INVESTIGATING THE ABILITY OF ACTIVE LEARNING METHOD FOR CHLOROPHYLL AND PIGMENT RETRIEVAL IN CASE-I WATERS USING SEAWIFS WAVELENGTHS

H. Taheri Shahraini [a,*] , S. Bagheri [b], F. Fell [c], A. Abrishamchi [a], M. Tajrishy [a], J. Fischer [d]

[a] Dept. of Civil Eng., Sharif University of Technology, Tehran, Iran - h_taheri@mehr.sharif.edu, (abrisham, tajrishy)@sharif.edu
[b] Dept. of Computer Eng., Sharif University of Technology, Tehran, Iran - Bagheri-S@sharif.edu
[c] Informus GmbH, Berlin, Germany - fell@informus.de
[d] Free University of Berlin, Institute for Space Science, Berlin, Germany - fischer@zedat.fu-berlin.de

Commission I, WG VI/4

KEY WORDS: Sea, Colour, Retrieval, Algorithm, Fuzzy logic, Multispectral, Database

ABSTRACT:

In order to evaluate and compare the performance of various ocean color chlorophyll or pigment algorithms, a data set containing in situ chlorophyll, pigment and remote sensing reflectance at the SeaWiFS wavelengths acquired at 919 stations has been compiled in the frame of the SeaBAM activities. 900 of these stations were located in case-1 waters and 19 stations in case II waters, respectively. Up to now many ocean chlorophyll models have been developed for modeling chlorophyll and pigment. In this paper, a new modeling method has been used for modeling SeaBAM data which has been named Active Learning Method (ALM). ALM has been innovated by Bagheri and Honda (1997a). The ALM is a fuzzy modeling method of an unknown system. This modeling method not only is similar to human logical thinking but also avoids mathematical complexity. The heart of this new method is fuzzy interpolation. The mentioned algorithm is based on searching for narrow paths on two-dimensional surfaces where each path approximates the relationship between two selected parameters from a multi-dimensional system. The results of ALM give RMSE=0.152 and R^2=0.932 for chlorophyll modeling with 900 data, RMSE=0.163 and R^2=0.930 for chlorophyll modeling with 919 data, and RMSE=0.159 and R^2=0.931 for pigment modeling with 919 data. The corresponding results for a selection of other chlorophyll or pigment retrieval schemes are:
OC4: RMSE=0.151 and R^2 =0.928 for chlorophyll modeling with 900 data, OC4: RMSE=0.156 and R^2 =0.932 for chlorophyll modeling with 919 data, ANN: RMSE=0.148 and R^2 =0.934 for chlorophyll modeling with 900 data and OP2: RMSE=0.171 and R^2 =0.917 for pigment modeling with 919 data.
In comparison with other retrieval models, the ALM model shows good performance among different chlorophyll and pigment models. According to different statistical and graphical evaluation criteria, ALM shows excellent results. Therefore, the chlorophyll and pigment concentrations can be retrieved from remote sensing reflectance at the SeaWiFS wavelengths using very simple model (ALM) and without mathematical complexity. The ALM can reach to more accuracy using dividing of the space of variables.

