September 2005
Vol. 2

ISSN: 021-6739

INTERNATIONAL JOURNAL OF

REMOTE SENSING AND EARTH SCIENCES

Published by
International Society of Remote Sensing and Sciences
IReSES

VERTICAL DISTRIBUTION OF CHLOROPHYLL-A BASED ON NEURAL NETWORK

TAKAHIRO OSAWA[1] CHAO FANG ZHAO[2], NUARSA I WAYAN [1] SWARDIKA, I KETUT[1] YASUHIRO SUGIMORI[1]

Abstract

An algorithm of estimating Vertical distribution of Chlorophyll-a (Chl-a) was evaluated based on Artificial Neural Networks (ANN) method in Hokkaido field in the northwest of Pacific Ocean. The algorithm applied to the data of SeaWiFS on OrbView-2 and AVHRR on NOAA off Hokkaido, has been applied on September 24, 1998 and September 28, 2001. Ocean color sensor provides the information of the photosynthetic pigment concentration for the upper 22% of the euphotic zone. In order to model a primary production in the water column derived from satellite, it is important to obtain the vertical profile of Chl-a distribution, because the maximum value of Chl-a concentration used to lie in the subsurface region. A shifted Gaussian model has been proposed to describe the variation of the chlorophyll-a (Chl-a) profile which consists of four parameters, i.e. background biomass (B_0), maximum depth of Chl-a (z_m), total biomass in the peak (h), and a measurement of the thickness or vertical scale of the peak (σ). However, these parameters are not easy to be determined directly from satellite data. Therefore, in the present study, an ANN methodology is used. Using in-situ data from 1974 to 1994 around Japan Islands, the above four parameters are calculated to derive the Chl-a concentration, sea surface temperature, mixed layer depth, latitude, longitude, and Julian days. The total of 6983 profiles of Chl-a and temperature are used for ANN. The correlation coefficients of these parameters are 0.79 (B_0), 0.73 (h), 0.76 (σ) and 0.79 (z_m) respectively. A site called A-line off Hokkaido is used to evaluate Chl-a concentration in each depth. After comparing with in-situ data and ANN model, the results show good agreement relatively. Therefore, the ANN method is applicable and available tool to estimate primary production and fish resources from the space.

Keywords : Ocean color, Chlorophyll-a (Chl-a), Vertical structure, Artificial Neural Networks (ANN).

I. Introduction

The study on the process of primary production effecting biological process is very important, especially to understand how it affects phytoplankton carbon fixation which influences the net CO_2 flux across the air-sea interface. The primary production depends on light availability and other environment factors such as temperature, nutrients, and the amount of phytoplankton present for photosynthesis (Morel, 1991).

Recently, the ocean-color remote sensing is an useful new tool for continuously monitoring the temporal and spatial variabilities of Chl-a concentration. The algorithms for estimating primary production have been widely developed (Platt and Sathyendranath, 1988; Falkowski *et al.*, 1998), while the global distribution of the primary production was also estimated (Longhust *et al.*, 1995, Antone and Morel, 1996; Behrenfeld and Falkowski, 1997a). Behrenfeld and Falkowski (1997) gave a summary for

1 Center for Remote Sensing and Ocean Science (CreSOS), Udayana University, Jimbaram Bukit Bali Indonesia.
2 Ocean University of Qingdao 5 Yushan Road Qingdao China.

analytic model and classified the primary productivity models into 4 different kinds, based on implicit levels of integrations, which include wavelength resolved models (WRM), wavelength integrated model (WIM), time integrated model (TIM) and depth integrated model (DIM).

The primary production quantity depends on the estimated vertical distribution of chlorophyll concentration in the water column. The vertical distribution patterns of chlorophyll concentration depend on seasons and regions. The chlorophyll maximum value is not always observed near or at the sea surface, but sometimes lies deeper than the bottom of the euphotic zone (Parsons *et a*l., 1984). In this case, the ocean color sensors cannot measure the chlorophyll maximum value (Gordon and McCluney, 1975). The solution of this problem was proposed by Platt *et a*l. (1988) with a Gaussian distribution model, which represents the vertical distribution of chlorophyll by four parameters. Morel and Berthon (1989) analyzed vertical distributions of the phytoplankton on the basis of measurements about 4000 points in all over the world, and chlorophyll concentration proportion occupied for the phytoplankton pigment concentration is from 65% to 80%. This condition can estimate all chlorophyll quantities of the photic region observed in the satellite. Platt's model can represent uniformity above and below the chlorophyll maximum value. Matsumura and Shiomoto (1993) added the gradient term to the Gaussian distribution. Kameda and Matsumura (1998) apply their techniques in the OCTS data in the sea around Japan.

