A MODIS-based estimation of chlorophyll a concentration using ANN model and in-situ measurements in the southern Caspian Sea

Salman Mahiny A.\(^1\); Fendereski F.\(^1\); Hosseini S. A\(^1\) & Fazli H.\(^2\)

\(^1\)Gorgan University of Agricultural Sciences and Natural Resources, P. O. Box: 386, Gorgan, Iran
\(^2\)Caspian Sea Ecology Research Centre (EACS), P. O. Box: 961, Sari, Iran.

[E-mail: fendereski_f@yahoo.com]

Received 27 June 2012; revised 28 November 2012

Chlorophyll-a data of the MODIS sensor with in-situ chlorophyll measurements from the southern Caspian Sea (SCS) is compared in the present study. Analysis showed an overestimation of chlorophyll-a concentration by MODIS in the area. Results also indicated a root mean square (RMS) log error of 39.4\%, for 53 coincident data points. An artificial neural network (ANN) with radial basis function was applied to the in-situ measurements and satellite imagery. It included physical-chemical properties of water as ancillary independent variables in the ANN procedure that enhanced the predictive capability of the model. Evaluation of the predictive capability of ANN approach was satisfying (RMS log error 18.9\%). Results showed retrieving chlorophyll-a concentration in the SCS from satellite is possible and will be improved through application of ANN and explanatory environmental parameters.

[Keywords: MODIS, Chlorophyll-a, ANNs, Southern Caspian Sea]

Introduction

The photosynthetic pigment, chlorophyll \(\alpha\) (chl- \(\alpha\)), has a universal distribution among all the photoautotrophic algae and cyanobacteria and is also widely used as a convenient index of phytoplankton biomass\(^1\). Thus, estimation of chl-\(\alpha\) concentration is critically integral to monitoring water quality\(^2\). Chlorophyll-\(\alpha\) concentration has also a significant economic effect in coastal and marine environments on fisheries resources and marine aquaculture development\(^3\). In all these cases, it is required to characterize and quantify large-scale temporal and spatial fluctuations in phytoplankton biomass and primary production\(^4\). It has been found that satellite imagery estimates are useful to study spatial and temporal variability of chlorophyll-\(\alpha\). These images provide regular and synoptic views of spatial distribution of chlorophyll concentration, which are unachievable by other means, and are ideally suited to cover the broad range of space and time scales associated with coastal and marine applications\(^5\).

Referring to the history of satellite data application to water bodies, we know that data from the Coastal Zone Color Scanner (CZCS) provided the first demonstration of the ability to observe the abundance and distribution of phytoplankton chlorophyll in the world’s ocean from space\(^6\). The scientific success of the CZCS mission induced the space agencies to develop a new generation of ocean color sensors. Moderate Resolution Imaging Spectro-radiometers (MODIS) detectors measure 36 spectral bands between 0.405 and 14.385 \(\mu\)m. Of the 36 spectral bands, 9 were customized for the ocean, with higher sensitivity and digitization bits than SeaWiFS. To date, only data products from MODIS/Aqua are deemed as science quality, while MODIS/Terra products are still regarded as provisional\(^7\).

These new generations of sensors are characterized by more detailed spectral information, higher spectral resolution, better calibration stability than the CZCS, and provide for the first time a global ocean daily coverage. At the same time, a parallel effort in the improvement of bio-optical and atmospheric correction algorithms has been carried out and several model-based estimation algorithms have been proposed. Most of the bio-optical algorithms estimating chlorophyll- \(\alpha\) or total pigment concentration (chlorophyll \(\alpha + \) pheopigments) from ocean radiance data have been empirically derived\(^8\). In this connection, results from the MODIS data processing algorithms are nearly identical to those from SeaWiFS. Generally, after atmospheric conditions
MAHINY et al.: ESTIMATION OF CHLOROPHYLL A CONCENTRATION

925
correction, band ratios between 443, 488, and 551-nm (OC3) are used to derive chlorophyll empirically. However, some studies have revealed a systematic and large overestimation of chlorophyll products for the MODIS and SeaWiFS algorithms (e.g. 9; 5; 10; 11). This observation has made scientists try develop new algorithms for each region to refine the results and enhance the associated performance.

