

# Application of Artificial Neural Network Models to the Estimation of Chlorophyll a Concentration in Lake Koyama, Tottori Prefecture, Japan

Sai, Koji

Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University

Harada, Masayoshi

Faculty of Agriculture, Kyushu University

Yoshida, Isao

Emeritus Professor, Tottori University

Hiramatsu, Kazuaki

Faculty of Agriculture, Kyushu University

他

<https://doi.org/10.5109/9330>

---

出版情報：九州大学大学院農学研究院紀要. 52 (2), pp.405-409, 2007-10-29. Faculty of  
Agriculture, Kyushu University

バージョン：

権利関係：

## Application of Artificial Neural Network Models to the Estimation of Chlorophyll *a* Concentration in Lake Koyama, Tottori Prefecture, Japan

Koji SAI<sup>1\*</sup>, Masayoshi HARADA, Isao YOSHIDA<sup>2</sup>,  
Kazuaki HIRAMATSU and Makito MORI

Laboratory of Drainage and Water Environment, Division of Regional Environment Science,  
Department of Bioproduction Environmental Sciences, Faculty of Agriculture,  
Kyushu University, Fukuoka 812–8581, Japan

(Received June 18, 2007 and accepted July 17, 2007)

An artificial neural network model with three-layer structure was applied to estimate chlorophyll *a* concentration in a lake located in Tottori Prefecture, Japan. First, input variables, which resulted in high calibration accuracy, were searched to examine the optimal network structure. The calibration accuracy was highest when input variables were set to TN, TP, DO, water temperature, solar radiation, air temperature, wind velocity, and Wedderburn number. This result means that the model incorporated the relationship between chlorophyll *a* concentration and the meteorological, hydraulic, and aquatic factors into the network structure.

The adaptability of the estimation of chlorophyll *a* concentration was examined. As a result, chlorophyll *a* concentration could not be sufficiently estimated. To improve estimation accuracy, the network structure was reconstructed by considering the time history of the variation of the meteorological and water quality data for the previous 24 hours and incorporating such data into the input variables. The result showed that the estimation accuracy was remarkably improved.

### INTRODUCTION

Closed water bodies such as lakes and inner bays suffer from various water environmental problems due to eutrophication. In a eutrophic lake, the profuse growth of phytoplankton has frequently occurred because the primary production of phytoplankton is extremely large. Phytoplankton blooming causes the deterioration of the water environment, such as the degradation of water clarity and damage to the aquatic ecosystem. To preserve water environment, it is a significant assignment to investigate the mechanism of phytoplankton blooming and establish the method for estimating the biomass of phytoplankton. Numerous studies have been done on the seasonal changes of biota of phytoplankton. Among them, several studies focused on the relationship between dominance of phytoplankton and aquatic factors. However, there has not been sufficient research on a quantitative evaluation of biomass of phytoplankton. Therefore, it is important to develop a numerical model that can estimate the biomass of phytoplankton.

Aquatic environmental factors, including the behavior of phytoplankton, were generally analyzed and estimated by an ecosystem model which incorporated the water flow analysis into the model (Nakata and Taguchi, 1982; Nakasone *et al.*, 2005). However, it is difficult to apply the model to the actual situation because the model contains numerous parameters and complicated

model equations for describing biochemical cycling in a lake. In recent years, an artificial intelligence technique has been actively applied to the evaluation and estimation of water environment. The artificial neural network model (ANNM), which possesses a superior ability for the pattern recognition, has been applied to the evaluation of the behavior of aquatic factors in rivers and other waterworks (Kondo, 1997; Hiramatsu, 1995; 1998).

At most of the closed water bodies in Japan, numerous weather items and water quality items such as dissolved oxygen, nutrients, and COD were observed continuously or periodically. The observed data contribute toward not only the analysis but also the estimation of water environment. Generally, the behavior of water environment intricately fluctuates with the interrelation between numerous water environmental factors. In this present paper, the fluctuation patterns of water environment extracted from the sufficient accumulated data have been utilized for estimation of water qualities. Specifically, ANNM, which is a nonlinear model and can extract the relationships between water environmental factors, was applied to estimate the behavior of water environments.

In this study, the feasibility of estimation of water environment in closed water bodies using ANNMs was examined. An object of estimation is chlorophyll *a*, which represents the amount of phytoplankton, in Lake Koyama, Tottori Prefecture, Japan. This paper proposes that the aquatic and meteorological data obtained by water environmental monitoring is used for estimation of water environment directly and actively.