The ANN is fundamentally useful to illustrate non-line type phenomenon (Simon, 1992). Although there are various techniques of realizing an equivalent function, the neural network used to have nonlinear input and output, even if date sets seldom consider those characteristics, the ANN is able to approximate a nonlinear relation. The ANN is applied to many environmental problems, i.e. weather forecasts or analysis of lake pollutions. In remote sensing field of ocean color, the ANN was applied to the analysis of the case-2 water in coastal area. Wind speeds in the ocean surface have been estimated using the relationship between backward scattering in each band and wind speeds.

The purpose of this paper is to estimate the chlorophyll biomass distribution in the water column around Japan with a shifted Gaussian distribution model (Platt *et al*, 1988) which combining ANN method relations between Gaussian parameters in the model and sea surface information, sea surface temperature, Chl-a. The results were evaluated with the same day *in-situ* data set of the distribution of Chl-a data with A-line off Hokkaido in September 24, 1998 and September 28, 2001 using Chl-a by SeaWiFS, sea surface temperature by NOAA -AVHRR , and mixed layer depth climatologically.

II. Materials and methods

2.1 *In-situ* data

In-situ data are used for a training and evaluation for ANN. The measurements of vertical distribution of Chl-a concentration (8649 points) collected by Japan Ocean Data Center (JODC) around Japan (Figure 1a) were applied. Figure 1b shows the number of collected *in-situ* data each month. The maximum number of *in-situ* data was observed in 1987 and the almost of data was collected during 1980s. The large numbers of measurements are observed in July and October, and the less number are observed in March and December. The observation points cover around Japan, however, in the coastal area the points are less especially around the Asia continent, then east area from 144E and south area from 24N are very few. The outputs of the ANN model were validated with the *in-situ* data off Hokkaido along the A-line, which is a long term line monitored by the National Fisheries Research Institute, Fisheries Research Agency. Cloud free SeaWiFS data were selected for this region in September 1998 and 2001. Figure 1b also indicates the number of stations and cruise locations for each cruise.

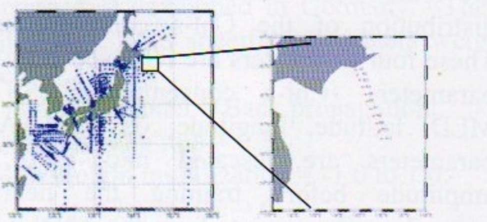

Figure 1. (a) In-situ ship measurement around Japan, (b) Sampling stations along line-A off Hokkaido, which is carried out by the National Fisheries Research Institute Fisheries Research Agency

2.2. Satellite observations

The spatial variability of Chl-a in the study area is observed with SeaWiFS (ver.2) data. The standard map of SeaWiFS Chl-a distribution analyzed from the Local Area Coverage (LAC) data for September 1998 and 2001 was obtained from the Distributed Active Archive Center (DAAC) of the NASA Goddard Space Flight Center. In SeaWiFS project, an algorithm of the pigment sum is derived from the Chl-a estimated from the ocean chlorophyll 4 (OC4) cubic polynomial functions (O' Reilly *et al.*, 2000) and Standard atmospheric correction (Gordon and Wang, 1994; Gordon, 1997). The spatial resolution is about 1.1 km in A-line cruise and the data set is observed at September 24, 1999 and September 28, 2002.

The SST distributions were also investigated using NOAA/AVHRR data for September 24, 1999 and September 28, 2001. The data sets of 2 km spatial resolution were obtained from Japan Marine Science and Technology Center (JAMSTEC). The AVHRR data were applied to MCSST method (McClain, 1985). Monthly mixed layer data were obtained from the NOAA/National Oceanographic Data Center (NODC).

III. Model Formulation

The biomass vertical distributions were determined by applying a Gaussian function (Platt, 1988; Morel, 1989). In this study, the observed values (8694 points) of vertical distribution of Chl-a concentration were applied to estimate vertical distribution of Chl-a concentration with ANN around Japan.

