessary when regression techniques that are robust to noise are to be used.

The method of analysing the error in prediction both across the range of prediction and over space has not been previously applied in similar studies and it has provided some information as to why the prediction is poor in some cases. It has also shown that the predictions by the linear regression and the neural network regression are affected by different factors in the data.

Table 6.6: The root mean squared error (RMSE) of the chlorophyll *a* and the correlation between the predicted and measured chlorophyll *a* are presented here for each subset of CASI bands and for both multiple linear regression and neural network regression. The difference between these values for the regression methods is calculated by subtracting the value for the neural network from that for the multiple regression.

FS method	subset	Multiple linear regression model RMSE	Multiple linear regression model corr'n†	Neural network regression model RMSE	Neural network regression model corr'n†	RMSE difference	corr'n† difference
HP	all 72	2.77	0.80	2.51	0.87	0.27	-0.07
ML	A	2.74	0.81	2.20	0.90	0.54	-0.09
ML	B	2.72	0.81	2.11	0.91	0.61	-0.10
ML	C	2.74	0.81	1.83	0.93	0.90	-0.12
ML	D	2.83	0.79	1.89	0.92	0.94	-0.13
ML	E	2.99	0.76	2.47	0.87	0.51	-0.11
ML	F	2.77	0.80	1.95	0.92	0.82	-0.11
ML	G	2.79	0.80	2.40	0.88	0.39	-0.08
ML	H	2.60	0.83	2.30	0.89	0.31	-0.06
ML	I	2.26	0.87	1.28	0.96	0.98	-0.09
ML	all 72	1.75	0.93	1.56	0.95	0.19	-0.02
NN	A	3.01	0.76	2.21	0.90	0.79	-0.14
NN	B	3.16	0.73	2.74	0.84	0.42	-0.11
NN	C	3.06	0.75	3.16	0.79	-0.10	-0.04
NN	D	3.64	0.62	3.29	0.81	0.35	-0.19
NN	E	2.84	0.79	2.38	0.88	0.46	-0.09
NN	F	2.65	0.82	2.19	0.89	0.47	-0.07
NN	G	2.82	0.79	2.11	0.90	0.71	-0.11
NN	H	2.68	0.82	2.29	0.89	0.39	-0.07
NN	I	2.62	0.82	1.63	0.94	1.00	-0.12
whole column $\mu^{\ddagger}$		2.77	0.80	2.23	0.89	0.55	-0.10
MLFS sets A-I $\mu^{\ddagger}$		2.72	0.81	2.05	0.91	0.67	-0.10
NNFS sets A-I $\mu^{\ddagger}$		2.94	0.77	2.44	0.87	0.50	-0.10

corr'n† = correlation, $\mu^{\ddagger}$ = mean

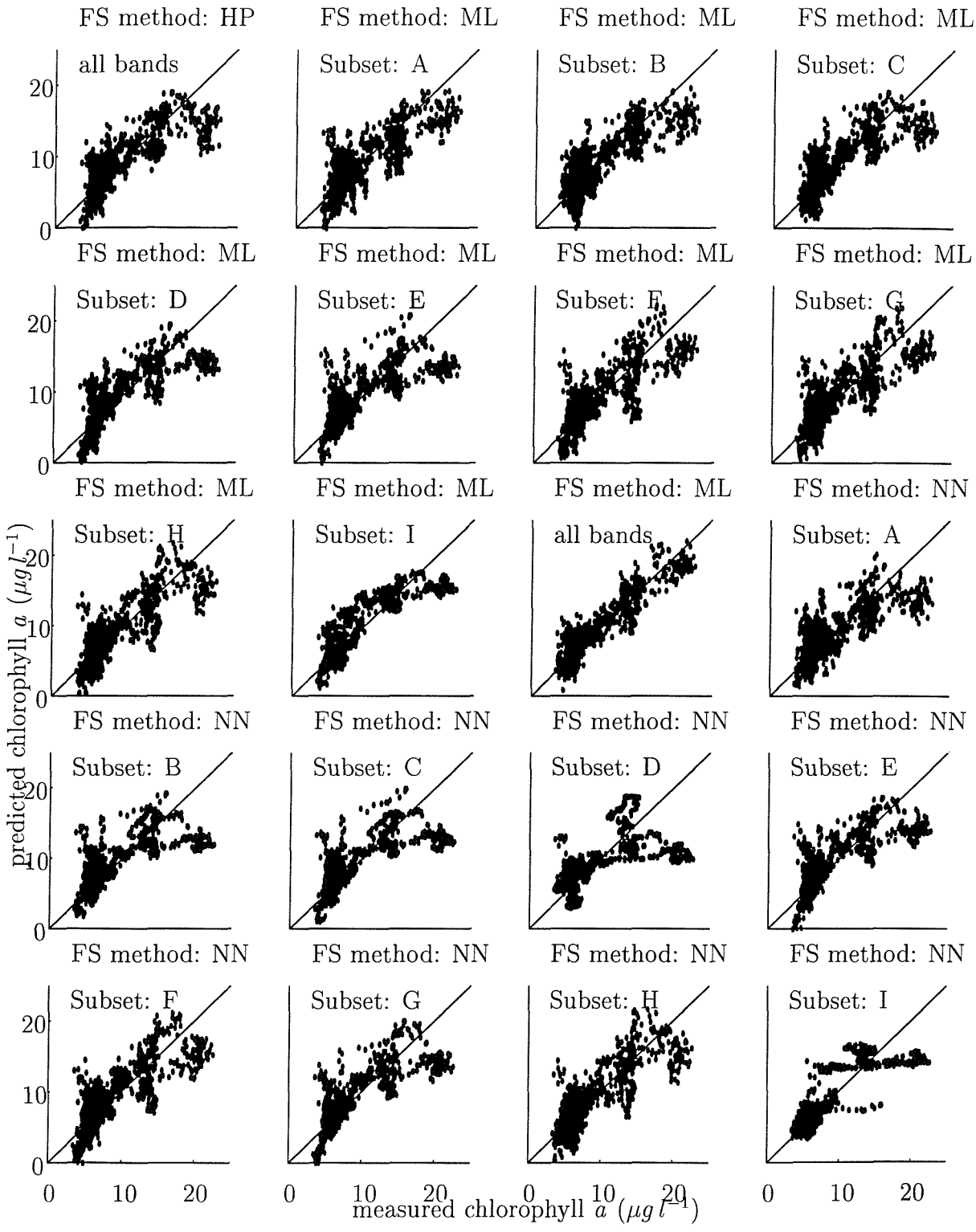


Figure 6.3: Plots of predicted values using multiple regression against the measured values for all the feature sets. Values can be assessed by their deviation from the one-to-one line. FS method = feature selection method, HP = hand picked, ML = multiple linear regression, NN = iterative neural network regression

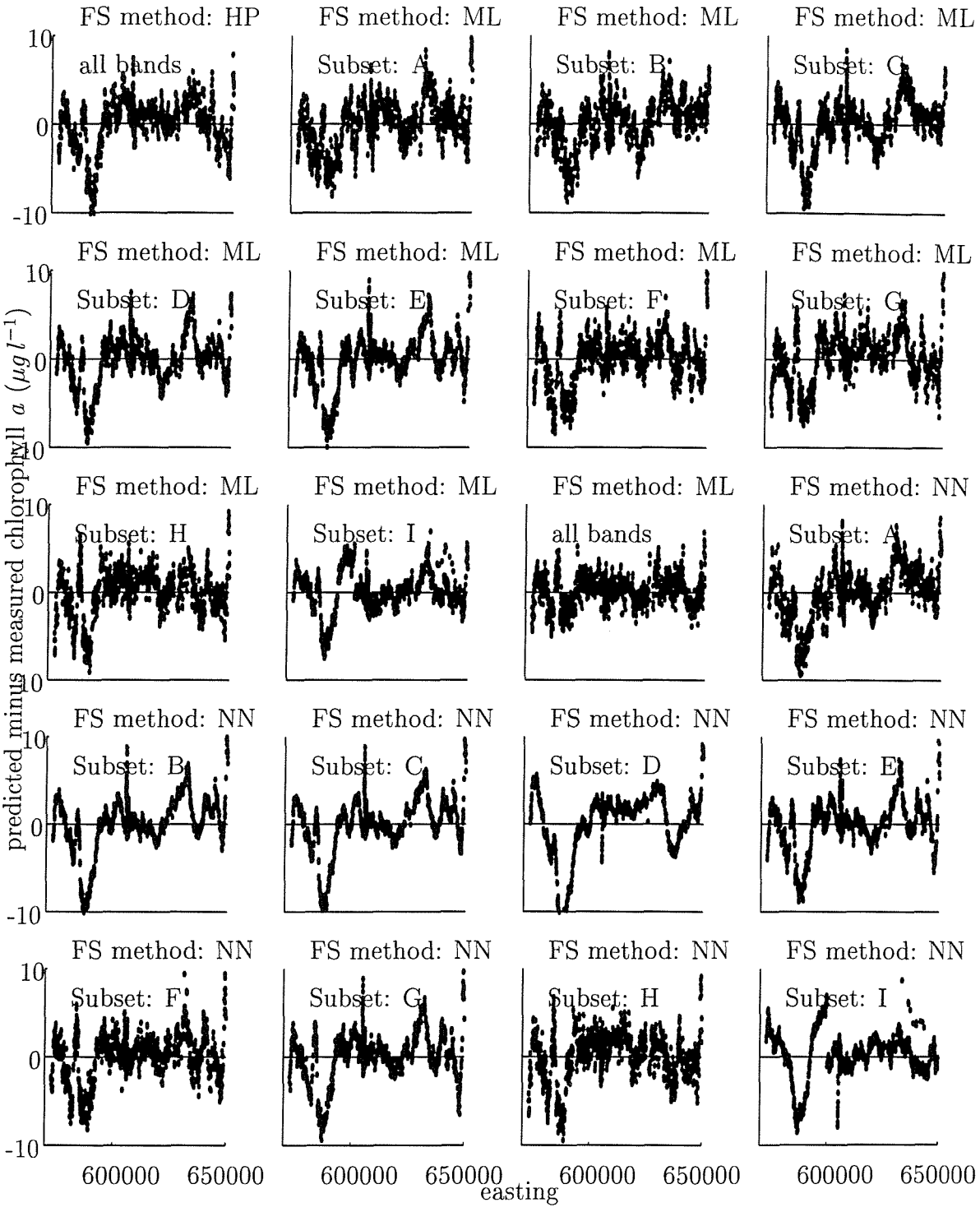


Figure 6.4: Plots of difference between predicted and measured values for multiple regression against the location along the cruise. Values can be assessed by their deviation from the line of no difference between predicted and measured values. FS method = feature selection method, HP = hand picked, ML = multiple linear regression, NN = iterative neural network regression

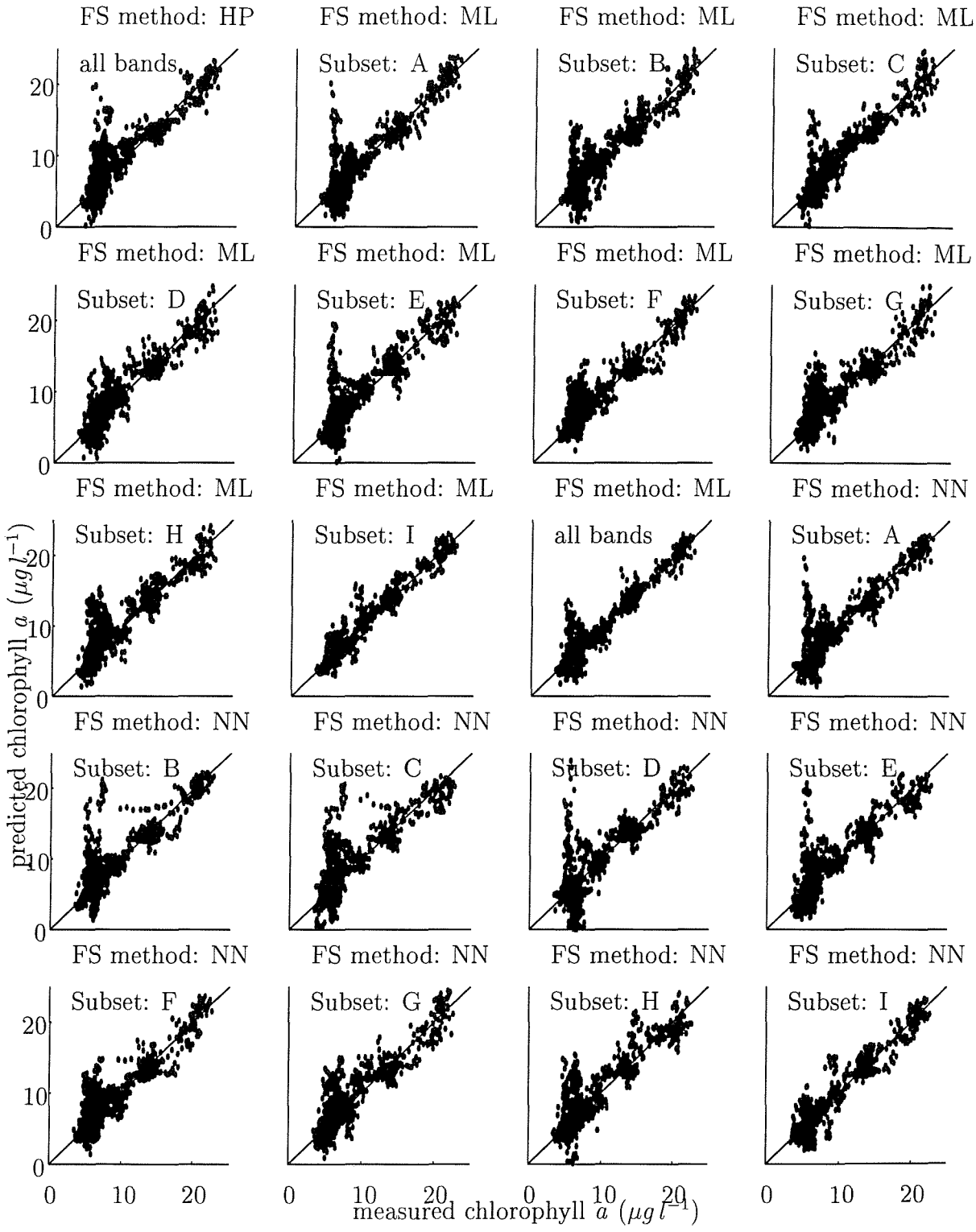


Figure 6.5: Plots of predicted values using neural network regression against the measured values for all the feature sets. Values can be assessed by their deviation from the one-to-one line. FS method = feature selection method, HP = hand picked, ML = multiple linear regression, NN = iterative neural network regression