### MATERIALS AND METHODS

#### Study area and outline of observation

Lake Koyama, which has a surface area of 6.8 km<sup>2</sup>,

<sup>1</sup> Laboratory of Drainage and Water Environment, Division of Regional Environment Science, Department of Bioproduction Environmental Sciences, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University

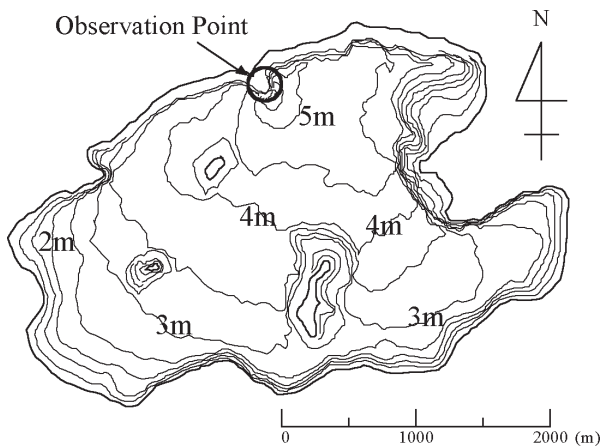
<sup>2</sup> Emeritus Professor, Tottori University

\* Corresponding author (E-mail: sai@bpes.kyushu-u.ac.jp)

an average depth of 2.8 m, a maximum depth of 6.5 m, and a storage capacity of 20 million m<sup>3</sup>, is a brackish lake located east of Tottori prefecture in Japan. The lake has five inflows and an outflow. The inflows are the Oide River, the Miyamaguchi River, the Yujiri River, the Nagara River, and the Fukui River; the outflow is the Koyama River. In the Koyama River, the intrusion of seawater occurs due to tides. At the mouth of the Koyama River, the sluice gate is constructed to regulate the amount of inflow of seawater, and salinity of lake water is controlled at 130 ppm for the irrigation period and 330 ppm for the non-irrigation period, respectively. Therefore, Lake Koyama is a low salinity, blackish lake in the irrigation period. The bottom topography of the lake inclines gently to the northeast, and its deepest area is the vicinity of the north shore, as shown in **Fig. 1**.

**1.** At its deepest, the deterioration of the water environment is remarkable, and the incidence of the oxygen-deficient water and phytoplankton blooming occurs frequently (Harada and Yoshida, 2001; Shimizu, 2002). Therefore, the authors conducted continuous observations and weekly observations of water quality and weather items at the deepest point from April 2004 to November 2005. Details of the observation items are shown in **Table 1**.

In weekly observations, water sampling was conducted at the surface layer (about 15 cm below the water surface) and at the bottom layer (about 15 cm



**Fig. 1.** Bottom topography of Lake Koyama and observation point.

**Table 1.** Observation items

Category	Items	Position	Interval
Weekly observation	Chlorophyll <i>a</i> TN TP	Surface layer and bottom layer	About 7 days
Continuous observation	EC	Surface layer and bottom layer	Ten minutes
	Salinity	Surface layer and bottom layer	
	DO	Surface layer and bottom layer	
	Water temp.	About 2 m above the water surface	
	Solar radiation	About 2 m above the water surface	
	Air temp.	About 5 m above the water surface	
	Wind velocity	About 5 m above the water surface	

from the bottom of the lake) between eleven a.m. and noon. The sampled water was analyzed for quality. In the continuous observation, the water quality and weather items were measured at ten minute intervals using an automatic data logger. In this observation, chlorophyll *a* concentration was analyzed by Spectrophotometric Measurement. In addition, the population and the biomass rate of phytoplankton were surveyed as biological research. Samples for determining phytoplankton composition were obtained by using plankton net, which has 65 μm mesh. The samples were fixed with 0.5% (v/v) buffered formalin and used for microscopic observation. For the continuous observed data, the values at one hour intervals were used in the numerical analysis, while it was measured at ten minute intervals.

**Artificial neural network model**

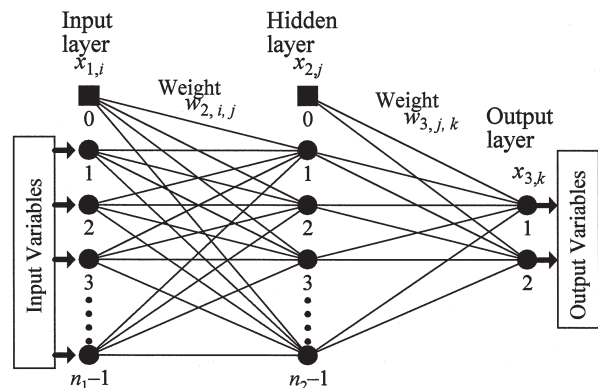
An artificial neural network model is a mathematical model that extracts the basic functions of biological neural networks. It is made by mutually interconnecting artificial neurons. In this study, a perceptron model is employed to the estimation of chlorophyll *a* concentration. The model has a multi-layer structure consisting of an input layer, a hidden layer, and an output layer, as shown in **Fig. 2**. The three-layer perceptron model possesses a superior ability for pattern recognition. The chlorophyll *a* concentration was estimated by this ability.

In a perceptron model composed of an input layer (*n*<sub>1</sub> units), a hidden layer (*n*<sub>2</sub> units), and an output layer (*n*<sub>3</sub> units), as shown in **Fig. 2**, outputs {*x*<sub>3,*k*</sub>; *k* = 1, *n*<sub>3</sub>} are computed as follows:

$$x_{2,j} = f \left[ \sum_{i=0}^{n_1-1} w_{2,i,j} x_{1,i} \right]; j = 1, n_2-1 \quad (1)$$

$$x_{3,k} = f \left[ \sum_{j=0}^{n_2-1} w_{3,j,k} x_{2,j} \right]; k = 1, n_3 \quad (2)$$

where *x*<sub>1,0</sub> (= 1) is the constant used for output computation of the hidden layer, *x*<sub>1,*i*</sub> {*i* = 1, *n*<sub>1</sub>-1} are the inputs to the hidden layer, *x*<sub>2,*j*</sub> {*j* = 1, *n*<sub>2</sub>-1} are the inputs to the output layer, *w*<sub>2,*i,j*</sub> {*i* = 1, *n*<sub>1</sub>-1, *j* = 1, *n*<sub>2</sub>-1} are the input-hidden weights used for output computation of



**Fig. 2.** The perceptron model with three-layer structure.

the hidden layer, and  $w_{3,jk}$  ( $j = 1, n_2 - 1, k = 1, n_3$ ) are the hidden–output weights used for output computation of the output layer. The equations (1) and (2) are used in the computations for the hidden and output layer outputs. Nonlinear transfer function  $f[y]$  is the sigmoid function defined by

$$f[y] = \frac{1}{1 + e^{-y}} \quad (3)$$

where  $\theta$  is the parameter that prescribes the response of  $f[y]$ . In this analysis, parameter  $\theta$  is provided as a constant ( $\theta = 1$ ) because it cannot be adjusted by the generalized delta law. The values of the desired output and input were normalized to a range of [0, 1].

In this study, the generalized delta law is used for network learning. First, the square–error between network outputs  $x_{3,j}$  and desired outputs  $t_{p,k}$  is defined by

$$E_p = \frac{1}{2} \sum_{k=1}^{n_3} (t_{p,k} - x_{3,k})^2 \quad (4)$$

Next, weights  $w$  and thresholds  $\theta$  are adjusted by transmitting the error from the output to the hidden layer and from the hidden to the input layer as

$$\Delta x_{3,k} = x_{3,k} (1 - x_{3,k}) (t_{p,k} - x_{3,k}) : k = 1, n_3 \quad (5)$$

$$\Delta x_{2,j} = x_{2,j} (1 - x_{2,j}) \sum_{k=1}^{n_3} w_{3,jk} \Delta x_{3,k} : j = 1, n_2 - 1 ; k = 1, n_3 \quad (6)$$

$$w_{3,jk}^{(M+1)} = w_{3,jk}^{(M)} + \Delta x_{3,k} \Delta x_{2,j} + a \{ w_{3,jk}^{(M)} - w_{3,jk}^{(M-1)} \} : j = 0, n_2 - 1 ; k = 1, n_3 \quad (7)$$

$$w_{2,i,j}^{(M+1)} = w_{2,i,j}^{(M)} + \Delta x_{2,j} \Delta x_{1,i} + a \{ w_{2,i,j}^{(M)} - w_{2,i,j}^{(M-1)} \} : i = 0, n_1 - 1 ; j = 1, n_2 - 1 \quad (8)$$

where  $\eta$  is the learning rate,  $a$  is the inertia coefficient, and  $w^{(M)}$  represents the weight calibrated  $M$  times. In Equations (7) and (8), the inertia term was added to avoid the despair of local minima and to accelerate the convergence (Hiramatsu *et al.*, 1995). The learning rate and the inertia coefficient were set at 5.0 and 0.9, respectively, by trial and error. The generalized delta law was conducted by the batch processing method. The weights and thresholds were finally determined by the calibration with 50,000 iterations.

## RESULTS AND DISCUSSION

### Examination of optimal network structure

To examine the optimal network structure, the input variables, which resulted in high calibration accuracy, were searched. From the viewpoint of efficient utilization of the data accumulated by water environmental observations, the basis network structure was constructed by using continuous observation items, which are possible to measure in short time intervals.

**Table 2.** Cases of input variables for calibration

Case	Input variables
Case 1	DO, water temp., solar radiation, air temp. and wind velocity.
Case 2	TN, TP, DO, water temp., solar radiation, air temp. and wind velocity.
Case 3	TN, TP, DO, water temp., solar radiation, air temp., wind velocity and Wedderburn number.

Furthermore, the weekly observation items, which relate to behavior of phytoplankton, were added into input variables of the basis network structure in order to improve the calibration accuracy.

In this section, calibrations of ANNMs were conducted using the input variables of the three cases shown in **Table 2**. First, the input variables of Case 1 were obtained from the continuous observations and the number of units in the input layer is  $n_1 = 7$ . Second, the input variables of Case 2 were obtained by adding nutrients (TN and TP), which affects the bleeding of phytoplankton, to Case 1 and its number of units in the input layer is  $n_1 = 11$ . Finally, the input variables of Case 3 were obtained by adding the Wedderburn number, which represents the strength of the stratification of the lake, to Case 2 and its number of units in the input layer is  $n_1 = 12$ . The Wedderburn number is defined as follows

$$W = \frac{\rho_2 - \rho_1}{\rho_1} \frac{gh}{U_*^2} \frac{h}{L} \quad (9)$$

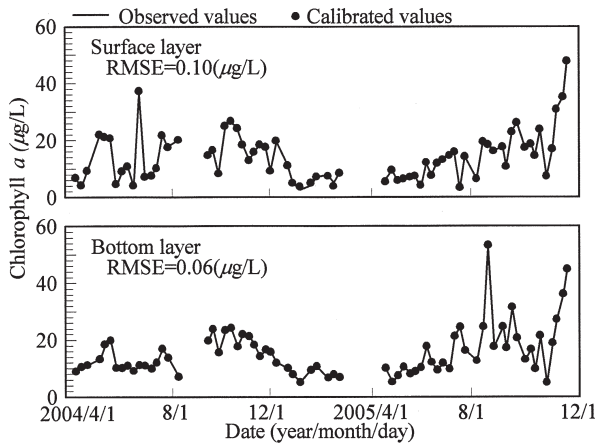
where  $\rho_1$  and  $\rho_2$  are the water density at the surface and bottom layers, respectively,  $g$  is acceleration due to gravity,  $h$  is the water depth,  $L$  is the horizontal length of the lake, and  $U_*^2$  is the friction velocity on the water surface caused by wind. In all cases, the output variables were chlorophyll  $a$  concentrations at the surface and bottom layers, and the number of the units in the output layer is  $n_3 = 2$ . The number of the units in the hidden layer is  $n_2 = 20$  in each case. The number of sets of calibration data was 70, which is the total of the weekly observations. The weekly observation items used in the following analysis were regarded as the values at noon. Therefore, for the continuous observation items in the input variables, the values observed at noon were used for the analysis.

The calibrated errors, which are the root mean square errors (RMSEs) between the observed values of chlorophyll  $a$  concentration, and the calculated values obtained after learning with 50,000 iterations, are shown in **Table 3**. As an example, the calibrated results of Case 3 are shown in **Fig. 3**. In the figure, the unwritten periods from 10<sup>th</sup> August to 15<sup>th</sup> September 2004 and from 22<sup>th</sup> February to 19<sup>th</sup> April 2005 are loss of the data due to instrument trouble. **Table 3** and **Fig. 3** reveal that the calibrated errors of each case were small, and network calibration was carried out sufficiently.

The generation of phytoplankton was affected by nutrient (especially, nitrogen and phosphorous) concentration in lakes. Also, the hydraulic factor affected the

**Table 3.** Calibrated errors obtained from Cases 1, 2, and 3

Case	Calibrated errors ( $\mu\text{g/L}$ )	
	Surface layer	Bottom layer
Case 1	0.49	0.51
Case 2	0.21	0.11
Case 3	0.10	0.06



**Fig. 3.** Calibrated results of Case 3.

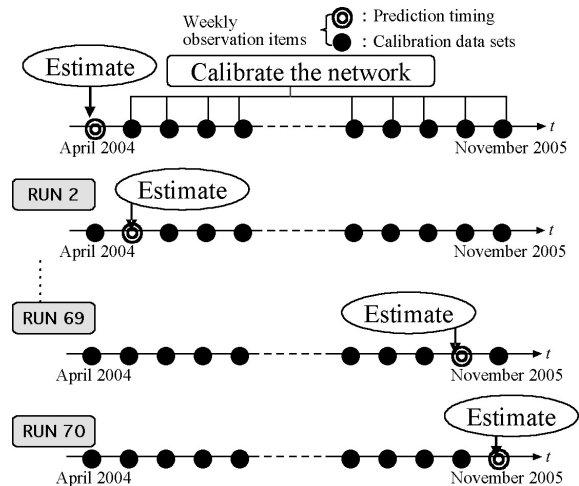
behavior of phytoplankton because the vertical movement of lake water is involved with the diffusion of water environmental factors such as nutrients and organic matter. The calibration accuracy of Case 3 is the highest of the three cases, denoting that the model incorporated the relationship between chlorophyll *a* concentration and the meteorological, hydraulic, and aquatic factors into the network structure.

**Estimation of chlorophyll *a* concentration**

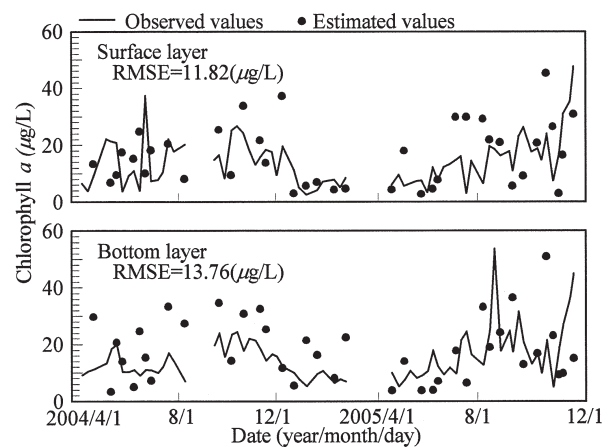
The feasibility of the estimation of chlorophyll *a* concentration was examined by a model with the input variables of Case 3. After calibrating the network using 69 sets selected from the 70 sets of calibration data, the estimated values of chlorophyll *a* concentration for another remaining set were calculated. By repeating this calculation 70 times, the estimation values for all weekly observations were calculated and compared with the observed results. The outline of this operation is shown in **Fig. 4**.

The estimated results and errors, which are the RMSEs between the calculated and observed values, are shown in **Fig. 5**. We can see the large deviations between the estimated and observed values in the figure. The estimated errors seem to exceed frequently  $10 \mu\text{g/l}$  at both the surface and bottom layers. This results show that the adopted model is insufficient to estimate chlorophyll *a* concentrations in the lake.

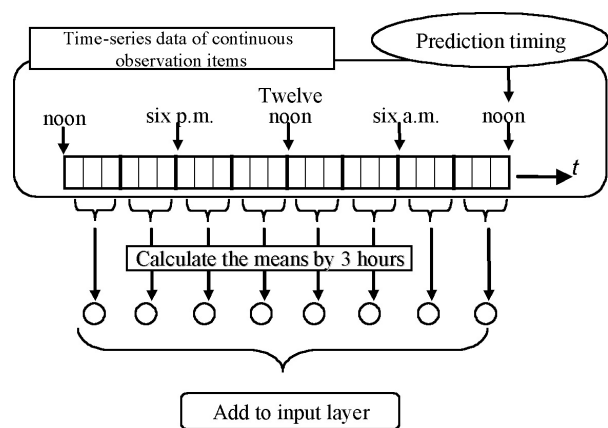
To improve estimation accuracy, ANNM was reconstructed by considering the time history of the variation of the meteorological and water quality data over the previous 24 hours and incorporating these data into the input variables. In this operation, the previous 24 hours of 8-fold mean-values of continuous observation items



**Fig. 4.** Outline of estimation of chlorophyll *a* concentration.



**Fig. 5.** Estimated results of Case 3.



**Fig. 6.** Outline of incorporating time history of continuous observation items to model.

were calculated, and these means were added to the input variables of Case 3, as shown in **Fig. 6**. This network structure is called Case 4 and has 69 units in its input layer. The units in the hidden and output layers were identical to Case 3.