A shifted Gaussian distribution (Fig 2.) was proposed using the Gauss parameters as follow (Platt, 1988):

$$chl(z) = chl(0) + \frac{h}{\sqrt{2\pi}\,\delta} \exp\left[-\frac{(z - zm)^2}{2\delta^2} \right] \quad (1)$$

where $\mathrm{Chl}(z)$ is the Chl-a concentration (mg.m-3) at each depth Z (m), $Chl(0)$ is the background Chl-a concentration at sea surface, h is the total chlorophyll above the background (m), δ standard deviation of Gaussian distribution, which controls the thickness of the chlorophyll maximum layer, and Zm is the depth of the chlorophyll maximum.

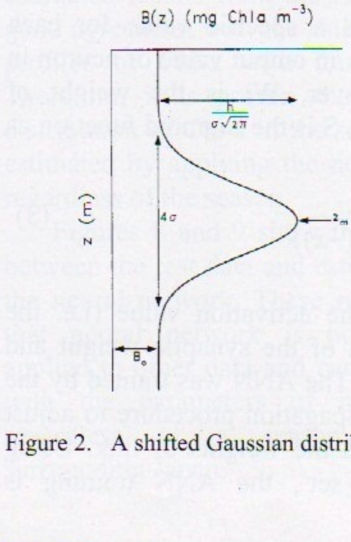

Figure 2. A shifted Gaussian distribution of Chl-a

Chl-a concentration at surface and sea surface temperature (SST) have been derived from SeaWiFS and NOAA/ AVHRR. The *in-situ* data of vertical profiles of Chl-a concentration and temperature for the ANN training and validations were collected by JODC. In this study, we assumed the mixed layer depth (MLD) as 0.5 degree from sea surface (Levitus, 1998).

3.1 The ANN Application

Using the ANN, the relationship between geophysical parameters and neurons can be established even if the relationship between data is not clear or non linear. The ANN adopts the feed-forward model (Error back propagation method, EBP), which is good for the optimization and the computation is extremely fast. The feed-forward model consists of input layers, output layers and one or some hidden layers between them. Each layer consists of several neurons, and each neuron is linked to neurons in the neighborhood layers with different weights and biases. The information from input layers is the feed-forwarded through processing neurons in hidden layer to the output neurons. The output value O of each neuron is calculated by:

$$O = S\left(-bias + \sum x_i w_i\right) \qquad (2)$$

where, *bias* is a specific value for each 'neuron', X_i is an output value of neuron in the former layer, Wi is the weight of incoming-link, S is the Sigmoid function as follows:

$$s(x) = \frac{1}{\left(1+\exp(-x)\right)} \qquad (3)$$

where, x is the activation value (i.e. the scalar product of the synaptic weight and input vector). The ANN was trained by the error back propagation procedure to adjust the biases and the weights of link. Using the 'training set', the ANN training is repeated until the sum squared error (SEE) being minimum. The SSE is defined as follows:

$$SEE = \sum_{sampleoutput} \sum (NN_{outpu_t} - true_{value})^2 \qquad (4)$$

3.2 Nonlinear analyses for parameter estimation

Four parameters must be known previously for estimating the vertical distribution of the Chl-a concentration. These four parameters are retrieved from 6 parameters (Chl-a concentration, SST, MLD, latitude, longitude, daytime). All parameters are rescaled into a [0,1] amplitude before training the neural networks. The outputs from the ANN, i.e. 4 parameters, are scaled back to its original units before plotting the results or performing the error calculations. All the neural network models presented in this paper have been trained using the EBP (Rumelhart *et al.*, 1986), which compound the unit bias nodes and the sigmoid activation functions. The neural network training has been carried out according to an early stopping strategy. Even though the early stopping term sometimes refers to a training procedure stopped after a given number of epochs, in this paper it is used to indicate a training procedure stopped as soon as the validation error begins to increase.

1) Preparations of parameters

The least squares method was applied to observe data of vertical distributions of Chl-a (8694 points) in order to estimate the four parameters for the shifted Gaussian parameters.

2) Model calculation

The data sets are separated randomly for two data sets, i.e. the training sets (6983) for teaching, and the test set (729) for validation during the ANN training.

