

Retrieval of Oceanic Chlorophyll Concentration Using Support Vector Machines

Haigang Zhan, Ping Shi, and Chuqun Chen

Abstract—This letter investigates the possibility of using a new universal approximator—support vector machines (SVMs)—as the nonlinear transfer function between oceanic chlorophyll concentration and marine reflectance. The SeaBAM dataset is used to evaluate the proposed approach. Experimental results show that the SVM performs as well as the optimal multilayer perceptron (MLP) and can be a promising alternative to the conventional MLPs for the retrieval of oceanic chlorophyll concentration from marine reflectance.

Index Terms—Oceanic chlorophyll, ocean color remote sensing, neural network, support vector machine (SVM).

I. INTRODUCTION

RETRIEVAL of chlorophyll concentration from ocean color observations requires a transfer function to convert satellite measurements into chlorophyll concentration. Statistical regression of radiance or reflectance versus chlorophyll concentration is the most popular approach to construct the transfer function. However, statistical regression has limitations because of the nonlinear relationship between radiance (or reflectance) and chlorophyll concentration. In recent years, more attention has been paid to the use of neural networks (NNs). The advantages of this approach are mainly due to its ability to approximate any nonlinear continuous function without *a priori* assumptions about the data. It is also more noise tolerant, having the ability to learn complex systems with incomplete and corrupted data. Different models of NNs have been proposed, among which, multilayer perceptrons (MLPs) with the backpropagation training algorithm are the most widely used. MLPs have been applied for retrieval of water constituent concentrations and optical properties in both case 1 and case 2 waters [1]–[7].

However, MLPs still suffer from some problems. First, the training algorithm may be trapped in a local minimum. The objective function of MLPs is very often extremely complex. The conventional training algorithms can easily be trapped in a local minimum and never converge to an acceptable error. In that case, even the training dataset cannot be fit properly. Second, it is generally a difficult task to determine the best architecture of

MLPs, such as the selection of the number of hidden layers and the number of nodes therein. Third, overfitting of the training dataset may also pose a problem. MLP training is based on the so-called empirical risk minimization (ERM) principle, which minimizes the error on a given training dataset. A drawback of this principle is the fact that it can lead to overfitting and thereby poor generalization [8], [9].

These problems can be avoided by using a promising new universal approximator, i.e., support vector machines (SVMs). SVMs have been developed by Vapnik [9] within the area of statistical learning theory and structural risk minimization (SRM). SVM training leads to a convex quadratic programming (QP) problem, rather than a nonconvex, unconstrained minimization problem as in MLP training; hence, it always converges to the global solution for a given dataset, regardless of initial conditions. SVMs use the principle of structural risk minimization to simultaneously control generalization and performance on the training dataset, which provides it with a greater ability to generalize. Furthermore, there are few free parameters to adjust, and the architecture of the SVM does not need to be found by experiments. SVM has been used to perform data merger from multiple ocean color sensors [10]. This letter is intended to explore the application of SVM for retrieval of oceanic chlorophyll concentration from marine reflectance and compare its performance with MLPs and statistical regression algorithms.

II. SVM FOR REGRESSION

SVMs were first developed to solve the classification problems, but recently they have been extended to the domain of regression approximation. In this section, we briefly introduce some basic ideas behind SVM for regression. A more detailed description of the technique can be found in [9] and [11].

Given the training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, $\mathbf{x}_i \in R^n$, $y_i \in R$ for the case of nonlinear regression, the SVM first maps \mathbf{x} into a high-dimensional feature space by using some nonlinear mapping φ , $\mathbf{z} = \varphi(\mathbf{x})$ and then constructs a linear model $f(\mathbf{x}, \mathbf{w})$ in this feature space

$$f(\mathbf{x}, \mathbf{w}) = \langle \mathbf{w} \cdot \varphi(\mathbf{x}) \rangle + b \quad (1)$$

where $\langle \mathbf{w} \cdot \varphi(\mathbf{x}) \rangle$ denotes the dot product between \mathbf{w} and $\varphi(\mathbf{x})$, \mathbf{w} is a vector in the feature space, and b is a constant. The goal of regression estimation is to find the best regression function by minimizing some loss function for all training data. One of the main characteristics of the SVM is that instead of minimizing the training error, it attempts to minimize the generalization error bound so as to prevent the phenomena of overfitting and achieve higher generalization performance. This

Manuscript received June 5, 2003; revised August 28, 2003. This work was supported by the South China Sea Institute of Oceanology (CAS) under Project LYQY200308 under the funds of knowledge innovation program, the National Natural Science Foundation of China under Project 40276049, and the Chinese Academy of Sciences under Project KZCX2-202.

The authors are with the Key Laboratory of Tropical Marine Environmental Dynamics (LED), South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China (e-mail: hgzhn@scsio.ac.cn; pshi@scsio.ac.cn; cqchen@scsio.ac.cn).

Digital Object Identifier 10.1109/TGRS.2003.819870

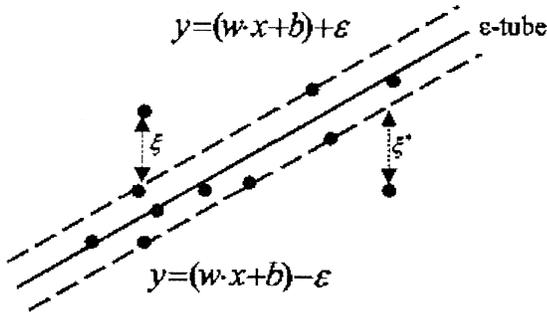


Fig. 1. Graphical depiction of an ε -tube and slack variables ξ_i, ξ_i^* .

generalization error bound is the combination of the training error (empirical risk) and a regularization term that controls the model complexity (structural risk). The first term is calculated by Vapnik's ε -insensitive loss function [9]

$$|y - f(\mathbf{x}, \mathbf{w})|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon, & \text{otherwise} \end{cases} \quad (2)$$

in which ε is the tolerance to error. This defines an ε -tube (Fig. 1) so that if the predicted value is within the tube the loss is zero, while if the predicted point is outside the tube, the loss is the magnitude of the difference between the predicted value and the radius ε of the tube; thus, SVM is more robust to small errors in the training data to the least square loss function used for MLPs. The Vapnik ε -insensitive loss function can be formally described by introducing nonnegative slack variables $\xi_i, \xi_i^*, i = 1, \dots, N$ to measure the deviation of training data outside the ε -insensitive zone (Fig. 1). The second term is a l_2 norm of the vector \mathbf{w} . Thus, SVM regression can be posed as a convex optimization problem as follows:

$$\begin{aligned} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{Subject to} & \begin{cases} y_i - \langle \mathbf{w} \cdot \varphi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^* \\ \langle \mathbf{w} \cdot \varphi(\mathbf{x}_i) \rangle + b - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n \end{cases} \end{aligned} \quad (3)$$

where C is a fixed regularization constant determining the tradeoff between the model complexity and the training error. This optimization problem can be solved through the technique of lagrange multipliers [9], [11], and the regression function is

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (4)$$

where $K(\mathbf{x}_i, \mathbf{x}) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}) \rangle$ is a kernel function that satisfies Mercer's condition [11], and α_i and α_i^* are the lagrange multipliers that obtained by solving the following QP problem

$$\begin{aligned} \min_{\alpha, \alpha^*} & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}_j) (\alpha_j^* - \alpha_j) \\ & + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{subject to} & \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (5)$$

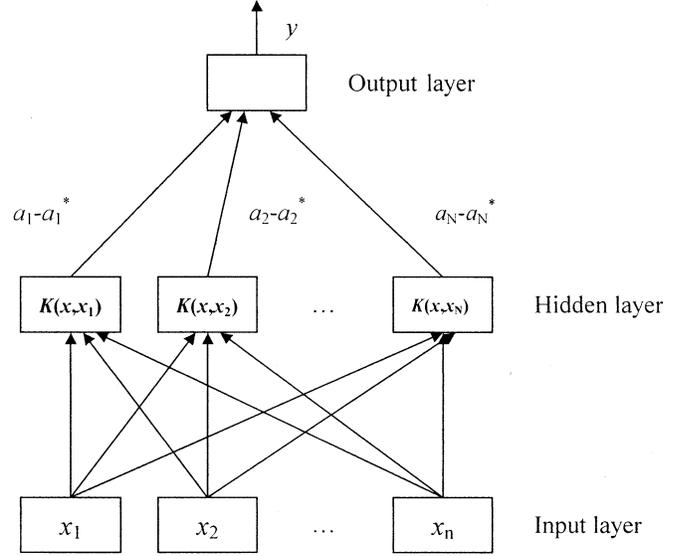


Fig. 2. Architecture of support vector machine.

and b can be obtained by exploiting the Karush–Kuhn–Tucker (KKT) conditions [9]

$$\begin{aligned} b &= y_i - \langle \mathbf{w}, \varphi(\mathbf{x}_i) \rangle - \varepsilon, \quad \text{for } \alpha_i \in (0, C) \\ b &= y_i - \langle \mathbf{w}, \varphi(\mathbf{x}_i) \rangle + \varepsilon, \quad \text{for } \alpha_i^* \in (0, C). \end{aligned} \quad (6)$$

Note that K is always positive definite. Then, (5) is a strictly convex problem and has a unique global minimum; thus, the problem of many local optima in the case of training, for example, an MLP, is avoided.

In representation (4), typically only some of the coefficients $(\alpha_i - \alpha_i^*)$ differ from zero, and the corresponding training data \mathbf{x}_i are called support vectors (SVs). These data are on the ε -tube border or outside the ε -tube, and the data inside the ε -tube do not contribute to the regression function. Generally, the larger the ε , the fewer the number of SVs and thus the sparser the representation of the solution. However, a larger ε can also decrease the approximation accuracy of the SVM. In this sense, ε is a tradeoff between the sparseness of the representation and closeness to the data.

Three commonly used kernel functions for nonlinear regressions are

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right) \quad (7)$$

$$K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}^T \mathbf{x}_i + 1)^p \quad (8)$$

$$K(\mathbf{x}_i, \mathbf{x}) = \tanh(\beta_0 \mathbf{x}^T \mathbf{x}_i + \beta_1). \quad (9)$$

Equation (7) is the radial basic function (RBF) kernel with width parameter σ ; equation (8) is the polynomial regression of degree p that will revert to the linear function when $p = 1$, and (9) is the two-layer sigmoid perceptron. The polynomial and RBF kernel functions always satisfy Mercer's condition, and the two-layer perceptron kernel function satisfies Mercer's condition only for some values of β_0 and β_1 [11].

As displayed in Fig. 2, the architecture of SVM is similar to that of MLP. However, their constructions are very different. In MLPs, determination of architecture, e.g., the number of hidden nodes depends on "trial and error." In contrast, in the two-layer

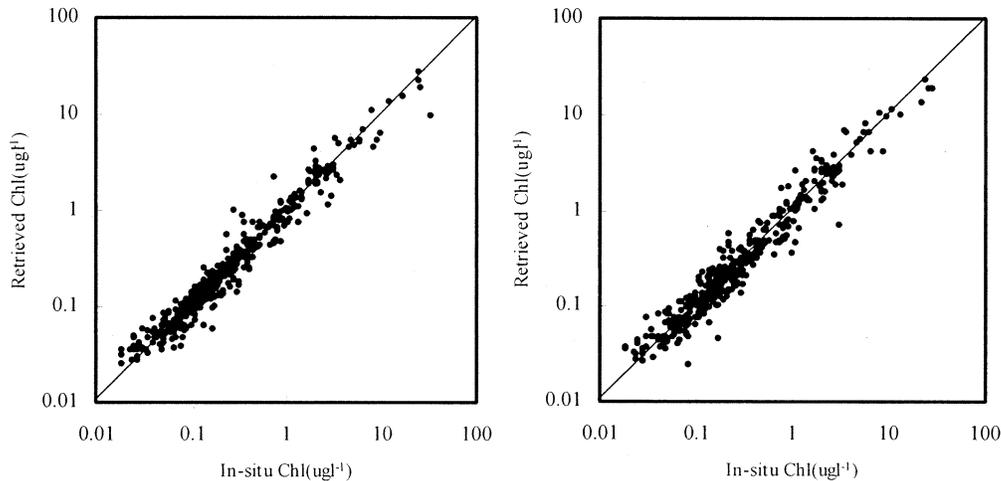


Fig. 3. Comparison of the SVM-derived versus *in situ* chlorophyll concentration on (a) training and (b) the validation dataset.

perceptron type of SVMs, the number of hidden nodes and their weight vectors are determined automatically by the number of SVs and their values, respectively.

III. RETRIEVAL EXPERIMENTS

A. Data Description and Preprocessing

To carry out an experimental analysis to validate the proposed approach, we considered an *in situ* dataset that was archived by the National Aeronautics and Space Administration Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Project as the SeaBAM dataset [12]. This dataset consists of coincident remote sensing reflectance (R_{rs}) at the SeaWiFS wavelengths (412, 443, 490, 510, and 555 nm) and surface chlorophyll concentration measurements at 919 stations around the U.S. and Europe. It is encompassed a wide range of chlorophyll concentration between 0.019 and 32.79 μgL^{-1} with a geometric mean of 0.27 μgL^{-1} . Most of the data are from Case 1 nonpolar waters, and about 20 data items collected from the North Sea and Chesapeake Bay should be considered as case 2 waters. Log-transferred R_{rs} values and chlorophyll concentrations are used as the inputs and output, respectively. The advantage of this transformation is that the distribution of transformed data will become more symmetrical and closer to normal. To facilitate training of SVMs, the values of each input and output were scaled into the range of $[-1 \ 1]$.

B. Training of the SVM

The training software used in our experiments is LIBSVM.¹ It is an integrated software package for support vector classification, regression, and distribution estimation and uses a modified sequential minimal optimization (SMO) algorithm to train the SVMs. The SMO algorithm breaks the large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop [13]. The

RBF kernel function was chosen because it is much more flexible than the two-layer perceptron and the polynomial kernel function. Consequently, it tends to perform best over a range of applications, regardless of the particulars of the data [14].

There are three free parameters, namely C , ε , and σ , that should be determined to find the optimal solution. In our experiments, we split the SeaBAM dataset into two subsets and used the split-sample validation approach to tune these free parameters. This split-sample validation approach estimates the free parameters by using one subset (training set) to train various candidate models and the other subset (validation set) to validate their performance [1], [2]. In order to ensure the representative of the datasets, the SeaBAM dataset was first arranged in increasing order of the chlorophyll concentrations and then, starting from the top, the odd order samples ($n = 460$) were picked up as the training set and the remaining samples ($n = 459$) were used as the validation set. We set $C = 15$, $\varepsilon = 0.03$, and $\sigma = 0.7$, because these values were found to produce the best possible results on the validation set by split-sample validation approach. After these parameters are fixed, the SVM automatically determines the number (how many SVs) and locations (the SVs) of RBF centers during its training.

C. Results

The performance of the SVM was evaluated using the same criteria as [2], namely, root-mean-square error (RMSE), coefficient of determination (R^2), and scatterplot of derived versus *in situ* chlorophyll concentrations. All results were based on log-transformed data. Fig. 3 displays the scatterplots of the SVM-derived versus the *in situ* chlorophyll (Chl) concentration on the training and the validation set. The RMSE for the training set is 0.122, and its R^2 is 0.958. The RMSE for the validation set is 0.138, and its R^2 is 0.946. The number of SVs is 288, which close to 60% of the training data. These SVs contains all the information necessary to model the nonlinear transfer function.

The performance of the SVM was compared with those of MLPs and SeaWiFS empirical algorithms. In order to allow for

¹LIBSVM: A library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

TABLE I
STATISTICAL RESULTS OF MLPS, SVM, AND EMPIRICAL ALGORITHMS ON THE VALIDATION SET

	RMSE							R ²						
MLP	Number of hidden nodes							Number of hidden nodes						
Trial	4	5	6	7	8	9	10	4	5	6	7	8	9	10
1	0.177	0.143	0.157	0.155	0.188	0.143	0.156	0.912	0.942	0.931	0.933	0.908	0.943	0.932
2	0.139	0.152	0.139	0.149	0.140	2.563	0.157	0.946	0.936	0.945	0.938	0.945	0.014	0.931
3	0.144	0.140	0.149	0.157	0.152	0.548	0.206	0.942	0.945	0.938	0.931	0.935	0.516	0.891
4	0.138	0.596	0.336	0.143	0.197	0.159	0.229	0.946	0.427	0.724	0.942	0.823	0.929	0.867
5	0.140	0.141	0.160	0.241	1.841	0.155	0.220	0.944	0.943	0.930	0.848	0.158	0.931	0.878
6	0.137	0.140	0.140	0.156	0.158	0.282	0.180	0.947	0.944	0.945	0.932	0.930	0.806	0.910
7	0.195	0.142	0.305	0.176	5.169	0.150	2.878	0.893	0.943	0.763	0.915	0.002	0.937	0.002
8	0.154	0.219	0.143	0.135	0.151	0.137	0.155	0.933	0.878	0.942	0.949	0.935	0.947	0.931
9	0.144	0.140	0.144	0.140	0.301	0.471	0.236	0.942	0.944	0.941	0.944	0.770	0.560	0.854
10	0.162	0.146	0.146	0.151	0.145	0.171	0.205	0.928	0.940	0.939	0.936	0.941	0.916	0.886
SVM	0.138							0.946						
OC2	0.172							0.919						
OC4	0.161							0.929						

