Artificial neural networks (ANN) based algorithms for chlorophyll estimation in the Arabian Sea

*Prakash Chauhan, P.V. Nagamani & Shailesh Nayak
Marine and Water Resources Group, Space Applications Centre, Ahmedabad 380015, India

*[E-mail: prakash@sac.isro.org]*

Received 1 November 2004; revised 31 May 2005

In-situ bio-optical measurements were collected during six ship campaigns in the north eastern Arabian Sea using SeaWiFS Multi-channel Profiling Radiometer (SPMR). An artificial neural network (ANN) based algorithms were constructed to estimate oceanic chlorophyll concentration using in-situ data. The different ANNs were obtained by systematic variations of architecture of input and hidden layer nodes for the Arabian Sea training data set. The performance of individual ANN-based pigment estimation algorithm was evaluated by applying it to the remote sensing reflectance data contained in validation data set. The performance of the most successful ANN was compared with commonly used empirical pigment algorithms. Compared to e.g. the SeaWiFS algorithms Ocean Chlorophyll-2 (OC2) and Ocean Chlorophyll-4 (OC4), the square of the correlation coefficient $r^2$ is increased from 0.69 for OC4, respectively 0.70 for OC2 to 0.96 for ANN algorithm. The RMS error of the estimated log-transformed pigment concentration dropped from 0.47 for OC2, respectively 0.41 for OC4 to 0.11 for ANN-based pigment algorithm.

[Key words: Artificial Neural Network (ANN), ocean colour, chlorophyll, Arabian Sea, algorithms]

[Int.Code: Int.Cl.7 G06N 3/02]

Introduction

Satellite remote sensing of ocean colour variables, such as concentrations of chlorophyll, suspended sediments and yellow substance have been now recognised as important tools for providing inputs for studying bio-geochemical cycles and oceanic primary production on a regional and global scale. A large number of ocean colour instruments such OCM, MODIS-AQUA/TERRA, MERIS and SeaWiFS are now available for continuously measuring the ocean colour of the global oceans from which these variables can be derived. More such instruments are scheduled for the future, so that the availability of global ocean colour data is assured for a long term. To derive regional or global fields of biogeochemical variables from satellite ocean colour data, fast and robust retrieval procedures are required.

Traditionally, empirical chlorophyll algorithms are used to generate chlorophyll images on a routine basis from satellite ocean colour data. Developers of empirical chlorophyll algorithms have employed regression analysis to fit in-situ field data to power function and polynomials. This approach has its limitation, however, in part because of the nonlinear nature of the transfer function between marine reflectance and chlorophyll concentration. Standard liner regression does not model nonlinear relations well except over small ranges, and nonlinear regression requires a priori knowledge of the nature of the nonlinear behaviour of transfer function, which is always not possible. The estimation of chlorophyll concentration from ocean-colour measurements is a parameter estimation problem, where a set of parameters $P = \{p_i, i=1,\ldots, I\}$ is estimated from a set of measurements $M = \{m_j, j=1,\ldots, J\}$. The functional relationship between measurements and parameters can be expressed as:

$$M = f(P) \quad \ldots (1)$$

By inverting equation (1), the set of parameters $P$ can be obtained from set of measurements $M$:

$$P = f^{-1}(M) \quad \ldots (2)$$

In the present context, $P$ represents the pigment concentration (mg m$^{-3}$), while $M$ is remote sensing reflectance $R_{rs}$ [sr$^{-1}$] in different spectral bands. In Eq (1), $f$ is the function describing the relationship between chlorophyll concentration and remote-sensing reflectance, which is a complex and nonlinear function. Since $f$ is complex and nonlinear it is
difficult to achieve an analytical inversion of this functional form. The traditional way to overcome this problem is to make assumptions on the functional form of $f^{-1}$ and then solve the Eq (2) by regression techniques\textsuperscript{1}. However, it is often difficult to find the appropriate functional form for $f^{-1}$, which has direct effect on the accuracy of the retrieved chlorophyll concentrations.