The estimated results and errors of Case 4 are



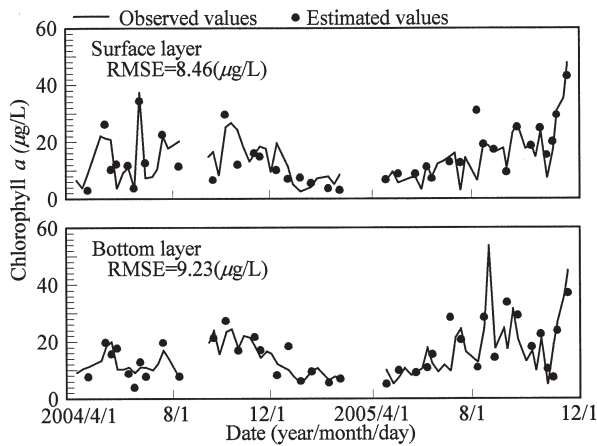


Fig. 7. Estimated results of Case 4.

shown in **Fig. 7**. The estimated results approximately agreed with the observed results, and the large deviations between the estimated and observed values, viewed at **Fig. 5**, improved remarkably in Case 4. We can conclude that it is valid for the estimation of chlorophyll *a* concentration using the ANNM to incorporate the history of the previous 24-hour data into the input layer of the network structure.

## CONCLUSIONS

In this study, estimation of chlorophyll *a* concentration was conducted by a perceptron model with a three-layer structure.

To produce the optimal network structure, the input variables, which resulted in high calibration accuracy, were searched. As a result, the calibration accuracy was highest when the input variables were set to TN, TP, DO, water temperature, solar radiation, air temperature, wind velocity and Wedderburn number. This result means that the obtained model incorporated the relationship between the chlorophyll *a* concentration and the meteorological, hydraulic and aquatic factors into

the network structure.

Next, chlorophyll *a* concentrations were estimated using above input variables. As a result, the ANNM could not estimate chlorophyll *a* concentrations sufficiently. To improve the estimation accuracy, ANNM was reconstructed by considering the time history of the variation of the continuous observation items. As a result, the estimated error remarkably decreased and the estimated results approximately agreed with the observed results.

## REFERENCES

- Harada, M. and I. Yoshida 2001 Fluctuation Characteristics of Dissolved Oxygen in Lake Koyama of Tottori Prefecture. *Journal of rainwater catchment systems*, **10(1)**: 29–35 (in Japanese with English abstract)
- Hiramatsu, K., S. Shikasho and K. Mori 1995 Application of Multi-Layered Perceptron Model to the Estimation of Chlorinity Variations in a Tidal River. *Transactions of the Japanese Society of Irrigation, Drainage and Reclamation Engineering*, **178**: 83–92 (in Japanese with English abstract)
- Hiramatsu K., S. Shikasho, K. Mori 1998 The use of genetic algorithm and artificial neural network for the short-term prediction of water-stages in a tidal river. *Proceedings of 11th Congress of Asia and Pacific Division of the International Association for Hydraulic Research (APD-IAHR)*, Yogyakarta, Sep. 8–10, pp. 587–595
- Kondo, T 1997 Nonlinear Steady State Modeling for River Quality by a Neural Network with a Self-Selection Ability of Network Structure. *Transactions of the Institute of Systems, Control and Information Engineers*, **10(6)**: 277–286 (in Japanese with English abstract)
- Nakasone, H., M. Kaburagi, H. Kuroda and T. Kato 2005 A Study on the Low DO Distribution in the Lake Kasumigaura. *Transactions of the Japanese Society of Irrigation, Drainage and Reclamation Engineering*, **239**: 471–479 (in Japanese with English abstract)
- Nakata, K. and K. Taguchi 1982 Numerical Simulation of Eutrophication Process in Coastal Bay by Eco-hydrodynamic Model –2– Coastal Bay Ecosystem Model. *Bulletin of the National Research Institute for Pollution and Resources*, **11(4)**: 47–69 (in Japanese with English abstract)
- Shimizu, T., I. Yoshida, M. Harada, K. Sitizyo and H. Takashiba 2002 Study on the formation of an oxygen deficient layer in Lake Koyama. *Proceedings of the International Agricultural Engineering Conference*, pp. 495–499