Investigating the feasibility of using neural networks to derive chlorophyll *a* prediction algorithms for case 2 waters

Isabel Sargent

April 13, 1999

Executive Summary

The measurement of chlorophyll *a* in the coastal zone of the United Kingdom is necessary for monitoring the state of the aquatic environment. Multispectral aerial imagery provides a wealth of information about this region but the algorithms that have been developed for estimating chlorophyll *a* have been found to be useful only for very local regions over short periods of time.

This study investigated the ability of neural networks to develop algorithms for chlorophyll *a* prediction and found:

- The neural network algorithms were significantly more successful than linear regression algorithms at predicting chlorophyll *a* in the same region as they were trained
- The relationship between water-leaving spectra and chlorophyll *a* is non-linear
- The FLH feature has the most linear relationship to chlorophyll *a*
- The blue wavelengths have a very non-linear relationship to chlorophyll *a* but hold a great deal of information about the chlorophyll *a* in water
- Non-linear algorithms performed better when applied to a new site

Although the algorithms were not successful at prediction chlorophyll *a* for different times or places, this study has found that neural network techniques are likely to provide the best means of producing the non-linear multiple regression algorithms necessary for the accurate prediction of chlorophyll *a* in the coastal zone.

Acknowledgments

This research was undertaken with the support of Dr Adrian Tatnall and Dr Martin Brown who have been invaluable over the last two and a half years. I am also particularly grateful to Dr Hugh Lewis for his advice and encouragement and for writing the neural network software. Thanks finally to Dr Alison Matthews and Dr Rebecca Allen for their guidance through the Environment Agency data archive.

Software used in this study:

Matlab 5.0	mathematical calculations
ERDAS Imagine 8.3	image processing
FLIERS SFT2.0[†]	neural network

[†] Available from http://www.isis.ecs.soton.ac.uk/research/projects/fliers/software

List of Figures

List of Tables

Glossary

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	Process of adjusting the weights backwards through a neural network during training
CASI	Compact Airborne Spectrographic Imager
Feed-forward	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
MLP	Multilayer Perceptron
NIR	Near Infrared

1 Introduction

Accurate estimates of chlorophyll *a* concentration are required for the assessment of coastal waters which are potentially subject to eutrophication. The EC Urban Waste Water Treatment Directive (91/271/EEC) requires the definition of these waters and the implementation of improved sewage effluent discharged to these regions. The UK is also a signatory to the Oslo and Paris Commission (OSPAR) strategy on hazardous substances, eutrophication and radioactive substances. This initially requires the definition of all areas which are potentially subject to eutrophication and will subsequently require comprehensive studies to be carried out in these areas.

Traditional laboratory analysis of water samples was found to give inadequate spatial and temporal density of sampling. The Environment Agency has therefore investigated the use of aerial remote sensing to measure chlorophyll *a* concentrations in coastal waters. Studies showed that accurate algorithms could be developed for specific sites, but that transferral of these algorithms to other geographical areas or different seasons introduced errors (Environment Agency 1997). This was considered to be due to inability of the regression techniques used to account for differing constituents within the water body and differences in atmospheric conditions. The Agency is continuing this investigation using a number of different techniques, one of which is described in this report.

1.1 Chlorophyll *a* in coastal waters

The waters around the coast of the United Kingdom have been classed as Case 2 waters, that is, waters for which the spectral signature is strongly affected by dissolved organic matter (DOM) or suspended sediment, or both (Gordon and Morel 1983). In addition to this, 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).

Traditionally the ratio of blue to green reflectance has been used to empirically determine chlorophyll *a* concentration (Gordon and Morel 1983; Kirk 1994). This ratio is a measure of the amount of absorption by chlorophyll *a* in the blue wavelengths compared to the region of little absorption in the green. Algorithms using this ratio work well for waters unaffected by sediment, DOM or a non-uniform atmosphere but have been found to operate poorly in Case 2 waters (Sathyendranath *et al.* 1989; Bukata *et al.* 1991).

For this reason other band difference and band ratio algorithms have been developed. At longer wavelengths, DOM absorption is minimized (Taylor and Smith 1967; Dekker 1993). Also, Quibell (1991) found that the effect of suspended sediment on water-leaving radiance at these wavelengths is equal. This has prompted the development of algorithms which utilize reflectance in the green, red and near-infrared wavelengths (Rundquist *et al.* 1996; Hoogenboom *et al.* 1998).

Neville and Gower (1977) suggested that alternative algorithms to those using band ratios could be derived using the peak in measured spectral response at 685 nm. This peak is probably 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). Not only does this peak increase with an elevation of chlorophyll concentration, but several studies have also found that the peak appears to move to longer wavelengths as chlorophyll increases (Gitelson 1992; Matthews 1994). The peak can be quantified by measuring its height above a baseline which

is drawn between two wavelengths or as the sum of the reflectance in the region of the peak (Gitelson *et al.* 1994).

Using the fluorescence peak to predict chlorophyll *a* has been found to have a number of drawbacks, however. Particularly problematic is the fact that attenuation of light by water and the atmosphere is much stronger at longer wavelengths resulting in a very weak fluorescence signal (Bricaud *et al.* 1995). As a result, the detected fluorescence signal originates only in the top 2 meters of the water (compared to the top 5 meters for shorter wavelengths) and hence is not a good indicator of chlorophyll *a* in the majority of the euphotic zone (Fischer and Kronfeld 1990). Surface reflectance has also been found to strongly affect this signal (Neville and Gower 1977).

Often the single feature algorithms have been found to hold too little information for accurate chlorophyll *a* detection and a number of authors have proposed that multiband algorithms may provide better results (Aiken *et al.* 1995). Sathyendranath *et al.* (1989) used a model of ocean colour to find a five waveband group which held almost as much information as the whole spectrum (also Wernand *et al.* 1997). New sensors with higher spectral resolution and more spectral channels provide a greater potential for success with multiband algorithms (Bricaud *et al.* 1995; Doerffer *et al.* 1995).

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. This is due to the non-linear way in which the optical signatures of the water constituents, and the optical properties of the water itself, interact (Sathyendranath *et al.* 1989; Fischer and Kronfeld 1990). Ideally, a non-linear regression model should be fitted to the data but standard non-linear regression methods require *a priori* knowledge of the data being modelled.

Neural networks present a method by which a non-linear function may be optimally modelled without any advance knowledge of the data structure. The remainder of this chapter will briefly discuss the basis for choosing neural networks for predicting chlorophyll *a* concentration and describe how one type, the Multilayer Perceptron, is used to model regression functions.

1.2 Using neural networks for chlorophyll *a* prediction

Over the last decade, neural networks have shown a great deal of potential in remote sensing applications as scene classifiers. These applications use the spectra of each pixel to determine, usually, the land cover in the image (e.g. Hepner *et al.* 1990; Benediktsson *et al.* 1990; Heermann and Khazenie 1992; Gopal and Woodcock 1996). Clark and Cañas (1993 and 1995) used neural networks to distinguish mineral spectra from mixtures of spectra.

More recently, land cover classifications have been performed which used neural networks to distinguish the spectral classes within so-called 'mixed pixels'. Initially, these networks were trained using pure examples of the class spectra (Civco 1993; Warner and Shank 1997) but better results have been achieved if examples of mixed spectra are use to train the network (Foody 1996; Foody *et al.* 1997).

Building on the success of these land-based uses of neural networks to derive regression algorithms, Keiner and Yan (1998) and Keiner and Brown (1999) have used neural networks to estimate chlorophyll *a* in coastal waters. These studies report a much better prediction of chlorophyll *a* in water than with using linear multiple regression methods.

1.3 The Multilayer Perceptron

The multilayer perceptron (MLP) is the neural network architecture most commonly chosen for remote sensing applications. Three stages are involved in its use: training, allocation and testing (Foody and Arora 1996). Training involves presenting patterns of spectral information (the inputs) and the corresponding desired outputs (for example, land cover type or chlorophyll *a* concentration) to the network so that it may 'learn' the underlying transfer function. During the allocation stage a new set of inputs are presented to the network for it to predict the corresponding outputs. The testing stage measures the accuracy of this prediction in order that the reliability of the algorithm may be assessed.

A single-layer network and a two-layer network were chosen for this study. 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 $\mu g/l$. The following paragraphs briefly describe how such a neural network works. For a more complete explanation, refer to Bishop (1995) and Haykin (1994).

The perceptron may be visualized as two layers of nodes interconnected with a layer of weights. In the single-layer network, the input and output layers of nodes are interconnected with one layer of weights. The multilayer perceptron (figure 1) has more than the one layer of weights. In the two layer MLP, a hidden set of nodes (h in figure 1) lies between the two layers of weights (w_1 and w_2 in figure 1).

Observed values, such as spectral radiance values, are entered into the network at the input nodes. If two bands of spectral data were to be used for the regression, the MLP would have two input nodes. These values are then passed to the next layer of nodes by being multiplied by the interconnecting weights. At each node in the next layer, the sum of all the weighted inputs is calculated (indicated by $\sum$ in figure 1). In a two-layer network, if there are i input nodes

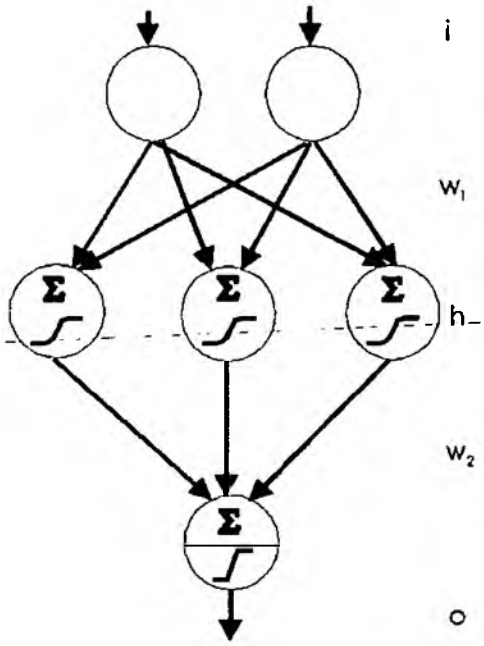


Figure 1: A two-layer MLP. i indicates the input nodes, in this case two features are being input to the network. w_1 is the first layer of weights by which the inputs are multiplied. The weighted inputs are then summed and passed over an activation function at h (in this case a sigmoid). The outputs from h are passed through the second layer of weights, w_2, and summed at the output node o. Finally, this value is passed through another activation function, in this case a linear transform, and the output value is the network's prediction of the values being determined.

and h hidden nodes, then each of the i inputs will be multiplied by h weights. Effectively, the i-dimensional input space is transformed to h-dimensional space.