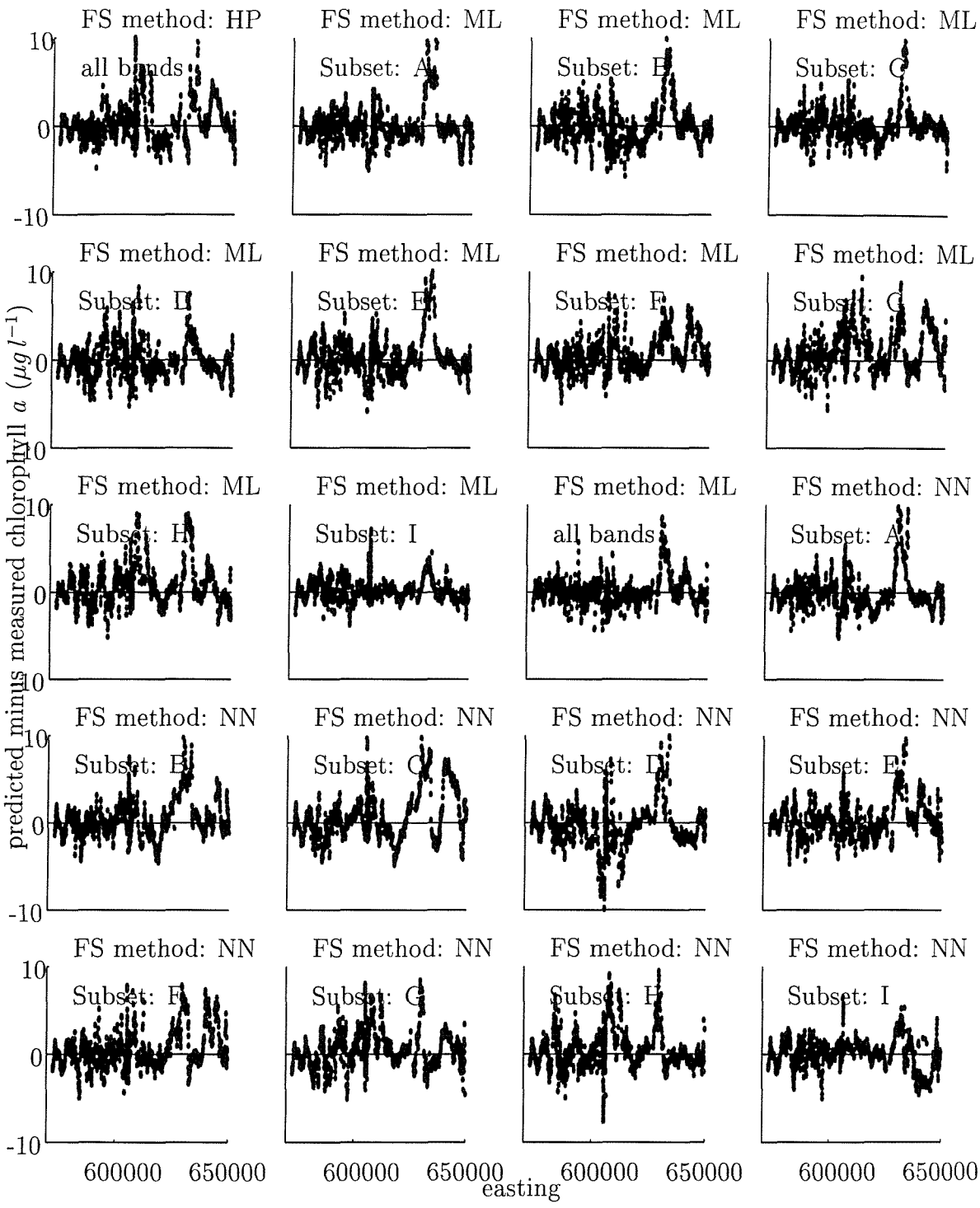


Figure 6.6: Plots of difference between predicted and measured values for multiple regression against the location along the cruise. Values can be assessed by their deviation from the line of no difference between predicted and measured values. FS method = feature selection method, HP = hand picked, ML = multiple linear regression, NN = iterative neural network regression

Chapter 7

Discussion

This research has investigated new techniques of predicting chlorophyll *a* from remotely sensed data where the spectral signal is significantly affected by environmental and sensor conditions. A typical set of remotely sensed and *in situ* data were obtained for this purpose and the approach taken was to use a very analytical method throughout the research. At each stage the findings were assessed and used to determine the direction that the next stage should follow. The following discussion draws all the findings of each stage of the research together to address the objectives set out at the beginning of the research.

The premise of this research was that the relationship between chlorophyll *a* and the detected spectral signal was non-linear and complex when affected by other environmental factors. A method by which non-linear models of varying complexity could be built easily was required. As demonstrated in chapter 3, neural networks were chosen as a method of regression because they are directly comparable to the more conventional method of multiple linear least-squares regression. The neural network technique is quite simple. However, it was also found to be time-consuming when more complex networks were being trained.

The choice of the neural network technique was validated in chapter 4 where it was

shown that models developed using the neural network technique always predicted more accurately than equivalent least-squares models. For example, the linear neural network model with the blue and green inputs predicted chlorophyll *a* with an RMS error of 4.11 $\mu g\, l^{-1}$ whereas the linear least-squares model predicted with an error of 34.78 $\mu g\, l^{-1}$. Furthermore, it was found that the different neural network architectures were simple to compare and thus the importance of different aspects of model development could by determined. The choice of the neural network technique was further supported in the final part of the research where the non-linear neural network models consistently produced more accurate predictions that were less affected by the known environmental influences than the linear multiple regression models with the same inputs.

This research aimed to determine if better algorithms could be developed for a 'typical' data set. Typically, imagery used to investigate chlorophyll *a* in the coastal zone is high spatial resolution airborne imagery. In the United Kingdom, CASI is often the instrument of choice to agencies investigating because its configuration is flexible. Although CASI has a very good noise level, over water the signal is low and so environmental and sensor effects can be strong. Both cluster analysis and principal component analysis demonstrated that, although the signal varied with chlorophyll *a* in the water, other environmental factors were having a strong, sometimes stronger effect on the signal.

Unfortunately, it was not possible to verify what factors in the environment caused these effects without the appropriate *in situ* measurements. Instead, a conjecture of the possible elements of the environment with which the principal components were correlated was made. Principal components 1 and 2 clearly were not related to chlorophyll *a*. An inspection of the imagery found that peaks in the first principal component corresponded to brighter regions in the water. The shape of the eigenvector corresponding to the second principal component indicated that CDOM may also be present in the scene. Because nothing was known about the distribution and spectral characteristics of these environmental factors is would have been unwise to try to correct for them.

A useful investigation used the information found in chapter 2 to determine whether

these factors could be shown to have a strong influence on the predictions. Two of the simple band ratios derived in the primary study were found to display filaments when plotted against chlorophyll *a*. An investigation into the nature of these filaments showed that they represented data points from different geographical regions. The blue-green ratio and the red-NIR ratio seemed to be affected by the changing spectral values recorded along the flight path as highlighted by the statistical analysis of the data (section 2.7.1) and cluster analysis of the data (section 2.7.3).

The data were also found to have strong sensor effects. Particularly strong was the noise that manifested with- and cross-the flight direction and the brightening of values at the edges of the image rows. The blue-green ratio, when plotted against chlorophyll *a* concentration was also clearly affected by the area of anomalous principal component values in the middle of image 1876 and this was probably due to data points having increased DN values as a consequence of being located close to the edge of the image. These effects were systematic and it would have been possible to correct for them, for example, each row in the un-geometrically corrected image could have had its pixels values adjusted according to some average of surrounding rows. Also, a model could have been fitted to the edge-brightening in the image and the brighter values adjusted accordingly. However, in accordance with the objective of the research to use a 'typical' data set, these corrections were not performed. Instead, the noise in the data was reduced by applying an averaging filter to the data.

The averaging filter also had the effect of mitigating errors in the co-location of data points. It is a simple technique that is commonly applied to data and may be considered an extension of the averaging performed by the sensor. Therefore, this study was undertaken with a data set containing realistic environmental effects and also several sources of error. However, only one basic pre-processing technique was applied to the data and so the results of this research are widely applicable to many such studies of chlorophyll *a* in Case 2 waters.

The investigation using simple least-squares linear regression confirmed that the usual method of developing chlorophyll *a* prediction models was not adequate for the present

data set. The linear and non-linear neural networks, however, tended to predict far more accurately. Although the more complex of these models were a little more accurate in prediction, the greater improvement in prediction was achieved with non-linear models. In the final development of models in chapter 6, the non-linear models again performed better than the linear models. However, it was apparent in the NNFS that the number of hidden nodes was not so vital in the development of an accurate model. Instead it was indicated that the optimal number of hidden nodes was determined by the weight initiation because the selected number of hidden nodes tended to vary with each training. Throughout this research it has become evident that non-linear regressions produce more accurate models but that the models do not need to be particularly complex. This is a particularly encouraging finding because it indicates that methods by which chlorophyll *a* prediction models are derived do not need to be overly computer intensive, rather attention for developing these models may be focused on other factors that will ensure a more accurate model.

One such factor, that was highlighted as very important throughout this research was the type and combination of feature that was to be input to the model. Much previous research into creating better chlorophyll *a* prediction models has concentrated on which features in the spectrum are best for the prediction. The primary study showed this to be a well-founded exercise as the most important factor in the accuracy of prediction was the spectral features used. Not only did an increase in the number of input features greatly reduce the error but particular features, in this case the blue band and the FLH, tended to be inputs to the more accurate models.

A deeper investigation into the features that were useful for predicting chlorophyll *a* concentration was undertaken, using three techniques of feature selection. The first of these used the common technique of selecting bands according to their known relationship to chlorophyll *a* and other environmental factors. This method selected bands that were different to the bands selected using the two automated methods which were based on the accuracy with which models predicted chlorophyll *a*. This indicated that certain bands, although alone being of little use to chlorophyll prediction, worked well in combination with other bands. It was a surprise that, when presented with all the

bands, the multiple regression feature selection (MLFS) determined that 34 were required for the accuracy of the prediction to remain statistically significant. With all these bands as inputs, both the linear and non-linear models produced very accurate predictions.

It is unfortunate that this research was unable to perform the NNFS with all the 72 bands. It was estimated that this would take several months to perform. The computational intensity of this technique is a limit to its more widespread use. However, the sets of features selected by NNFS in this research have provided a much greater understanding of the nature of the chlorophyll *a* to spectra relationship. It was clearly found that, where non-linear models may be developed, the blue region of the spectrum should be utilised.

It was clear that the regions of the spectrum that were most suitable for predicting chlorophyll *a* were determined by the type of model used. In the primary study, it was found that the blue, green and red image bands tended to perform better with the non-linear models whereas the FLH feature performed well within linear models and tended to reduce the training time required for non-linear neural networks, indicating that these were also converging on a linear solution.

As in the primary study, it was found that certain bands were more relevant to certain types of model. The linear MLFS tended to select green and red bands. The non-linear feature selection (NNFS) consistently selected blue bands. This was unexpected because these bands had been shown to be the most noisy, even after the two-dimensional Gaussian filter had been applied to the image (section 2.7.5). Previous studies have removed noisy bands (e.g. Benediktsson *et al.* 1995) but this may not have been well-founded. This indicated either that the neural network method was far more robust to noise in the data or that the blue wavebands had a very non-linear relationship to chlorophyll *a* concentration and so would not produce good predictions in a linear model. Another consideration was that, along with band 72 (913 nm), these noisy bands may have ensured that the function did not over-train. The work of Gross *et al.* (1999) showed clearly that the neural network required noise in the training

data to enable a good generalisation. This validates the choice of feature selection algorithms because alternative algorithms that do not base selection on the accuracy of prediction are more likely to select features for which generalisation of the model is poor. It is likely that blue wavebands were consistently selected for the neural network because of a combination of all these factors. Firstly that useful information was held by these bands but it had a non-linear relationship to chlorophyll *a* concentration and secondly that the noise in the data allowed the noise-robust neural network to extract this information without over-fitting to the training data.