1. INTRODUCTION

Chlorophyll as the most important photosynthetic pigment is an indicator of the amount of phytoplanktonic biomass (Falkowski, 1994). Monitoring of the amount of phytoplankton in water bodies is of high interest, because of its different ecologic, economic and sanitary effects (Liew and Kwoh, 2003). The costs of traditional in situ sampling and analysis are high and also time consuming. Large spatio-temporal variations of chlorophyll concentration in water bodies are observed, therefore the traditional methods are not always appropriate, because these methods only provide information at a limited number of sampling points each time of measurement (Harma et al., 2001). Remote sensing of water can be a useful tool for monitoring of water bodies (Yang et al., 2000) and can monitor the spatio-temporal variations of chlorophyll concentration in water bodies.
The water quality monitoring using remote sensing data should be performed by specific sensors referred to as ocean color sensors. The ocean color sensors have high spectral resolution and signal to noise ratio (SNR) (Liew and Kwoh, 2003). These sensors usually have more bands than earth observation sensors in the visible range. Operational satellite ocean color research began in October 1978 with the Coastal Zone Color Scanner (CZCS) sensor aboard the Nimbus-7 satellite. Phytoplanktons are the major contributor to ocean color in offshore water (Evans and Gordon, 1994), and CZCS could monitor and quantify the phytoplankton chlorophyll concentrations in the open oceans (Gordon et al., 1983). The SeaWiFS sensor is the first of a new generation of ocean color sensors. Table 1 shows its characteristics.
In January 1997, NASA convened a working group (SeaWiFS Bio-optical Algorithm Mini-Workshop; hereinafter referred to as SeaBAM) whose primary goal was the identification of chlorophyll-a (C) and chlorophyll-a + phaeopigments (C+P) algorithms suitable for operational use by SeaWiFS (Firestone and Hooker, 1998). The existence of large database is necessary for development of empirical models and evaluation of their performances. For this purpose, the data from different source have been compiled and combined. This new data set has been named SeaBAM data set and contains 919 concurrent chlorophyll-a concentration (C), pigment concentration (C+P) and spectral

* Corresponding author

remote sensing reflectance ($R_{rs}(\lambda)$) data. 900 data of this database are considered to be located in case I waters. The SeaBAM database covers a chlorophyll range of 0.03 – 30 (mg/m^3). The central wavelengths for spectral remote sensing reflectance are 412, 443, 490, 515, 555 and 670 nm. The SeaBAM data set is freely available via internet (http://seabass.gsfc.nasa.gov). The details about the data sources, the methods used to combine and process data from different sources and processes have been presented in O'Reilly et al. (1998), Maritorena et al. (2002) and O'Reilly and Maritorena (2002).

Central wavelengths of visible bands (nm)	412, 443, 490, 510, 555, 670
Satellite	Sea Star
Spatial Resolution (m)	1100
Swath Width (Km)	2800

Table 1. Characteristics of SeaWiFS

Many different chlorophyll-a and pigment empirical models have been tuned to the SeaBAM data set (see e.g. O'Reilly et al. (1998), O'Reilly and Maritorena (2002) or Zhang et al. (2003). Characteristics and performance of the different methods are presented in Table 2.

All of the models in Table 2 are chlorophyll-a algorithms and their slope, intercept (when a line is fitted to the log transformed of observed and modeled chlorophyll-a) and bias error values (between the log transformed of observed and modeled chlorophyll-a) are 1.0, 0.00 and 0.00, besides OP2 algorithm which is pigment retrieval scheme with slope, intercept and bias values of 0.996, 0.012 and 0.014, respectively. Slope, intercept and bias error values for the ANN algorithm by Zhang et al. (2003) have not been presented and are unknown.

In this paper, a new method (Active Learning Method (ALM)) is presented for modeling of SeaBAM data, which is similar to human logical thinking, and avoid of mathematical complexity.

In section 2, Active Learning Method is described briefly. Some evaluation criteria are presented in section 3 for comparison between ALM and other algorithms. Results of chlorophyll and pigment modeling of SeaBAM dataset using ALM and discussion about the results has been presented in section 4. Also the results of this new modeling method will be compared with the results of other appropriate models in section 4. Finally, a brief conclusion has been presented in section 5.

2. ACTIVE LEARNING METHOD

Active Learning Method (ALM) is a fuzzy modeling method which has been developed by Bagheri and Honda (1997a).

The main idea behind ALM is the assumption that human learning (or any modeling procedure) forms by considering the behavior of complicated systems as collection of simple systems which are single input single output (SISO) systems. This modeling method not only is similar to human logical thinking but also avoids mathematical complexity.