3) The ANN training and test

The ANN training was determined by the bias and weight for each parameters, such as bias and weight. Using the 'training set', the ANN training was done until minimum of training. The ANN training plays two roles. First as a generation of weight for the neural network, which means that training net protects the over trading by the validation using the 'test set'. The other as an optimization between input parameters and output parameters. Each of these sets was then applied to the Stuttgart Neural Network Simulator (SNNS version 4.2) developed in Germany. The following SNNS specific parameters were used:

ANN model = Back propagation;
Learning rate = 0.2;
Weight initialization = -1.0 to 1.0;
Input layer size = 6;
Hidden layer 1 size = 50;
Output layer size = 4.

4) Application to satellite data

The trained ANN is applied to the Chl-a, and the satellite data of SeaWiFS are used to retrieve the vertical profiles of Chl-a concentration.

IV. Results

Figure 3 shows the example of vertical distributions of Chl-a represented by the shifted Gaussian distribution function with the presumed value using the least squares method each month around the Japan.

Figure 4 shows the correlation coefficient between *in-situ* data of Chl-a and Gaussian distribution profile. The shifted Gaussian distribution were estimated from the Chl-a vertical distribution of the in-situ monitoring data and 85% of the data set was able to show as the shifted Gaussian distribution with the correlation coefficient of 0.9.

The vertical distribution of Chl-a is greatly divided into three patterns. They are 1) the Chl-a becomes a maximum near surface or surface, 2) the maximum Chl-a is found in mid-water, and 3) the Chl-a is a perpendicular distribution from the surface. These condition is well suitable for the Gaussian distribution regardless of time, and location (latitude and longitude), respectively.

Figure 5 shows the error between the estimates from the ANN for each parameter and data for testing the Gaussian distribution. The correlation coefficient of B0, h, σ and Zm showed 0.94, 0.72, 0.73, and 0.81 respectively.

Figures 6 and 7 show some profiles of the shifted Gaussian distribution in Sanriku area. In general, the results from the ANN method are a little over estimated compared with those from the least squared method. It was described from estimate and data for the evaluation from the neural network. For example, the Gaussian distribution, off Sanriku. The correlation coefficient (R) between measured value and Gaussian distribution is high, i.e. 0.92 (Figure 6), and 0.97 (Figure 7). Generally, it is proven that the Chl-a maximum is observed in the sub-surface layer. The Chl-a exist in Sanriku area and near 300m water depth. It is proven that A-line can be estimated with vertical distribution of Chl-a from ANN.

When the data sets were used as input into the ANN (not used in this study), the distribution of correlation coefficient has been understood to obtain the shifted Gaussian distribution for evaluation, which means that the shifted Gaussian distribution can be estimated. The estimated results from the ANN present a good agreement with the estimated vertical profiles of Chl-a with four parameters. Therefore, it is proven that the vertical distribution of Chl-a in the sea area can be estimated by applying the neural network, regardless of the season.

Figures 8 and 9 show the relationship between the test data and estimated data of the neural network. These results indicate that neural network is possible to be applied to other data and can be estimated with the parameters of input layer 6 regardless of a season and space surrounding Japan.

A test site which is called the A-line was prepared around Hokkaido Island in Japan (Figure 1a) and selected two observations.