The different types of model were also found to respond differently to the error in the data. By comparing the spatial distribution of error with the conjected sources of error, as represented by the first and second principal components and the estimate of tidal flow between sampling, it was possible to identify the causes of poor predictions. For example, it was apparent that the error in the linear models was frequently concentrated around a region of the cruise track around 585000 m east. This corresponded to a marked change in values for the second principal component. However, the neural network models did not show any increase in error for the same data. Both the models were affected by some influence around 630000 to 640000 m which could be correlated with a combination of changes in value in both the first and second principal components.

Because it has been shown that a non-linear model is better for determining the concentration of chlorophyll *a* in the water, it follows that other constituents do not linearly vary with the spectral values. Therefore, it was difficult to determine which environmental factors caused each of these regions of error. However, this technique of qualitatively investigating error was a valuable method which would be worth extending to use with a more comprehensive set of *in situ* data.

The visual assessment of the correlation between measured and predicted chlorophyll *a* concentrations was particularly useful to determine the reliability of the models. Often, although the RMS error and correlation coefficients for some models was found to be good, the data were clearly not distributed evenly about the line of measured

chlorophyll *a*=predicted chlorophyll *a*. This was particularly true for the linear models and should be born in mind when considering the quantitative assessments of their goodness of prediction. The corresponding values for the non-linear models fitted much better to this 1:1 line. This is another simple analysis technique that was found to have a great deal of value when assessing the goodness of prediction models.

This technique also highlighted another issue worthy of discussion. For the non-linear model a large over-prediction was frequently evident for measured chlorophyll *a* concentrations of between 5 and 7 $\mu g\, l^{-1}$. Such an over-prediction was evident in a few of the 1:1 plots in the primary study, however the error in prediction for the models on chapter 6 was quite considerable - up to 25 $\mu g\, l^{-1}$. This was thought to have been caused by the change in the method by which the training data were selected between the primary and secondary study. The method that was used in the primary study involved simply taking a random sample from the data as the training set. This resulted in a training set with much the same distribution as the testing set. Because a large proportion of the data were found in a short range of chlorophyll *a* concentrations, it was thought that this selection biased the regression towards this range of values by effectively providing *a priori* information about the data distribution to the neural network (Foody *et al.* 1995a; Foody *et al.* 1995b). Therefore, the training data sets used in the subsequent research were selected evenly from throughout the range of chlorophyll *a* concentrations. This did not bias the predictions towards certain values, however it was thought that not enough training examples were provided to characterise the relationship between chlorophyll *a* concentration and spectral values and this was noted particularly in the neural network models' prediction of a chlorophyll *a* concentration of about 6 $\mu g\, l^{-1}$. Clearly it was important to characterise the data well in the training set to ensure a good prediction of test data chlorophyll *a* concentration.

Another aspect of the research which may have resulted in a greater accuracy was to refine the neural network parameters. Only the neural network inputs, number of hidden nodes and training time were varied during the course of this research. The use of the neural network was therefore kept simple and the results achieved are attainable by any researcher in the field of Ocean Colour.

An assessment of the mean squared error achieved during training found that this was clearly declining up until the training was stopped. The validation error was used as an indicator of optimal training times in the primary study but it was found later that although the validation error tended to increase after an initial trough, it always began to decline again beyond 2000 iterations.

Because the procedure involved 10 networks working on the same problem it was likely that the training error achieved during the feature selection was perhaps optimal, given the neural network parameters. One network in the 10 achieved the lowest error because its initial weight conditions were the nearest to optimal. When testing the neural network models, the errors were greater than during the NNFS. The *one* neural network being trained during this testing stage probably did not start with the near-optimal initial weight conditions that the one neural network in *ten* feature selection network did. Therefore, this network probably did not reach the global minima in weight space at the time that the training was stopped. Better results may have been achieved by training for longer, training several times and choosing the network with the lowest error or adding a momentum term which would allow the training algorithm to find other minima in weight space and thus perhaps the global minima. Again, to keep the procedure simple, the networks were all trained for what was considered a reasonable number of iterations. Other parameters that could have been altered were the activation function (a tanh was used here), the number of layers (there were two in this research) and the learning rate (which was always set to 0.001). However, a low training error does not necessarily indicate a good testing error because there is a danger of over-fitting the model to the training data. This research found that the most straight-forward use of neural networks allow well-generalised models for predicting chlorophyll *a* in Case 2 waters to be developed.

This research was not able to apply these results to another data set of a different region or a different season. It is likely that the relationship between chlorophyll *a* and spectral information will change under different conditions (Kutser *et al.* 1995). However, the neural network was crucial in the development of a method for investigating this relationship that shall be useful in future work to extend knowledge about using remote

sensing to detect chlorophyll a in Case 2 waters.

Chapter 8

Conclusions

8.1 Summary of thesis

In previous ocean colour studies, a number of spectral features have been used to determine the concentration of chlorophyll *a* in water. In open ocean water, a linear model using a blue-green band ratio has been found to be adequate. However, in waters affected by environmental factors other than chlorophyll *a* other features from the spectrum have been used. More recently models that require more than one spectral feature as inputs have been developed and a few studies have used neural networks in the development of non-linear chlorophyll *a* prediction models. The research presented here has built on these previous works by investigating the value of using neural networks to develop chlorophyll *a* prediction models where the error in the data was large and difficult to correct for - a typical data set.

The various sources of this error were investigated and it was found that a number of factors had a greater influence on the spectral values than did chlorophyll *a*. The data were also found to be highly affected by noise. However, it was confirmed that the chlorophyll *a* signal was present in the spectral measurements and a qualitative assessment was made of the main sources of error in the data using statistical, cluster

and principal component analyses and calculation of the presence of error due to tidal motion of the water. This investigation resulted in a clear understanding of the sources of error that were likely to be encountered when predicting chlorophyll *a* from these data according to both the spectral and geographic regions in the data.

Single- and multi-layer perceptrons were chosen to derive chlorophyll *a* prediction models. This type of neural network was well used and understood and could be directly compared to simple and multiple linear regression because the same least-squares technique was used to fit the models to the data. This error minimisation technique assumes that the output error has a normal distribution. The ability to develop and compare several models of varying linearity and complexity at the same time was useful to the investigation of the nature of the relationship between chlorophyll *a* and spectral information. Early on the need for a selection of model inputs was recognised.

The data were divided randomly into training, validation and testing sets. Chlorophyll *a* prediction models were developed using the standard linear regression and neural network techniques with the training and validation sets. The spectral features for these models were selected using information from previous studies and from relationships that had been found within the data. These models were then assessed using a number of methods to determine their validity and their ability to predict chlorophyll *a* in the testing data set. The linearity, complexity and inputs to the models were also compared to determine which were the most important factors in the accuracy of the predictions.

Building on the findings of this investigation, the technique was refined to investigate the spectral dependence of the relationship between chlorophyll *a* and spectral information. The training data were chosen this time to be distributed evenly about the range of the chlorophyll *a* concentrations. A standard technique of selecting spectral bands was compared to two feature selection methods that were closely related to the regression techniques to be used. The resulting sets of features were then regressed on chlorophyll *a* to produce further chlorophyll *a* prediction models. Both the results of the feature selection and the prediction of the models were assessed to identify which

spectral regions were the most useful given the model to be used and the environmental conditions of the scene. The error in the model was used to determine which environmental factors were likely to result in the most error and with which models.

This work has made several contributions to research in the field of ocean colour remote sensing. A number of simple techniques have been demonstrated that extract valuable information from a data set. For example, the spatial locations of environmental substances that are likely to strongly affect the remotely sensed signal may be estimated using cluster analysis and principal component analysis. Also, by investigating the architecture and the manner of training of a simple neural network, details about the relationship between the spectral input data and the chlorophyll *a* concentration output data may be inferred.

Two techniques that have proved particularly valuable were those of neural network regression and automated feature selection. The former technique allowed the investigation of the linearity and complexity of the relationship between chlorophyll *a* and remotely sensed spectra to be thoroughly investigated. The latter showed that there may be better sets of spectral features for chlorophyll *a* prediction than those that are indicated by the knowledge of the spectral signal of chlorophyll *a*.

Therefore, although many of the findings of this research may only relate to the *Norfolk 30/05/96* data, they do indicate a need for a more thorough investigation of ocean colour data. Furthermore, the techniques presented here may be applied to any new data set being analysed. The conclusions of any such investigation will then be of great value to the ocean colour community.

8.2 Further work

This research was undertaken in a sequential and modular fashion with the results at each stage being used to determine the method for the next. Although the final conclusions of the current research are laid out in the next section, the success of this

research indicates that there remains a great deal of potential for further investigations in this area. The areas in which the research could develop further are the data used in the research, the type of input features used and the refinement of the neural network method. These topics are addressed under their respective headings below.

8.2.1 The data set

This research only covered one site at one date. There is a great deal of evidence that different chlorophyll *a* prediction algorithms are required for each site and season and so this research should be extended to include different conditions. There were a number of aspects to the data for which it was difficult to ascertain the accuracy of the data. There were just enough *in situ* measurements of chlorophyll *a* for the calibration of the fluorometer data but a more comprehensive set of calibration points would have allowed greater confidence in this calibration. There were also uncertainties with the co-location of some data points, especially when there was a difference of several hours between *in situ* measurement and overflight. Ambiguities such as these are to be expected in real data, however where a number of cruises and over-flights have been performed for several different seasons, it should be possible to select enough of the more reliable data to provide a very useful data set for the development of this research.

Where it is necessary to derive unique empirical chlorophyll *a* predictions models for each site, the derivation of reliable models is usually restricted by the number of locations for which both spectral and chlorophyll *a* measurements are available. The research presented here used the maximum number of samples available for training the regression algorithms. However, it would be of value to determine the accuracy which may be achieved using fewer data points and which methods of determining the model parameters (conventional or neural network) are most suited to research with a limited data set.

Another source of spectral measures are space-borne data. Space-borne sensors have necessarily a much coarser spatial or spectral resolution (or both). It would be worth-

while investigating how the results from this research 'scale-up' to the resolution of sensors such as SeaWiFS. Such an investigation would be of relevance to the design of future space-borne sensors, specifically aimed at viewing coastal regions.

8.2.2 The spectral features

Initially, many different types of feature were investigated for this research. The original bands were used but the value of other more complex bands is recognised. In section 1.3 band ratios and differences and spectral derivatives were discussed. These features have shown more potential than the simple image bands in many studies and so should be useful inputs to a neural network model.

With access to a spatial data set, it seemed pertinent to investigate the potential of using spatial information as an input to chlorophyll *a* prediction models. Since the 1970s, a great deal of interesting research into the spatial variability of phytoplankton has been undertaken. By comparing the spatial variation of chlorophyll *a* to other tracers of ocean currents such as temperature or suspended sediment, it has been noticed that at certain spatial scales the variance spectra of the two factors diverge or, similarly, the autocorrelation between them declines (Denman 1976; Steele and Henderson 1979; Gower *et al.* 1980; Campbell and Esaias 1985; Strutton *et al.* 1997). This occurs when phytoplankton growth or zooplankton grazing is at a greater rate than the dissipation of energy by eddies (Denman and Platt 1976; Davis *et al.* 1991).

In waters where the spatial distribution of phytoplankton is thought to be a consequence of currents the spatial distribution of tracers which are easier to detect, such as suspended sediments, may be used as an aid to determination the distribution of chlorophyll *a*. Under those conditions when the spatial signal of phytoplankton can be separated from that of tracers in the water, the spatial information in the data should be investigated for use in chlorophyll *a* prediction where suspended sediments strongly affect the water-leaving signal.

Neural networks are ideally suited to the input of spatial information (Miller *et al.* 1995). For example, multidimensional input arrays can be used to account for spatial and spectral dimensions in the data. This may remove the need for an averaging window around the data points to correct for noise.

It was noted in the primary study that bands that were close to one another may have been more useful in the prediction of chlorophyll *a* because they were similarly affected by environmental influences. This was not investigated further due to a need to restrict the size of data sets from which features were selected. However, it is worth investigating in future research as the relative location of ideal bands for predicting chlorophyll *a* is of great relevance to ocean colour research and to the design of ocean colour sensors in the future.

8.2.3 The neural network

There are a great many techniques being developed in the field of neural networks which may be applicable to this field of research. It is important that the subject area be fully understood before applying any new method, to ensure that it is suitable to the problem. The research presented here has highlighted a number of areas in which the current feed-forward/back-propagation method could be developed. For example, it was thought that the final networks did not achieve a global error minimum. Future studies could simply train the network for longer, or utilise parameter optimisation algorithms other than the gradient descent algorithm used here.

Another issue raised in this research is the noise level of the data. It was considered possible that the noise in the data improved the generalisation during training. Providing noise explicitly in the input has been suggested (Sietsma and Dow 1991; Bishop 1995) and certainly could prove useful in this case. The explicit addition of noise in the input would therefore allow the information value of bands selected by the feature selection to be better investigated.