Algorithm	Type*-Wavelength (nm)	N	R^2	RMSE
OC1d	Cubic-490, 555	919	0.918	0.172
OC2	MCP-490, 555	919	0.918	0.172
OC3d	MCP-443, 52, 555	919	0.928	0.161
OC4	MCP-443, 490, 510, 555	919	0.932	0.156
OP2	MCP-490, 555	919	0.917	0.171
ANN	NN-443, 490, 510, 555	900	0.934	0.148
OC2	MCP-490, 555	900	0.918	0.162
OC4	MCP-555, 443, 490, 510, 555	900	0.928	0.151

Table 2. Statistical results for chlorophyll and pigment algorithm tuned to SeaBAM data (O'Reilly et al., 1998; Zhang et al., 2003; O'Reilly and Maritorina, 2002)

For performing ALM, all of the measured data need to be projected on x-y planes, where x and y are inputs and outputs, respectively. Figure 1a shows the exerted projection process on the normalized $\log(R_{rs}(443)/ R_{rs}(555))$ data of SeaBAM data set.

For each projected plane, a narrow path shall be extracted. This operation is performed using a specific fuzzy interpolation and curve fitting technique referred to as Ink Drop Spread (IDS) method (Bagheri and Honda, 1997a). This method searches for continuous possible paths on the interpolated data points on each plane. In this method, we assume that each data point on each data plane is a light source, which has a cone shape illumination pattern. As the distance from these light sources increases, their illumination pattern will interfere and generate new bright areas. Figure 1b shows the exerted IDS algorithm on Figure 1a. We assume the cones are fuzzy membership functions for data points. Then the narrow path is extracted by applying the center of gravity method on the IDS results (Figure 1c).

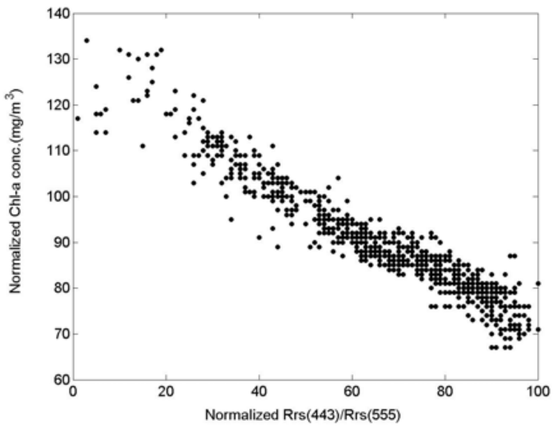
The spread functions, which show the amount of spread of data on each plane resulting from the effects of other variables, can be calculated using a method presented in Bagheri and Honda (1999). Then the output of the system can be calculated by equation 1.

$$y = \frac{\left[\frac{1}{a_1} f_1(x_1) + \frac{1}{a_2} f_2(x_2) + \frac{1}{a_n} f_n(x_n)\right]}{\left(\frac{1}{a_1} + \frac{1}{a_2} + \ldots + \frac{1}{a_n}\right)} \qquad (1)$$

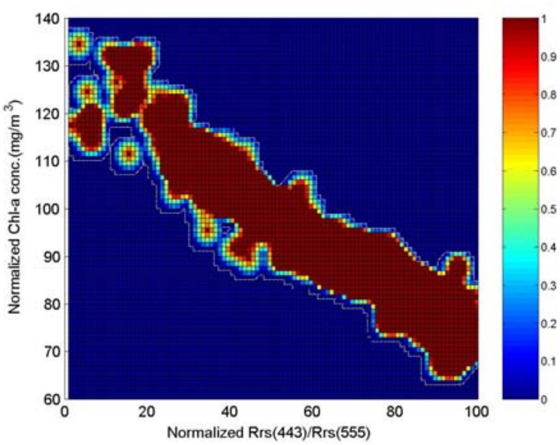
* Cubic: Cubic polynomial ($C = 10^{(a_0 + a_1 R + a_2 R^2 + a_3 R^3)}$), NN: feed forward neural network trained using back propagation method, MCP: Modified Cubic Polynomial ($C = 10^{(a_0 + a_1 R + a_2 R^2 + a_3 R^3)} + a_4$).

Where y = the output of system (function)
$x_1, x_2, \ldots, x_n$ = inputs of the system (variables)
$f_1, f_2, \ldots, f_n$ = the narrow path functions for plane x-y for each variable
$a_1, a_2, \ldots, a_n$ = spread values.