this comparison, the training and validation data were preprocessed in a similar manner for SVM and MLP uses, and the results of MLPs and SeaWiFS empirical algorithms were based on the same validation set as was used for the SVM. There are a large number of factors that control the performance of MLPs, such as the number of hidden layers, the number of hidden nodes, activation functions, epochs, weights initialization methods, and parameters of the training algorithm. It is a difficult task to obtain an optimal combination of these factors that produces the best retrieval performance. We used MLPs with one hidden layer and tan-sigmoid activation, and we trained them using the Matlab Neural Network Toolbox 4.0 with the Levenberg–Marquardt algorithm. The epoch was set to 500, and other training parameters were set to the default values of the software. The training process was run ten times with different random seeds for the number of hidden nodes from four to ten. The statistical results of the MLPs, the SVM, and the SeaWiFS algorithm OC2 and OC4 on the validation set are reported in Table I. Several results can be found from this table. First, the performance of the SVM is as good as the optimal MLP solution. There are only two trials in which the RMSE of the best MLP is slightly smaller than that of the SVM. Second, the optimal number of hidden nodes is difficult to determine because it varies with different weights initialization. Third, large errors occurred in some trials due to the training algorithm being trapped in a local minimum. Finally, the SVM and the best MLP with different weights initialization outperform the SeaWiFS empirical algorithms.

IV. CONCLUSION

The use of SVMs in retrieval of oceanic chlorophyll concentration was studied in this letter. Experiments on SeaBAM dataset demonstrated that the performance of SVMs were comparable in accuracy to the best MLP. Advantages of SVMs

over MLPs include the existence of fewer parameters to be chosen, a unique, global minimum solution, and high-generalization ability. The proposed method seems to be a promising alternative to the conventional MLPs for modeling the non-linear transfer function between chlorophyll concentration and marine reflectance.

It is worthy to note that SVM generalization performance, unlike those of conventional neural networks such as MLPs, does not depend on the dimensionality of the input space. The SVM can have good performance even in problems with a large number of inputs and, thus, provide a way to avoid the curse of dimensionality [9]. This makes it attractive for case 2 waters, since more spectral channels are needed for retrieval of some parameters in such waters. Further research will be carried out to validate the performance of SVMs for inverse problem in case 2 waters.

ACKNOWLEDGMENT

The authors would like to thank the SeaWiFS Bio-optical Algorithm Mini-Workshop (SeaBAM) for their on-line SeaBAM data (http://seabass.gsfc.nasa.gov/seabam/bioopt_workshop.html) and C.-C. Lin and C.-J. Lin for their software package LIBSVM.

REFERENCES

- [1] L. E. Keiner and X. H. Yan, "A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery," *Remote Sens. Environ.*, vol. 66, pp. 153–165, 1998.
- [2] L. E. Keiner and C. W. Brown, "Estimating oceanic chlorophyll concentrations with neural networks," *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 189–194, 1999.
- [3] H. Schiller and R. Doerffer, "Neural network for estimating of an inverse model-operational derivation of case II water properties from MERIS data," *Int. J. Remote Sens.*, vol. 20, no. 9, pp. 1735–1746, 1999.
- [4] D. Buckton and E. O'Mongain, "The use of neural networks for the estimation of oceanic constituents based on MERIS instrument," *Int. J. Remote Sens.*, vol. 20, no. 9, pp. 1841–1851, 1999.

- [5] Z. P. Lee, M. R. Zhang, K. L. Carder, and L. O. Hall, "A neural network approach to deriving optical properties and depths of shallow waters," in *Proc. Ocean Optics XIV*, S. G. Ackleson and J. Campell, Eds. Washington, DC, 1998.
- [6] A. Tanaka, T. Oishi, M. Kishino, and R. Doerffer, "Application of the neural network to OCTS data," in *Proc. Ocean Optics XIV*, S. G. Ackleson and J. Campell, Eds. Washington, DC, 1998.
- [7] L. Gross, S. Thiria, R. Frouin, and B. G. Mitchell, "Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration," *J. Geophys. Res.*, vol. 105, no. C2, pp. 3483–3495, 2000.
- [8] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Networks*, vol. 10, pp. 988–1000, Sept. 1999.
- [9] ———, *The Nature of Statistical Learning Theory*, 2nd ed. New York: Springer-Verlag, 2000.
- [10] E. J. Kwiatkowska and G. S. Fargion, "Merger of ocean color data from multiple satellite missions within the SIMBIOS project," in *Proc. SPIE*, vol. 4892, 2002, pp. 168–182.
- [11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [12] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103, no. C11, pp. 24 937–24 953, 1998.
- [13] J. Platt, "Fast training of SVM's using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999.
- [14] A. J. Smola, "Learning with kernels," Ph.D. thesis, GMD, Birlinghoven, Germany, 1998.