A number of neural networks were trained in this research. Although some were found to give better predictions than others, all the networks trained in the secondary study predicted to within a reasonable accuracy. Rather than selecting the best neural network to predict chlorophyll *a* concentration in a future study, it may be expedient to combine the 'expertise' of all the networks in a committee of networks (Bishop 1995). Such a network has as inputs the predictions from several other networks. In this way the curse of dimensionality is avoided because each network only deals with a small amount of data. The accuracy of prediction may also be increased and furthermore the resulting network may predict chlorophyll *a* concentration in new data with greater accuracy. This technique may be applied to feature selection such that all the available image bands may be selected from.

It is beyond the scope of this research to investigate the many different neural network techniques that may be of use in producing chlorophyll *a* prediction algorithms. However, the basic techniques demonstrated here have shown a great deal of potential. This indicates that there is value in exploring further the new developments in the field of neural networks.

8.3 Conclusions

The approach taken in the research presented here is highly analytical. All aspects of the regression of chlorophyll *a* on spectral data have been investigated thoroughly. This includes the quality of the data, the linearity and complexity of the models used and the causes of error. Performing such a complete examination of the regression has not previously been undertaken in the field of ocean colour. Furthermore, applying the two automatic methods of feature selection is a new approach to the problem of designing robust models for the prediction of chlorophyll *a*. The method by which neural networks were employed to investigate the nature of the chlorophyll *a*-spectra relationship is original in remote sensing as a whole. Prior to this research, the data used with neural networks has not been typical of that used by bodies interested in

detecting chlorophyll *a* in Case 2 waters. However, this research uses a very typical set of data.

Not only were these techniques entirely novel, but they have enabled a valuable contribution to the knowledge in the field of ocean colour. The following pages summarise these findings according to the aims of this investigation into the *Norfolk 30/05/96* data.

1. The separability of the chlorophyll *a* spectral signature in the presence of other environmental factors and sensor noise
    - The spectral signal from chlorophyll *a* may be separated from those of the atmosphere, suspended sediment and CDOM
    - However it was apparent that the signal could not fully be separated using a linear technique
2. The severity of the contaminating signals and noise
    - Other environmental factors and sensor noise have a much stronger influence on the detected spectral signal than does chlorophyll *a*
    - One spectral feature is not adequate to characterise the change in chlorophyll *a* under these circumstances
3. The applicability of neural networks to predicting chlorophyll *a* concentrations from spectra contaminated by other environmental factors and sensor noise
    - Neural networks provide a convenient method by which non-linear multiple regression may be achieved
    - All stages of the training of neural-networks provide useful information about the nature of the chlorophyll *a*-spectra relationship
4. A comparison of linear regression and neural network regression techniques
    - Neural networks almost always outperformed linear regression models

- The neural network was, however, more time-consuming to train
- The neural network tended to require fewer bands to produce a better prediction model
- The multiple linear regression models tended to be more strongly affected by environmental factors whereas the neural network models tended to be more strongly affected by the data set used for training

5. The nature of the relationship between chlorophyll *a* and different regions of the spectrum
   - FLH has a linear relationship to chlorophyll *a*
   - The blue region of the spectrum is noisy and has a non-linear relationship to chlorophyll *a*
   - Other regions of the spectrum did not show a strong non-linear relationship to chlorophyll *a* but by the increase in performance with fewer bands, it was inferred that some extra information was available to the non-linear models from throughout the spectrum
6. The identification of the most appropriate spectral regions for the prediction of chlorophyll *a* concentration
   - Where possible, a selection of features that is based on the goodness of prediction using those features should be used to highlight useful features in the data that may not be uncovered by conventional methods
   - Features with a linear relationship to chlorophyll *a* such as FLH are useful for linear regression-derived models
   - The blue spectral region contains a lot of information about chlorophyll *a* concentration but is only useful if a non-linear model that is robust to noise may be derived
7. Other findings
   - It is important to use the most suitable features for the type of prediction model to be used

- The use of several types of model assessment techniques (qualitative and quantitative) is desirable to gain a full understanding of the performance of chlorophyll *a* prediction models and sources of error

The techniques and results presented here have shown that reliable chlorophyll *a* prediction algorithms can be developed even for regions previously thought to be too contaminated by other spectral signatures.

Appendix A

The CASI enhanced spectral bandset

Band	Lower limit (nm)	Centre (nm)	Upper limit (nm)	Width (nm)
1	401.6	405.7	409.8	8.2
2	408.6	412.7	416.8	8.2
3	415.6	419.7	423.8	8.2
4	422.6	426.7	430.8	8.2
5	429.7	433.8	437.9	8.2
6	436.7	440.8	444.9	8.2
7	443.7	447.8	451.9	8.2
8	450.8	454.9	459	8.2
9	457.8	461.9	466	8.2
10	464.9	469	473.1	8.2
11	471.9	476	480.1	8.2
12	479	483.1	487.2	8.2
13	486	490.1	494.2	8.2
14	493.1	497.2	501.3	8.2
15	500.1	504.3	508.5	8.4
16	507.1	511.3	515.5	8.4
17	514.2	518.4	522.6	8.4

continued on next page

continued from previous page

Band	Lower Limit (nm)	Centre (nm)	Upper Limit (nm)	Width (nm)
18	521.3	525.5	529.7	8.4
19	528.3	532.5	536.7	8.4
20	535.4	539.6	543.8	8.4
21	542.5	546.7	550.9	8.4
22	549.6	553.8	558	8.4
23	556.7	560.9	565.1	8.4
24	563.8	568	572.2	8.4
25	570.9	575.1	579.3	8.4
26	578	582.2	586.4	8.4
27	585.1	589.3	593.5	8.4
28	592.2	596.4	600.6	8.4
29	599.3	603.5	607.7	8.4
30	606.5	610.7	614.9	8.4
31	613.6	617.8	622	8.4
32	620.7	624.9	629.1	8.4
33	627.9	632.1	636.3	8.4
34	635	639.2	643.4	8.4
35	642.1	646.3	650.5	8.4
36	649.3	653.5	657.7	8.4
37	656.4	660.6	664.8	8.4
38	663.6	667.8	672	8.4
39	670.7	674.9	679.1	8.4
40	677.9	682.1	686.3	8.4
41	685.1	689.3	693.5	8.4
42	692.2	696.4	700.6	8.4
43	699.4	703.6	707.8	8.4
44	706.6	710.8	715	8.4
45	713.8	718	722.2	8.4

continued on next page

continued from previous page				
Band	Lower Limit (nm)	Centre (nm)	Upper Limit (nm)	Width (nm)
46	720.9	725.1	729.3	8.4
47	728.1	732.3	736.5	8.4
48	735.3	739.5	743.7	8.4
49	742.5	746.7	750.9	8.4
50	749.7	753.9	758.1	8.4
51	756.9	761.1	765.3	8.4
52	764.1	768.3	772.5	8.4
53	771.3	775.5	779.7	8.4
54	778.6	782.8	787	8.4
55	785.8	790	794.2	8.4
56	793	797.2	801.4	8.4
57	800.2	804.4	808.6	8.4
58	807.5	811.7	815.9	8.4
59	814.7	818.9	823.1	8.4
60	822	826.2	830.4	8.4
61	829.2	833.4	837.6	8.4
62	836.5	840.7	844.9	8.4
63	843.7	847.9	852.1	8.4
64	851	855.2	859.4	8.4
65	858.3	862.5	866.7	8.4
66	865.5	869.7	873.9	8.4
67	872.8	877	881.2	8.4
68	880.1	884.3	888.5	8.4
69	887.4	891.6	895.8	8.4
70	894.7	898.9	903.1	8.4
71	902	906.2	910.4	8.4
72	909.3	913.5	917.7	8.4

Appendix B

Calculating the tidal motion of the sampled parcels of water

```
>> % TIME:
>> time  = [1000 1100 1200 1300 1400 1500 1600 1700 1800]';
>>
>> % SPEED IN KNOTS:
>> speed = [1.5  1.4  1.1  0.6  0.0  0.9  1.6  1.9  1.6 ]';
>>
>> % DIRECTION FROM NORTH = 0:
>> dirn  = [285  285  270  270  270  105  105  105  105 ]';
>>
>> % SPEED IN M/S:
>> speedms = speed*0.514;
>>
>> % CONVERSION TO RADIANS:
>> c = asin(0.5)/30;
>> dirnc = dirn*c;
>>
>> % DISTANCE EAST PER SECOND:
>> E = speedms.*sin(dirnc);
>>
>> % DISTANCE NORTH PER SECOND:
>> N = speedms.*cos(dirnc);
>>
>> % SECONDS IN AN HOUR:
>> hour = 3600;
>>
>> % SECONDS IN A MINUTE:
>> min  = 60;
>>
```

```
>> % HIGH WATER WAS AT 0849 HOURS:
>> HW = 8*hour + 49*min;
>>
>> % INITIATE MATRIX OF VELOCITIES OF THE CHLOROPHYLL SAMPLES:
>> ChlVel  = zeros(2300,2);
>>
>> % INITIATE MATRIX OF VELOCITIES OF THE SPECTRAL SAMPLES:
>> SpecVel = zeros(2300,2);
>>
>> % INITIATE MATRIX OF DIFFERENCE IN VELOCITIES BETWEEN
>> % CHLOROPHYLL AND SPECTRAL SAMPLES
>> VelDiff = zeros(2300,2);
>>
>> % PRODUCE MATRICES OF VELOCITIES BETWEEN SAMPLINGS:
>> for i = 1:2300
Cindex = round((ChlTime(i)-HW)/hour)-1;
Sindex = round((SpecTime(i)-HW)/hour)-1;
ChlVel(i,:) = [E(Cindex),N(Cindex)];
SpecVel(i,:) = [E(Sindex),N(Sindex)];
%
% DETERMINE WHICH SAMPLES WERE THE FIRST AND LAST AT EACH
% DATA POINT:
start = min([ChlTime(i),SpecTime(i)]);
stop  = max([ChlTime(i),SpecTime(i)]);
%
% FOR EACH SECOND, CALCULATE THE RELATIVE MOTION OF THE
% SAMPLED PARCELS OF WATER
for s = start:stop
VelDiff(i,1) = VelDiff(i,1) + E(round((s-HW)/hour)-1);
VelDiff(i,2) = VelDiff(i,2) + N(round((s-HW)/hour)-1);
end
end
>>
>> % THE DISTANCE BETWEEN THE LOCATIONS OF THE SAMPLED
>> % PARCELS OF WATER AT THE TIME OF THE SECOND MEASUREMENT
>> DistDiff = sqrt(VelDiff(:,1).^2 + VelDiff(:,2).^2);
```

Appendix C

The output of the multiple linear regression feature selection method

The following pages show the output from the multiple linear regression feature selection method described in chapter 5. Each table gives the results for one run of the function.

Table C.1: Explanation of headings of tables C.2-D.9

table value	meaning	
number of inputs	the number of inputs being included in the current configuration	
number of hidden nodes	the number of hidden nodes being included in the current configuration	
number of weights	the number of weights being included in the current configuration (current value/previous value)	
mean squared error	the error after the current band/hidden node has been rejected	
numerator degrees of freedom	= the difference between the degrees of freedom for the full and reduced model	used to calculate L and F_α
denominator degrees of freedom	= the degrees of freedom for the full model (current value/previous value)	
L	The test statistic	
F_α	Value from the f-distribution corresponding to a significance level of 95 %	
accept or reject?	Did the statistic determine to accept or reject the reduced configuration?	
rejected band	If the statistic determine that the reduced input configuration should be accepted, which band had been removed?	
double lines	indicate the final chosen configuration	
remaining bands	The bands left in the final configuration	