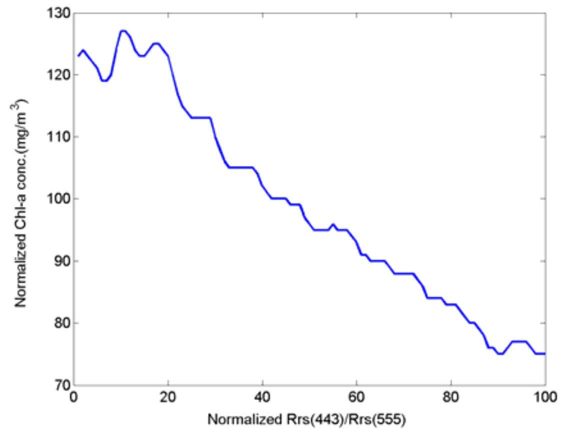
If the result of the first step was not acceptable, in the next step, the space of variables can be divided to more parts and a rule base will be generated for modeling. This dividing method will continue until the given model generates acceptable results. For more information, refer to Bagheri and Honda (1997a, 1997b, 1998 and 1999).
In this study, we do not use dividing of space of variables and will investigate the performance of only the first step of ALM.



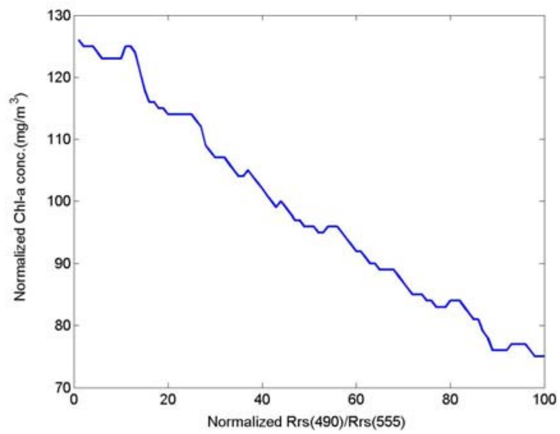
(a)



(b)



(c)



(d)

Figure 1. a) The normalized initial point data, b) Exerted IDS method on figure 1a data, c) Narrow path for $R_{rs}(443)/R_{rs}(555)$, and d) Narrow path for $R_{rs}(490)/R_{rs}(555)$
(In x-axes, Log-transformed data have been normalized)

	Evaluation
	Statistical
Regression slope	1 ± 0.01
Regression intercept	0 ± 0.01
Bias	0 ± 0.01
R^2	> 0.9
RMS	< 0.185
Negative estimates	None
	Graphical
Scatter	Linear distribution; few outliers (model: in situ, < 5:1 and > 1:5)
Quantile-quantile	Linear; data overlap the 1:1 line; no discontinuities
Relative frequency	Congruency with in situ data

Table 3. Criteria for model evaluation (Log-transformed data) (O'Reilly et al., 1998)

3. EVALUATION CRITERIA

Some statistical and graphical criteria have been defined by O'Reilly et al. (1998) and other researchers for evaluating the performances of different models. Table 3 lists these statistical and graphical criteria. These criteria should be applied on log-transformed data. In a quantile-quantile (q-q) plot, the ordered model data will be drawn against the ordered in situ data. For more information about this plot drawing method refer to Chambers et al. (1983).

4. RESULTS AND DISCUSSION

All of the programs for ALM modeling have been written and run using MATLAB 6.5.

SeaBAM chlorophyll-a and pigment concentrations and spectral remote sensing reflectance data have been used in this study. In this problem, output of system is chlorophyll-a or pigment concentrations and variables are $\log(R_{rs}(412)/R_{rs}(555))$, $\log(R_{rs}(443)/R_{rs}(555))$, $\log(R_{rs}(490)/R_{rs}(555))$ and $\log(R_{rs}(510)/R_{rs}(555))$. The ALM method has been applied to different combinations of these variables and the results have shown that the best results are observed when using $\log(R_{rs}(443)/R_{rs}(555))$ and $\log(R_{rs}(490)/R_{rs}(555))$ as variables. The initial models using these variables are depicted in equations 2 and 3.