In the case of the linear, one-layer network, the sum of the weighted inputs is then passed through an activation function. This may be either a linear function, if a continuous output (such as concentration) is desired, or a threshold function if a binary output (as in classification problems) is required. The resulting values are the output of the network.

For the two-layer case, the h inputs to the hidden set of nodes are scaled by an activation function. Usually this is a non-linear function such as a sigmoid or tanh. The scaled outputs of these nodes are then passed on to the next set of nodes via another set of weights, summed at each node and then passed through another set of activation functions to form the output to this network.

Initially, the network needs to be trained with patterns of known inputs and outputs. In this study, the network was trained with spectral features as inputs and measured chlorophyll *a* concentration as outputs. This is achieved by feeding the input information forward, as described above and then assessing the error in the output by comparing it with the measured values. The error can then be 'back-propagated' through the network and the weights changed according to the magnitude of the error. 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 multilayer perceptron has a very simple operation, however, its use can be complicated by the many parameters which may be adjusted to achieve a better performance. These parameters are necessary because the neural network does not make any *a priori* assumptions about the data. The most basic parameters are summarized in table 1. It has also been shown in some studies that the weights at the initialization of training can have a significant effect on the performance of the network (Paola and Schowengerdt 1997) and so a number of runs of each network may be necessary to ensure that the optimum performance has been achieved for that set of parameters.

Parameter	Possible values	Purpose
number of layers	1,2,3, ...	Control the complexity of the functions being modelled. One layer results in a linear function, increasing the number of layers results in increasingly complex, non-linear functions
number of inputs	1,2,3, ...	Can be any number of measurements, or combinations of measurements
number of hidden nodes	1,2,3, ...	Control the complexity of the functions being modelled
number of iterations	1, ... ,10, ... , 1000, ...	Generally, the training error decreases with the number of iterations. Too few iterations, and the network will not have converged well. However, if training continues for too long, the network may over-fit the function to the training data causing a loss of generalisation of the function
activation function	linear, sigmoid, tanh, threshold	Control the value of the output from each node
learning rate η	$0 < \eta < 1$	controls the magnitude of each weight change
momentum term μ	$0 < \mu < 1$	changes the effect of the learning rate depending on the change in error between iterations

Table 1: The main parameters used for setting up a neural network.

If all possible parameters were to be adjusted and tried in combination with all other variations of parameters then several thousand runs of the network would be required. For this reason, this study made a number of assumptions about the effect of changing each of the parameters based on a few preliminary trainings of the networks. Details of these assumptions and the training of the multilayer perceptron are given in the section 5.2.

2 Aims and objectives

The aim of this study is to investigate the potential for using neural networks to derive the concentration of chlorophyll *a* in coastal waters. The specific objectives are as follows:

1. To investigate the ability of neural network-derived algorithms to predict chlorophyll *a* from spectral information using a set of *in situ* chlorophyll *a* measurements as training data
2. To investigate the robustness of these algorithms in the same coastal region over time
3. To investigate the robustness of these algorithms in other locations around the UK coast

3 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, three locations were chosen for this study which will be referred to as *Norfolk 31/05/96*, *Norfolk 11/08/96* and *Holderness 19/08/96*.

The first set comprised three 72 band CASI images in *enhanced spectral mode* (see appendix A). These had been flown over the north Norfolk coast on 31st May 1996. Concurrent with these overflights, a cruise took underway fluorometer measurements and several samples of water along the cruise track. The water samples were then processed in the laboratory to determine the chlorophyll *a* concentration of the water in $\mu g/l$. These data were used to develop the algorithms as well as to test them.

The second data set, also of the north Norfolk Coast, had two 72 band CASI images flown on the 11th August 1996. Four coincident water samples were also taken and processed for chlorophyll *a* concentration. These data allowed the assessment of the validity of the algorithms over time.

The third data set contained two 72 band CASI images of the Holderness Coast for 19th August 1996. There were 17 water samples taken during the simultaneous cruise which were processed for chlorophyll *a* concentration. These data were used the assess the portability of the algorithms around the coast.

All three data sets conform to the quality standards set out by the Environment Agency which are detailed in (Environment Agency 1997). Table 2 summarizes the data sets.

Data set	Images	Number of fluorometer measurements	Number of laboratory samples
Norfolk 31/05/96	imag1875 imag1876 imag1877	3142	17
Norfolk 11/08/96	imag2334 imag2335	—	4
Holderness 19/08/96	imag2356 imag2369	—	17

Table 2: The three data sets used in this study

3.1 Atmospherically corrected data

Atmospherically corrected formats of the above images were considered for this study. These data were processed using the atmospheric correction algorithms developed by Plymouth Marine Laboratory as part of the COAST project (Aiken *et al.* 1995). This requires subsampling the 72 channel imagery to 15 channels of up to 25 nm width before applying the algorithm. It was found that the fluorescence peak around 685 nm became more prominent as a result of this correction. However, the data were noisy with many pixels having zero values in one or more bands. A previous study using this data had found a slightly reduced accuracy for chlorophyll

a algorithms developed from it (Environment Agency 1997). After some consideration it was decided that only the original imagery would be used for the study, since this would allow us to explore whether algorithms that did not require the extra stage of atmospheric correction could be developed.

4 Compiling the data sets for the study

To train and test the algorithms, sets of spectra and their corresponding chlorophyll *a* concentrations needed to be compiled. The fluorometer readings were to be used as the measure of chlorophyll *a* concentration for the *Norfolk 31/05/96* data. This required calibrating the fluorometer using the 17 laboratory samples as is described in section 4.1. All the data sets then had the spectral information extracted from the images at the site of each chlorophyll *a* data point (section 4.2 and section 4.3).

4.1 Calibration of the fluorometer data

The Through-flow Fluorometer measured solar-induced fluorescence. This needed to be calibrated to chlorophyll *a* in $\mu g/l$ 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 course of the cruise.

A fluorometer reading was found for each laboratory datum using the time fixes provided with the 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. The geographical position at which the water sample had been taken was then assumed to be the same as the corresponding fluorometer reading.

A number of different buffers were used around the calibration points to determine if a better correlation between measured chlorophyll *a* and fluorometer readings could be gained. Buffer sizes of 50, 100, 150, 200 and 250 metres were used (see appendix B) but it was found that increasing the buffer around the data point decreased the correlation. Therefore the fluorometer was calibrated using only those fluorometer readings that lie at the locations of the calibration data points.

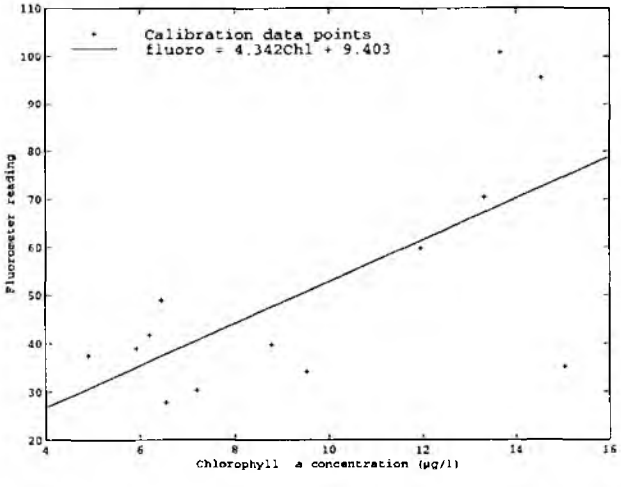


Figure 2: Calibration of the fluorometer data using the laboratory water samples. The calibration data set and the fluorometer data have a correlation coefficient of 0.6640 which, for 13 data points, is significant at the 10% level.

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

$$fluoro = 4.342Chla + 9.403 \quad (1)$$

was found where *fluoro* indicates the fluorometer reading and *Chla* indicates the chlorophyll *a*

concentration in $\mu g/l$. Inverted, this produces the calibration

$$Chla(\mu g/l) = 0.230 fluoro - 2.163 \quad (2)$$

as illustrated in figure 2. This relationship was used to calibrate all 3142 data points to chlorophyll *a* in $\mu g/l$. For the rest of the study, for the *Norfolk 31/05/96* data set, the fluorometer readings will be referred to as chlorophyll *a* measurements.

4.2 Extraction of spectral information

A number of preliminary tests were performed to determine the most appropriate method for extracting the reflectance values corresponding to each chlorophyll *a* measurement. Three possible methods were proposed:

1. Extract the reflectance value in the pixel nearest to the position of the fluorometer reading
2. Extract the median of each band within a window of a predetermined size
3. Extract the weighted mean of each band within a window of a predetermined size

It is usual to use some form of buffer when extracting information from images of water scenes because current and tidal effects cause uncertainty in the true location of the chlorophyll *a* measurement in the image. However, the study 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; Yoder *et al.* 1987).

Assessing the validity of various window sizes for the latter two methods was considered very computer-intensive and so the median pixel value and the weighted mean value was calculated for five bands and four window sizes only for image 1876. The weighted mean method used a two-dimensional Gaussian curve to weight the pixel values thus biasing the mean to the values at the centre of the window. The correlations for both methods and for each band and window size are shown on appendix C.

Generally as the window size increases, there is an increase in the correlation of chlorophyll *a* measurements with spectral data obtained from the window. This is consistent with the findings of the Environment Agency in their study using this data (Environment Agency 1997). There is also a slight improvement with the use of the weighted mean method. It was therefore determined that the spectral data to be used in the rest of this study would be extracted from the imagery using the weighted mean method and a window of 21 by 21 pixels - this is equivalent to about 200 meters on the ground.

4.2.1 Norfolk 31/05/96

Using the weighted mean method described above, the value of each band was calculated for each chlorophyll *a* measurement. The three sets of spectral values, corresponding to the three images, were then concatenated. Because the images overlap, some data points were removed by giving priority to the data which had a shorter time interval between the overflight and the *in situ* sampling. This resulted in pixels from image 1875 having a higher priority to those in image 1876 where they overlap, and image 1876 having priority over image 1877.

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 2354 chlorophyll *a* measurements (calibrated fluorometer readings) with corresponding spectral readings in each CASI band. These were randomly separated into three sets - a training set on which to create the algorithms, a validation set to be used in training the MLP and a testing set to test the accuracy of the algorithms. These sets are summarized in table 3.

4.2.2 Norfolk 11/08/96

Of the four chlorophyll *a* data points available for this site, only one was found to lie in the regions covered by the images. Therefore, each pixel in the images of this set was processed using the weighted mean method. The algorithms would be validated by applying them to the images as a whole and making a qualitative assessment of the resulting chlorophyll *a* images. The statistical properties of this data set are assumed to be similar to those of the four data samples for this location. The chlorophyll *a* concentration for this site is very low, and appears to have a very low range (table 3).