Table C.2: Results of multiple linear regression feature selection for subset A

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	7.8582						
8	8.3069	1	191	10.9057	3.92	reject	37

remaining bands = 1, 10, 19, 28, 37, 46, 55, 64

Table C.3: Results of multiple linear regression feature selection for subset B

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	10.0222						
7	9.2138	1	191	-15.4058	3.92	accept	2
6	8.2862	1	192	-19.3302	3.92	accept	47
6	9.0222	1	193	17.1433	3.92	reject	56

remaining bands = 11, 20, 29, 38, 56, 65

Table C.4: Results of multiple linear regression feature selection for subset C

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	10.3042						
7	8.4811	1	191	-33.7927	3.92	accept	3
6	8.0964	1	192	-8.7102	3.92	accept	48
6	8.7006	1	193	14.405	3.92	reject	57

remaining bands = 12, 21, 30, 39, 57, 66

Table C.5: Results of multiple linear regression feature selection for subset D

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	15.4476						
7	14.4094	1	191	-12.8366	3.92	accept	4
6	12.002	1	192	-32.0786	3.92	accept	58
5	12.054	1	193	0.83641	3.92	accept	40
5	13.427	1	194	22.0979	3.92	reject	67

remaining bands = 13, 22, 31, 49, 67

Table C.6: Results of multiple linear regression feature selection for subset E

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	13.4522						
7	12.1312	1	191	-18.7562	3.92	accept	59
6	11.5614	1	192	-9.0185	3.92	accept	23
5	11.6565	1	193	1.5879	3.92	accept	32
4	11.8353	1	194	2.9759	3.92	accept	68
4	14.6441	1	195	46.2774	3.92	reject	50

remaining bands = 5, 14, 41, 50

Table C.7: Results of multiple linear regression feature selection for subset F

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	10.7614						
7	9.6848	1	191	-19.1077	3.92	accept	24
6	8.965	1	192	-14.27	3.92	accept	60
5	8.9355	1	193	-0.63682	3.92	accept	69
4	8.8472	1	194	-1.9153	3.92	accept	15
4	15.7375	1	195	151.867	3.92	reject	6

remaining bands = 6, 33, 42, 51

Table C.8: Results of multiple linear regression feature selection for subset G

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	12.1074						
7	10.0241	1	191	-32.8652	3.92	accept	16
6	9.8354	1	192	-3.6139	3.92	accept	61
5	9.8813	1	193	0.89999	3.92	accept	25
5	11.2105	1	194	26.0969	3.92	reject	70

remaining bands = 7, 34, 43, 52, 70

Table C.9: Results of multiple linear regression feature selection for subset H

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	10.3819						
7	10.39	1	191	0.14937	3.92	accept	62
6	10.4044	1	192	0.26587	3.92	accept	71
6	11.4719	1	193	19.8018	3.92	reject	26

remaining bands = 8, 17, 26, 35, 44, 53

Table C.10: Results of multiple linear regression feature selection for subset I

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
8	14.1003						
7	14.1033	1	191	0.040584	3.92	accept	36
6	13.9915	1	192	-1.5225	3.92	accept	27
5	13.0093	1	193	-13.5487	3.92	accept	63
5	13.34	1	194	4.9321	3.92	reject	72

remaining bands = 9, 18, 45, 54, 72

Table C.11: Results of multiple linear regression feature selection of all available bands

number of inputs	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F	accept or reject?	rejected band
72	18.8939						
71	8.5934	1	127	-69.2372	3.92	accept	33
70	5.0387	1	128	-52.9487	3.92	accept	61
69	4.3718	1	129	-17.0734	3.92	accept	49
68	3.6299	1	130	-22.0597	3.92	accept	72
67	3.0919	1	131	-19.4165	3.92	accept	2
66	2.9112	1	132	-7.7139	3.92	accept	9
65	2.7403	1	133	-7.8082	3.92	accept	41
64	2.6087	1	134	-6.4338	3.92	accept	12
63	2.5582	1	135	-2.6153	3.92	accept	47
62	2.4719	1	136	-4.5865	3.92	accept	70
61	2.357	1	137	-6.3674	3.92	accept	21
60	2.2725	1	138	-4.9489	3.92	accept	60
59	2.2634	1	139	-0.55633	3.92	accept	32
58	2.2571	1	140	-0.3907	3.92	accept	20
57	2.2467	1	141	-0.64909	3.92	accept	71
56	2.0794	1	142	-10.573	3.92	accept	3
55	1.856	1	143	-15.3637	3.92	accept	50
54	1.7883	1	144	-5.254	3.92	accept	69
53	1.7363	1	145	-4.2189	3.92	accept	30
52	1.7158	1	146	-1.7193	3.92	accept	31
51	1.7031	1	147	-1.0886	3.92	accept	25
50	1.704	1	148	0.080471	3.92	accept	42
49	1.6908	1	149	-1.1546	3.92	accept	7
48	1.6976	1	150	0.60398	3.92	accept	46
47	1.7012	1	151	0.31953	3.92	accept	4
46	1.7085	1	152	0.64912	3.92	accept	5
45	1.6677	1	153	-3.6575	3.92	accept	8
44	1.6828	1	154	1.3977	3.92	accept	37
43	1.6918	1	155	0.82722	3.92	accept	68
42	1.6778	1	156	-1.2858	3.92	accept	17
41	1.6783	1	157	0.040629	3.92	accept	36
40	1.6879	1	158	0.90951	3.92	accept	22
39	1.701	1	159	1.2296	3.92	accept	55
38	1.7276	1	160	2.5052	3.92	accept	13
37	1.7598	1	161	2.9995	3.92	accept	10
36	1.7014	1	162	-5.3748	3.92	accept	56
35	1.6401	1	163	-5.8726	3.92	accept	53
34	1.6493	1	164	0.92307	3.92	accept	48
34	1.7029	1	165	5.354	3.92	reject	23

remaining bands = 1, 6, 11, 14, 15, 16, 18, 19, 23, 24, 26, 27, 28, 29, 34, 35, 38, 39, 40, 43, 44, 45, 51, 52, 54, 57, 58, 59, 62, 63, 64, 65, 66, 67

Appendix D

The output of the neural network feature selection method

The following pages show the output from the neural network feature and model selection function described in chapter 5. Each table gives the results for one run of the function. See table C.1 for an explanation of the headings for these tables.

Table D.1: Results of neural network feature selection for band subset A

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.06						
8	17	171/181	0.058	10	29/19	-0.065	2.38	accept	
8	16	161/171	0.057	10	39/29	-0.072	2.18	accept	
7	16	145/161	0.052	16	55/39	-0.193	1.84	accept	64
7	15	136/145	0.05	9	64/55	-0.23	2.04	accept	
6	15	121/136	0.049	15	79/64	-0.08	1.75	accept	46
6	14	113/121	0.049	8	87/79	-0.136	2.02	accept	
5	14	99/113	0.044	14	101/87	-0.603	1.75	accept	37
5	13	92/99	0.051	7	108/101	2.387	2.09	reject	
4	14	85/99	0.063	14	115/101	3.06	1.75	reject	

remaining bands = 1, 10, 19, 28, 55

Table D.2: Results of neural network feature selection for band subset B

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.057						
8	17	171/181	0.051	10	29/19	-0.209	2.38	accept	
8	16	161/171	0.049	10	39/29	-0.066	2.18	accept	
7	16	145/161	0.046	16	55/39	-0.171	1.84	accept	65
7	15	136/145	0.046	9	64/55	0.044	2.04	accept	
6	15	121/136	0.046	15	79/64	-0.042	1.75	accept	11
6	14	113/121	0.051	8	87/79	1.045	2.02	accept	
5	14	99/113	0.04	14	101/87	-1.269	1.75	accept	56
5	13	92/99	0.043	7	108/101	0.982	2.09	accept	
4	13	79/92	0.04	13	121/108	-0.601	1.75	accept	47
4	12	73/79	0.041	6	127/121	0.434	2.1	accept	
3	12	61/73	0.041	12	139/127	0.149	1.75	accept	38
3	11	56/61	0.04	5	144/139	-1.177	2.21	accept	
2	11	45/56	0.071	11	155/144	10.436	1.75	reject	
3	10	51/56	0.04	5	149/144	-0.014	2.21	accept	
2	10	41/51	0.063	10	159/149	8.935	1.83	reject	
3	9	46/51	0.043	5	154/149	2.761	2.21	reject	
2	10	41/51	0.071	10	159/149	11.611	1.83	reject	

remaining bands = 2, 20, 29

Table D.3: Results of neural network feature selection for band subset C

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.058						
8	17	171/181	0.054	10	29/19	-0.128	2.38	accept	
8	16	161/171	0.061	10	39/29	0.384	2.18	accept	
7	16	145/161	0.04	16	55/39	-0.824	1.84	accept	66
7	15	136/145	0.044	9	64/55	0.528	2.04	accept	
6	15	121/136	0.043	15	79/64	-0.04	1.75	accept	48
6	14	113/121	0.043	8	87/79	0.035	2.02	accept	
5	14	99/113	0.048	14	101/87	0.718	1.75	accept	57
5	13	92/99	0.048	7	108/101	-0.132	2.09	accept	
4	13	79/92	0.047	13	121/108	-0.123	1.75	accept	12
4	12	73/79	0.051	6	127/121	1.72	2.1	accept	
3	12	61/73	0.05	12	139/127	-0.225	1.75	accept	39
3	11	56/61	0.053	5	144/139	1.768	2.21	accept	
2	11	45/56	0.092	11	155/144	9.341	1.75	reject	
3	10	51/56	0.04	5	149/144	-6.99	2.21	accept	
2	10	41/51	0.093	10	159/149	19.434	1.83	reject	
3	9	46/51	0.058	5	154/149	13.073	2.21	reject	
2	10	41/51	0.092	10	159/149	19.022	1.83	reject	

remaining bands = 3, 21, 30

Table D.4: Results of neural network feature selection for band subset D

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.1						
8	17	171/181	0.083	10	29/19	-0.321	2.38	accept	
8	16	161/171	0.092	10	39/29	0.314	2.18	accept	
7	16	145/161	0.075	16	55/39	-0.445	1.84	accept	31
7	15	136/145	0.076	9	64/55	0.004	2.04	accept	
6	15	121/136	0.065	15	79/64	-0.566	1.75	accept	67
6	14	113/121	0.064	8	87/79	-0.166	2.02	accept	
5	14	99/113	0.065	14	101/87	0.037	1.75	accept	13
5	13	92/99	0.075	7	108/101	2.256	2.09	reject	
4	14	85/99	0.067	14	115/101	0.276	1.75	accept	58
4	13	79/85	0.069	6	121/115	0.409	2.17	accept	
3	13	66/79	0.069	13	134/121	-0.01	1.67	accept	22
3	12	61/66	0.08	5	139/134	4.582	2.21	reject	
2	13	53/66	0.133	13	147/134	9.638	1.67	reject	

remaining bands = 4, 40, 49

Table D.5: Results of neural network feature selection for band subset E

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.069						
8	17	171/181	0.064	10	29/19	-0.143	2.38	accept	
8	16	161/171	0.077	10	39/29	0.613	2.18	accept	
7	16	145/161	0.063	16	55/39	-0.452	1.84	accept	32
7	15	136/145	0.071	9	64/55	0.765	2.04	accept	
6	15	121/136	0.06	15	79/64	-0.614	1.75	accept	68
6	14	113/121	0.056	8	87/79	-0.681	2.02	accept	
5	14	99/113	0.061	14	101/87	0.497	1.75	accept	14
5	13	92/99	0.065	7	108/101	1.087	2.09	accept	
4	13	79/92	0.071	13	121/108	0.656	1.75	accept	50
4	12	73/79	0.076	6	127/121	1.566	2.1	accept	
3	12	61/73	0.102	12	139/127	3.677	1.75	reject	
4	11	67/73	0.072	6	133/127	-1.215	2.1	accept	
3	11	56/67	0.109	11	144/133	6.265	1.75	reject	
4	10	61/67	0.071	6	139/133	-0.191	2.1	accept	
3	10	51/61	0.096	10	149/139	4.807	1.83	reject	
4	9	55/61	0.077	6	145/139	1.919	2.1	accept	
3	9	46/55	0.094	9	154/145	3.492	1.88	reject	
4	8	49/55	0.074	6	151/145	-0.894	2.1	accept	
3	8	41/49	0.102	8	159/151	7.209	1.94	reject	
4	7	43/49	0.075	6	157/151	0.279	2.1	accept	
3	7	36/43	0.092	7	164/157	5.054	2.01	reject	
4	6	37/43	0.067	6	163/157	-2.848	2.1	accept	
3	6	31/37	0.094	6	169/163	10.957	2.1	reject	
4	5	31/37	0.079	6	169/163	5.161	2.1	reject	
3	6	31/37	0.086	6	169/163	7.92	2.1	reject	

remaining bands = 5, 23, 41, 59

Table D.6: Results of neural network feature selection for band subset F

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.095						
8	17	171/181	0.087	10	29/19	-0.157	2.38	accept	
8	16	161/171	0.084	10	39/29	-0.089	2.18	accept	
7	16	145/161	0.077	16	55/39	-0.224	1.84	accept	69
7	15	136/145	0.064	9	64/55	-0.994	2.04	accept	
6	15	121/136	0.066	15	79/64	0.14	1.75	accept	24
6	14	113/121	0.063	8	87/79	-0.495	2.02	accept	
5	14	99/113	0.068	14	101/87	0.466	1.75	accept	15
5	13	92/99	0.069	7	108/101	0.374	2.09	accept	
4	13	79/92	0.066	13	121/108	-0.378	1.75	accept	51
4	12	73/79	0.077	6	127/121	3.167	2.1	reject	
3	13	66/79	0.1	13	134/121	4.738	1.67	reject	

remaining bands = 6, 33, 42, 60

Table D.7: Results of neural network feature selection for band subset G

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.104						
8	17	171/181	0.102	10	29/19	-0.038	2.38	accept	
8	16	161/171	0.109	10	39/29	0.177	2.18	accept	
7	16	145/161	0.092	16	55/39	-0.367	1.84	accept	16
7	15	136/145	0.098	9	64/55	0.349	2.04	accept	
6	15	121/136	0.086	15	79/64	-0.524	1.75	accept	34
6	14	113/121	0.092	8	87/79	0.771	2.02	accept	
5	14	99/113	0.081	14	101/87	-0.726	1.75	accept	52
5	13	92/99	0.097	7	108/101	2.707	2.09	reject	
4	14	85/99	0.087	14	115/101	0.52	1.75	accept	70
4	13	79/85	0.088	6	121/115	0.109	2.17	accept	
3	13	66/79	0.126	13	134/121	4.067	1.67	reject	
4	12	73/79	0.084	6	127/121	-0.809	2.1	accept	
3	12	61/73	0.136	12	139/127	6.449	1.75	reject	
4	11	67/73	0.089	6	133/127	1.104	2.1	accept	
3	11	56/67	0.125	11	144/133	4.918	1.75	reject	
4	10	61/67	0.086	6	139/133	-0.649	2.1	accept	
3	10	51/61	0.119	10	149/139	5.366	1.83	reject	
4	9	55/61	0.094	6	145/139	2.077	2.1	accept	
3	9	46/55	0.124	9	154/145	5.253	1.88	reject	
4	8	49/55	0.089	6	151/145	-1.33	2.1	accept	
3	8	41/49	0.127	8	159/151	8.105	1.94	reject	
4	7	43/49	0.095	6	157/151	1.768	2.1	accept	
3	7	36/43	0.13	7	164/157	8.21	2.01	reject	
4	6	37/43	0.11	6	163/157	4.053	2.1	reject	
3	7	36/43	0.122	7	164/157	6.361	2.01	reject	