$$C = \frac{\left[\frac{1}{a_1} f_1\left(\log\frac{R_{rs}(443)}{R_{rs}(555)}\right) + \frac{1}{b_1} g_1\left(\log\frac{R_{rs}(490)}{R_{rs}(555)}\right)\right]}{\left(\frac{1}{a_1}+\frac{1}{b_1}\right)} \tag{2}$$

$$C+P = \frac{\left[\frac{1}{a_2} f_2\left(\log\frac{R_{rs}(443)}{R_{rs}(555)}\right) + \frac{1}{b_2} g_2\left(\log\frac{R_{rs}(490)}{R_{rs}(555)}\right)\right]}{\left(\frac{1}{a_2}+\frac{1}{b_2}\right)} \tag{3}$$

Where f_1, g_1, f_2 and g_2 = narrow path functions
a_1, b_1, a_2 and b_2 = spread values
C = chlorophyll-a concentration
C+P = pigment (chlorophyll-a+ Phaeopigments) concentration.

Figure 1c and Figure 1d show f_1 and g_1, when modeling chlorophyll-a using 919 SeaBAM data. When only 900 data are used for modeling, the functions (f_1, g_1, f_2 and g_2) and spread values (a_1, b_1, a_2 and b_2) will slightly change.

The statistical results of modeling 919 and 900 SeaBAM data using ALM are presented in Table 4.

As one can see, slope, intercept and bias of the linear regression between the in situ log(C) and modeled log(C) are out of the range specified in Table 3. Figure 2 shows the linear regression results on the scatter plot of modeled and observed data. Linear shifting of models are necessary in these conditions. According to the regression slope and intercept values, linear functions have been added to the initial models (Equations 2 and 3). Therefore, the final models can be written as equations 4 and 5.

$$C = \left(\frac{\left[\frac{1}{a_1} f_1\left(\log\frac{R_{rs}(443)}{R_{rs}(555)}\right) + \frac{1}{b_1} g_1\left(\log\frac{R_{rs}(490)}{R_{rs}(555)}\right)\right]}{\left(\frac{1}{a_1}+\frac{1}{b_1}\right)} - I_1 \right) \Bigg/ S_1 \tag{4}$$

$$C+P = \left(\frac{\left[\frac{1}{a_2} f_2\left(\log\frac{R_{rs}(443)}{R_{rs}(555)}\right) + \frac{1}{b_2} g_2\left(\log\frac{R_{rs}(490)}{R_{rs}(555)}\right)\right]}{\left(\frac{1}{a_2}+\frac{1}{b_2}\right)} - I_2 \right) \Bigg/ S_2 \tag{5}$$

Where I_1, I_2 = the intercepts
S_1, S_2 = the slopes of the linear regressions, which have been presented in Table 4.

Algorithm	ALM-chla	ALM-pigment	ALM-chla
N	919	919	900
R^2	0.930	0.931	0.932
RMSE	0.157	0.154	0.147
Slope	0.931	0.927	0.933
Intercept	-0.043	-0.032	-0.043
Bias	0.0036	0.0007	0.0034

Table 4. Statistical results of ALM model for chlorophyll and pigment modeling

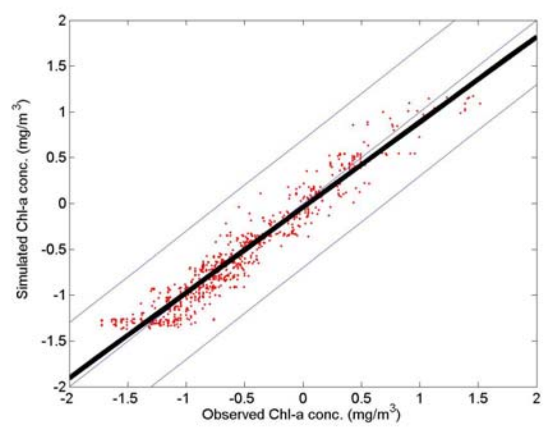
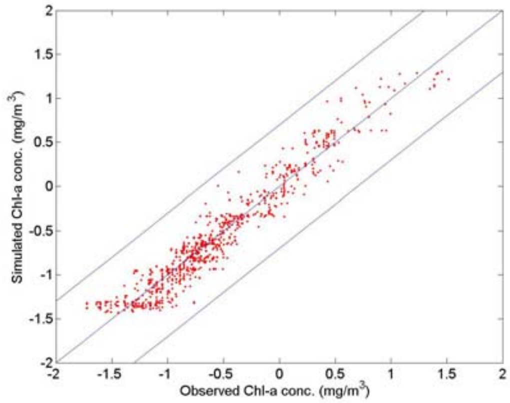


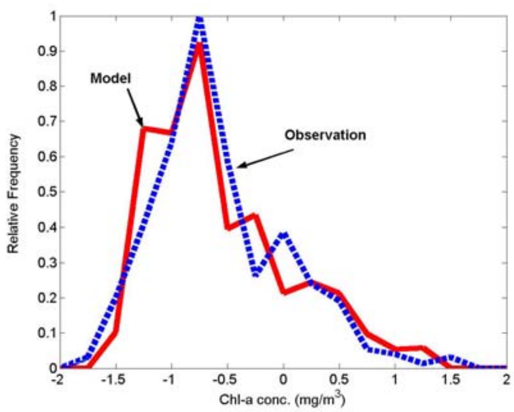
Figure 2. Simulated chlorophyll-a data using ALM versus observed data (points), with interpolated line (thick line), simulation: observation= 5:1, 1:5 and 1:1 (narrow lines) (All data are Log-transformed)

The statistical results of final models have been shown in Table 5. According to Table 3 and Table 5, the statistical results of final model are satisfactory.

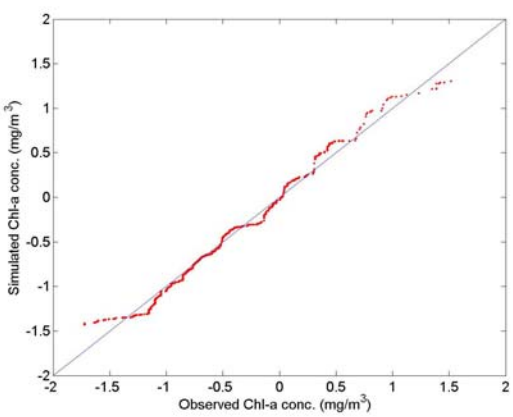
Figures 3a-3c show the graphical results of final ALM model for modeling chlorophyll-a using 919 SeaBAM data.



(a)



(b)



(c)

Figure 3. a) Simulated chlorophyll-a data using ALM after linear shifting versus observed data(points), simulation: observation= 5:1, 1:5 and 1:1 (narrow lines), b) Q-Q diagram of ALM results after linear shifting, c) Relative frequency distribution of ALM results after linear shifting and observation data.

According to Figures 3a-3c, the graphical results are acceptable and satisfy the graphical conditions (Table 3). However, the q-q diagram does not overlap the 1:1 line for very small chlorophyll-a concentrations.

For better evaluation of performance of ALM, its results are compared with other appropriate models. Tables 2, 4 and 5 are used to perform this comparison.

Algorithm	N	R^2	RMSE
ALM-chla	919	0.930	0.163
ALM-pigment	919	0.931	0.159
ALM-chla	900	0.932	0.152

Table 5. Statistical results of ALM model after linear shifting (for all of the states, slope=1, intercept=0.0 and bias=0.00)

The comparison between ALM and other models shows that the ALM has better statistical results than OC1d, OC2 and has similar results as compared to the OC3d model in chlorophyll-a modeling. When 919 SeaBAM data are used for modeling, the OC4 shows better statistical results than ALM but when Case-I water data (900) are used for modeling, the ALM shows better statistical results. Also, the ALM has better statistical results than OP2 for pigment modeling. However, the ANN model shows better R^2 and RMSE values than ALM after linear shifting (Table 5), but we have no information about the slope, intercept and bias of the ANN model. A comparison between ANN and ALM without linear shifting (Table 4) shows similar statistical results.

In addition, the ALM not only is very simple modeling method, without mathematical complexity and similar to human modeling method, but also can be considered as one of the best chlorophyll-a and pigment modeling methods.

5. CONCLUSIONS

Active Learning Method (ALM) not only is similar to human logical thinking and avoid of mathematical complexity, but also can model chlorophyll-a and Pigment using SeaBAM data set.

All statistical and graphical results of ALM are satisfactory. In comparison with other successful models, ALM shows very good results. According to these results, ALM can be considered as a very suitable method for ocean color constituent retrieval in case I waters. This method can estimate the chlorophyll-a and pigment concentrations using the spectral remote sensing reflectance ratios ($\log(R_{rs}(443)/R_{rs}(555))$, $\log(R_{rs}(490)/R_{rs}(555))$). Therefore, this modeling method has the ability for chlorophyll-a and pigment retrieval of local and global water bodies using SeaWiFS wavelengths.

REFERENCES

Bagheri, S. and N. Honda, 1997a. A new method for establishing and saving fuzzy membership functions. In: *13th Fuzzy Symposium*, Toyama, Japan, pp. 91-94.

Bagheri, S. and N. Honda, 1997b. Outlines of a living structure based on biological neurons for simulating the active learning method. In: *7th Intelligent Systems Symposium*, Sapporo, Japan, pp. 183-188.

Bagheri, S. and N. Honda, 1998. Fuzzy prediction, a method for adaptation. In: *14th Fuzzy Symposium*, Gifu, Japan, pp. 317-320.

Bagheri, S. and N. Honda, 1999. Recursive fuzzy modeling based on fuzzy interpolation. *Journal of Advanced Computational Intelligence*, 3(2), pp. 114-125.

Chambers, J. M., W. S. Cleveland, B. Kleiner and P. A. Tukey, 1983. *Graphical Methods for Data Analysis*. Wadsworth, Belmont, 395 p.

Evans, R. and H. R. Gordon, 1994. Coastal zone color scanner 'system calibration': A retrospective examination. *Journal of Geophysical Research*, 99(C4), pp. 7293-7307.

Falkowski, P. G., 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis Research*, 39(3), pp. 235-258.

Firestone, E. R. and S. B. Hooker, 1998. SeaWiFS prelaunch technical report series final cumulative index. Tech. Memo., TM-1998-104566 4-8, Vol. 43, NASA, USA.

Gordon, H. R. , D. K. Clark, J. W. Brown, O. B. Brown, R. H. Evans and W. W. Broenkow, 1983. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates. *Applied Optics*, 22(1), pp. 20-36.

Harma, P., J. Vepsalainen, T. Hannonen, T. Pyhalahti, J. Kamari, K. Kallio, K. Eloheimo and S. Koponen, 2001. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland. *The Science of the Total Environment*, 268(1-3), pp. 107-121.

Liew, S. C. and L. K. Kwoh, 2003. Monitoring algal blooms from space: possibilities and limitations. In: *Workshop on Red Tide Monitoring in Asian Coastal Waters*, Tokyo, Japan, pp. 1-3.

Maritorena, S., J. O'Reilly and B. D. Schieber, 2002. " The SeaBAM Evaluation Data Set", Greenbelt, USA. http://seabass.gsfc.nasa.gov/seabam/pub/maritorena_oreilly_schieber/ (accessed 10 Nov. 2004)

O'Reilly, J., S. Maritorena, B. Mitchell, D. Siegel, K. Carder, S. Garver, M. Kahru, and C. McClain, 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103(C11), pp. 24937-24953.

O'Reilly, J. and S. Maritorena, 2002. "SeaBAM algorithm Evaluation". Narragansett, USA. http://seabass.gsfc.nasa.gov/seabam/pub/oreilly_maritorena/ (accessed 10 Nov. 2004)

Yang, M. D., R. M. Sykes and C. J. Merry, 2000. Estimation of algal biological parameters using water quality modeling and SPOT satellite data. *Ecological Modelling*, 125, pp. 1-13.

Zhang, T., F. Fell, Z. S. Liu, R. Preusker and J. Fischer, 2003. Evaluating the performance of artificial neural network techniques for pigment retrieval from ocean color in Case I waters. *Journal of Geophysical Research*, 108(C9), Art. No. 3286.