There are many ways to evaluate possible relationships between dependent and independent variables. Among them, artificial neural network (ANN) approach provides the means to flexibly model nonlinear relationships. Artificial neural networks have great capacity in predictive modeling, and during ANN modeling all the characters describing the unknown situation must be presented to the trained ANN, and the identification (prediction) is then given. Due to these characteristics, the ANN approach was used in this study.

Present study consists of three different process of study. Firstly, to deploy MODIS satellite imagery for assess the chlorophyll-a in the SCS. Secondly, to assess the relationship between satellite recorded measurements and in situ data and thirdly, to achieve better results in chlorophyll estimation from satellite imagery using the ANN approach.

Materials and Methods

Study area is located between 36.33º to 47.07º N and 45.43º to 54.20º E. Sampling took place aboard the Research Vessel Guilan from the 14th to the 27th October, 2008, covering 60 stations distributed along 11 transects as shown in Fig. 1. Bottom depth ranged from 16 to 840 m.

At each station, data were measured in various depths by a portable chlorophyll fluorometer emission joined with CTD sensor. The instrument was used for temperature, pH, dissolved oxygen, conductivity, salinity, turbidity and chlorophyll a concentration. To compare chlorophyll-a concentration from ocean-color sensor with that measured in the field, chlorophyll images were downloaded (Level-2 Aqua MODIS, 1 km resolution, in HDF format) from NASA GSFC (http://oceancolor.gsfc.nasa.gov). Attempts were made to correctly read the imagery data and in the end an ENVI module written specifically for this purpose was employed. Figure 2 shows the distribution of chlorophyll a in the southern Caspian Sea based on satellite images for the date of sampling. In situ data were coincident (occurring within the same time) and collocated (occurring within a single MODIS level2 pixel). Due to poor visibility caused by clouds, 7 out of 60 stations were eliminated.

Both in situ and MODIS data were logarithmically transformed (base 10) before comparison.

Comparison was made by computation of root mean square (RMS) log error as:

\[
\text{RMS} = \sqrt{\frac{\sum \left[ \log(S) - \log(I) \right]^2}{n}} \times 100
\]

where

- \( S \) = Satellite chlorophyll
- \( I \) = in situ chlorophyll
- \( n \) = number of samples

RMS is an estimate of the error of the satellite data set and indicates the covariance between data set and in situ observations. This statistical measure provides information on the performance and uncertainty of the satellite chlorophyll data set. The decision to logarithmically transform chlorophyll data before statistical evaluation is based on the natural distribution of ocean chlorophyll, which is lognormal.

![Fig. 1—Caspian Sea and the transects for CTD stations](image1)

![Fig. 2—MODIS Aqua Level 2 chlorophyll a concentration in the southern part of the Caspian Sea for the sampling time](image2)
Stations are divided into two regions: coastal areas (defined as bottom depth >100 m) and offshore areas, and performed separate RMS log error analysis for each region. In addition, applied ANN with satellite chlorophyll, temperature, dissolved oxygen, pH and turbidity as inputs and measured chlorophyll α as output target of the model. To select input data for the model, we used statistical correlation analysis, through which we decided to leave out two input variables, i.e. conductivity and salinity.

ANNs have recently become the focus of much attention, largely because of their wide range of applicability and the power with which they can treat complicated problems. ANNs can identify and learn correlated patterns between input data sets and corresponding target values. After training, ANNs can be used to predict the output of new independent input data. Neural networks are inherently more flexible than conventional empirical models, as they are able to approximate virtually every multivariable function, provided that enough data are available for their training and that their structure is adequate. There are several examples in which different environmental parameters have been used to train the network: Scardi used ANN to predict phytoplankton primary production using surface biomass (satellite derived chlorophyll concentration), irradiance, and water temperature as inputs and phytoplankton primary production as outputs of the model; Stathakiset al. deployed ANN to forecast wheat yield using soil moisture content, above ground biomass, storage organs biomass, and remote sensing information in the form of the NDVI (Normalized Difference Vegetation Index) as input and the yield as output of the model.