remaining bands = 7, 25, 43, 61

Table D.8: Results of neural network feature selection for band subset H

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.118						
8	17	171/181	0.112	10	29/19	-0.099	2.38	accept	
8	16	161/171	0.113	10	39/29	0.033	2.18	accept	
7	16	145/161	0.11	16	55/39	-0.067	1.84	accept	71
7	15	136/145	0.1	9	64/55	-0.547	2.04	accept	
6	15	121/136	0.104	15	79/64	0.158	1.75	accept	17
6	14	113/121	0.098	8	87/79	-0.509	2.02	accept	
5	14	99/113	0.077	14	101/87	-1.356	1.75	accept	26
5	13	92/99	0.078	7	108/101	0.214	2.09	accept	
4	13	79/92	0.073	13	121/108	-0.483	1.75	accept	62
4	12	73/79	0.081	6	127/121	2.173	2.1	reject	
3	13	66/79	0.121	13	134/121	6.079	1.67	reject	

remaining bands = 8, 35, 44, 53

Table D.9: Results of neural network feature selection for band subset I

number of inputs	number of hidden nodes	number of weights	mean squared error	numerator degrees of freedom	denominator degrees of freedom	L	F_α	accept or reject?	rejected band
8	18	181	0.065						
8	17	171/181	0.060	10	29/19	-0.138	2.380	accept	
8	16	161/171	0.065	10	39/29	0.268	2.180	accept	
7	16	145/161	0.056	16	55/39	-0.341	1.840	accept	36
7	15	136/145	0.067	9	64/55	1.165	2.040	accept	
6	15	121/136	0.056	15	79/64	-0.687	1.750	accept	45
6	14	113/121	0.062	8	87/79	1.019	2.020	accept	
5	14	99/113	0.055	14	101/87	-0.665	1.750	accept	18
5	13	92/99	0.057	7	108/101	0.323	2.090	accept	
4	13	79/92	0.053	13	121/108	-0.470	1.750	accept	63
4	12	73/79	0.053	6	127/121	-0.230	2.100	accept	
3	12	61/73	0.062	12	139/127	1.893	1.750	reject	
4	11	67/73	0.050	6	133/127	-0.962	2.100	accept	
3	11	56/67	0.062	11	144/133	2.830	1.750	reject	
4	10	61/67	0.054	6	139/133	1.623	2.100	accept	
3	10	51/61	0.064	10	149/139	2.574	1.830	reject	
4	9	55/61	0.052	6	145/139	-0.925	2.100	accept	
3	9	46/55	0.063	9	154/145	3.394	1.880	reject	
4	8	49/55	0.049	6	151/145	-1.392	2.100	accept	
3	8	41/49	0.064	8	159/151	5.801	1.940	reject	
4	7	43/49	0.053	6	157/151	2.159	2.100	reject	
3	8	41/49	0.063	8	159/151	5.288	1.940	reject	

remaining bands = 9, 27, 54, 72

Bibliography

Abbott, M. R, O. B Brown, R. H Evans, H. R Gordon, K. L Carder, F. A Mller-Karger, and W. E Esaias, 1994. Ocean color in the 21st century: A strategy for a 20-year time series. Technical Report NASA Technical Memorandum 104566, Vol. 17, NASA Goddard Space Flight Center, Maryland.

Adams, J. B, M. O Smith, and P. E Johnson, 1986. Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site. *Journal of Geophysical Research* **91**((B8)), 8098–8112.

Aiken, J, S Hudson, G Moore, and H Bottrell, 1995a. Future development of airborne remote sensing techniques. Technical report, Plymouth Marine Laboratory. A preliminary report to the National Rivers Authority.

Aiken, J, G. F Moore, C. C Trees, S. B Hooker, and D. K Clark, 1995b. The SeaWiFS CZCS-type pigment algorithm. Technical Report NASA Technical Memorandum 104566, Volume 29, NASA Goddard Space Flight Center, Maryland.

Antoine, D and A Morel, 1999. A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing one. *International Journal of Remote Sensing* **20**(9), 1875–1916.

Atkinson, P. M, M. E. J Cutler, and H. G Lewis, 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing* **18**(4), 917–935.

Babey, S. K and C. D Anger, 1989. A compact airborne spectrographic imager (CASI). In *Proceedings of the IEEE International Geoscience Remote Sensing*

Symposium '89, pp. 1028–1031. IEEE.

Bagheri, S, M Stein, and R Dios, 1998. Utilty of hyperspectral data for bathymetric mapping in a turbid estuary. *International Journal of Remote Sensing* **19**(6), 1179–1188.

Barne, J. H, C. F Robson, S. S Kaznowska, J. P Doody, and N. C Davidson (Eds.), 1995. *Eastern England: Flanborough Head to Great Yarmouth*, Volume Region 6 of *Coasts and seas of the United Kingdom*. Peterborough Joint Nature Conservation Committee.

Benediktsson, J. A and J. R Sveinsson, 1997. Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing* **18**(4), 727–740.

Benediktsson, J. A, J. R Sveinsson, and K Arnason, 1995. Classification and feature extraction of AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing* **33**(5), 1194–1205.

Benediktsson, J. A, P. H Swain, and O. K Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing* **28**(4), 540–550.

Bishop, C. M, 1995. *Neural Networks for Pattern Recognition*. Oxford University Press.

Bizzi, S, O Arino, and P Goryl, 1996. Operational algorithm to correct the along track and across track striping in the JERS-1 OPS images. *International Journal of Remote Sensing* **17**(10), 1963–1968.

Bolgrien, D. W, R. C Wrigley, R. A Armstrong, and A. S Brooks, 1995. Absorption spectra for chlorophyll, particles and dissolved organic carbon in Green Bay, Lake Michigan. In Environmental Research Institute of Michigan (1995), pp. 163–168.

Bricaud, A, A Morel, and V Barale, 1995. MERIS potential for ocean colour studies in the open ocean. In P. J Curran and Y. C Robertson (Eds.), *Remote Sensing in Action*, pp. 133–140. Remote Sensing Society: RSS, Nottingham.

Brown, D. E, P. J Shaw, S Vittone, and K Weise, 1994. A comparison of inductive

modeling techniques for pediatric decision making. In *IEEE International conference on systems, man and cybernetics. Humans, information and technology.*, Volume 1, pp. 919–924. IEEE.

Buckton, D, E O'Mongain, and S Danaher, 1999. The use of Neural Networks for the estimation of oceanic constituents based on the MERIS instrument. *International Journal of Remote Sensing* **20**(9), 1841–1851.

Bukata, R. P, G. P Harris, and J. E Bruton, 1974. The detection of suspended solids and chlorophyll *a* utilizing multispectral ERTS-1 data. In *Proceedings of the 2nd Canadian Symposium on Remote Sensing*, pp. 551–564.

Bukata, R. P, J. H Jerome, K. Y Kondratyev, and D. V Pozdnyakov, 1991. Satellite monitoring of optically-active components of inland waters: an essential input to regional climate change impact studies. *Journal of Great Lakes Research* **17**(4), 470–478.

Campbell, J. W and W. E Esaias, 1985. Spatial patterns of temperature and chlorophyll in Nantucket Shoals from airborne remote sensing data, May 7-9, 1981. *Journal of Marine Research* **43**, 139–161.

Carder, K. L, S. K Hawes, K. A Baker, R. C Smith, R. G Steward, and B. G Mitchell, 1991. Reflectance model for quantifying chlorophyll *a* in the presence of productivity degredation products. *Journal of Geophysical Research* **96**(C11), 20599–20611.

Carder, K. L, R. G Steward, G. R Harvey, and P. B Ortner, 1989. Marine humic and fulvic acids: Their effects in remote sensing of ocean chlorophyll. *Limnology and Oceanography* **34**(1), 68–81.

Civco, D. L, 1993. Artificial neural network for land-cover classification and mapping. *International Journal of Remote Sensing* **7**(2), 173–186.

Civco, D. L and Y Waug, 1994. Classification of multispectral, multisource spatial data using artificial neural networks. In *Proceedings of the 1994 Annual ASPRS/ACSM Convention*, Volume I, pp. 123–133. American Society for Photogrammetry and Remote Sensing/American Congress on Surveying and Map-

ping.

Clark, C and A Cañas, 1995. Spectral identification by artificial neural network and genetic algorithm. *International Journal of Remote Sensing* **16**(12), 2255–2275.

Clarke, G. L, G. C Ewing, and C. J Lorenzen, 1970. Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration. *Science* **167**, 1119–1121.

Comrey, A. L, 1973. *A first course in factor analysis.* Academic Press.

Cooley, W. W and P. R Lohnes, 1971. *Multivariate Data Analysis.* John Wiley and Sons.

Cosandier, D, T Ivanco, and S Mah, 1992. The geocorrection and integration of the global positioning system with the Compact Airborne Spectrographic Imager. In *Proceedings of the 15th Canadian Symposium on Remote Sensing*, Volume 1, pp. 385–390.

Cover, T. M, 1974. The best two independent measurements are not the two best. *IEEE Transactions on Systems, Man and Cybernetics* **SMC-4**(1), 116–117.

Cracknell, A. P, 1999. Remote sensing techniques in estuaries and coastal zones - an update. *International Journal of Remote Sensing* **19**(3), 485–496.

Csillag, F, L Pásztor, and L. L Biehl, 1993. Spectral band selection for the characterization of salinity status of soils. *Remote Sensing of Environment* **43**, 231–242.

Cunningham, A, 1996. Variability of *in vivo* chlorophyll fluorescence and its implications for instrument development in bio-optical oceanography. *Scientia Marina* **60**(Suppliment 1), 309–315.

Davis, C. S, G. R Flierl, P. H Wiebe, and P. J. S Franks, 1991. Micropatchiness, turbulence and recruitment in plankton. *Journal of Plankton Research* **49**, 109–151.

Dekker, A. G, 1993. *Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing.* Ph. D. thesis, Free University, Amsterdam.

Dekker, A. G, T. J Malthus, and E Seyham, 1991. Quantitative modeling of inland

water quality for high resolution MSS systems. *IEEE Transactions on Geoscience and Remote Sensing* **29**(1), 89–95.

Dekker, A. G, T. J Malthus, and M. M Wijnen, 1992b. Spectral band location for remote sensing of turbid and/or eutrophic waters. In *Proceedings of the First Thematic Conference on Remote Sensing for Marine and Coastal Environments: Emerging Technologies and Systems*, Volume 2, pp. 955–971. Environmental Research Institute of Michigan: ERIM, USA.

Dekker, A. G, T. J Malthus, M. M Wijnen, and E Seyhan, 1992a. The effect of spectral bandwidth and positioning on the spectral signature analysis of inland waters. *Remote Sensing of Environment* **41**, 211–225.

Denman, K. L, 1976. Covariability of chlorophyll and temperature in the sea. *Deep-Sea Research* **23**, 539–550.

Denman, K. L and T Platt, 1976. The variance spectrum of phytoplankton in a turbulent ocean. *Journal of Marine Research* **34**, 593–601.

Doerffer, R, 1981. *Factor analysis on ocean colour interpretation*, pp. 339–345. Volume 13 of Gower (1981).

Doerffer, R and J Fischer, 1994. Concentrations of chlorophyll, suspended matter and gelbstoff in case II waters derived from satellite Coastal Zone Colour Scanner data with inverse modelling methods. *Journal of Geophysical Research* **99**, 7457–7466.

Doerffer, R, J Fischer, M Stösel, C Brockmann, and H Grassl, 1989. Analysis of Thematic Mapper data for studying the suspended matter distribution in the coastal area of the German Bight (North Sea). *Remote Sensing of Environment* **28**, 61–73.

Doerffer, R, K Sørensen, and J Aiken, 1999. MERIS potential for coastal zone applications. *International Journal of Remote Sensing* **20**(9), 1809–1818.

Downing, J. P, M Meybeck, J. C Orr, R. R Twilley, and H.-W Scharpenseel, 1993. Land and water interface zones. *Water, Air and Soil Pollution* **70**, 123–137. Published as Wisniewski, J and Sampson, R N, 1993. *Terrestrial Biospheric*

Carbon Fluxes: Quantification of Sinks and Sources of CO_2. Kluwer Academic Press.

Dyer, K. R and T. J Moffat, 1998. Fluxes of suspended matter in the East Anglian plume Southern North Sea. *Continental Shelf Research* **18**(11), 1311–1331.

Eklundh, L and A Singh, 1993. A comparative analysis of standardised and unstandardised principal component analysis in remote sensing. *International Journal of Remote Sensing* **14**(7), 1359–1370.

Elashoff, J. D, R. M Elashoff, and G. E Goldman, 1967. On the choice of variables in classification problems with dichtomous variables. *Biometrika* **54**, 668–670.