4.2.3 Holderness 19/08/96

Using the weighted mean method described above, the value of each band was calculated for each chlorophyll *a* measurement. Twelve of the 17 chlorophyll *a* points lay within the images. Again there was quite an overlap and so spectral data was chosen from the image whose flight time is nearest to that of each chlorophyll *a* sample. All samples were within two hours of the overflights. This data set also has low chlorophyll *a* values and a low range of values (table 3).

set	number of data points	Chlorophyll *a* $\mu g/l$ minimum	maximum	mean	standard deviation
Norfolk 31/05/96					
training	687	3.15	22.62	8.86	4.63
validation	484	3.14	22.16	9.02	4.76
testing	1183	2.91	22.60	8.98	4.73
Norfolk 11/08/96	*based on the four chlorophyll a measurements made at this location*				
testing	—	1.07	1.83	1.52	0.34
Holderness 19/08/96					
testing	12	0.71	3.15	1.77	0.78

Table 3: Statistics of the data sets which will be used to train and assess the algorithms.

4.3 Extracting a subset of spectral features

The volume of spectral information was reduced by picking out a number of spectral features from which to develop the algorithms. 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 4.3), their correlation as ratios with chlorophyll *a* (appendix E) in the training set, and their correlation with each other (appendix D) and also bands close to known features in the spectrum were considered. This resulted in a subset of eight bands (table 4) being extracted from the training, validation and testing sets.

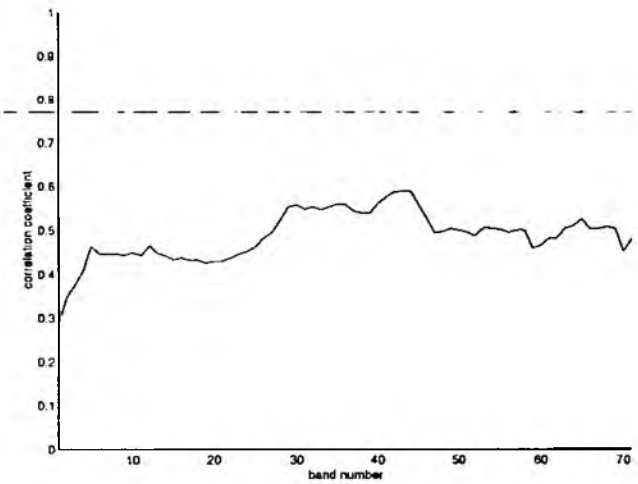


Figure 3: Correlation of bands with Chlorophyll *a* for the training set. With 687 data points, all correlations above ~0.1 may be considered significant. The highest correlations are for red wavelengths, especially around 690 nm. A trough in correlation is evident in the green wavelengths, bands 15 - 28. The lower correlation in bands 1 to 4 may be due to the high noise level evident here.

Band number	Wavelength † (*nm*)	Comments
6	441 (blue)	Close to chlorophyll *a* absorption peak and has one of the highest correlations with chlorophyll *a* in the blue.
20	540 (green)	Close to green reflectance peak and is 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 gives the highest correlation with chlorophyll *a* of whole training set.
29	604 (red-green)	Ratio of band 27 and 29 gives the highest correlation with chlorophyll *a* of 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		

Table 4: The subset of eight bands chosen for algorithm development. There are numerous methods available to allow an intelligent choice of spectral features. This method was based on characteristics of the training data set as well as on an understanding of chlorophyll *a* spectra.

The fluorescence feature near 685 nm was also extracted. The peak is not very clear in the CASI spectra but a drop in reflectance is almost always evident from band 40 (682 nm) to band 41 (689 nm). The fluorescence line height (FLH) feature is usually measured above a baseline described by two bands either side of the peak. Several baselines were tried and the correlation coefficient of the FLH measurement with chlorophyll *a* was used to determine which FLH measurement for

use in algorithm development (table 5).

Name of FLH measure	reference	Baseline wavelengths†		Baseline Bands		Correlation coefficient
Gitelson92	(Gitelson 1992)	675	730	39	47	0.7282
Gitelson94a	(Gitelson *et al.* 1994)	650	715	36	45	0.5308
Gitelson94b	(Gitelson *et al.* 1994)	670	730	38	47	0.7227
FischerK90a	(Fischer and Kronfeld 1990)	645	725	35	46	0.5646
FischerK90b	(Fischer and Kronfeld 1990)	645	670	35	38	0.6671

†see appendix A for full details

Table 5: Trials of different FLH measurements. Fluorescence height was measured at band 40 above a straight baseline described by two bands in the spectrum.

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

5 Training the algorithms

The algorithms were produced using the *Norfolk 31/05/96* training data with the testing data kept aside to test the algorithms. The neural network also required the validation set to determine the value of some of the parameters in its architecture.

5.1 Linear regression analysis

Using the subset of bands defined in section 4.2 several sets of band ratios were produced. A simple linear regression using the least squares method was produced for each ratio. This was then inverted to produce the algorithm which would predict chlorophyll *a* concentration from the ratios. The resulting linear agorithms are summarized in table 6 and illustrated in figure 4.

Feature	Description	Correlation with Chl *a*	Gradient	Intercept
blue-green ratio	$\frac{band\ 6}{band\ 20}$	0.1273	489.1263	-484.8864
red-NIR ratio	$\frac{band\ 41}{band\ 47}$	0.1608	584.2633	-746.2736
green-red ratio 1	$\frac{band\ 39}{band\ 20}$	0.4987	554.9866	-303.2480
green-red ratio 2	$\frac{band\ 29}{band\ 27}$	0.7698	480.7465	-439.2500
red-red ratio	$\frac{band\ 41}{band\ 39}$	0.7362	795.4905	-632.0558
FLH	see section 4.3	0.7282	0.9920	-19.3429

Table 6: The least-squares approximation of the relationship $feature = f(Chl)$ is calculated (this is illustrated in figure 4). This is then inverted to find the relationship in the form $Chl = Gradient * feature + Intercept$ which best predicts chlorophyll *a* from the spectral information from the training data.

5.2 Training the MLP

Training a multilayer perceptron can prove complicated because there are so many adjustments that can be made to the network to improve the training. A few preliminary trainings were performed and tested against the validation set (section 4.2) and assumptions were made on the outcomes as to the best activation functions and learning rate to be used throughout the course of this study. For the two-layer network, it was determined that a tanh activation function was to be used at the hidden nodes and a linear function used at the output nodes. For both types of network, the learning rate would remain 0.001 through the study. It was also verified that the value of the initial weights does not effect the training of the network and so each network only needed to be initialized once.

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

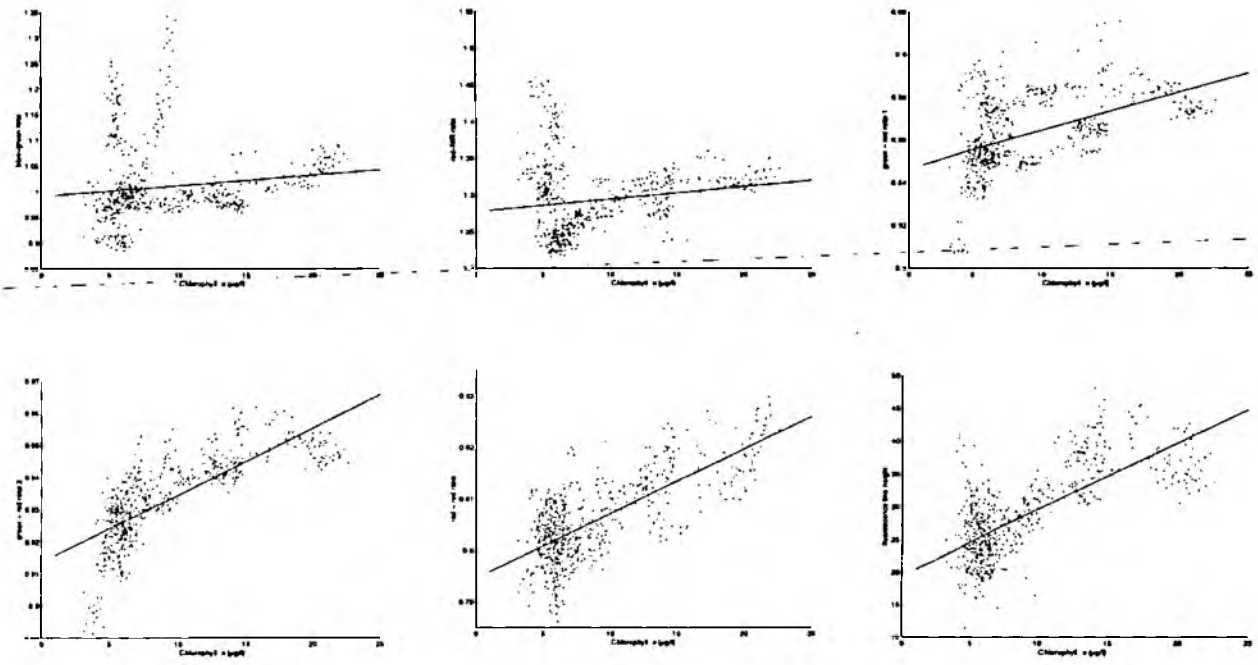


Figure 4: These figures show the value of the six spectral features used for the linear regression analysis plotted against chlorophyll *a* concentration. The straight line illustrates the line of least-squares fit to the data.

network. These features were: bands 6, 20, 41 and the FLH measurements. 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 validation data set was then used to estimate the ideal number of iterations for training the networks. At the end of each iteration the first 100[1] of the validation data points were passed through the network and the output error of this calculated. Because only the training error is back-propagated through the network, it is expected that the validation error will decrease for a number of iterations, then increase when the network begins to over-train. The number of iterations at which the validation error is minimum was then be 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 algorithms developed using a neural network model, to those using linear regression.

[1]By dividing the data using a random method (section 4.2), the statistical properties of the training, testing and validation sets were very similar. A subset of the validation set was chosen arbitrarily for training the network so that, if the network was to reduce the training error by predicting the mean chlorophyll *a* for every pattern, an increase in error would be evident for the validation set.

6 Results

The algorithms 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 which is calculated as:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}})^2}} \quad (3)$$

where x_i is the ith point in the testing set, $\hat{x}_i$ is the prediction for this point, $\bar{x}$ is the mean of chlorophyll *a* measurements in the testing set, $\bar{\hat{x}}$ is the mean of the predictions of these measurements and n is the number of data in the testing set. This measurement is the normalized covariance of the predicted and desired values and increases with the 'similarity' of the corresponding values. The second measure is the Root Mean Squared Difference (RMSDIF) which is calculated as:

$$\varepsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \quad (4)$$

This measurement increases with the error of the prediction. If the mean value were always predicted the RMSDIF would be equal to the standard deviation of the desired values.

6.1 Testing with the Norfolk 31/05/96 test set

The linear and neural network algorithms were applied to the 1183 testing data points and the correlation and RMSDIF between the predictions and the measured values were calculated.