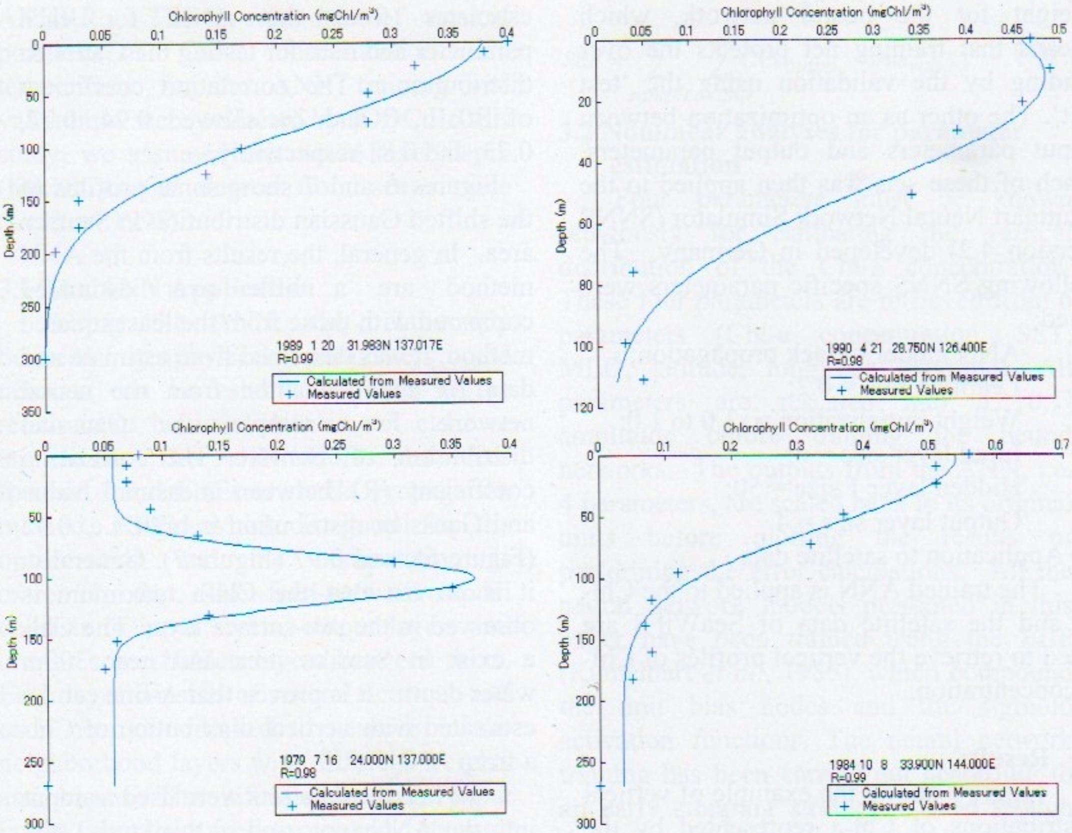

Figure 3. The Least square method applied to a shifted Gaussian distribution

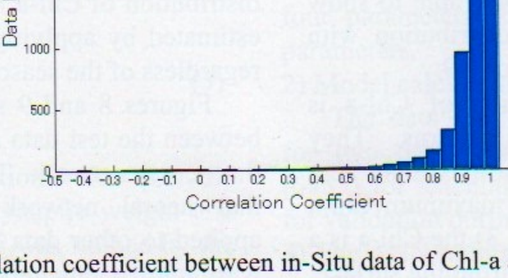

Figure 4. The Correlation coefficient between in-Situ data of Chl-a and Gaussian distribution profile