Environment Agency, 1997. Calibration of CASI imagery for high concentrations of chlorophyll-*a* in turbid waters. Technical report, National Centre for Environmental Data and Surveillance.

Environmental Research Institute of Michigan, 1995. *Proceedings of the Third Thematic Conference on Remote Sensing for Marine and Coastal Environments: Technology and Applications.* Environmental Research Institute of Michigan: ERIM, USA.

Estrada, M, C Marrasé, and J Salat, 1996. *In vivo* fluorescence/chlorophyll *a* ratio as an ecological indicator in oceanography. *Scientia Marina* **60**(Suppliment 1), 317–325.

Farrington, G. A, H. J Hoogenboom, A. G Dekker, and T. J. M Malthus, 1994. Understanding spectral derivatives of water reflectance from high resolution airborne imagery. In *First International Airborne Remote Sensing Conference and Exhibition*, Volume III, pp. 46–55. Strasbourg, France, 11-15 September.

Ferrari, G. M, N Hoepffer, and M Mingazinni, 1996. Optical properties of the water in a deltaic environment: Prospective tool to analyse satellite data in turbid waters. *Remote Sensing of Environment* **58**, 69–80.

Fischer, J, 1985. On the information content of multispectral radiance measurements over an ocean. *International Journal of Remote Sensing* **6**(5), 773–786.

Fischer, J, R Doerffer, and H Graßl, 1986. Factor analysis of multispectral radiances over coastal and open ocean water based on radiative transfer calculations. *Applied Optics* **25**(3), 448–456.

Fischer, J and U Kronfeld, 1990. Sun-stimulated chlorophyll fluorescence 1: Influence of oceanic properties. *International Journal of Remote Sensing* **11**(12), 2125–2147.

Fischer, J and P Schlüssel, 1990. Sun-stimulated chlorophyll fluorescence 2: Impact of atmospheric properties. *International Journal of Remote Sensing* **11**(12), 2149–2162.

Fisher, P, 1997. The pixel: a snare and a delusion. *International Journal of Remote Sensing* **18**(3), 679–685.

FLIERS, 1999. Fuzzy Land Information from Environmental Remote Sensing final report. Eu environment and climate programme, topic 3.1.1. contract number env4-ct96-0305, Collabration of: The University of Leicester, UK; The University of Southampton, UK; Joint Research Centre, European Commission, Italy; VTT Automation, Espoo, Finland; The Aristotolian University of Thessaloniki, Greece.

Foody, G. M, 1996. Relating the land-cover composition of mixed pixels to artificial neural network classification output. *Photogrammetric Engineering and Remote Sensing* **62**(5), 491–499.

Foody, G. M, R. M Lucas, P. J Curran, and M Honzak, 1997. Non-linear mixture modelling without end-members using an artificial neural network. *International Journal of Remote Sensing* **18**(4), 937–953.

Foody, G. M, M. B McCulloch, and W. B Yates, 1995a. Classification of remotely-sensed data by an artificial neural- network - issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing* **61**, 391–401.

Foody, G. M, M. B McCulloch, and W. B Yates, 1995b. The effect of training set size and composition on artificial neural- network classification. *International Journal of Remote Sensing* **16**(9), 1707–1723.

Foschi, P. G and D. K Smith, 1997. Detecting subpixel woody vegetation in digital imagery using two artificial intelligence approaches. *Photogrammetric Engineering and Remote Sensing* **63**(5), 493–500.

Fraser, R. N, 1998a. Hyperspectral remote sensing of turbidity and chlorophyll *a* amoung Nebraska Sand Hills Lakes. *International Journal of Remote Sensing* **19**(8), 1579–1589.

Fraser, R. N, 1998b. Multispectral remote sensing of turbidity among Nebraska Sand Hills lakes. *International Journal of Remote Sensing* **19**(15), 3011–3016.

Fraser, R. S, S Mattoo, E.-N Yeh, and C. R McClain, 1997. Algorithm for atmospheric and glint corrections of satellite measurements of ocean pigment. *Journal of Geophysical Research-Atmospheres* **102**(D14), 17107–17118.

Fukunaga, K and W. L. G Koontz, 1970. Application of the Karhunen-Loève expansion of feature selection and ordering. *IEEE Transactions on Computers* **C-19**(4), 311–318.

Giannini, L, 1981. *Differential spectroscopy for the coastal water quality identification by remote sensing*, pp. 395–402. Volume 13 of Gower (1981).

Gieskes, W. W and G. W Kraay, 1983. Unknown chlorophyll *a* derivatives in the North Sea and the tropical Atlantic Ocean revealed by HPLC analysis. *Limnology and Oceanography* **28**(4), 757–766.

Gitelson, A, 1992. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing* **13**(17), 3367–3373.

Gitelson, A, M Mayo, Y. Z Yacobi, A Parparov, and T Berman, 1994. The use of high-spectral-resolution radiometer data for detection of low chlorophyll concentrations in Lake Kinneret. *Journal of Plankton Research* **16**(8), 993–1002.

Goodin, D. G, L Han, R. N Fraser, D. C Rundquist, W. A Stebbins, and J. F Schalles, 1993. Analysis of suspended solids in water using remotely sensed high resolution derivative spectra. *Photogrammetric Engineering and Remote Sensing* **59**(4), 505–510.

Gordon, H. R, 1981. *A preliminary assessment of the Nimbus-7 CZCS atmospheric correction algorithm in a horizontally inhomogeneous atmosphere*, pp. 257–265. Volume 13 of Gower (1981).

Gordon, H. R, 1997. Atmospheric correction of ocean color imagery in the earth observing system era. *Journal of Geophysical Research-Atmospheres* **102**(D14), 17081–17106.

Gordon, H. R, D. H Clark, J. L Mueller, and W. A Hovis, 1980. Phytoplankton pigments from the nimbus-7 coastal zone color scanner: Comparisons with surface measurements. *Science* **210**, 63–66.

Gordon, H. R, D. K Clark, J. W Brown, O. B Brown, and R. S Evans, 1982. Satellite measurement of the phytoplankton pigment concentration in the surface waters of a warm core Gulf Stream ring. *Journal of Marine Research* **40**, 491–502.

Gordon, H. R, T Du, and T Zhang, 1997. Remote sensing of ocean color and aerosol properties: resolving the issue of aerosol absorption. *Applied Optics* **36**(33), 8670–8684.

Gordon, H. R and A. Y Morel, 1983. *Remote assessment of ocean colour for interpretation of satellite visible imagery: A review.* Number 4 in Lecture Notes on Coastal and Estuarine Studies. Springer-Verlag, New York.

Gould, R. W and R. A Arnone, 1997. Remote sensing estimates of inherent optical properties in a coastal environment. *Remote Sensing of Environment* **61**, 290–301.

Gower, J. F. R (Ed.), 1981. *Oceanography from Space*, Volume 13 of *Marine Science*. Plenum Press.

Gower, J. F. R and G Borstad, 1981. *Use of the in vivo fluorescence line at 685 nm for remote sensing surveys of surface chlorophyll a*, pp. 329–338. Volume 13 of Gower (1981).

Gower, J. F. R, K. L Denman, and R. J Holyer, 1980. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. *Nature* **288**, 157–159.

Gower, J. F. R, R Doerffer, and G. A Borstad, 1999. Interpretation of the 685 nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *International Journal of Remote Sensing* **20**(9), 1771–1786.

Gower, J. F. R, S Lin, and G. A Borstad, 1984. The information content of different optical spectral ranges for remote chlorophyll estimation in coastal waters. *International Journal of Remote Sensing* **5**(2), 349–364.

Great Britain Hydrographic Department, 1996. *Admiralty tide tables: European waters including the Mediterranean Sea*, Volume 1. Hydrographer of the Navy.

Great Britain Ministry of Defence Navy Hydrographic Department, 1976. *Tidal stream atlas: North Sea southern portion*, Volume NP 251. Hydrographic Department, Taunton.

Green, A. A, M Berman, P Switzer, and M. D Craig, 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing* **26**(1), 65–74.

Gross, L, S Thiria, and R Frouin, 1999. Applying artificial neural network methodology to ocean color remote sensong. *Ecological Modelling* **120**, 237–246.

Gross, L, S Thiria, R Frouin, and B. G Mitchell, 2000. Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *Journal of Geophysical Research* **105**(C2), 3483–3495.

Han, L, 1997. Spectral reflectance with varying suspended sediment concentration in clear and algae-laden waters. *Photogrammetric Engineering and Remote Sensing* **63**(6), 701–705.

Han, L and D. C Rundquist, 1994. The reponse of both surface reflectance and the underwater light field to various levels of suspended sediments: Preliminary results. *Photogrammetric Engineering and Remote Sensing* **60**(12), 1463–1471.

Han, L, D. C Rundquist, L. L Liu, R. N Fraser, and J. F Schalles, 1994. The spectral response of algal chlorophyll in water with varying levels of suspended sediment. *International Journal of Remote Sensing* **15**(18), 3707–3718.

Hanelt, D, 1996. Photoinhibition of photosynthesis in marine macroalgae. *Scientia Marina* **60**(Suppliment 1), 243–248.

Haykin, S, 1994. *Neural Networks: A Comprehensive Foundation.* Macmillan, New York.

Heermann, P. D and N Khazenie, 1992. Classification of multispectral remote sensing data using a back propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing* **30**(1), 81–88.

Hepner, G. F, T Logan, N Ritter, and N Bryant, 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing* **56**(4), 469–473.

Hill, C, 1998. The validation of TSM and chlorophyll a semi-empirical, universal, coastal colour algorithms; in the Bristol Channel and Norfolk Wash Coast. Third year dissertation, University of Southampton.

Holm-Hansen, O, C. J Lorenzen, R. W Holmes, and J. D. H Strickland, 1965. Fluorometric determination of chlorophyll. *Journal du Conseil. Conseil Permanent International pour l'Exploration de la Mer* **30**(1), 3–15.

Hoogenboom, H. J, A. G Dekker, and I. A Althuis, 1998. Simulation of AVIRIS sensitivity for detecting chlorophyll over coastal and inland waters. *Remote Sensing of Environment* **65**(3), 333–340.

Justice, C. O, B. L Markham, S. R. G Townshend, and Kennard, 1989. Spatial degredation of satellite data. *International Journal of Remote Sensing* **10**(9), 1539–1561.

Karhunen, J, E Oja, L Wang, R Vigário, and J Joutsensalo, 1997. A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks* **8**(3), 486–504.

Keiner, L. E and C. W Brown, 1999. Estimating oceanic chlorophyll concentrations with neural networks. *International Journal of Remote Sensing* **20**(1), 189–194.

Keiner, L. E and X.-H Yan, 1998. A neural network model for estimating sea surface

chlorophyll and sediments form Thematic Mapper imagery. *Remote Sensing of Environment* **66**(2), 153–165.

Keller, I and J Fischer, 1998. Details and improvements of the radiometric calibration procedure of the Compact Airborne Spectrographic Imager (CASI). In Schaepman, Schläpfer, and Itten (1998), pp. 81–88.

Kirk, J. T. O, 1994. *Light and Photosynthesis in Aquatic Ecosystems* (Second ed.). Cambridge University Press.

Kondratyev, K. Y, D. V Pozdnyakov, and L. H Pettersson, 1998. Water quality remote sensing in the visible spectrum. *International Journal of Remote Sensing* **19**(5), 957–979.

Kutser, T, H Arst, T Miller, L Käärmann, and A Millius, 1995. Telespectrometrical estimation of water transparency, chlorophyll-*a* and total phosphorus concentration and Lake Peipsi. *International Journal of Remote Sensing* **16**(16), 3069–3085.

Land, P. E and J. D Haigh, 1996. Atmospheric correction over case 2 waters with an iterative fitting algorithm. *Applied Optics* **35**(27), 5443–5451.

Lathrop, R. G and T. M Lillesand, 1986. Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan. *Photogrammetric Engineering and Remote Sensing* **52**(5), 671–680.

Lazar, M, Z Ben-Avraham, and E Ben-Dor, 1998. Comprehensive comparison of atmospheric corrections of "CASI" hyperspectral images over water - a case study. In Schaepman, Schläpfer, and Itten (1998), pp. 97–103.

Lee, J. B, A. S Woodyatt, and M Berman, 1990. Enhancement of high spectral resolution remote-sensing data by noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing* **28**(3), 295–304.

Lee, Z, K. L Carder, S. K Hawes, R. G Steward, T. G Peacock, and C. O Davis, 1994. Model for the interpretation of hyperspectral remote-sensing reflectance. *Applied Optics* **33**(24), 5721–5732.

Lewis, H, 1997. FLIERS, Fuzzy Land Information from Environmental Remote Sensing: Project details. `http://www.geog.le.ac.uk/fliers`.

Lewis, H. G, M Brown, A. R. L Tatnall, M Nixon, and J Manslow, 1998. Data analysis and empirical classification in FLIERS. Technical Report ISIS-3-98, Department of Electronics and Computer Science, University of Southampton.

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

Mackenzie, F. T, J. M Bewers, R. J Charlson, E. E Hofmann, G. A Knauer, J. C Kraft, E.-M Nöthig, B Quack, J. J Walsh, M Whitfield, and R Wollast, 1991. *What is the importance of ocean margin processes in global change?*, pp. 433–454. In Mantoura, Martin, and Wollast (1991).

Malki, H. A and A Moghaddomjoo, 1991. Using the Karhunen-Loève transformation in the back-propagation algorithm. *IEEE Transactions on Neural Networks* **2**(1), 162–165.

Malthus, T, 1997. An introduction to underwater light processes. In A. P Cracknell and E. S Rowan (Eds.), *Physical processes in the coastal zone: computer modelling and remote sensing. Proceedings of the forty ninth Scottish Universities Summer School in Physics*, pp. 233–247. Scottish Univerisities Summer School in Physics: SUSSP Publications, Edinburgh and Institute of Physics Publishing, Bristol.

Malthus, T. J and A. G Dekker, 1995. First derivative indices for the remote sensing of inland water quality using high spectral resolution reflectance. *Environment International* **21**(2), 221–232.

Malthus, T. J, C. J Place, S Bennet, and S North, 1996. An evaluation of the Airborne Thematic Mapper sensor for monitoring inland waters. In D. N. M Donaghue and Y Zong (Eds.), *Remote Sensing Science and Industry*, pp. 317–324. Remote Sensing Society: RSS, Nottingham.

Mantoura, R. F. C, J.-M Martin, and R Wollast (Eds.), 1991. *Ocean margin processes in global change*. Wiley.

Maritorena, S, J O'Reilly, and B. D Schieber, 1997. The SeaBAM evaluation data set. Technical memo, NASA, Goddard Space Flight Center, `http://seabass.gsfc.nasa.gov/seabam/bioopt_workshop.html`.

Mather, P. M, 1996. *Computer processing of remotely sensed images. An introduction.* John Wiley and Sons.

Matthews, A, 1994. *High Resolution Spectral Remote Sensing of Phytoplankton in the Coastal Zone.* Ph. D. thesis, Department of Oceanography, University of Southampton.

Matthews, A. M and S. R Boxall, 1994. Novel algorithms for the determination of phytoplankton concentration and maturity. In *Proceedings of the Second Thematic Conference on Remote Sensing for Marine and Coastal Environments*, Volume I, pp. 173–180. Environmental Research Institute of Michigan: ERIM, USA.

Mertes, L. A. K, M. O Smith, and J. B Adams, 1993. Estimating suspended sediment concentrations in surface waters of the amazon river wetlands from landsat images. *Remote Sensing of Environment* **43**, 281–301.

Miller, D. M, E. J Kaminsky, and S Rana, 1995. Neural-network classification of remote-sensing data. *Computers and Geosciences* **21**(3), 377–386.

Miller, R. L, M Giardino, B. A McKee, J. F Cruise, G Booth, R Rovansek, D Muirhead, W Cibula, K Holladay, R. E Pelletier, W Hudnall, C Bergeron, J Ioup, G loup, and G Love, 1995. Processes and fates of sediments and carbon in Baratana Bay, LA. In Environmental Research Institute of Michigan (1995), pp. 233–244.

Mitchell, J. F. B, C. A Wilson, and W. M Cunnington, 1987. On CO_2 climate sensitivity and model dependence of results. *Quarterly Journal of the Meteological Society* **113**, 293–322.

Mitchelson, E. G, N. J Jacob, and J. H Simpson, 1986. Ocean colour algorithms from the Case 2 waters of the Irish Sea in comparison to algorithms from Case 1 waters. *Continental Shelf Research* **5**(3), 403–415.

Mittenzwey, K.-H, S Ullrich, A. A Gitelson, and K. Y Kondratiev, 1992. Deter-

mination of chlorophyll *a* of inland waters on the basis of spectral reflectance. *Limnology and Oceanography* **37**(1), 147–149.

Moore, G. K, 1978. Satellite surveillance of physical water-quality characteristics. In J. J Cook (Ed.), *Proceedings, 12th International Symposium on Remote Sensing of Environment*, pp. 445–462. Center of Remote Sensing Information and Analysis: Environmental Research Institute of Michigan.

Morel, A, 1980. In-water and remote measurements of ocean color. *Boundary-Layer Meteorology* **18**(2), 177–201.

Morel, A, 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 waters). *Journal of Geophysical Research* **93**(c9), 10749–10768.

Morel, A and L Prieur, 1977. Analysis of variations in ocean color. *Limnology and Oceanography* **22**(4), 709–722.

Morel, A. Y and H. R Gordon, 1980. Report of the working group on water color. *Boundary-Layer Meteorology* **18**, 343–355.

Myers, R. H, 1986. *Classical and modern regression with applications* (Second ed.). PWS-KENT.

NASA, 1982. The marine resources experiment program (MAREX). Technical Report 87368, NASA/Goddard Space Flight Center, Greenbelt, Maryland.

Neville, R. A and J. F. R Gower, 1977. Passive remote sensing of phytoplankton via chlorophyll a fluorescence. *Journal of Geophysical Research* **82**(24), 3847–3493.

Nieke, B, W. F Vincent, J.-C Therriault, L Legendre, J.-F Berthon, and A Condal, 1997. Use of a ship-borne laser fluorosensor for remote sensing of chlorophyll *a* in a coastal environment. *Remote Sensing of Environment* **60**, 140–152.

Novo, E. M, J. D Hanson, and P. J Curran, 1989. The effect of sediment type on the relationship between the reflectance and suspended sediment concentration. *International Journal of Remote Sensing* **10**(7), 1283–1289.

Otsu, N, 1984. Karhunen-Loeve line fitting and a linearity measure. In *Proceedings of the 7th International Conference on Pattern Recognition*, pp. 486–489. IEEE.

Paola, J. D and R. A Schowengerdt, 1995a. A detailed comparison of backpropagation neural-network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing* **33**, 981–996.

Paola, J. D and R. A Schowengerdt, 1995b. A review and analysis of backpropagation neural networks for classification of remotely-sensed multispectral imagery. *International Journal of Remote Sensing* **16**(16), 3033–3058.

Pérez-Ruzafa, A, J Gilabert, A Bel-Lan, V Moreno, and J. M Gutiérrez, 1996. New approach to chlorophyll *a* determination in shallow coastal waters by remote sensing. *Scientia Marina* **60**(Suppliment 1), 19–27.

Peters, S. W. M, A. G Dekker, and P Keller, 1998. From end-user requirements to sensor requirements. In Schaepman, Schläpfer, and Itten (1998), pp. 215–219.

Piramuthu, S, 1996. Feature selection and neuro-fuzzy systems. *Neural Network World* **6**(2), 201–208.

Price, J. C, 1990. On the information content of soil reflectance spectra. *Remote Sensing of Environment* **33**, 113–121.

Prieur, L and S Sathyendranath, 1981. An optical classification of coast and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnology and Oceanography* **26**(4), 671–689.

Quenzel, H and M Kaestner, 1981. *Masking effect of the atmosphere in remote sensing of chlorophyll*, pp. 365–370. Volume 13 of Gower (1981).

Quibell, G, 1991. The effect of suspended sediment on reflectance from freshwater algae. *International Journal of Remote Sensing* **12**(1), 177–182.

Rast, M, L Bézy, and S Bruzzi, 1999. The ESA Medium Resolution Imaging Spectrometer MERIS - a review of the instrument and its mission. *International Journal of Remote Sensing* **20**(9), 1681–1702.

Reinerman, P. N and K. L Carder, 1995. Monte Carlo simulation of the atmospheric point-spread function with an application to correction for the adjacency effect.

Applied Optics **34**(21), 4453–4471.

Richards, J. A, 1995. *Remote sensing digital image analysis. An introduction.* (Second ed.). Springer-Verlag.

Roger, R. E, 1996. Principal Components transform with simple, automatic noise adjustment. *International Journal of Remote Sensing* **17**(14), 2719–2727.

Rumelhart, D. E, G. E Hinton, and R. J Williams, 1986. Learning representations by back-propogating errors. *Nature* **323**, 533–536.

Rumelhart, D. E and J. L McClelland (Eds.), 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*, Volume 1: Foundations. MIT Press.

Rundquist, D. C, L Han, J. F Schalles, and J. S Peake, 1996. Remote measurement of algal chlorophyll in surface waters: The case for the first derivative of reflectance near 690 nm. *Photogrammetric Engineering and Remote Sensing* **62**(2), 195–200.

Sager, G and R Sammler, 1968. *Atlas der Gezeitenströme für die Nordsee, den Kanal and die Irische See.* Seehydrographischer Dienst der DDR. (summarised in Barne *et al.* 1995).

Sathyendranath, S, L Prieur, and A Morel, 1989. A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *International Journal of Remote Sensing* **10**, 1373–1394.

Schaepman, M, D Schläpfer, and K. I Itten (Eds.), 1998. *Proceedings of the First EARSel Workshop on Imaging Spectroscopy.* European Association of Remote Sensing Laboratories.

Schiller, H and R Doerffer, 1999. Neural network for emulation of an inverse model - operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing* **20**(9), 1735–1746.

SCOR-Unesco, 1966. Determination of photosynthetic pigments in sea-water. Monographs on oceanographic methodology 1, SCOR-Unesco.

Sietsma, J and R. J. F Dow, 1991. Creating artificial nerual networks that generalize. *Neural Networks* **4**, 67–79.

Singh, A and A Harrison, 1985. Standardised principal components. *International Journal of Remote Sensing* **6**(6), 883–896.

Singh, R. P, C Olbert, C Lindermann, M Schaale, and R Furrer, 1997. Atmospheric monitoring with a spectrographic imager. *International Journal of Remote Sensing* **18**(5), 1183–1188.

Smetacek, V, U Bathmann, E.-M Nöthig, and R Scharek, 1991. *Coastal eutrophication: Causes and consequences*, pp. 251–279. In Mantoura, Martin, and Wollast (1991).

Steele, J. H and T. W Henderson, 1979. Spatial patterns in North Sea plankton. *Deep Sea Research* **26A**, 955–963.

Steppe, J. M, K. W Bauer, and S. K Rogers, 1996. Intergrated feature and architecture selection. *IEEE Transactions on Neural Networks* **7**(4), 1007–1014.

Strutton, P. G, J. G Mitchell, and J. S Parslow, 1996. Non-linear analysis of chlorophyll a transects as a method of quantifying spatial structure. *Journal of Plankton Research* **18**(9), 1717–1726.

Strutton, P. G, J. G Mitchell, and J. S Parslow, 1997. Using non-linear analysis to compare the spatial structure of chlorophyll with passive tracers. *Journal of Plankton Research* **19**(10), 1553–1564.

Sturm, B, 1981. *Ocean colour remote sensing and quantitative retrieval of surface chlorophyll in coastal waters using Nimbus CZCS data*, pp. 267–279. Volume 13 of Gower (1981).

Talcott, J. C, 1995. Remotely sensed data as applied in the coastal water predictive visibility model. In Environmental Research Institute of Michigan (1995), pp. 273–285.

Tanis, F. J and D. V Pozdnyakov, 1995. Evaluation of proposed coastal ocean colour algorithms for the retrieval of bio-pigments and suspended sediment in the Great Lakes. In Environmental Research Institute of Michigan (1995), pp. 140–150.

Tassan, S, 1981. *The influence of wind in the remote sensing of chlorophyll in the sea*, pp. 371–375. Volume 13 of Gower (1981).

Tassan, S and M Ribera d'Alcalá, 1993. Water quality monitors by Thematic Mapper in coastal environments. A performance analysis of local bio-optical algorithms and atmospheric correction procedures. *Remote Sensing of Environment* **45**, 177–191.

Taylor, J. E and R. C Smith, 1967. Spectroradiometric characteristics of natural light under water. *Journal of the Optical Society of America* **57**(5), 595–601.

Townshend, J. R. G, 1984. Agricultural land-cover discrimination using thematic mapper spectral bands. *International Journal of Remote Sensing* **5**(4), 681–698.

Warner, T. A and M. C Shank, 1997a. An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks. *Photogrammetric Engineering and Remote Sensing* **63**(11), 1285–1294.

Warner, T. A and M. C Shank, 1997b. Spatial autocorrelation analysis of hyperspectral imagery for feature selection. *Remote Sensing of Environment* **60**, 58–70.

Warner, T. A, K Steinmaus, and H Foote, 1999. An evaluation of spatial autocorrelation feature selection. *International Journal of Remote Sensing* **20**(8), 1601–1616.

Wernand, M. R, S. J Shimwell, and J. C De Munck, 1997. A simple method of full spectrum reconstruction by a five-bands approach for ocean colour applications. *International Journal of Remote Sensing* **18**(9), 1977–1986.

White, H, 1992. *Artificial Neural Networks. Approximation and learning theory.* Blackwell.

Williams, J. J, J. D Humphery, P. J Hardcastle, and D. J Wilson, 1998. Field observations of hydrodynamic conditions and suspended particulate matter in the southern North Sea. *Continental Shelf Research* **18**(11), 1215–1233.

Wollast, R, 1991. *The coastal organic carbon cycle: fluxes, sources and sinks*, pp. 365–381. In Mantoura, Martin, and Wollast (1991).

Wollast, R, 1993. *Interactions in estuaries and coastal waters*, Chapter 14. Wiley.

Wrigley, R. C, R. E Slye, S. A Klooster, R. S Freedman, M Carle, and L. F McGregor, 1992. The airborne ocean color imager - system description and image- processing.

Journal of Imaging Science Technology **36**(5), 423–430.

Yentsch, C. S and D. W Menzel, 1963. A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. *Deep-Sea Research* **10**, 221–231.