In recent years, artificial neural network (ANN) based techniques have been found to be well suited for solving the nonlinear problem as they allow to approximate the nonlinear relationship between observations and target parameters without explicitly knowing their functional dependence\textsuperscript{2}. Artificial neural network approach has several distinct advantages compared to other techniques such as principal component analysis. First, development of an inversion algorithm is not needed for the retrieval of ocean-colour constituents\textsuperscript{3}. Second, after a neural network has been created, processing time for parameter retrieval is known and is comparatively very short\textsuperscript{4}. Neural networks have been used in a number of geophysical applications, but it is recently that attempts have been made to estimate optically-active ocean variables with the help of neural network based algorithms\textsuperscript{5,6}.

In this paper, we have used ANN based methodology for the estimation of the pigment concentration in Case 1 waters of the northeastern Arabian Sea, adjoining to the west coast of India (Fig.1). Inputs to the analysis are the in-situ spectral remote-sensing reflectance and chlorophyll concentration data collected in the study area. A comparison of the chlorophyll pigment retrieval using neural network algorithms and with that from conventional empirical algorithms is also presented.

**Materials and Methods**

Bio-optical data obtained during cruises conducted in the north eastern Arabian Sea has been used for this study. This data set contains remote-sensing reflectance ($R_m$) and normalized water leaving radiance ($L_{uw}$) at seven wavelengths (i.e. 412, 443, 490, 510, 555, 670 and 780 nm) and in-situ chlorophyll concentration at the stations where $R_m$ and $L_{uw}$ were measured. Sea-truth collection campaigns were conducted onboard the Research Vessel ORV-Sagar Kanya and FORV-Sagar Sampada as a part of IRS-P4 OCM calibration-validation experiments during April 2000 to February 2003. The study area within the Arabian Sea covers the region between 13°-23° N latitude and 65° - 72°E longitude, with most of the sampling stations representing Case 1 waters. In-situ data collected during four cruises in three different seasons has been used in the present work. Table 1 shows the summary of the data collection campaigns and number of data points collected. Total numbers of 159 remote sensing reflectance spectra were measured during these campaigns.

All the optical data were processed according to the SeaWiFS optical data processing protocols\textsuperscript{7}. The downwelling irradiance just below the sea surface [$E_d(0',\lambda)$] and the upwelling radiance just below the sea surface [$L_{uw}(0',\lambda)$] were calculated by performing a least squares fit for the statistical regression of depth ($z$) versus $\log(E_d)$ and $\log(L_{uw})$, respectively, and projecting the best-fit curves to zero depth (i.e. sea surface). The upwelling radiance just below the ocean surface, [$L_{uw}(0',\lambda)$], was converted to water leaving radiance, $L_{uw}(\lambda)$, using $L_{uw} = 0.54 L_{uw}(0')$, where 0.54 is a mean coefficient summarizing the effect of internal reflection of the upwelling flux during transmission through the air-sea interface. The remote sensing
Reflectance, $R_{rs}(\lambda)$, the ratio of water leaving radiance to the downwelling irradiance just above the sea surface $E_d(0^+,\lambda)$, was computed using the following equation:

$$R_{rs}(\lambda) = \frac{L_w(\lambda)}{E_d(0^+,\lambda)} \quad \ldots (3)$$

where $E_d(0^+,\lambda)$ was calculated from $E_d(0^-,\lambda)$, using the following equation:

$$E_d(0^+,\lambda) = \frac{E_d(0^-,\lambda)}{0.96} \quad \ldots (4)$$

where 0.96 is the transmittance across the air-sea interface, assuming a normal incidence angle. Fluorometric profiles of chlorophyll were obtained using the Atlantic Wetstar fluorometer.

### Results and Discussion

The chlorophyll concentration ranged between 0.07 and 1.6 mg m$^{-3}$ with a mean value of 0.39 mg m$^{-3}$, which is relatively higher than the global ocean mean value of chlorophyll (0.19 mg m$^{-3}$) reported by Antoine et al.$^9$. The standard deviation for this data set was computed to be 0.25 mg m$^{-3}$. It was observed that chlorophyll concentration values in the north eastern Arabian Sea were in middle ranges representing mesotrophic water conditions, with relatively few samples having concentration higher than 1 mg m$^{-3}$. Figure 2 shows the relationship between measured chlorophyll concentration and remote-sensing reflectance values in different spectral bands.

### Neural networks

In the present study, we use a multilayer perceptron (MLP) type of ANN to derive the functional relationship between the spectral remote sensing reflectance and the chlorophyll concentration. A ‘multilayer feed-forward’ artificial neural network was trained using the data set described in Table 1.