Architecture used in our study was a feed-forward neural network, where the neurons were grouped into layers. All connections are feed-forward; that is, they allow information transfer only from an earlier layer to the next consecutive layers. Neurons within a layer are not connected, and neurons in non adjacent layers are not connected. Input signals are presented to the network via an input layer. The nodes in the input layers do not process input signals but pass them to one or more hidden layers, where the actual processing is done via a system of weighted connections. The hidden layers then link to an output layer, which provides the outputs of the network.

Network utilized in this study was radial basis function (RBF). RBF networks are special feed-forward networks that have a single hidden layer. Activation functions of the neurons in the hidden layer are radial basis functions, while the neurons in the output layer have simple linear activation functions. Radial basis functions are a set of predominantly nonlinear functions such as Gaussian functions that are built up into one function. Each Gaussian function responds only to a small region of the input space where the Gaussian is centered. Fig. 3 depicts our feed-forward neural network that has three layers: an input layer, a hidden layer, and an output layer. As shown in Fig. 3, the input layer consisted of 5 neurons; \( x_1, x_2, \ldots, x_5 \). In Fig. 3 and also in Table 1, it is have shown that the parameters used as inputs to the network are temperature, dissolved oxygen, pH and turbidity. These variables were shown to be ecologically related to the phytoplankton biomass. We also utilized satellite chlorophyll data as another input to the model. It is used in-situ measurements of chlorophyll concentration as output of our network; i.e. \( y \) in Fig. 3.

Activation function used in this network, which transforms the activation level of a neuron into an output signal was Gaussian function.

Table 1 shows environmental variables used as input to the ANN model. Reliability of the model estimations was examined using RMS log error of the observed chlorophyll concentrations and model predictions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Ave.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (C°)</td>
<td>22.45</td>
<td>1.25</td>
</tr>
<tr>
<td>DO (ppm)</td>
<td>7.72</td>
<td>0.36</td>
</tr>
<tr>
<td>pH</td>
<td>8.24</td>
<td>0.1</td>
</tr>
<tr>
<td>Turbidity (FTU)</td>
<td>2.97</td>
<td>2.46</td>
</tr>
<tr>
<td>Satellite chl-α (mg/m³)</td>
<td>2.78</td>
<td>0.89</td>
</tr>
<tr>
<td>Chl-α (mg/m³)</td>
<td>1.21</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Fig. 3—Schematic representation of the RBF adopted for the retrieval of Chl-α. RBF architecture consists of five inputs, ten hidden units, and one output.
Results

Field measurements of chlorophyll-a concentration varied from 0.65 to 2.17 (mg/m$^3$) compared to MODIS Aqua Level 2 chlorophyll data (1.4 to 4.86 (mg/m$^3$)). Comparison of in situ and MODIS chlorophyll values along all 11 transects showed MODIS overestimated chlorophyll in the area. As can be seen in Table 2, RMS log error for this comparison was 39.4% and the average error was 1.59.

Results from separating coastal stations and using only offshore ones showed RMS log error of 38.9% for offshore stations (a total of 31 stations) compared to 39.4% RMS log error observed for all stations, indicating no significant difference between the two in this study.

Reliability of the model estimations was examined using RMS log error of observed chlorophyll concentrations and model predictions. Table 2 shows the RMS log error and average error, after applying the ANNs to the test data (18.9% and 0.47, respectively).