6.1.1 Linear regression

Table 7 gives the correlation and RMSDIF of the chlorophyll *a* values predicted by the linear algorithms.

Feature	ρ	ε
blue-green ratio	0.1594	34.7824
red-NIR ratio	0.1625	27.8989
green-red ratio 1	0.4993	8.3975
green-red ratio 2	0.7641	4.0862
red-red ratio	0.7613	4.2775
FLH	0.7600	4.2703

Table 7: The correlation and RMSDIF for the Norfolk 31/05/96 testing set. ρ is the correlation, ε the RMSDIF.

There quite a distinction between the performances of the first and second three algorithms. The blue-green, red-NIR and green-red 1 ratios have not performed well, the first two having error values which indicate that predicted chlorophyll *a* is well outside the range of measured values. The second green-red 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 $\mu g/l$) and good correlations. These algorithms are based on features which are close to each other in the spectrum. It is possible that these ratios performed better than the more conventional ratios because the influence of the atmosphere, DOM and suspended sediment is approximately uniform over the shorter spectral range (Neville and Gower 1977).

Looking at the plots of the spectral features against chlorophyll *a* concentration (figure 4) it is clear to see why a good prediction was not found for some of the conventional band ratios. The blue-green ratio displays three distinct 'filaments' in this figure and the red-NIR ratio has two of these. The relationship between the ratio and chlorophyll *a* is therefore extremely non-linear. These filaments can be associated with precise regions in the coastal zone (appendix F). This illustrates the very site-specific nature of algorithms based in these measures.

6.1.2 Neural network models

The number of iterations to which each of the 60 networks was trained to is given in appendix G. 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. This indicates that FLH has a near-linear relationship with chlorophyll *a*.

The testing set was passed through the trained networks and the outputs compared to the measured chlorophyll *a* values. Appendix H gives the correlation and RMSDIF for all 60 networks. Three parameters were varied during the training and testing of the neural network - the number of inputs, the features passed into the network and the number of hidden nodes. Table 8 summarizes how varying these parameters affected the output of the network.

The top left table shows how the average correlation and RMSDIF change with the number of input nodes. There is a very clear increase in correlation and decrease in error with more input features, indicating that each input added more useful information.

The top right table compares the number of hidden nodes. Where no nodes were used (a one-layer network), the predictions are not as good. However, increasing the number of nodes in a two-layer network does not appear to have had a strong effect on the prediction. From this we may infer that, although the relationship between the spectral information and chlorophyll *a* is non-linear, it is not particularly complex.

The two bottom tables indicate the effect that particular bands have on the outcome of the model, and whether certain bands compliment each other in the model. The greatest difference here is that band 20 does not appear to contribute as much to the model as the three other bands. Band 6 in combination with FLH or with band 41 appears to produce much better models than any other combinations of features.

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

Three of the trained networks were used to predict the chlorophyll *a* the whole images of the Norfolk 31/05/96 data (appendix I. The range of chlorophyll *a* predictions is acceptable for most of the images. Overestimates are likely in the north of the region, possibly where edge

Number of inputs	$\sum$	ρ	ϵ
1	16	0.6359	3.7394
2	24	0.7447	3.2135
3	16	0.8567	2.3947
4	4	0.9167	1.7514

Number of hidden nodes	$\sum$	ρ	ϵ
0	15	0.6802	3.5424
2	15	0.7743	2.9587
6	15	0.7875	2.8177
10	15	0.7862	2.8329

Input includes band	$\sum$	ρ	ϵ
6	32	0.8096	2.5928
20	32	0.7536	3.0115
41	32	0.7902	2.8166
FLH	32	0.8251	2.7367

Input includes bands	$\sum$	ρ	ϵ
6 20	16	0.8335	2.4185
6 41	16	0.8777	2.1167
6 FLH	16	0.8834	2.1492
20 41	16	0.7957	2.7426
20 FLH	16	0.8318	2.6463
41 FLH	16	0.8401	2.5582

$\sum \equiv$ The number of network's results that the mean correlation and RMSDIF were calculated from, $\rho \equiv$ Mean correlation coefficient, $\epsilon \equiv$ Mean RMSDIF

Table 8: These tables summarize the goodness of the neural network predictions for Norfolk 31/05/96. The top left table finds the mean correlation and RMSDIF for networks with a given number of inputs. The top right table finds these means for networks with a given number of hidden nodes. The bottom left table finds the means for all the networks which have a particular spectral feature (a band or FLH) as an input and the lower left table finds these means for all the networks for which a combination of two spectral features are found as inputs.

brightening has effected the images. The two more complex networks appear a little more sensitive to noise in the images whereas MLP 22b (two inputs, 2 hidden nodes) has produces a smooth chlorophyll *a* prediction.

6.2 Portability over time - the Norfolk 11/08/96 test set

Because only one measurement of chlorophyll *a* could be obtained for this set, the algorithms were applied to the images. The three best linear regressions (green-red 2 ratio, red red ratio and FLH) were compared to three of the neural networks (MLP 22b, MLP 36c and MLP 410a). The resulting images are shown in appendix J.

6.2.1 Linear regression

The ranges of values predicted by the linear algorithms were:
Green-red ratio = < -90 *to* > 200 $\mu g/l$
Red-red ratio = 1 *to* 101 $\mu g/l$
FLH = < -10 *to* > 200 $\mu g/l$
The negative values are clearly inaccurate and the ranges of chlorophyll *a* values are much higher than those measured at the location. These predictions are not reliable.

6.2.2 Neural network models

The ranges of values predicted by the neural networks are:
MLP 22b = 2 *to* 42 $\mu g/l$
MLP 36c = 1 *to* 27 $\mu g/l$
MLP 410a = -10 *to* 21 $\mu g/l$
The predictions are probably too high for the region but are more realistic than the linear regression predictions. The images do show some interesting structure which may result form chlorophyll *a* in the water. Patches of high or low chlorophyll *a* are quite clear. This structure is visible in the CASI imagery as brighter patches in the water, particularly in the green wavelengths.

6.3 Portability over space - the Holderness 19/08/96 test set

The 12 Holderness data points were applied to all the linear regression algorithms and all the trained neural networks and the correlation and RMSDIF between the predicted and measured values were calculated.

6.3.1 Linear regression

Table 9 summarizes the results of this test.

Feature	ρ	ε
blue-green ratio	-0.8459	26.9219
red-NIR ratio	0.7586	123.9834
green-red ratio 1	0.7965	42.6846
green-red ratio 2	0.5142	15.5455
red-red ratio	0.1128	8.3482
FLH	0.7718	32.3012

Table 9: The correlation and RMSDIF for the Holderness 19/08/96 testing set. ρ is the correlation, ε the RMSDIF.

Although there are some good correlations between predicted and measured values, the error values are all poor, especially considering the standard deviation for this set is only $0.78 \mu g/l$. The linear regressions have predicted values ranging over several orders of magnitude.

The two lowest RMSDIFs result from the green-red 2 and red-red ratios - the two ratios chosen because they had a good correlations with chlorophyll *a* in the *Norfolk 31/05/96* training set. The FLH algorithm performs quite well but consistently over-estimates chlorophyll *a*, suggesting that the algorithm may be rescaled to produce a better chlorophyll *a* estimation. Again, the better performance was achieved by the last three algorithms although the predictions are too poor for the algorithms to be considered reliable.

6.3.2 Neural network models

Table 10 summarizes the correlation and RMSDIF of the neural network tests.

Number of inputs	$\sum$	ρ	ε
1	16	0.3071	13.6284
2	24	0.2088	14.9294
3	16	0.2962	14.0766
4	4	0.2889	13.6898

Number of hidden nodes	$\sum$	ρ	ε
0	15	0.6083	21.8699
2	15	0.1132	11.3943
6	15	0.1882	11.7698
10	15	0.1450	12.0557

Input includes band	$\sum$	ρ	ε
6	32	0.1251	14.0519
20	32	0.2332	14.3325
41	32	0.3154	14.7292
FLH	32	0.3818	14.0544

Input includes bands	$\sum$	ρ	ε
6 20	16	0.1510	14.0648
6 41	16	0.1900	14.1764
6 FLH	16	0.2058	14.0860
20 41	16	0.3763	14.4324
20 FLH	16	0.3897	14.0228
41 FLH	16	0.3224	14.3762

$\sum \equiv$ The number of network's results that the mean correlation and RMSDIF were calculated from, $\rho \equiv$ Mean correlation coefficient, $\varepsilon \equiv$ Mean RMSDIF

Table 10: These tables summarize the goodness of the neural network predictions for Holderness 19/08/96.

The number of inputs appears to have a slight effect on the reliability of the prediction - slightly lower RMSDIFs were achieved by networks with one or four inputs.

The best errors are achieved when a network with 2 or 6 hidden nodes is used. Again, this highlights the non-linear nature of the relationship between spectral information and chlorophyll *a* concentration. The high errors resulting from using the linear networks are associated with high correlations, as with the linear regression. The error actually increases as the number of hidden nodes increases above 2, possibly because the more complex models that result are over-fitted to the training data and that a better generalisation was achieved with the two-layer, two hidden node network.

No real conclusions can be drawn from the last two tables in table 10. There is a very slight improvement when using blue or FLH as an input, but this is tiny compared to the magnitude of the error.

Looking at the individual results for each network, the three best networks (mlp12d, mlp16a and mlp110a) are all produced using only one input (either FLH or blue). As with the linear regression, these algorithms may be more robust to the effects of sediment and the atmosphere through the spectrum because they do not rely on the relationship between spectrally-distant bands.

Overall, the predictions made of this data set were poor. One possible reason is because the chlorophyll *a* concentration of the site is below that of the training data, so the algorithms are attempting to extrapolate the values. Another possible reason may be that the effects of suspended sediment, DOM and the atmosphere are different to those at the training site, so the

algorithms are unable to correct for them.

7 Discussion

The neural networks proved a simple way of combining several features and of modelling non-linear regression functions. Although the networks were rather complicated to initiate, once trained they proved efficient at predicting chlorophyll *a* concentration. Moreover, the results of the multiple runs allowed some insights into the nature of the data being modelled.

From the improved performance with the two-layer networks, we may conclude that a non-linear relationship does exist between spectral qualities and chlorophyll *a*, which explains the poor performance of the linear regression functions. The networks also provided information about which features are useful for chlorophyll *a* prediction.

As has been found in several studies, the FLH measure gave good predictions for the linear regressions. In fact, the relationship between this and chlorophyll *a* seemed to be 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 was supported by the short training time required by networks with FLH as an input. Of the other linear regressions, those ratios produced from spectrally-close wavebands gave better performances.