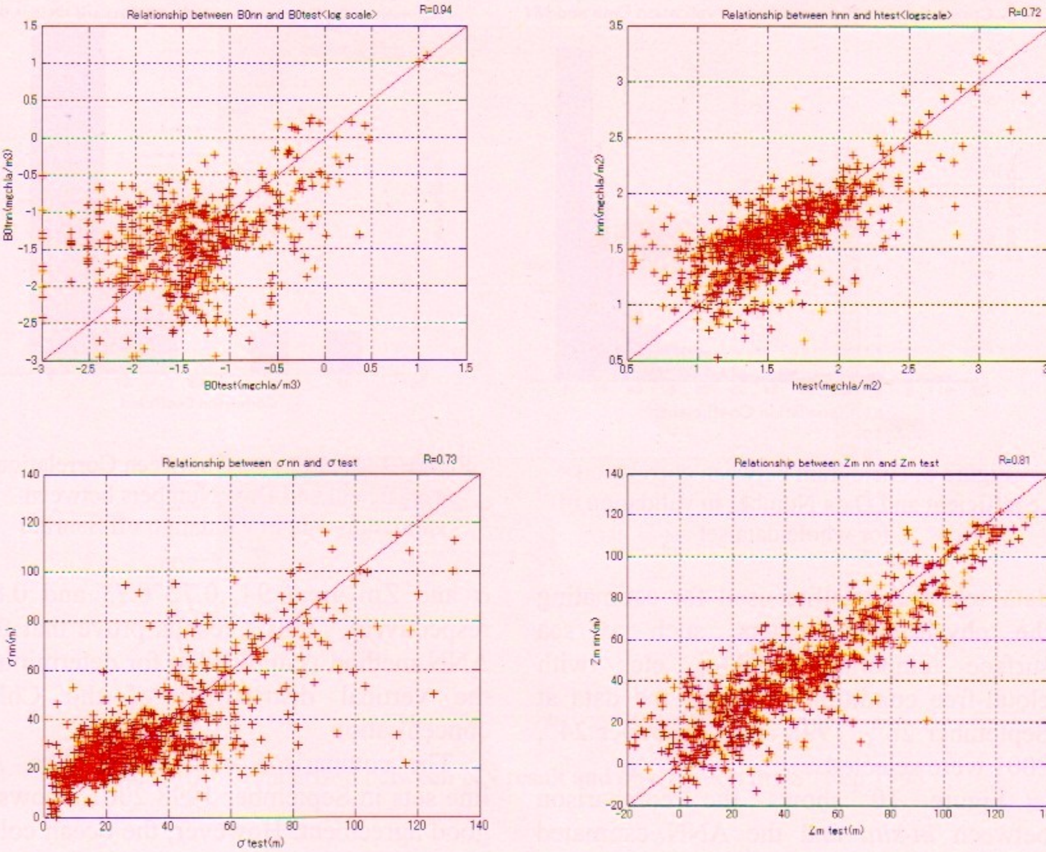

Figure 5. Relationship between validation data and NN in each (a) B0, (b) h, (c) σ, (d) Zm

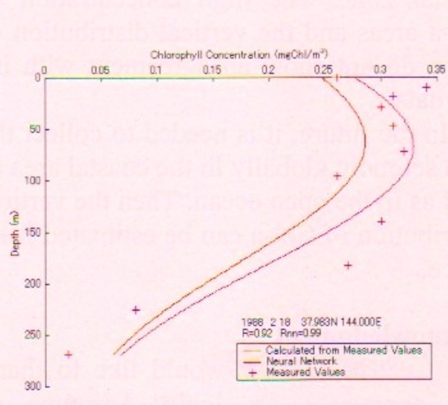

Figure 6. Comparison between in-situ, LM and NN off Sanriku in February

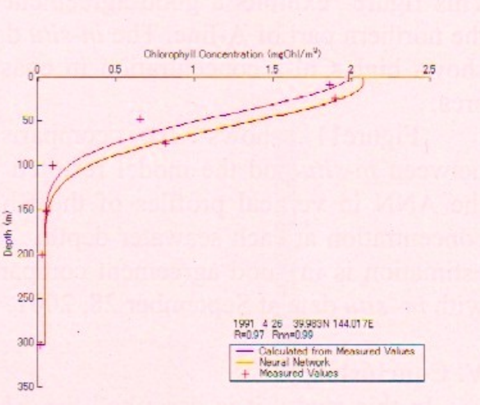

Figure 7. Comparison between in-situ, LMS and NN off Sanriku in April

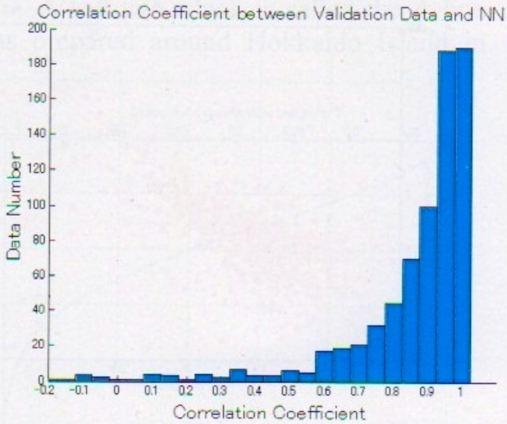

Figure 8. Histogram between correlation coefficient and Data Number in validation off for whole data set

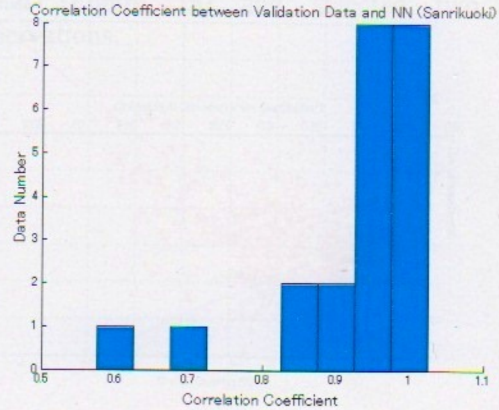

Figure 9 Histogram of between Correlation coefficient and Data Numbers between validation data in validation off Sanriku

data sets and satellite used for estimating the physical parameters, such as sea surface temperature, Chl-a, etc., with cloud-free conditions. Finally, the data at September 28th , 1998 and September 24th, 2001 were selected.

Figure 10 shows the comparison between *in-situ* and the ANN estimated cross section of the Chl-a concentration along the A-line in September 28, 1998. This figure exhibits a good agreement in the northern part of A-line. The *in-situ* data shows high Chl-a concentration in coastal area.

Figure11 shows the comparison between *in-situ* and the model resulted by the ANN in vertical profiles of the Chl-a concentration at each seawater depth. The estimation is in good agreement compared with *in- situ* data at September 28, 2001.

V. Conclusion

In this study, it is described the ANN method to solve the vertical distribution of Chl-a retrieved from sea surface conditions. The correlation coefficients of the training data sets for the four parameters of b0, h, σ and Zm are 0.78, 0.73 0.76 and 0.79 respectively. The correlation coefficients of the parameters for the test data sets of b0, h, σ and Zm are 0.94 ,0.72 0.73 and 0.81 respectively. This results prove that the ANN method is applicable for determining the vertical distribution of the Chl-a concentration.

The comparison with *in-situ* data, A-line sets in September 1998 2001, shows a good agreement. However, the ocean color sensor cannot be observed the high concentration of Chl-a especially in the coastal zone. The high concentration of Chl-a areas and the vertical distribution of Chl-a do not show an agreement with in-situ data.

In the future, it is needed to collect the data set more globally in the coastal area as well as in the open ocean. Then the vertical distribution of Chl-a can be estimated from space.

Acknowledgment

The authors would like to thank and appreciate Dr Ichio Asanuma of EORC/ NASDA for providing AVHRR data; the SeaWiFS project team in GSFC/NASA and the PODAAC team in JPL/NASA for providing a global AVHRR MCSST data. This work was also partly supported by China NSF No. 49976035.