Discussion

Typically, coastal and inland water bodies are classified as Case 2 water bodies, where agents other than phytoplankton such as suspended inorganic particles and/or dissolved organic matter (and perhaps even a bottom effluent in shallower areas) make a significant contribution to the optical properties. In Case 2 waters, these agents vary independently of phytoplankton and each other$^{10}$, optical properties of Case 2 waters are very different from those of Case 1 waters$^{17}$. Attempts are made to separate coastal stations (case 2) to see if the observed bias would be reduced using only offshore ones (case 1)$^{11}$. Results showed no significant difference between the two in this study.

Our objective was also to investigate whether the bias observed in the previous results could be reduced using ANNs approach. A brief comparative look at the results in Table 2 shows that using the ANN model along with environmental variables gives a good prediction of chlorophyll concentration from MODIS imagery.

The large size of the Caspian Sea requires space borne sensors to monitor phytoplankton pigment concentration at adequate time and space scales.$^{18}$ Common feature of the Caspian Sea is high freshwater inflow, with the coastal areas mainly classified as case 2 waters. In such waters, as opposed to case 1 waters, variability of seawater optical properties are influenced not just by phytoplankton and the materials associated with it, but also by other substances independent of phytoplankton. Standard algorithms for processing of satellite ocean color data, derived mainly from case1 waters, break down in optically-complex case 2 waters. Standard algorithms are generally based on regression equations meaning definite relationship between the chlorophyll $\alpha$ and yellow substance (eroded materials and pollutants) absorption coefficients. If seawater absorption increases due to enhanced content of yellow substance brought by rivers, the algorithm attributes the additional absorption to increasing chlorophyll concentration and overestimates the latter.

Regional processing algorithms for the Caspian Sea were developed on the basis of field data. Like the northern part, the southern part of the Caspian Sea is influenced by the runoff of the Kura River and the rivers flowing down the Iranian coast. Part of Volga waters also reach southern areas and affect the bio-optical characteristics in the region. Formation and variability of the bio-optical characteristics in each sub-region are mainly determined by internal processes, although the interrelationships between the sub-regions certainly exist.$^{19}$ It seems the algorithms explaining relationship between satellite imagery and in situ data are based on field records from the northern part of the Sea. To date, no official report on such studies is available for the southern part of the Caspian Sea. We suggest that the developed regional algorithms may not work well in the southern sub-region of the Caspian Sea. Our study provides a basis for such comparisons while directly retrieving chlorophyll concentration from satellite data, based on other environmental data collected in the field.

This work can be regarded as a basis for developing a regional algorithm for our region and like many other marine ecosystems in the world, proposing a local algorithm for the SCS is also recommended.

In this study, we used Aqua MODIS chlorophyll data. Further studies need to be done using imagery from other available satellites sensors for achieving

<table>
<thead>
<tr>
<th>Table 2—RMS Log Error and Bias between In Situ Data and the Aqua MODIS and ANN Prediction Chlorophyll Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS Log Error %</td>
</tr>
<tr>
<td>In Situ vs. Satellite Chl-$\alpha$ Data</td>
</tr>
<tr>
<td>In Situ vs. Predicted Chl-$\alpha$ Using ANN</td>
</tr>
</tbody>
</table>
better estimations. In addition, we used in situ measurements as input variables to the model. It was clear to us that for more accurate retrieving of chlorophyll concentration from satellite data, we required more accurate environmental data. This can be done using systematic sampling efforts guided by initial assessment of information acquired from satellite imagery. We also suggest that applicability of satellite derived environmental data (e.g., SST) and some stable parameters (e.g., bathymetry) need to be examined.

Conclusion

Moderate agreement between MODIS and in situ chlorophyll data, together with satisfying results from ANN model wherein ecologically important parameters were also included, reveals the applicability and necessity of further studies on extracting chlorophyll-a from satellite imagery in the area. For this, suggest simultaneous sea sampling and application of modern techniques like ANN. Also should not neglect developing a local algorithm for this purpose in the area.

Acknowledgements

Authors thank the Captain and crew members of the R/V Guilan and the Caspian Sea Ecological Research Centre.

References