The performance of the blue band was more of a surprise. Many recent studies have avoided this part of the spectrum because of the interference of DOM and the atmosphere and this study certainly found that the blue-green linear algorithm was a very poor predictor. However, this input provided good predictions as an input to non-linear regressions. This non-linear behaviour was highlighted by the plots of blue-green ratios against chlorophyll *a*.

The neural networks performed better than the linear regression algorithms for the *Norfolk 11/08/96* data. Non-linear algorithms developed in this manner therefor show some promise with predicting chlorophyll *a* at a site over time. However, none of the algorithms showed much promise for predicting chlorophyll *a* at a different site (Holderness). This is to be expected since the inputs chosen for training the algorithms, and the training data itself, was all based on the characteristics of a very different site.

Several points can be noted of the performance of the algorithms when applied to the *Holderness 19/08/96* data. Firstly, the algorithms that performed the best only had one input. This may be similar to the finding that linear regressions with spectrally-close bands gave better results. It is likely that environmental factors, such as sedimentary and atmospheric conditions were quite different at the Holderness site. Hence, the algorithms that perform best are those that are more robust to differing effects of these factors across the spectrum.

In general, the more complex non-linear algorithms performed best when predicting chlorophyll *a* for data for the same place and time as the data on which it was trained. However, less complex but still non-linear algorithms performed better when making predictions for new data sets.

7.1 Future work

The major drawback of this study was the shortage of suitable data from other sites. To develop this research further, datasets containing imagery and *in situ* measurements of chlorophyll *a* for several different regions would be required.

The choice of features and the non-linearity of the algorithm were found to be the dominant factors in the algorithms' effectiveness. The relationship between features and chlorophyll *a* changed between sites, however, so using a single site to develop an algorithm for the whole coast is unlikely to be successful. Clearly, more portable algorithms would be best developed using data from many different environments.

With data that are more representative of the whole UK coast, it is recommended that a comprehensive analysis of spectral features should be undertaken to determine which are most suitable for an algorithm which will estimate chlorophyll *a* for the whole set. The method is chosen for this should take into account the non-linearity of the relationship found between spectral features and chlorophyll *a*. Neural networks themselves make possible several methods of non-linear feature extraction such as non-linear principle component analysis (Bishop 1995) and the comparison of the performance of networks trained with different combinations of features as demonstrated in this study.

Whatever method is chosen for feature selection, it is unlikely to be exhaustive because so many spectral features are possible with the enhanced spectral mode CASI imagery. It will therefore be important that the resulting algorithm makes the most of those features chosen. In this study, several algorithms have been developed and compared to each other. A future study could use a neural network to combine several algorithms, including those produced by linear regression and other techniques, in a *committee* network. This utilizes the ability of a neural network to automatically weight and assign reliability measures to algorithms and produces an algorithm with better generalisation than any of its component algorithms (Bishop 1995).

This study has found that there is potential for improved chlorophyll *a* prediction algorithms to be developed using neural networks and has provided a useful basis for further research into this subject.

8 Conclusions

A number of conclusions may be drawn from this study:

1. There is non-linearity in the relationship between spectral information and chlorophyll *a* in water. Hence, linear algorithms will not predict chlorophyll *a* well for all but very short ranges of chlorophyll *a*.
2. An increase in the number of input features to the algorithm provides a better prediction of chlorophyll *a* and so multiple regression algorithms should be considered where possible.
3. Neural networks provide a simple means of developing non-linear, multiple regression algorithms.
4. The reliability of algorithms over time and space could not be satisfactorily assessed, however preliminary results appear to demonstrate that algorithms should be developed using many data from differing environmental conditions to ensure a good generalisation.
5. The FLH feature gave good predictions of chlorophyll *a* and, although its relationship to chlorophyll *a* is non-linear, it was the most suitable of all the features when only linear modelling is possible.
6. The blue wavelengths should not be ignored. Blue appeared to have a very non-linear relationship to chlorophyll *a*, but this relationship was strong and the blue waveband enabled some good predictions of chlorophyll *a*.
7. With an adequate provision of suitable image and *in situ* data, there is great potential for using neural networks in future studies of chlorophyll *a* around the coast of the UK

References

Aiken, J., S. Hudson, G. Moore, and H. Bottrell (1995, April). 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 (1995). The SeaWiFS CZCS-type pigment algorithm. Technical Report NASA Technical Memorandum 104566, Volume 29, NASA Goddard Space Flight Center, Maryland.

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 in Geoscience and Remote Sensing 28*(4), 540–550.

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

Bricaud, A., A. Morel, and V. Barale (1995). MERIS potential for ocean colour studies in the open ocean. See Curran and Robertson (1995), pp. 133–140.

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.

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

Clark, C. and A. Cañas (1993). Spectral identification by artificial neural network. In *Towards Operation Applications*. Remote Sensing Society: RSS, Nottingham.

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

Curran, P. J. and Y. C. Robertson (Eds.) (1995). *Remote Sensing in Action*. Remote Sensing Society: RSS, Nottingham.

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

Doerffer, R., K. Sørensen, and J. Aiken (1995). MERIS: Potential for coastal zone application. See Curran and Robertson (1995), pp. 166–175.

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.

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

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. and M. K. Arora (1996). Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications. *Pattern Recognition Letters 17*, 1389–1398.

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.

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.

Gopal, S. and C. E. Woodcock (1996). Remote sensing of forest change using artificial neural networks. *IEEE Transactions on Geoscience and Remote Sensing 34*(2), 398–404.

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.

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.

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.

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

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

Matthews, A. (1994). *High Resolution Spectral Remote Sensing of Phytoplankton in the Coastal Zone.* Ph. D. thesis, Department of Oceanography, University of Southampton.

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.

Paola, J. D. and R. A. Schowengerdt (1997). The effect of neural network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering and Remote Sensing 63*(5), 535–544.

Quibell, G. (1991). The effect of suspended sediment on reflectance from freshwater algae. *International Journal of Remote Sensing 12*(1), 177–182.

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.

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.

Steele, J. H. and T. W. Henderson (1979). Spatial patterns in North Sea plankton. *Deep Sea Research 26A*, 955–963.

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.

Warner, T. A. and M. C. Shank (1997). An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks. *Photogrammetric Engineering and Remote Sensing 63*(11), 1285–1294.

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.

Yoder, J.-A., C. R. McClain, J. O. Blanton, and L.-Y. Oey (1987). Spatial scales in CZCS chlorophyll imagery of the Southeastern US continental shelf. *Limnology and Oceanography 32*, 929–941.

Appendices

A The CASI enhanced spectral bandset

Band number	Lower Limit (nm)	Centre (nm)	Upper Limit (nm)	Band 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
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

Continued on following page

Band number	Lower Limit (nm)	Centre (nm)	Upper Limit (nm)	Band Width (nm)
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
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

B The fluorometer calibration data

			Size of buffer (*meters*)		250		200		150		100		50	
n[⊗]	Easting	Northing	Laboratory Chlorophyll *a* ($\mu g/l$)	Fluorometer reading at data point	$N^{\dagger}$	$\mu^{\ddagger}$	$N^{\dagger}$	$\mu^{\ddagger}$	$N^{\dagger}$	$\mu^{\ddagger}$	$N^{\dagger}$	$\mu^{\ddagger}$	$N^{\dagger}$	$\mu^{\ddagger}$
1	637916	334383	8.79	39.78	10	39.10	8	39.25	6	39.32	4	39.25	2	39.08
2	627234	342133	7.19	30.34	11	31.37	9	31.49	7	31.67	4	31.69	2	31.06
3	609448	347257	5.92	38.97	12	38.80	9	39.04	6	39.48	4	39.29	2	39.40
4	589185	349454	14.54	95.57	12	96.86	10	97.08	8	97.17	4	97.31	2	96.28
8	561126	348043	15.04	35.29	23	34.73	17	34.89	13	34.82	8	34.73	3	35.17
9	568869	346080	11.95	59.85	8	60.37	7	60.58	6	60.80	4	60.65	3	60.73
10	585398	349595	13.66	100.90	10	98.92	8	99.13	7	99.53	5	99.34	3	99.47
11	605690	347793	4.90	37.51	7	37.49	6	37.38	5	37.31	3	37.21	2	36.98
12	632086	339701	9.52	34.20	6	32.05	5	32.49	4	33.01	3	32.73	2	33.07
13	651182	318615	6.55	27.80	6	29.40	5	28.89	4	28.43	3	28.00	2	27.71
15	605171	345716	13.33	70.46	26	64.86	23	67.98	20	71.37	17	73.00	8	72.35
16	605306	346226	6.20	41.73	18	41.30	14	40.30	9	40.45	5	41.27	3	41.63
17	606449	350076	6.46	48.98	8	49.98	7	50.16	5	49.38	4	49.61	2	49.98
			correlation	0.6640		0.6454		0.6559		0.6637		0.6637		0.6648

[†] Number of fluorometer readings, [‡] Mean fluorometer reading, [⊗] Data point number

Comparison of different buffer sizes around calibration data points. The correlation between mean fluorometer readings within the buffer and the chlorophyll *a* measurement improves with smaller buffer size.

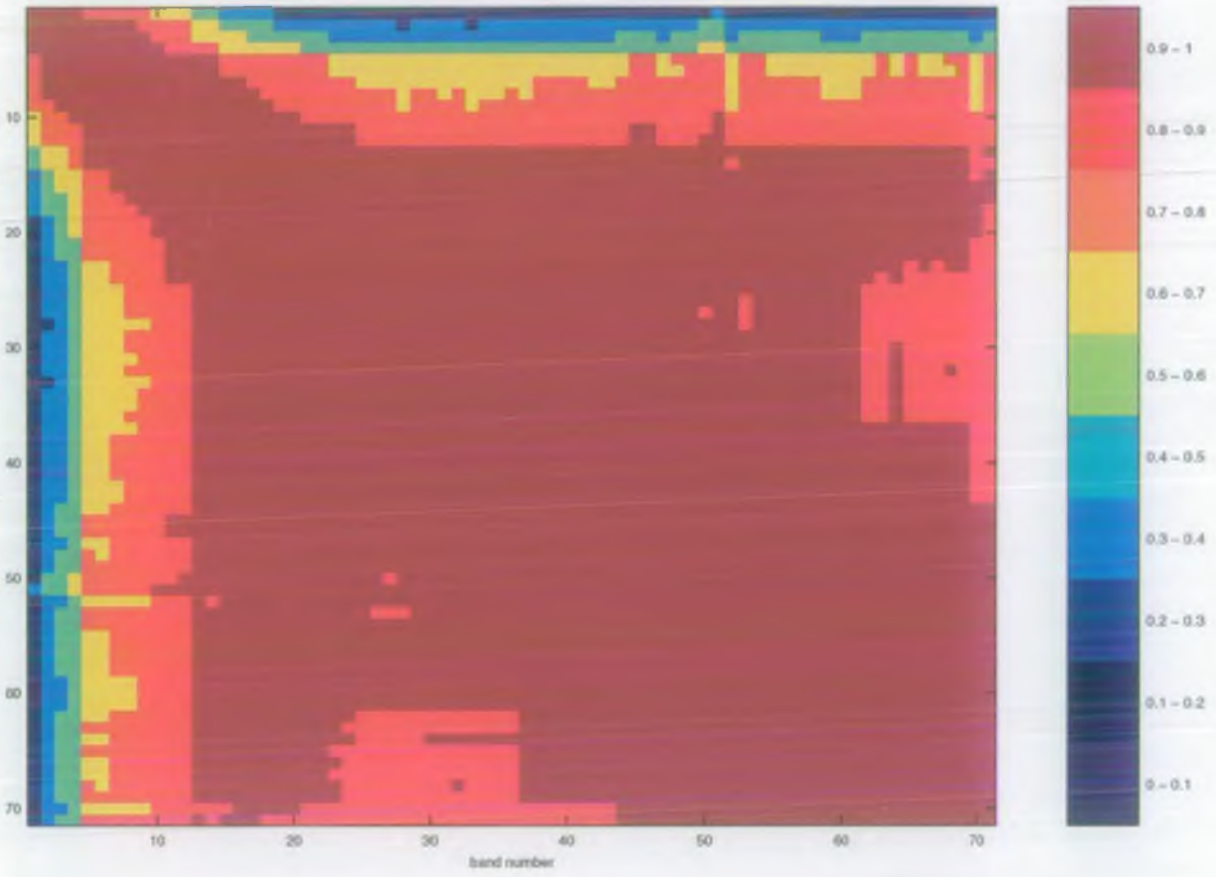
C Comparison of different buffer sizes around fluorometer data points

	Median for CASI band:					Weighted mean for CASI band:					
$\Sigma^{\dagger}$	15	20	30	40	60	15	20	30	40	60	$\mu^{\ddagger}$
1	0.5080	0.5323	0.6863	0.6852	0.4318						
7	0.6497	0.6276	0.7116	0.7252	0.5354	0.6502	0.6305	0.7114	0.7267	0.5410	0.6509
11	0.6427	0.6350	0.7088	0.7249	0.5462	0.6483	0.6339	0.7103	0.7265	0.5468	0.6523
15	0.6396	0.6386	0.7068	0.7262	0.5593	0.6477	0.6352	0.7096	0.7270	0.5515	0.6542
21	0.6413	0.6431	0.7057	0.7258	0.5575	0.6463	0.6364	0.7090	0.7273	0.5540	0.6546
$\mu^{\otimes}$			0.6526					0.6535			

[†] Size of sides of window in pixels (one pixel has an edge of approximately 20m). [‡] Mean correlation coefficient for each window size. [⊗] Mean correlation coefficient for all window sizes for each method of averaging values in the window

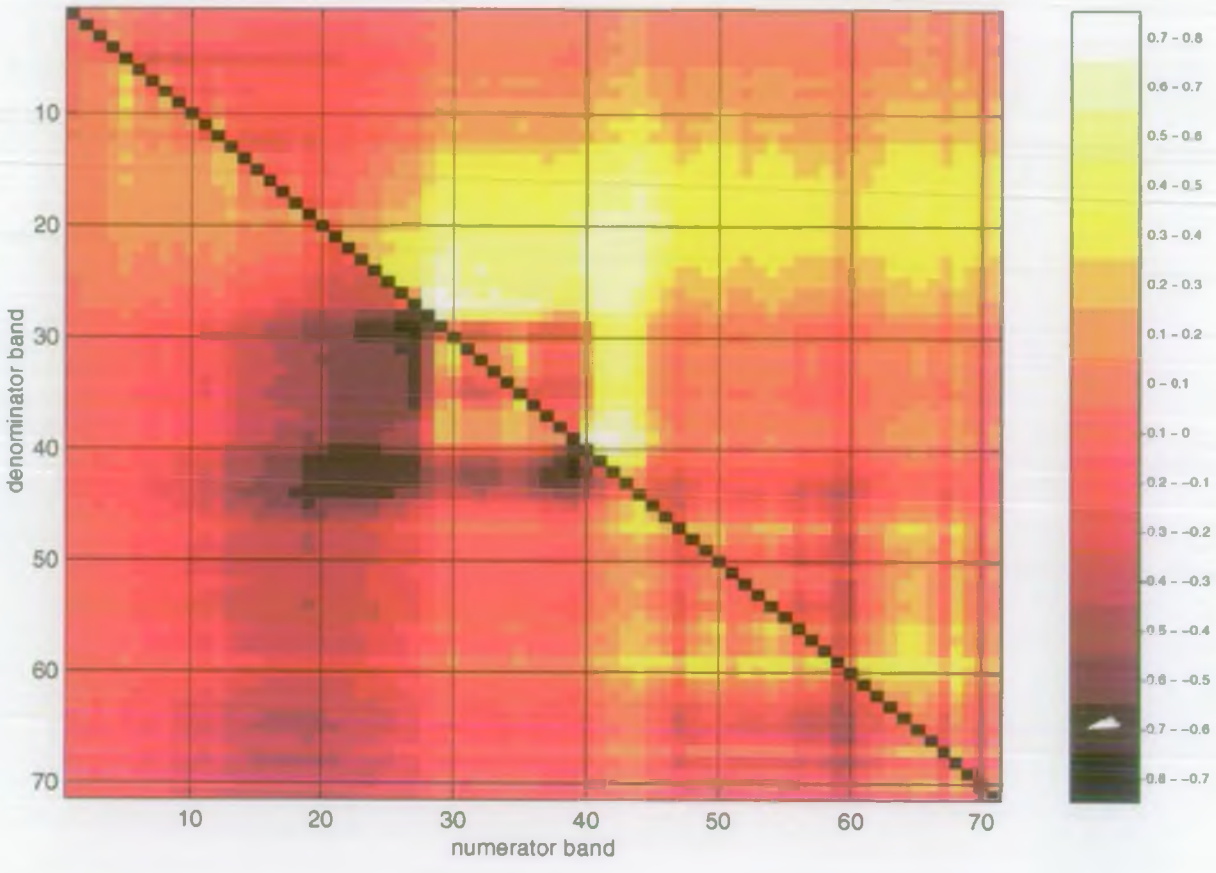
Correlation coefficients of different sized windows using two different methods of combining the values within the window. The first method finds the median value within the window, the second finds the mean of the values which are weighted using a Gaussian function. Using the values within a window of any size improves the correlation and generally there is a slight improvement with larger window sizes. The Gaussian weighting also gives a slightly better correlation overall.

D Correlation of training band with each other

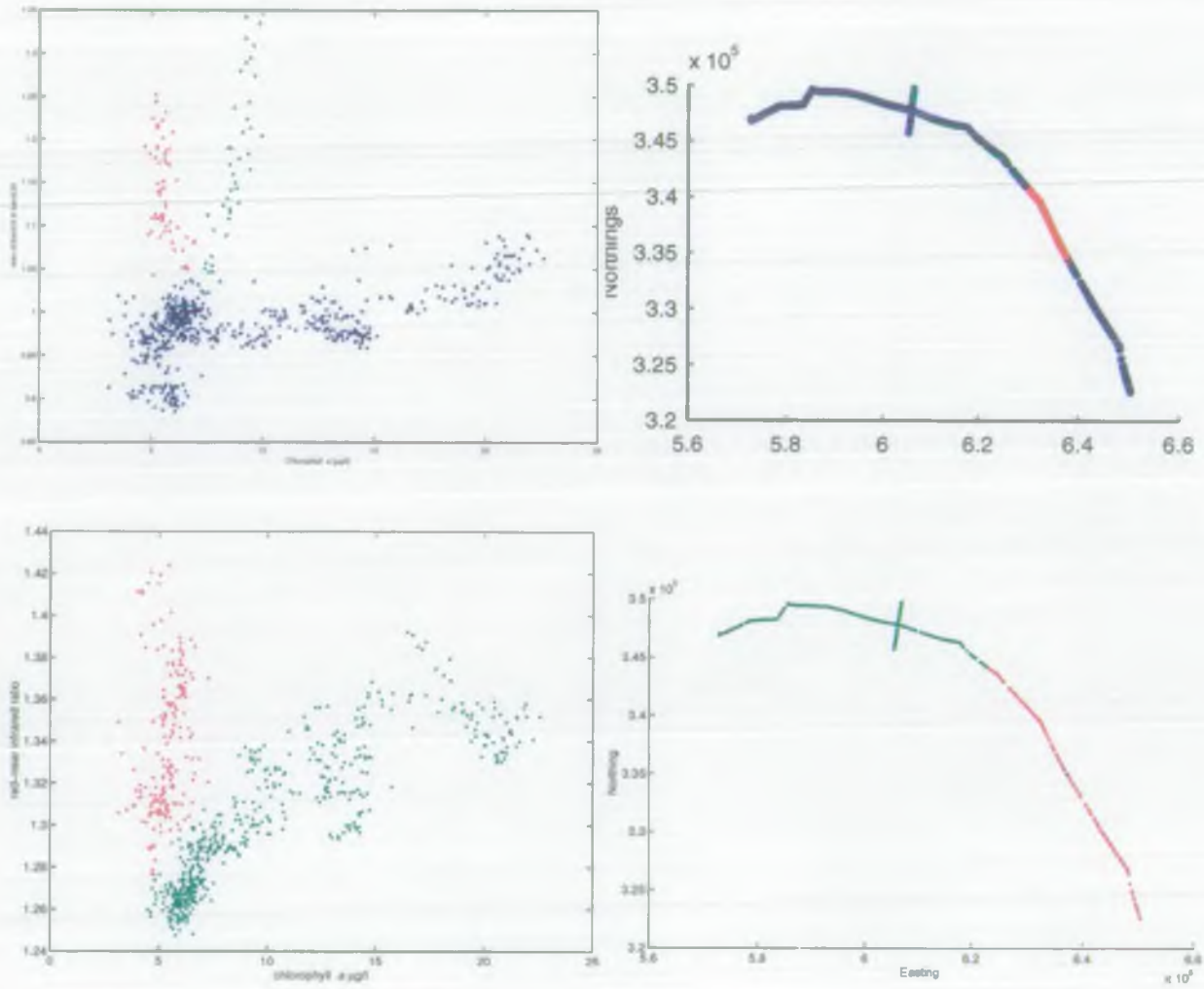


E Correlation of training band ratios with Chlorophyll *a*

Correlations are for ratios in the *Norfolk 31/05/96* training data set only.



F The different water bodies within the Norfolk region



G MLP training mean squared error

MLP name	Input bands	Number of hidden nodes	Training MSE at 2000	Validation MSE at 2000	Lowest validation MSE at:	Training MSE at this point	Validation MSE at this point
mlp1a	6	0	17.1570	3.3830	13	17.2314	3.1208
mlp1b	20	0	17.4713	5.3474	10	18.0630	4.3608
mlp1c	41	0	14.3462	5.3040	16	14.3744	5.2350
mlp1d	FLH	0	10.0575	10.9864	7	13.2581	7.6730
mlp12a	6	2	0.6105	0.0664	172	0.6591	0.0366
mlp12b	20	2	0.7282	0.3039	110	0.7171	0.2227
mlp12c	41	2	0.4263	0.4388	10	0.6403	0.2274
mlp12d	FLH	2	0.3633	0.4804	7	0.5442	0.4456
mlp16a	6	6	0.6219	0.1056	115	0.6630	0.0338
mlp16b	20	6	0.7402	0.3239	131	0.7164	0.2184
mlp16c	41	6	0.4257	0.4366	8	0.6448	0.2211
mlp16d	FLH	6	0.3618	0.4910	6	0.5097	0.4526
mlp110a	6	10	0.6192	0.0983	115	0.6655	0.0332
mlp110b	20	10	0.7288	0.3031	99	0.7183	0.2217
mlp110c	41	10	0.4256	0.4351	6	0.6847	0.2269
mlp110d	FLH	10	0.3618	0.4893	4	0.5774	0.4591
mlp2a	6,20	0	16.9167	3.6551	14	16.9654	3.4699
mlp2b	6,41	0	14.3617	5.1752	14	14.4894	4.7835
mlp2c	6,FLH	0	9.9126	8.8655	9	11.2301	6.9917
mlp2d	20,41	0	10.5211	9.6166	12	13.4466	4.4229
mlp2e	20,FLH	0	10.0518	10.8472	8	12.1053	7.8029
mlp2f	41,FLH	0	9.8724	9.2292	9	11.2888	7.6961
mlp22a	6,20	2	0.6067	0.1025	463	0.6330	0.0735
mlp22b	6,41	2	0.2488	0.0618	$\geq$2000	0.2488	0.0618
mlp22c	6,FLH	2	0.3192	0.2434	8	0.5094	0.2328
mlp22d	20,41	2	0.3764	0.3987	8	0.6726	0.2214
mlp22e	20,FLH	2	0.3430	0.4880	8	0.5188	0.3244
mlp22f	41,FLH	2	0.2910	0.4152	5	0.4857	0.2894
mlp26a	6,20	6	0.4769	0.0684	1780	0.4984	0.0634
mlp26b	6,41	6	0.2310	0.0598	1351	0.2369	0.0587
mlp26c	6,FLH	6	0.2208	0.0634	$\geq$2000	0.2208	0.0634
mlp26d	20,41	6	0.3721	0.3799	5	0.7023	0.2223
mlp26e	20,FLH	6	0.3184	0.5135	5	0.5617	0.3434
mlp26f	41,FLH	6	0.2782	0.3973	4	0.4854	0.3006
mlp210a	6,20	10	0.4907	0.0881	$\geq$2000	0.4907	0.0881
mlp210b	6,41	10	0.2283	0.0586	$\geq$2000	0.2283	0.0586
mlp210c	6,FLH	10	0.2243	0.0727	$\geq$2000	0.2243	0.0727
mlp210d	20,41	10	0.3727	0.3644	5	0.7058	0.2158
mlp210e	20,FLH	10	0.3189	0.5102	4	0.5248	0.3261
mlp210f	41,FLH	10	0.2820	0.4218	4	0.5509	0.3027
mlp3a	6,20,41	0	9.4785	6.2830	15	12.4302	3.3058
mlp3b	6,20,FLH	0	9.8235	8.6167	9	11.1597	6.8509
mlp3c	6,41,FLH	0	9.8632	8.7347	10	10.7658	7.4115
mlp3d	20,41,FLH	0	8.7713	11.5411	9	10.9406	8.2164
mlp32a	6,20,41	2	0.1624	0.1262	$\geq$2000	0.1624	0.1262
mlp32b	6,20,FLH	2	0.2052	0.1264	$\geq$2000	0.2052	0.1264
mlp32c	6,41,FLH	2	0.1621	0.1017	$\geq$2000	0.1621	0.1017
mlp32d	20,41,FLH	2	0.2832	0.4344	7	0.5264	0.2718
mlp36a	6,20,41	6	0.1432	0.1136	$\geq$2000	0.1432	0.1136
mlp36b	6,20,FLH	6	0.1341	0.1478	518	0.1803	0.1035
mlp36c	6,41,FLH	6	0.1164	0.0803	$\geq$2000	0.1164	0.0803
mlp36d	20,41,FLH	6	0.2548	0.3638	4	0.5330	0.2613
mlp310a	6,20,41	10	0.1322	0.0876	$\geq$2000	0.1322	0.0876
mlp310b	6,20,FLH	10	0.1340	0.1658	631	0.1908	0.1408
mlp310c	6,41,FLH	10	0.1090	0.0583	$\geq$2000	0.1090	0.0583
mlp310d	20,41,FLH	10	0.2577	0.3532	4	0.5253	0.2688
mlp4a	6,20,41,FLH	0	8.2217	8.1801	10	10.2596	7.2131
mlp42a	6,20,41,FLH	2	0.1279	0.0989	$\geq$2000	0.1279	0.0989
mlp46a	6,20,41,FLH	6	0.0813	0.0459	$\geq$2000	0.0813	0.0459
mlp410a	6,20,41,FLH	10	0.0694	0.0279	1640	0.0746	0.0269

H MLP testing correlation and error for Norfolk 31/05/96

MLP name†	Correlation	RMSDIF
mlp1a	0.4919	4.1362
mlp1b	0.4635	4.2796
mlp1c	0.611	3.7548
mlp1d	0.76	3.5307
mlp12a	0.6277	3.6828
mlp12b	0.5701	3.8896
mlp12c	0.6732	3.6924
mlp12d	0.7667	3.397
mlp16a	0.6259	3.6916
mlp16b	0.5709	3.8882
mlp16c	0.657	3.703
mlp16d	0.766	3.2639
mlp110a	0.6248	3.6961
mlp110b	0.5689	3.8956
mlp110c	0.6346	3.8236
mlp110d	0.7626	3.5049
mlp2a	0.4998	4.1108
mlp2b	0.6105	3.7746
mlp2c	0.7675	3.2252
mlp2d	0.6659	3.6343
mlp2e	0.7611	3.3637
mlp2f	0.7686	3.2352
mlp22a	0.6455	3.614
mlp22b	0.8789	2.284
mlp22c	0.7701	3.2432
mlp22d	0.629	3.7893
mlp22e	0.7637	3.2889
mlp22f	0.7915	3.1753
mlp26a	0.7306	3.3269
mlp26b	0.8845	2.2447
mlp26c	0.8959	2.129
mlp26d	0.6074	3.8786
mlp26e	0.7617	3.4378
mlp26f	0.7797	3.1578
mlp210a	0.7554	3.2873
mlp210b	0.8929	2.1876
mlp210c	0.8934	2.1556
mlp210d	0.5959	3.8876
mlp210e	0.7588	3.2988
mlp210f	0.7656	3.3938
mlp3a	0.7009	3.472
mlp3b	0.7698	3.213
mlp3c	0.7693	3.1549
mlp3d	0.7781	3.1821
mlp32a	0.9278	1.7694
mlp32b	0.9175	1.9356
mlp32c	0.9305	1.7569
mlp32d	0.7752	3.3213
mlp36a	0.9347	1.6871
mlp36b	0.923	1.8451
mlp36c	0.9526	1.462
mlp36d	0.7564	3.3377
mlp310a	0.9398	1.618
mlp310b	0.9246	1.8115
mlp310c	0.9538	1.4498
mlp310d	0.7528	3.2994
mlp4a	0.785	3.0689
mlp42a	0.9464	1.541
mlp46a	0.9669	1.2117
mlp410a	0.9685	1.1838

† refer to appendix G

I Testing with Norfolk 31/05/96 imagery

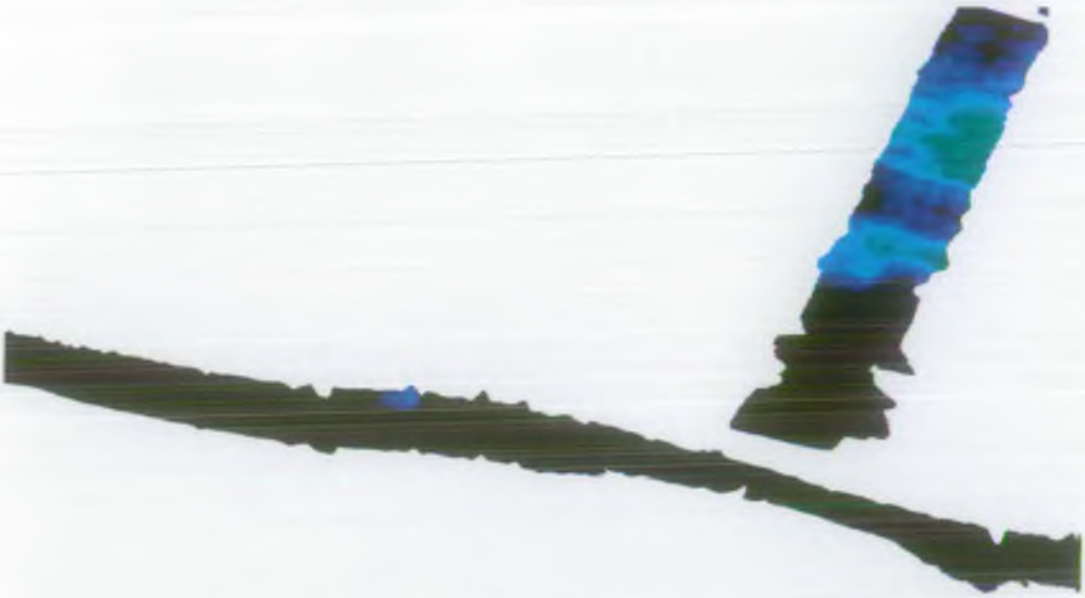
Three of the trained neural networks were used to produce the following three images. These neural networks increase in complexity:
MLP 22b with two inputs (bands 6 and 41) and two hidden layers
MLP 36c with three inputs (bands 6 and 41 and FLH) and six hidden nodes
MLP 410a with four inputs (bands 6, 20 and 41 and FLH) and 10 hidden nodes.

