Journal of Hydroscience and Hydraulic Engineering
Vol. 17, No.2 November, 1999, 51-57

APPLYING ARTIFICIAL NEURAL NETWORKS IN CHLOROPHYLL TIME SERIES SIMULATION IN WATER ENVIRONMENT

By

Y. SHEN
Osaka University, Suita City, Osaka, Japan

K. NAKATSUJI
Osaka University, Suita City, Osaka, Japan

SYNOPSIS

For modeling and forecasting ecological processes of our environment, difficulties arise due to the stochastic nature of observation data, the limits in understanding complicate natural system and the interaction between various factors in modeling. The new "black box" model ANN (Artificial Neural Network) is among our reasonable solutions. From this viewpoint, the present study deals with the techniques of artificial neural network applications, data set analysis, and optimization techniques. The aim of the analysis is to obtain useful practical information from the available data-sets and to try to minimize the effects of error and uncertainty in the forecasting process. Various tests are included in this work on the problems of Chlorophyl concentration simulation in Osaka Bay of Japan.

INTRODUCTION

With the high performance digital computers widely applied in the field of hydraulic and environmental simulation, besides traditional numerical models, many new robust and efficient techniques are developed and applied; e.g. new global optimization algorithms, artificial neural networks (ANN), fuzzy systems and so on.

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way of biological nervous systems, such as the brain, processes information. It is composed of a large number of highly interconnected processing elements (neurons) working together to solve specific problems. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true for ANNs as well.

Artificial neural networks contain the "input layer" (distributes input vector information to the units in the next layer through links), the "output layer" (outputs of the network which are compared to the target values during training), and the "hidden layers" (which have no direct connection with the outside world). Figure 2 (A) shows an example of

ANN structure while Figure 1(B) shows a typical ANN neuron. Neurons receive information (X_1, X_2...X_n) from outside model/ANN (the initial inputs) or from the interconnections, and transfer information to the nearby layer units through links. Each link has an associated "weight", which controls the propagation of information in the network and represents the adaptability of the ANN during the learning process. Here incoming signal is transformed to the output signal through Y=f(X) which stands for the threshold function (i.e. sigmoid function). During training, the optimization algorithm updates the network weights (by back propagation) so as to minimize the error function, which is the measure of the difference between the network output and the target values. The detail of the processes has been discussed by Aleksander et.al (1990) and Master (1994). ANNs, like people, learn by examples.

ANNs have been applied to an increasing number of real-world problems of considerable complexity. Their most important advantage is in solving problems that we do not have an algorithmic solution for or for which an algorithmic solution is too complex to be found. Applying ANN does not need a priori knowledge of the underlying process of natural system. Thus, the quality of the resulting model depends largely on the quality of the data we use. Shen et. al (1998) applied the ANN (MLP neural networks developed by Delft Hydraulics Inst., The Netherlands) to simulate water quality time series for Lake Veluwe in The Netherlands, and obtained good results.

The application included in the present study is, simulation of chlorophyl concentration by total nitrogen, total phosphorus, water transparency and water temperature. Our main attention is on the seasonal variation of water quality distribution in Osaka Bay together with correlation analysis for different ecological data series. The application is so far on the stage of preliminary research and it may form a basis for further investigation.

APPLICATION: SIMULATION OF CHLOROPHYLL-A TIME SERIES IN OSAKA BAY

Osaka Bay is an oval-shaped semi-enclosed bay with 58 km in length and 30 km in width, open to the Seto Inland Sea (shown in Fig. 2). As reported by Yamane and Nakatsuji (1996), strong seasonal variation of water quality data (include chlorophyll concentration) has been observed in the bay area based on their data analysis. Due to the inflow from Yodo River (northeast of the bay), the chlorophyl levels in the eastern area of the bay remains high throughout the year. The chlorophyl concentration level at the bay head to the east of the tidal front becomes extremely high in August, while there appears strong stratification.

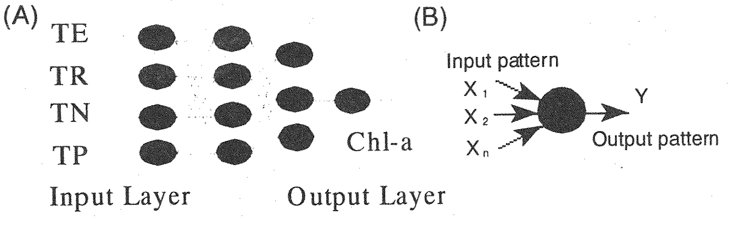

Figure 1. Structure of the ANN applied in this case study.

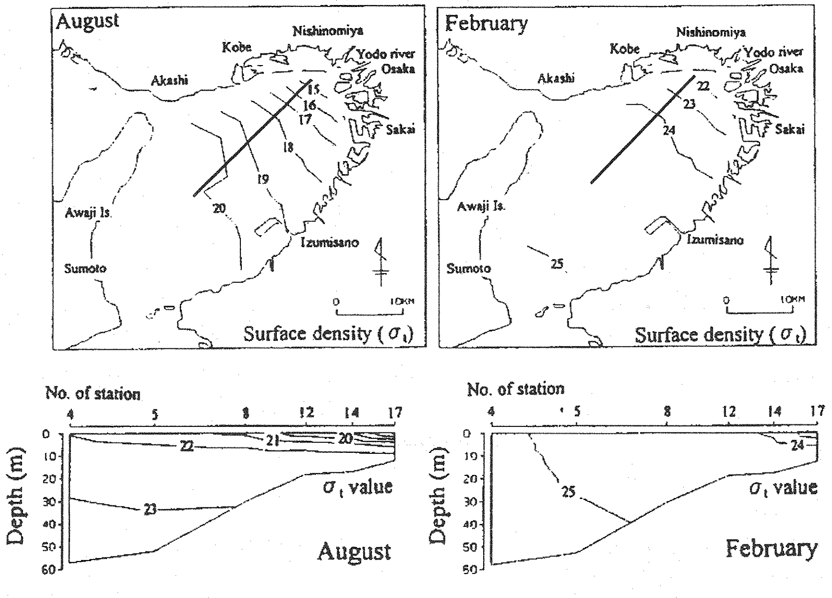

Figure 2. Horizontal and vertical distribution of density in summer and winter. (vertical distribution is shown at cross section in the upper figure). Yamane and Nakatsuji 1997.

Before this case study, intensive research had been carried out by the researchers of Osaka University on characteristics of the water quality in Osaka Bay. As indicated by Yamane and Nakatsuji (1996), Osaka Bay may be divided in four different areas shown in Fig 3 (a) and (b). In their research, with the "clustering method" the authors analyzed the water quality characteristics (including PH value, water temperature, water transparency, COD and DO etc.) in both summer and winter seasons. In the present study, we try to test the performance of ANN in representing the water surface chlorophyll-a concentration time series based on other water quality data sets (i.e. nitrogen and phosphorous). In our tests, we apply the monthly surface water quality data of Osaka Bay provided by Osaka and Hyogo prefectural governments. Four test sites in the eastern Osaka Bay have been selected. They are: observation stations 60101, 60102, 60202 and 60302, as shown in Fig. 3, which are represented as sites 101, 102, 202 and 302, respectively. Sites 101, 102 and 302 locate in Area I, Area II, and Area III respectively, while site 202 locates in Area II in summer, and in Area III in winter. The two objectives for our tests are, firstly to test the performance of the ANN in representing the seasonal variation of the Chlorophyll-a in the eastern bay head; and secondly to test that if the disturbance on the learning objects (ANN inputs) may affect the chlorophyll-a simulation performance of ANN.

In the present case study, we use the ANNs developed by Y. Danon to conduct the tests, with nonlinear sigmoid function as the activation function. The ANNs learn by adjusting the weights (with the back-propagation algorithm) to minimize the error (Root mean square). Masters (1994) gives a detail description of the process in each node. The structure of the ANN used in this case study is shown in Fig. 1 (A).

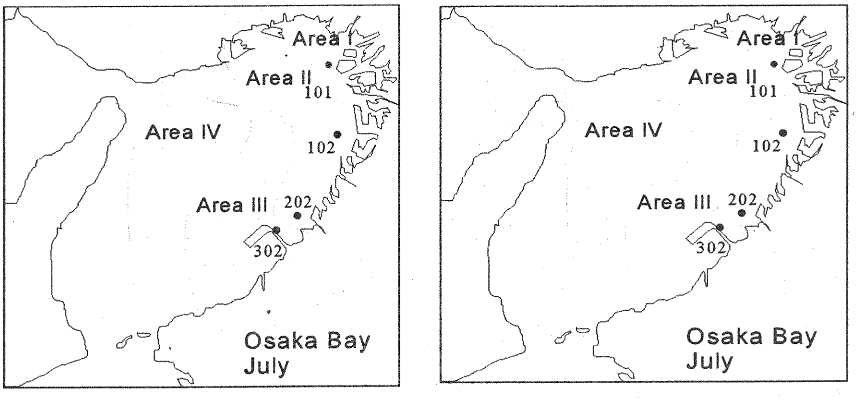

Figure 3 Test sites, Osaka Bay and division due to water quality characteristics in winter and summer seasons

The inputs and outputs of the 4-layer ANN are listed in Table 1 (which are all in the form of monthly records). The Inputs include: (1) TE (water temperature), (2) TN (total nitrogen), (3) TP (total phosphorus), and (4) TR (transparency). Among the output and four inputs, for TN, TP and Chl-a, we provide the measured concentration time series. For the training of ANN, in sites 102, 103, and 104 (shown in Figure 3), we choose the 1986-1992 surface Chlorophyll-a time series (monthly data, 60 samples). Whereas, for site 101, we use data of 1985 together with 1987-1990 (also 60 samples in total) due to the missing of its data in 1986. For testing, we use data of 1991-1992 (24 samples) for verification for all test sites. Fig. 4 shows the computational results of ANN tests.

Input 1	TE: Water Temperature	[°C]
Input 2	TN: Total Nitrogen	[mg/l]
Input 3	TP: Total Phosphorus	[mg/l]
Input 4	TR: Transparency	[m]
Output	Chl-a: Chlorophyl-a	[μg/l]

Table 1 Inputs/ output of the ANNs (application 1)

Fig. 4 shows the training and testing of ANNs in simulation of Chlorophyll-a concentration time series in the test sites 101, 102, 202, 302. In fig. 4, the graphics on the left show the patterns (60 patterns every site, five-year's monthly data, from every April to March next year) for the four sites for training, while the graphics on the right show the patterns for testing (24 patterns every site, two-year's monthly data). In the training of ANN, a process like "model calibration", from the tests in all the four sites we obtained good fittings between the model outputs and observation data. On the other hand, for the testing sessions, results in the four stations made difference gratuitously from North to South. From sites 302 (Area III), 202 (Area II & III), 102 (Area II) and 101 (Area I), the closer the site's distance from the bay head, the better the simulation/testing results that we obtained. As a kind of black box model with

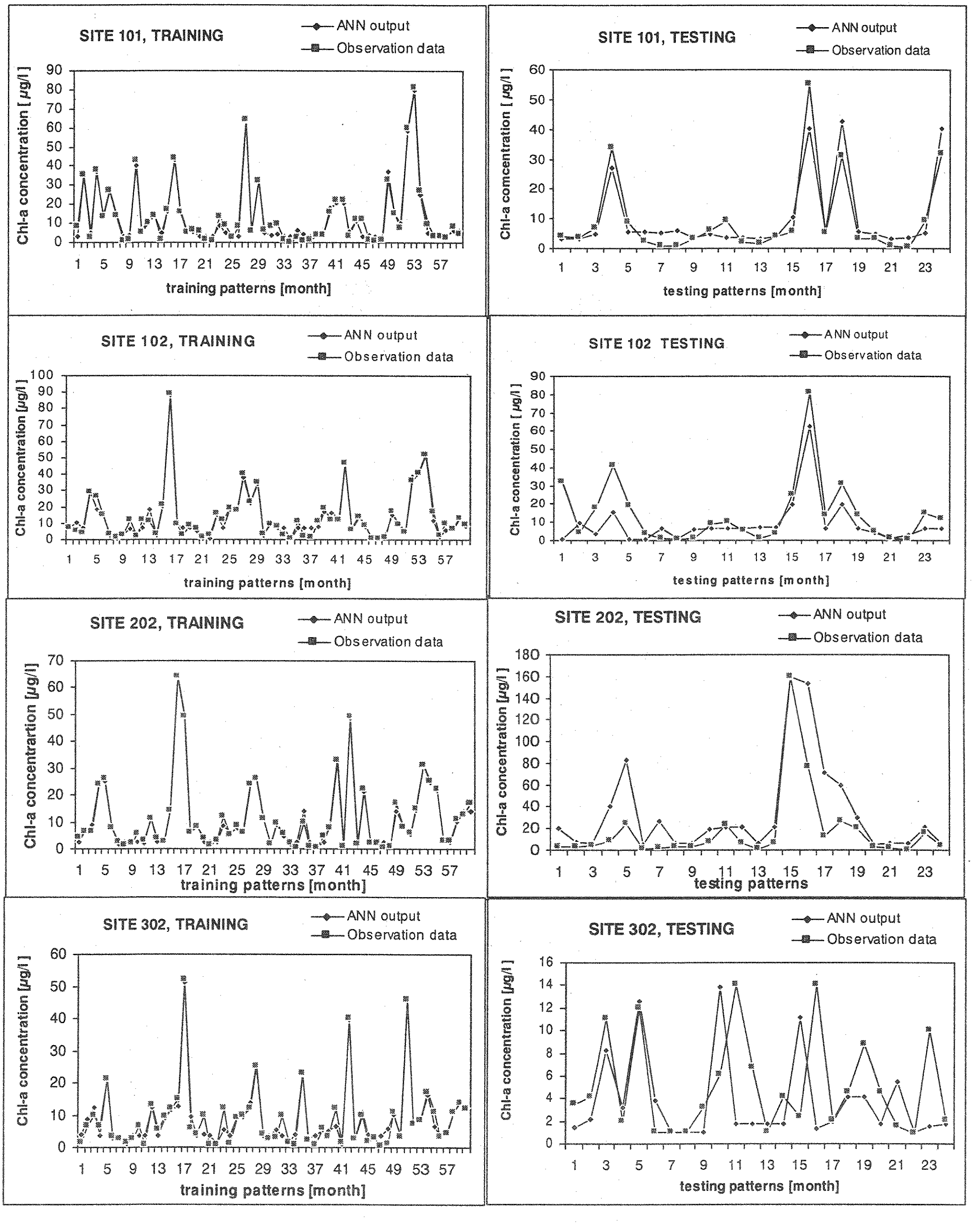

Figure 4 Training and testing of ANNs in simulation of Chlorophyll-a concentration time series in the test sites 101, 102, 202, 302.

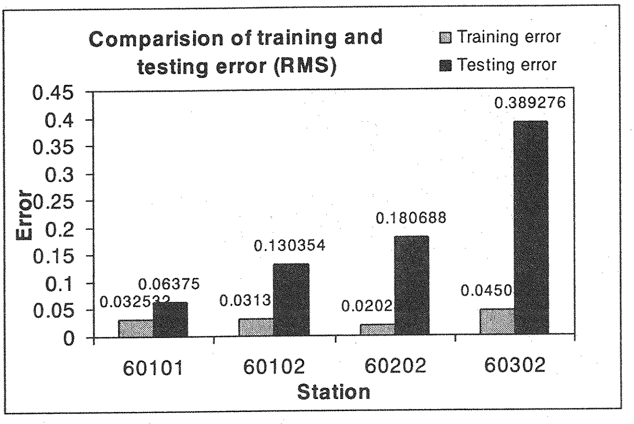

Figure 5 Comparison of training and testing error.

complicate structure, ANNs always perform well in the training session, however, in the testing session of ANN, the results may be different due to the ANN's capability and the inter-correlation between the training data and the testing data. Fig.5 summarizes the comparison of training and testing errors (RMS) for the four test locations. For testing errors, as the error values increase from Area I to Area III, the performances of the "verification" become poor. As illustrated in Fig. 4, among the "testing parts" from tests in four locations, the chlorophyll-a time series in site 101 (Area I), especially the peak values, are reasonably reproduced. In site102 and 202 (Area II), we observe some underestimation and overestimation in the peak values. In site 302 (Area III), it seems almost all of the peaks are captured in ANN output, however most peaks are somehow delayed. Such "delay", causes the very high testing error (RMS) in site 302 (station 60302), as shown in fig.5.

On hydrodynamic and water quality view points (i.e., tide, flow, stratification, and COD concentration distribution), the northeastern part (bay head area) stays in relatively "stable"condition (with less affection from tidal flow) compared with other parts of Osaka Bay. In addition, as reported by Yamane et. al (1997), the inflows from the rivers cause stratification in the top 5-meter layer in the bay-head area while in southern and western parts of the bay the water is well mixed. Figure 1 also shows the horizontal and vertical distribution of density in summer and winter in Osaka Bay. Since the rate of biological process is different in fresh water (surface water in bay head) and salty water (well-mixed water in the southern bay), the production rates of chlorophyll also make difference. This fact may explain the reason of the difference among the "verification" results in the four locations. Nevertheless, from results in Fig. 4, we find that ANNs give reasonable results in simulating Chl-a time series by "learning" from TE, TN, TP and TR, especially in the bay head area where the surface water is dominated by the river inflows and the hydrodynamic/ecological disturbance is not so strong.

CONCLUSION

In this study, experiments of artificial neural networks (ANN) on simulation of Chlorophyll-a time series have been conducted in different areas in Osaka Bay that were classified by their water quality characteristics (TE, TN, TP and TR). It was shown that, without any information about the physical processes in the ecological system, ANNs, the kind of black-box models which purely on the basis of historical data, is capable to reproduce accurately the dynamics of the behavior of the indicators describing this system. On the other hand, in the present study, the computation results show that the simulation performance of ANN varied according the test site's stratification situation, since the rate of Chlorophyll production make difference in fresh water and salty water. It is shown that the correlation between ANN and the learning object of ANN may affect the performance of the ANN model. In addition, from the present study, we find that in order to get more accurate computation results by ANN, more frequent sampling is needed for further environmental monitoring, especially in summer seasons while the Osaka Bay obtains large amounts of nutrients from the rivers.

AKNOWLEDGEMENT

We want to thank Dr. N. Yamane (CTI. Consultant Ltd, Japan) for providing the water quality data. In addition, we would also like to thank Dr. Dong Jin Han in Osaka University for his advice and comments on ecological process.

REFFERENCES

1) Aleksander, I. and H. Morton : An introduction to neural computing. Chapman and Hall, London. 1990.
2) Kuczera, G. : Improved parameter inference in catchment Models, Water Resour. Res., 19(5), 1151-1172, Oct. 1983.
3) Masters, T. : Practical neural network recipes in C++, Chapter 3 & 4, 1994.
4) Osaka Prefectural Fisheries Experimental Station (1973-1992). Regular observation in shallow sea, Bull. Osaka Pref. Fish. Exp. Stat. (in Japanese)
5) Raman, H. and N. Sunilkumar : Multivariate modeling of water resources time series using artificial neural networks. Hydrol. Sci. 40(2), April, 1995.
6) Solomatine, D. P. : Two strategies of adaptive cluster covering with descent and their comparison to other algorithms. J. of Global optimization, 1997.
7) Shen, Y., H.V.D. Boogaard and D. P. Solomatine: Improving performance of Chlorophyl concentration time series simulation with artificial neural networks. Annual Journal of Hydraulic Engineering, JSCE, April, 1998.
8) Yamane, N., K. Nakatsuji, H. Kurita and K. Muraoka: Field data collection and modeling for verification of an ecosystem model in Osaka Bay, Japan. Estuarine and coastal Modeling, Proceedings of the conference ASCE, Oct., 1997.
9) Yamane, N., K. Nakatsuji, Seasonal variation of density structure in Osaka Bay. Hydrodynamics, 1996 Balkema, Rotterdam, ISBN90 5410 8606

(Received March 15, 1999 ; revised July 2, 1999)