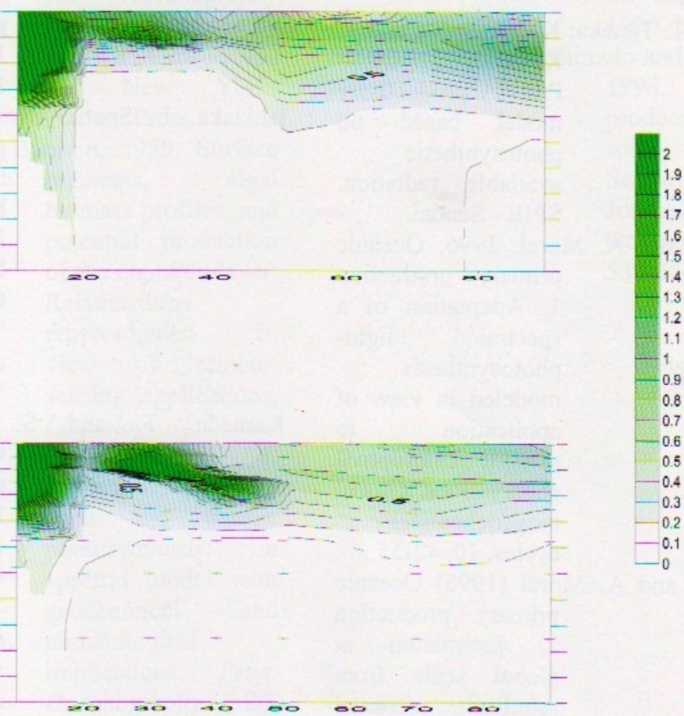

Figure 10. Comparison between NN result and observation in 28th Sep 1998

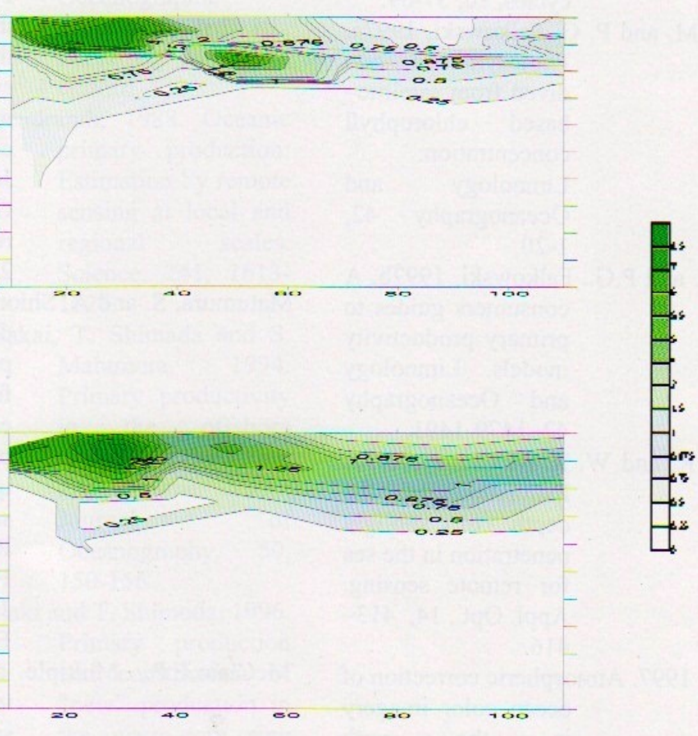

Figure 11. Comparison NN and observation data in 24th September 2001

References

Asanuma I.; T. Tanaka; K. Matumoto and T. Kawano, 2000. Primary productivity model based on photosynthetic available radiation. SPIE - Sendai.

Antoine, D. and A. Morel, 1996. Oceanic primary production 1. Adaptation of a spectral light-photosynthesis modeled in view of application to satellite chlorophyll observations. Global biogeochemical cycles, 10, 42-55.

Antoine, D. and A. Morel (1996) Oceanic primary production 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. Global biogeochemical cycles, 10, 57-69.

Behrenfeld, M. and P. G. Falkowski, 1997a. Photosynthetic rate dived from satellite–based chlorophyll concentration. Limnology and Oceanography 42, 1-20.

Behrenfeld, M. and P.G.. Falkowski, 1997b. A consumers guides to primary productivity models. Limnology and Oceanography 42, 1479-1491.

Gordon, H. R. and W. R. McCluney, 1975. Estimation of the depth of sunlight penetration in the sea for remote sensing. Appl. Opt., 14, 413–416.

Gordon, H.R., 1997. Atmospheric correction of ocean color imagery in the earth observing system journal of geophysical research 102(D4), pp. 17091-17106

Ishizaka. J., Spatial distribution of Primary production off Sanriku, Northwestern Pacific, during Spring Estimated by Ocean color and Temperature Journal of Oceanography. Vol. 54, 553 - 564

Kameda, T. and S. Masumura, 1998. Chlorophyll biomass off Sanriku, Northwestern Pacific, estimated by Ocean Color and Temperature Scanner (OCTS) and a vertical distribution model. Journal of Oceanography. 54, 509-516

Lewis, M.R.; J.J. Cullen and T. Platt, 1983. Phytoplankton and thermal structure in the upper ocean: consequences of no uniformity in chlorophyll profile. Journal of Geophysical research.88, 2565-2570

Matumura, S. and A. Shiomoto, 1993. Vertical distribution of primary productivity function ( ) for the estimation of primary productivity using satellite remote sensing. Bull. Nat. Res Inst Far Seas Fish.,30 227-270 (in Japanese with English abstract).

McClain,E.P., Multiple atmospheric-window techniques for satellite-derived sea surface temperatures, In: Oceanography from space, edited by J. F. R. Gower, 73-85, New York: Plenum Press.

Morel, A., and J.F. Berthon, 1989. Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications, *Limnology and Oceanography,* 34, 1545-1562.

Morel, A., 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. *Prog. Oceano*graphy ., 26, 263–306.

Parsons, T. R.; M. Takahashi and B. Hargrave, 1984. Biologic Oceanographic Processes. 3rd ed., Pergamon Press, Oxford, 330.

Platt, T and Sathyendranth, 1988. Oceanic primary production: Estimation by remote sensing at local and regional scales. Science, 241, 1613-1620

Shiomoto, A., K. Sakai, T. Shimada and S. Matumura, 1994. Primary productivity in the offshore Oyashio in the spring and summer 1990. Journal of Oceanography. 50, 150-156

Shiomoto, A.; K. Sasaki and T. Shimoda, 1996. Primary production and contribution of "new" production in the warm-core ring and the cold streamer off Sanriku in May 1990. Lamer, 34, 1-9.

Shiomoto, A.; S. Hashimoto and T. Murakami, 1998. Primary productivity and solar radiation off Sanriku in May 1997, Journal of Oceangraphy.54, 539-544.