(a) MLP 22b at 2000 iterations

(b) MLP 36c at 2000 iterations

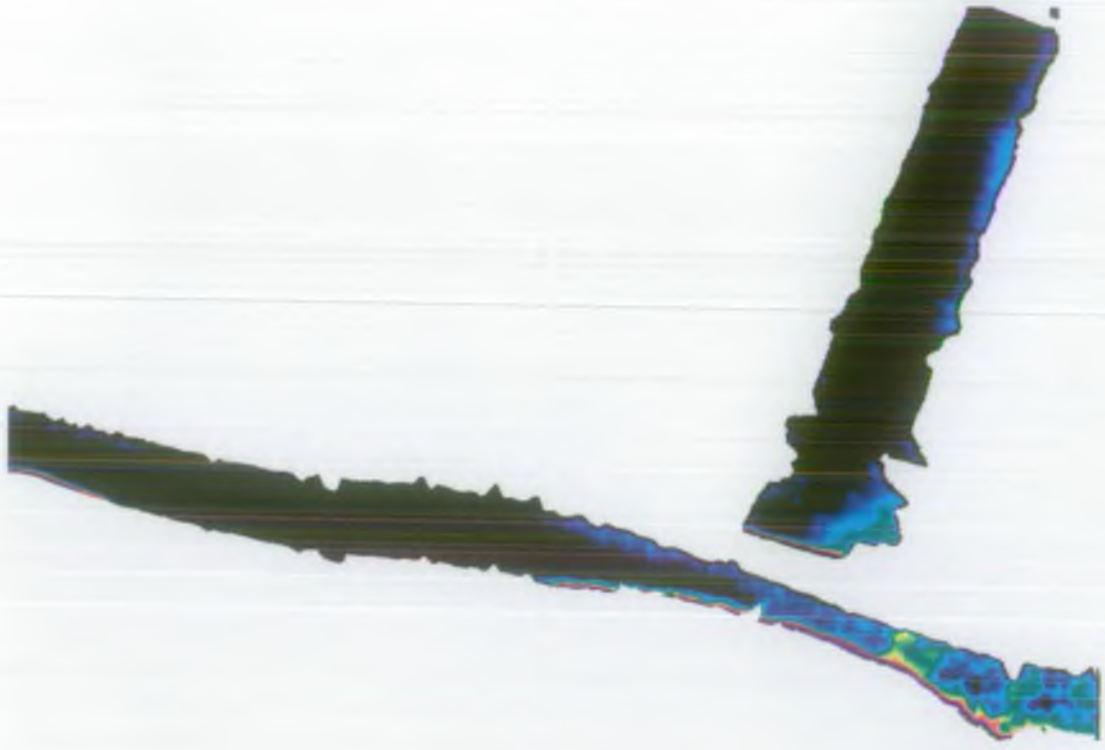
(c) MLP 410a at 1640 iterations

J Testing with Norfolk 11/08/96 imagery

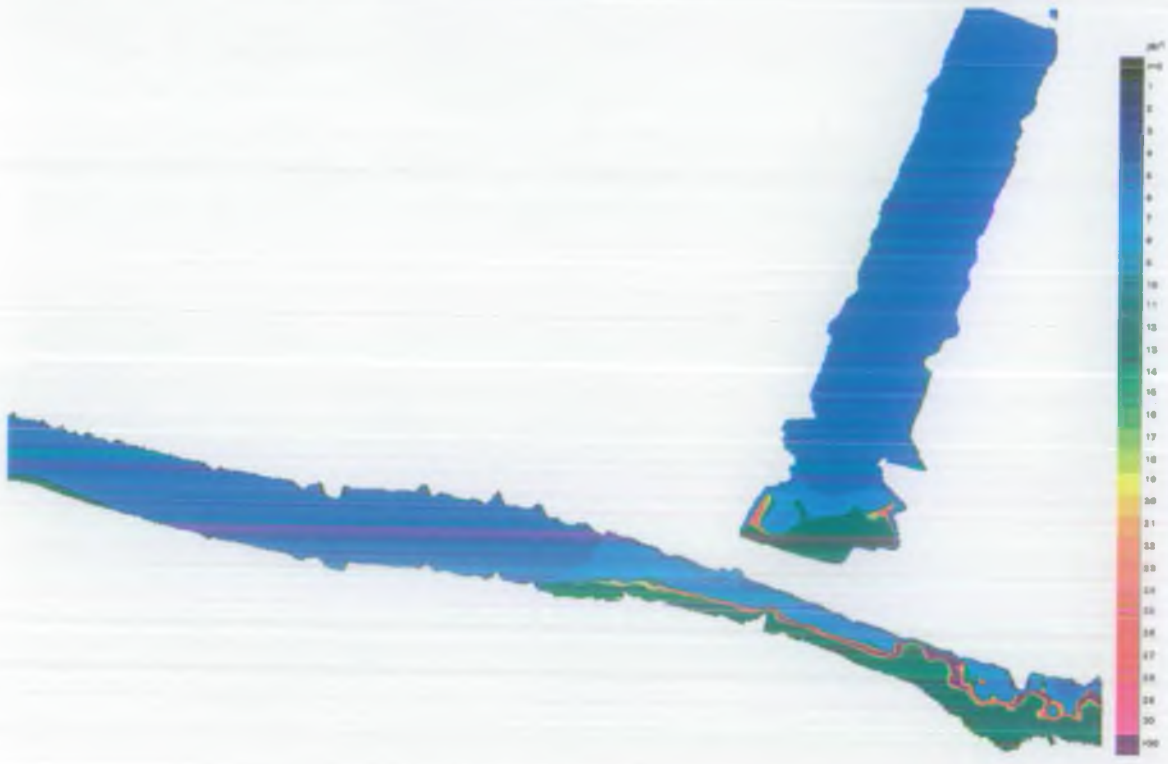


(a) Green-red ratio 2 linear regression

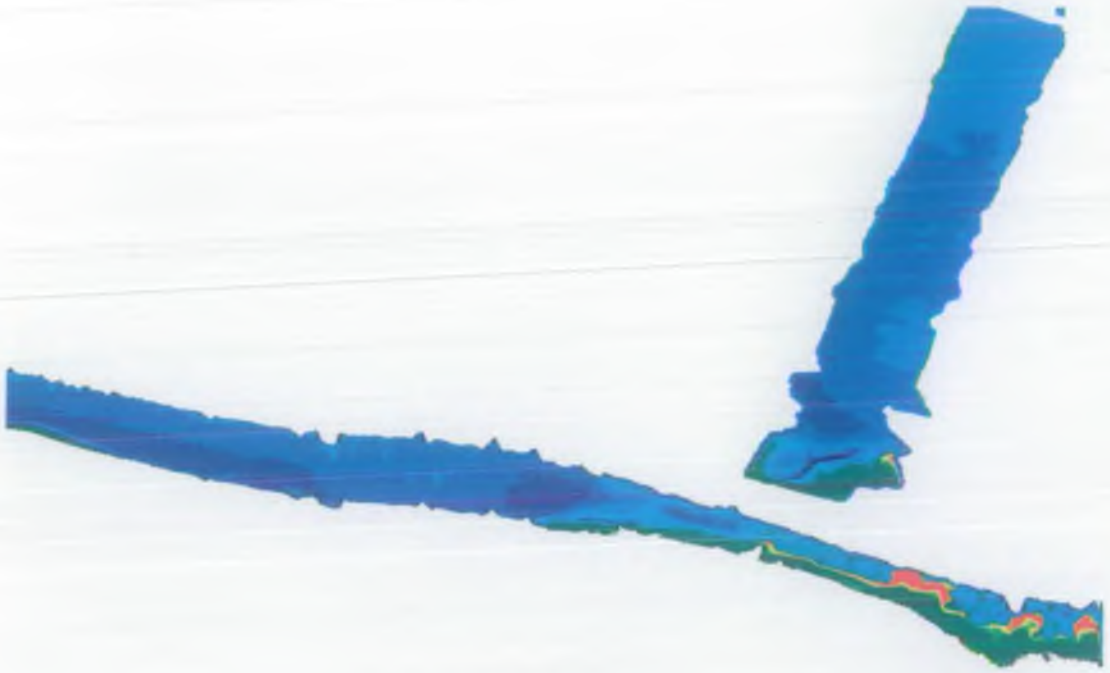
(b) Red-red ratio linear regression



(c) FLH linear regression



(d) MLP 22b at 2000 iterations



(e) MLP 36c at 2000 iterations

(f) MLP 410a at 1640 iterations

K MLP testing correlation and error for Holderness 19/08/96

MLP name†	Correlation	RMSDIF
mlp1a	-0.0416	20.0204
mlp1b	0.2214	21.8871
mlp1c	0.6731	25.9261
mlp1d	0.7718	19.0196
mlp12a	0.2039	10.1033
mlp12b	-0.2163	12.2833
mlp12c	0.2607	11.7600
mlp12d	0.6733	9.5640
mlp16a	0.1798	9.1556
mlp16b	-0.1666	11.0668
mlp16c	0.4105	12.8977
mlp16d	0.7378	10.4710
mlp110a	0.1452	8.6018
mlp110b	-0.1853	11.0810
mlp110c	0.4624	13.9542
mlp110d	0.7831	10.2627
mlp2a	0.0584	21.8042
mlp2b	0.6534	25.2623
mlp2c	0.7434	20.8475
mlp2d	0.7372	23.3738
mlp2e	0.7658	19.8851
mlp2f	0.7614	21.8844
mlp22a	-0.0238	12.2422
mlp22b	0.0238	11.8202
mlp22c	0.1006	11.0806
mlp22d	0.0279	11.8039
mlp22e	0.2137	11.7998
mlp22f	0.1920	11.6167
mlp26a	-0.7344	15.2366
mlp26b	0.0237	11.5753
mlp26c	-0.1168	13.4522
mlp26d	0.3011	12.5611
mlp26e	0.5819	11.8194
mlp26f	0.4673	12.8340
mlp210a	-0.7732	11.8033
mlp210b	-0.6213	11.2505
mlp210c	0.1037	14.7085
mlp210d	0.3657	13.7228
mlp210e	0.5845	13.0308
mlp210f	0.5764	12.8911
mlp3a	0.7256	23.0442
mlp3b	0.7471	20.1116
mlp3c	0.7532	21.8135
mlp3d	0.7846	21.5967
mlp32a	-0.0240	11.4302
mlp32b	0.0867	10.6448
mlp32c	0.0239	11.9357
mlp32d	0.1770	11.5012
mlp36a	-0.0173	10.6606
mlp36b	0.1078	11.0955
mlp36c	0.2153	10.8469
mlp36d	0.4165	12.4516
mlp310a	0.9318	10.2745
mlp310b	0.1757	11.9305
mlp310c	-0.8034	12.1493
mlp310d	0.4385	13.7382
mlp4a	0.7704	21.5725
mlp42a	-0.0219	11.3278
mlp46a	0.4160	10.4227
mlp410a	-0.0087	11.4360

† refer to appendix G