### Table 1 — Data sources and characteristics of the data set

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Location</th>
<th>Time Period</th>
<th>Number of points</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK-152</td>
<td>Arabian Sea</td>
<td>March 30- April 13, 2000</td>
<td>21</td>
<td>Profiles of upwelling radiance, downwelling irradiance, fluorometric chlorophyll-a</td>
</tr>
<tr>
<td>SK-171</td>
<td>Arabian Sea</td>
<td>November 3-18, 2001</td>
<td>59</td>
<td>Profiles of upwelling radiance, downwelling irradiance, fluorometric chlorophyll-a</td>
</tr>
<tr>
<td>SK-186</td>
<td>Arabian Sea</td>
<td>January 3-20, 2003</td>
<td>65</td>
<td>Profiles of upwelling radiance, downwelling irradiance, fluorometric chlorophyll-a</td>
</tr>
<tr>
<td>FORV-212</td>
<td>Arabian Sea</td>
<td>Feb. 28-Mar. 5, 2003</td>
<td>13</td>
<td>Profiles of upwelling radiance, downwelling irradiance, fluorometric chlorophyll-a</td>
</tr>
</tbody>
</table>

Fig. 2—Relationship between chlorophyll concentration and normalised water leaving radiance in (a) 412 nm, (b) 443 nm, (c) 490 nm, (d) 510 nm and (e) 555 nm for the data collected in the Arabian Sea.
consists of an input layer, at least one hidden layer and one output layer. Signals are fed to the input layer. From there, they propagate forward through any hidden layer(s) and then arrive at the output layer. During this propagation the mapping from an input signal or vector to the output vector takes place. Figure 3 shows a simple neural network with one input layer having nodes corresponding to the number of spectral bands used, one hidden layer and one output layer. At each hidden node \( j \), the inputs pass through a weighted sum to obtain an output vector \( y_j \) as:

\[
y_j = \sum_{i=1}^{n} w_{ij} x_i + b_j
\]

... (5)

where \( x_i \) are the inputs, \( w_{ij} \) are the weights associated with each input/node connection, and \( b_j \) is the bias associated with node \( j \). This sum is used in a nonlinear activation function as:

\[
z_j = g(y_j) = \tanh(y_j)
\]

... (6)

For this ANN, the hyperbolic tangent was used as activation function, but other sigmoid functions may also be employed. The output of the hidden layer, \( z_j \), acts as the input to the final node. This ANN architecture has been found useful for function approximation as in modeling the transfer function between remotely-sensed reflectance inputs and \textit{in situ} chlorophyll\(^{10} \).

**Training data set**

Neural networks ‘learn’ like the human brain, to do simple tasks. A training set of data is presented to the network, whose inputs and correct outputs are known, and the network uses the training set to adjust its own weights into a pattern that will produce the correct output for a wide range of given inputs. When training is complete, the network’s response is relatively insensitive to small variations in the input. This allows the network to generalize, i.e. to disregard noise and detect the underlying patterns. In this study we have partitioned the bio-optical data set collected in the Arabian Sea into two components, a) the training data set and b) validation or test data set. The remote-sensing reflectance patterns along with chlorophyll concentration were arranged in an increasing order of chlorophyll concentration. Every third data point was picked up to form validation data set. In this case out of 159 reflectance spectra, 106 spectra were used as training data set and remaining 53 spectra were used as validation data set.

**Performance evaluation of ANN based pigment algorithms**

The estimation of the chlorophyll pigment concentration using remote-sensing data makes use of information contained in the water-leaving radiance data. In order to determine which and how many spectral bands are best suited for pigment estimation, four combinations of the input data were used. Table 2 shows the different combinations of the input spectral bands used in this study along with number of neurons used in the hidden layer. The comparative performance of all four combinations was evaluated in terms of two error measures. First, the root mean square error (RMSE) is defined as:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [\log_{10}(Chl_i^{\text{D}}) - \log_{10}(Chl_i^{\text{M}})]^2}
\]

... (7)

where, \( Chl \) represents the pigment concentration, and the superscripts \( D \) and \( M \) indicate derived and

<table>
<thead>
<tr>
<th>Input</th>
<th>Neurons in hidden layer</th>
<th>Training data (N = 104)</th>
<th>Validation data (N = 53)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>R²</td>
<td>RMSE</td>
</tr>
<tr>
<td>( R_s 412, R_{s 443}, R_{s 490}, R_{s 510}, R_{s 555} )</td>
<td>10</td>
<td>0.10 0.96</td>
<td>0.11 0.96</td>
</tr>
<tr>
<td>( R_s 443, R_{s 490}, R_{s 510}, R_{s 555} )</td>
<td>10</td>
<td>0.13 0.93</td>
<td>0.14 0.93</td>
</tr>
<tr>
<td>( R_s 443, R_{s 490}, R_{s 515} )</td>
<td>10</td>
<td>0.19 0.90</td>
<td>0.18 0.90</td>
</tr>
<tr>
<td>( R_s 490, R_{s 510}, R_{s 555} )</td>
<td>10</td>
<td>0.25 0.89</td>
<td>0.23 0.89</td>
</tr>
</tbody>
</table>

![Fig. 3—Architecture of the multilayer perceptron used in this study.](image-url)
measured values and secondly, the square of the Pearson’s correlation coefficient $r^2$. The pigment concentrations were log-transformed before calculating the RMSE and correlation coefficient. Table 2 shows the performance of all the four input combination for training data set as well as for validation data set. It is clear from Table 2 that all the four input combinations of the remote sensing reflectance values have performed well for both the training as well as the validation data sets with RMSE lesser than 0.23 and $r^2$ better than 0.89. However, the input data set with maximum number of spectral bands (i.e. five spectral bands) has outperformed the input reflectance combinations those contain only three spectral bands.

Comparison of ANN based algorithms to empirical algorithms

The developed ANN based algorithms were further compared to the most successful of the empirical algorithms complied by O’Reilly et al.\(^1\). In the present study two most popular algorithms namely Ocean Chlorophyll 2 (OC2) and Ocean Chlorophyll 4 (OC4) have been used for the comparative analysis. Both these algorithms make use of band ratios of blue to green spectral bands. The mathematical formulation of these algorithms is as follows;

\[
OC2_{-Chl} = 10^{(a_0 + a_1 R + a_2 R^2 + a_3 R^3)} + a_4 \quad \ldots (8)
\]

where, \(R = \log_{10}(R_{490}/R_{555})\) and \(a = [0.3410, -3.001, 2.811, -2.0410, -0.040]\)

\[
OC4_{-Chl} = 10^{(a_0 + a_1 R + a_2 R^2 + a_3 R^3)} + a_4 \quad \ldots (9)
\]

where, \(R = \log_{10}([R_{443}>R_{490}>R_{555}]/R_{555})\) and \(a = [0.47,-3.85,4.53,-2.44,-0.0414]\)

Figure 4 shows scatter plots between in-situ measured chlorophyll and OC2, OC4 and best of the ANN algorithm (i.e. the one with five spectral band input) derived chlorophyll concentrations, respectively. For the 53 validation points the OC2 algorithm yielded a RMSE of 0.47 and $r^2$ value of 0.70. OC4 algorithm returned a RMSE of 0.41 and $r^2$ of 0.69. The results of both these algorithms were outperformed by all the four ANN algorithms shown in Table 2. Comparing to the OC2 and OC4 and the most successful ANN algorithm, the square of the correlation coefficient $r^2$ is increased from 0.69 (OC4), respectively, 0.70 (OC2) to 0.96 (ANN). The root mean square error of the retrieved log-transformed pigment concentration dropped from 0.47 for OC2, respectively, 0.41 for OC4 to 0.11 for the ANN based pigment estimation scheme.

When compared to existing empirical algorithms like OC2 and OC4 the ANN based algorithms outperformed them, yielding better $r^2$ and RMSE values. However, the validity of such algorithms needs to be tested for other regions of the Arabian Sea, not covered in the data collection sampling plan, as the data which has been used to develop these algorithms may not represent the bio-optical conditions prevailing elsewhere in the Arabian Sea. In future it is planned that the ANN based algorithms will be tested and validated with satellite sensors such as OCM for their accuracy and consistency.

References


2 Thiria S, Mejia C, Badran F, & Crepon, M, A neural network approach for modeling nonlinear transfer functions:


