Journal of Hydroscience and Hydraulic Engineering
Vol. 26, No.2 November, 2008, 27-39

NEW APPROACH FOR TIME SERIES ESTIMATION OF POLLUTANT LOAD IN RIVERS BY USING OPTICAL SENSOR MEASUREMENT AND ARTIFICIAL NEURAL NETWORK

By

Minghuan Liu

Dept. of Civil Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan

Tadaharu Ishikawa

Dept. of Environmental Science and Technology, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan

Kenji Yoshimi

Dept. of Environmental Science and Technology, Tokyo Institute of Technology, Midori-ku, Yokohama, Japan

and

Kentaro Kudo

IDEA Consultants, Inc., Tsuzuki-ku, Yokohama, Japan

SYNOPSIS

This study proposes a new type monitoring technique of pollutant load in rivers: Optical characteristics of river water are monitored by a multi-item optical device. The relation between the sensor signals and the water qualities obtained from occasional sample analysis is modeled by an Artificial Neural Network (ANN). Next, the time series of pollutant load can be produced from the optical signals. Field experiments were conducted in seven rivers flowing into Lake Kasumigaura. The ANN model trained by the data obtained in 2005 successfully produced the time series of pollutant load observed in 2006 and 2007. The ANN model works well in watersheds of different land use conditions if it is trained by the data obtained in each river. Furthermore, there is a possibility that an ANN model constructed in a river basin can be tentatively applied to other river basins where data of water sampling analysis are not sufficient to calibrate the ANN model if the land use conditions are similar to some extent.

INTRODUCTION

Over the past several decades, many closed water bodies such as lakes, reservoirs, and inner bays became eutrophicated by pollutant load generated by human activities in the watersheds. While the load from point sources has been reduced by establishing laws and regulations in recent years, non-point sources such as agricultural lands, livestock and poultry farms, urban areas and roads, still discharge polluted effluents. Accordingly, monitoring and regulating the latter is a major difficulty in solving the eutrophication problem (1). Because pollutant loads from non-point sources rapidly increase during floods, water quality analysis of monthly or even weekly sampled water does not have sufficient time resolution (2), so it is necessary to develop techniques of continuous or high frequent measurements of water quality.

Robotic *in situ* laboratory is installed in some rivers for high frequent water analysis, but it is not widely available because it is expensive and maintenance is costly. On the other hand, monitoring of some water quality items is done by using immersed-type optical sensors: concentration of particulate phosphorus can be estimated by using turbidimeters (3-5); a fluorometer for measurement of chlorophyll-a (*Chl-a*) is already commercialized and widely used in field studies (6); and dissolved *COD* is measured by using a *UV* meter in trial (7). At present, however, the items that can be measured in this way are limited.

As a result, the major indices for eutrophication such as *COD*, *TN* (total nitrogen) and *TP* (total phosphorus) are estimated by using so-called *L-Q* method (2, 8, 9). The method assumes the pollutant load (*L*) as a single function of the river flow rate (*Q*) being based on a simple fact that pollutant load increases during a flood event. Needless to say, the *L-Q* equation is a totally empirical correlation which depends on the characteristics of each river. Many researchers have examined the errors of this method and have concluded that the degree of accuracy is not very high (3). However, the strong point of *L-Q* method is that it can estimate a time series of load without much effort, and accordingly, it is still used widely in practice.

The purpose of this study is to increase the possibility of water quality monitoring with optical sensors based on an extensional idea of *L-Q* method, which will be described in the next section. Field data of the study was collected in seven rivers flowing into Lake Kasumigaura for three years. Artificial neural network (ANN) was employed to formulate and determine comprehensive empirical correlations between the signals from optical sensors and the loads of *COD*, *TN* and *TP* with the data obtained in the first year. The performance of ANN models was examined by using the data of successive two years.

BASIC IDEAS

Generally speaking, any measurement techniques employ some correlations between what we want to know and what we can measure. The correlations can be classified into two categories. One is "universal correlations supported by scientific theories and evidence". For example, a standard measurement technique of material strain is based on the correlation between the deformation rate of a gauge and the change of its electric characteristics. The classical thermometer uses a correlation between the mercury's temperature and its volume, and dissolved oxygen is measured by using the correlation between the intensity of electric current and the amount of oxygen in the electrolytic liquid. Such measurement techniques are reliable, but the measurable items available for field measurements are limited.

The main point of this study is to utilize the "peculiar correlations at each site of measurement" as much as possible in order to estimate a time series of water quality. Let us assume the following ideal situation (See Fig. 1): There are two water sources in a water basin, say A and B. We focus on two items of water quality α and β, whose concentrations are C_α and C_β, respectively. The water from A has high C_α and low C_β, and the water from B has an opposite characteristics, low C_α and high C_β. When the ratio of Water-A and Water-B changes in the process of rain runoff, the water quality observed at a downstream station may move along the straight line as shown in the figure. By doing this, we can "estimate" C_α by measuring C_β, even if we cannot measure C_α directly.

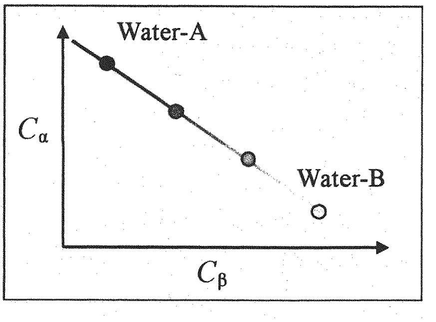

Fig. 1 Diagram of the two water sources

Extending the above considerations, the authors intend to develop a method of water quality monitoring by using the signals from optical sensors (10, 11): If the water quality item β has some optical characteristics such as pigment or fluorescence, the signals from optical sensors may have a peculiar correlation with C_α as well as C_β. As a result, the pollutant load of the item α (L_α) may have a peculiar correlation with the signals and the river flow rate. In a practical situation, the correlation may be comprehensive and nonlinear among multiple factors. In this study, an artificial neural network is employed to formulate and determine the comprehensive peculiar correlations.

METHOD DESCRIPTION

Extension of L-Q method based on peculiar correlation

The so-called *L-Q* equation is a typical peculiar correlation. It assumes the pollutant load (L) as a single function of the river flow rate (Q) as shown in Eq. (1). The concrete form of the equation is constructed by means of field data obtained at each station.

$$L = f_1(Q) \tag{1}$$

Although the accuracy of the estimation is questioned by many researchers, the *L-Q* method is still used widely in practice because it can estimate the time series of L without much effort.

As pollutant load is a product of concentration (C) and river flow rate (Q), Eq. (1) can be written in the following form:

$$L = C(Q) \times Q \tag{2}$$

The equation expresses that the pollutant concentration is a single function of Q. On the other hand, many observations provide evidence that the peak of particulate substances appears in the increasing phase of a flood. This fact means that L cannot be expressed only by Q.

In view of this matter, we attempt to introduce other factors to construct a new peculiar correlation.

$$L = f_2(Q; X_1, X_2, X_3, \cdots) \tag{3}$$

where X_1, X_2, $\cdots$ are indices of water quality whose time series can be obtained easily. Eq. (1) is the simplest form of Eq. (3) obviously. As aforementioned, the signals from optical sensors were used as the indices in this study. In general, however, they can be other measurable items such as electric conductivity, temperature difference between water and air, and so on.

The counterpart of Eq. (2) can be written as follows:

$$L = C(X_1, X_2, \cdots) \times Q \tag{4}$$

where Q is dropped from the term of C, because the dependency of C on X_is is more essential and reasonable in a sense of scientific reasoning than the indirect dependency on Q. The authors already compared the performances of Eq. (3) and Eq. (4) in a previous paper (12), and found that the results from Eq. (4) are slightly better than those from Eq. (3). Therefore, Eq. (4) is adopted in this paper.

Table 1 Basic information and measurement periods of the experiment sites in this study

River	Area (km^2)	Annual mean flow rate (m^3/s)	Land use (%)						Measurement period
			Paddy	Cropland	Forest	Urban	Grassland	Other use*	
Koise	144.6	3.59	19.5	18.6	49.4	7.0	3.7	1.8	6/1/2005~12/1/2005& 5/18/2006~12/31/2007
Sonobe	71.6	1.46	15.3	41.9	23.8	13.1	1.3	4.6	
Hokota	39.5	0.93	11.3	53.9	24.2	7.5	1.8	1.3	9/15/2006~12/31/2007
Ono	144.8	2.22	14.0	32.8	18.2	21.5	3.9	9.6	
Sakura	333.0	7.70	28.1	16.9	34.6	11.7	4.1	4.6	
Seimei	25.0	0.61	20.0	29.1	22.7	18.5	4.5	5.2	
Tomoe	113.2	2.11	15.9	46.8	21.6	11.0	1.1	3.6	

*Other uses include water surface, wasteland and other.

In this study, the loads of *COD*, *TN* and *TP* were considered as L in the equation. X_1 and X_2 are two kinds of signals from an optical device (Compact-CLW: ALEC Electronics). X_1 is the intensity of back scattered light whose wavelength is 880 nm, and the converted output of this signal is so-called turbidity (*Tb*). X_2 is the intensity of fluorescent light of 680~1000 nm from the target water induced by the emission light of 470 nm. It has been said that X_2 has a correlation with the chlorophyll-a (*Chl-a*) obtained from water analysis, but they are not identical because the correlation depends on the species of algae, and measurements are also affected by the existence of other fine particles in water.

Artificial Neural Network (ANN)

Recently, ANN has been widely used for empirical modeling of hidden dynamics in the environment (13-16). It is said that ANN belongs to a class of data driven approach whereas conventional statistical methods are model driven (17). In other words, ANN is more flexible than conventional statistic methods because it can grasp complex relations among environmental data, or, in a sense, ANN is able to simplify the procedure of statistical analysis of the complex relations (18).

The software named "Predict" supplied by the Neuralware Company was adopted in this study in order to model the peculiar correlation in the form of Eq. (4). The Predict was designed based on "Cascade-Correlation Learning Architecture" which begins with a minimal network, then automatically trains and adds new hidden units one by one, creating a multi-layer structure (19). One of the benefits of this method is that the network retains the structure even if the training set changes, and that it requires no back-propagation of error signals through the connection of the network. In our case, therefore, even if the condition of river basin gradually changes, the model can be easily improved by adding training data obtained from recent measurement (12).

STUDY AREA AND EXPERIMENTAL SETUP

Studied sites

Lake Kasumigaura is the second largest lake in Japan which is located to the northeast of Tokyo. The average depth is only 4 meters and the water surface area is about 220 km^2. The area of the watershed is 2157 km^2. The lake is eutrophic due to pollutant load mainly from non-point sources in the watershed. The experiment was carried out in seven rivers flowing into the lake. Fig. 2 shows the location of measurement sites, and Table 1 shows the areas and land use classifications of the river basins.

Table 2 Water sampling numbers of the seven rivers

River	Year	COD	TP	TN
Koise	2005	50	50	35
	2006	35	35	29
	2007	65	65	65
Sonobe	2005	50	50	28
	2006	37	37	34
	2007	50	50	50
Others*	2006	21	21	21
	2007	48	48	48

* Others include the Hokota, Ono, Sakura, Seimei and Tomoe River.

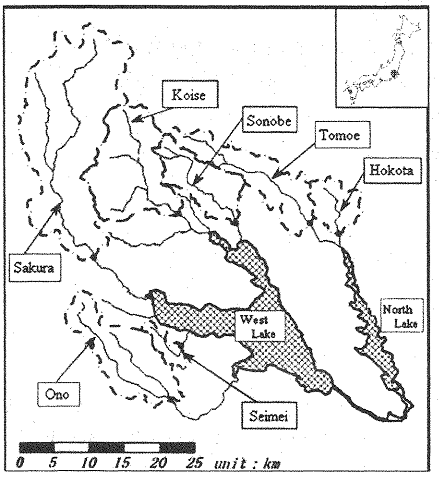

Fig. 2 Location of experiment sites

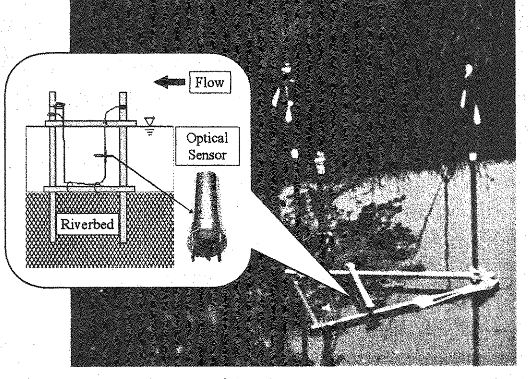

Photo 1 Experimental setup

Optical sensor measurement

An immersed-type optical device (Compact-CLW) which integrates two optical sensors was used to monitor river water. The specifications of the device were described in the previous chapter. A wiper is installed at the end of the device to clean the windows where the excitation and fluorescent lights pass through at regular intervals. The device was placed at the low channel of each river about 1 meter under the water surface of normal condition (see Photo 1). The measurement interval was 10 minutes. The experiments started in the year 2005 at the Koise River and the Sonobe River, and in 2006 at other five other river locations (see Table 1).

Water sampling and chemical analysis

The water for chemical analysis was sampled in both storm condition and low flow condition. The frequency of the latter was once or twice in a month accompanied with the maintenance of the optical device. The major items of chemical analysis were *COD*, *TP* and *TN*. They are analyzed by using the potassium permanganate acidic method, the potassium peroxidisulfate decomposition – molybdenum blue method and the potassium peroxidisulfate decomposition – E. cadmium reduction method respectively. The number of water samples is shown in Table 2.

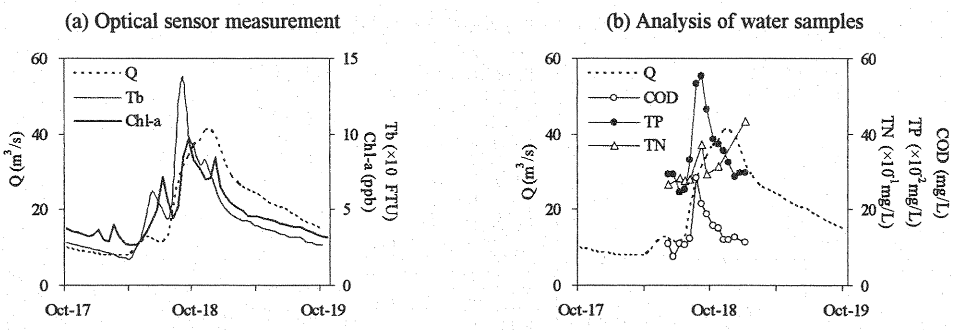

Fig. 3 Observations during a flood event at October 17th-18th, 2005 in the Koise River

River flow rate

Time series of water surface levels were provided by the Water Information System (supported by Ministry of Land, Infrastructure and Transport, Japan) except for the Sonobe River where the authors collected the data of water surface level by using the HOBO Water Level Loggers (Onset Company). The time series of water levels were converted to river flow rate by an *H-Q* curve for each river supplied by the Kasumigaura River Office. The annual average flow rates of the rivers are shown in Table 1 (20).

RESULTS AND DISCUSSION

Experiment results

Fig. 3(a) shows the signals from the optical sensors and river flow rate during a flood event of the Koise River in the year 2005. The solid line shows *Tb* while the bold line shows *Chl-a*, and the broken line represents the river flow rate. It can clearly be seen that the fluctuations of *Tb* and *Chl-a* were different from that of river flow rate.

Fig. 3(b) shows the results of water chemical analysis of the major items during this flood event. There was a time lag between the peak of river flow rate and the peaks of pollutant concentrations which was mostly caused by the particulate forms of pollutants which had a correlation with the *Tb* in Fig. 3(a).

Long-term applicability of ANN model

Data collected from the Koise River, where the measurement period was the longest and the volume of data was the largest among the seven rivers, were used to discuss the long-term application of peculiar correlation models constructed by ANN in the form of Eq. (4). Independent model was trained respectively for *COD*, *TP* and *TN* as shown in Fig. 4. The total data were divided into two parts: the data obtained in 2005 for training the model and the data obtained in the years 2006 and 2007 for evaluating the performance of the model.

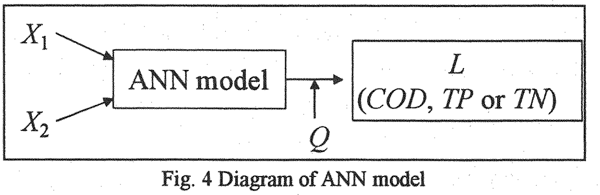

Fig. 4 Diagram of ANN model

Because the ANN model is built on a group of empirical and nonlinear relations among input, hidden and output neurons, it is difficult to express it with a simple equation of explicit form. However, it is possible to examine its response characteristics by supplying artificial data, which are arranged to change systematically, into the ANN model.

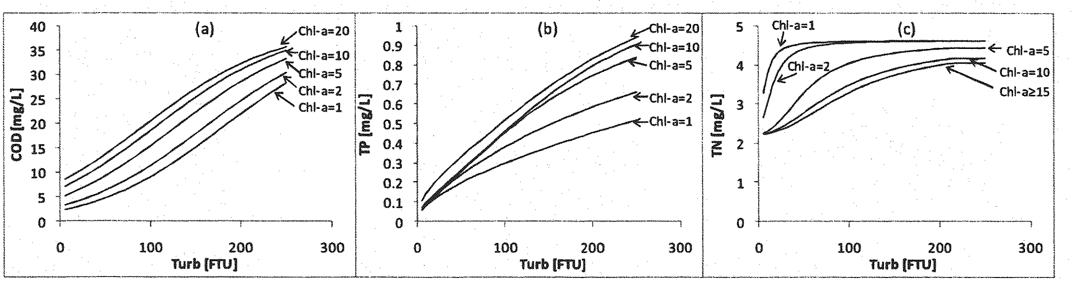

Fig. 5 Examination of the ANN models of *COD*, *TP* and *TN* of the Koise River

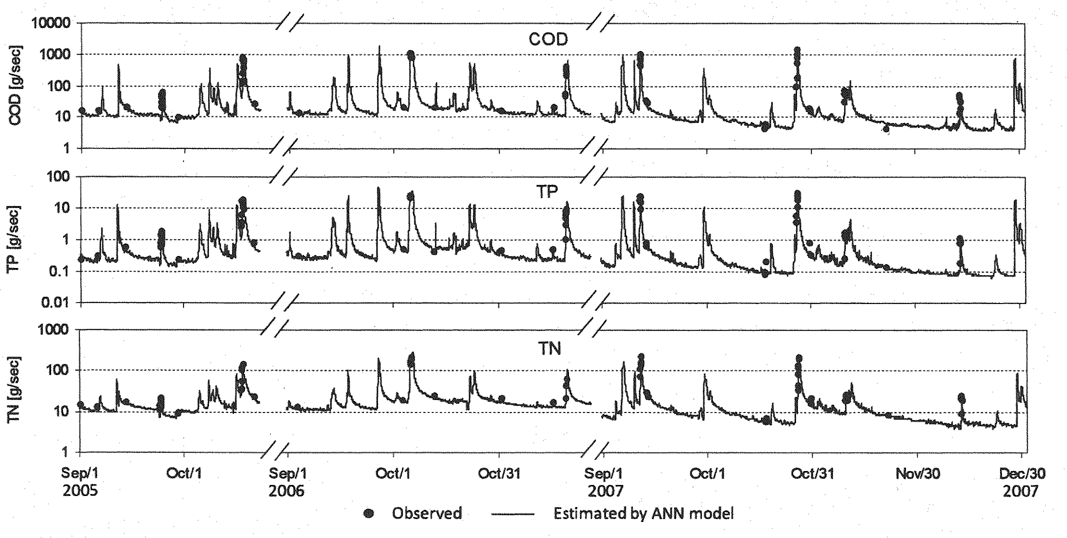

Fig. 6 Time series of pollutant fluxes of the Koise River estimated by the ANN models

Fig. 5 shows the results of the examination for the case of Koise River. *COD* (Fig. 5(a)) has positive correlations with *Chl-a* and *Tb*, and becomes a monotonous function at high *Chl-a*. *TP* (Fig. 5(b)) has the same tendency as *COD*, probably because particulate components are dominant for these two items during flood events. On the other hand, the response of the ANN model for *TN* (Fig. 5(c)) has a complicated tendency. In general, the response of ANN is not always possible to interpret because it is an empirical expression of peculiar correlations.

The time series of the pollutant fluxes of *COD*, *TP* and *TN* estimated by the ANN models are shown in Fig. 6. The solid lines are values estimated by the ANN models, and the dots show the results of chemical analysis. In spite of slight discrepancies in some parts, the general agreement seems to be good even though they were deduced from a model that was calibrated only by the data of the year 2005.

The detailed variations of the same data during four flood events are enlarged in Fig. 7. On the other hand, Fig. 8 compares the observed and estimated water qualities (concentrations) for the same flood events. The broken lines show river flow rate. The scales and patterns of the hydrograph and pollute-graph in the figures varied from one flood to another. The peak of river flow rate was preceded by the peak of pollutant concentration in most cases as shown in Fig. 8. Such phenomena cannot be described by the conventional *L-Q* method. This fact showed the possibility that the introduction of the proposed measurement method can improve the estimation accuracy of pollutant load remarkably.

Nevertheless, every field measurement based on peculiar relations will fail if site conditions change. Therefore, the ANN models should be inspected with new data at some intervals. As is shown in Fig. 6 the model trained with the

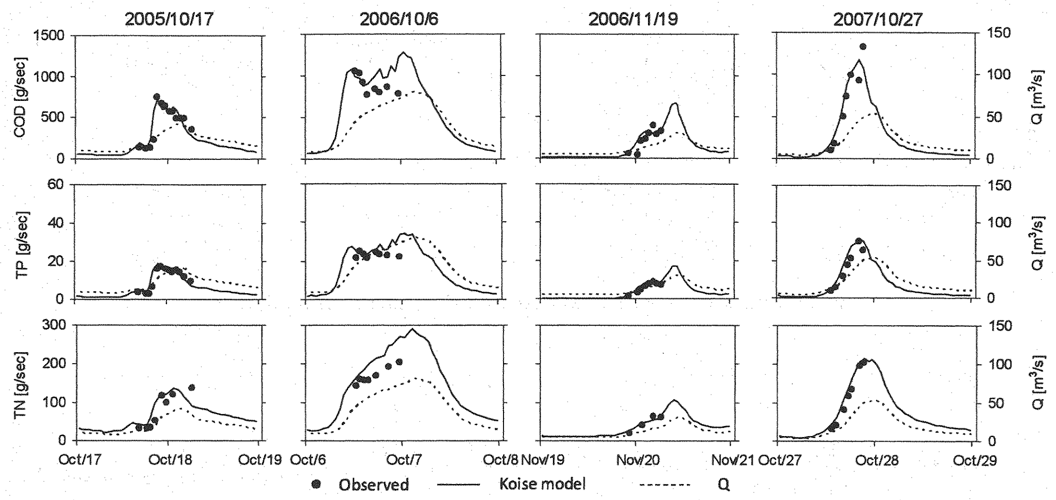

Fig. 7 Enlarged figures of pollutant fluxes of four flood events of the Koise River

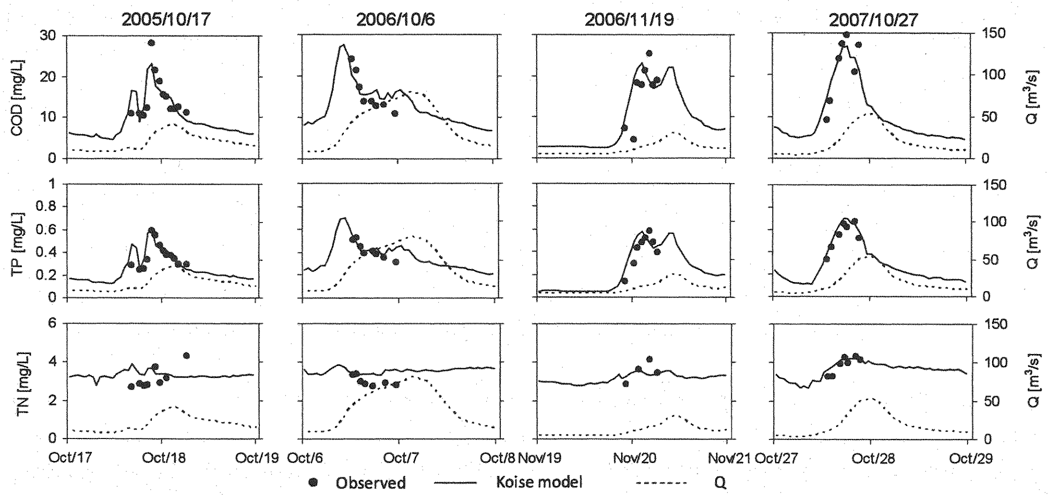

Fig. 8 Enlarged figures of pollutant concentration of four flood events of the Koise River

data obtained in the year 2005 was still effective for the years 2006 and 2007, making very frequent calibration unnecessary. However, the criterion of its frequency is a matter to be examined in the future.

Practicality for application to river basins of different characteristics

ANN models were constructed for the seven rivers listed in Table 1 in order to assess the model practicality to the river basins with different characteristics. The areas and land use classifications of these river basins are very different. The field experiment in these rivers was carried out from September, 2006 except for the Sonobe River where the period was same as the Koise River. The data were also divided into two parts: the data obtained in the year 2006 for training the ANN model and the data of the year 2007 for evaluating the model performance. For the Sonobe River, the data of the year 2005 were used for training and the data of the years 2006 and 2007 were used for evaluating.

Time series of *COD*, *TP* and *TN* fluxes estimated by means of the ANN models are shown in Fig. 9. To save the

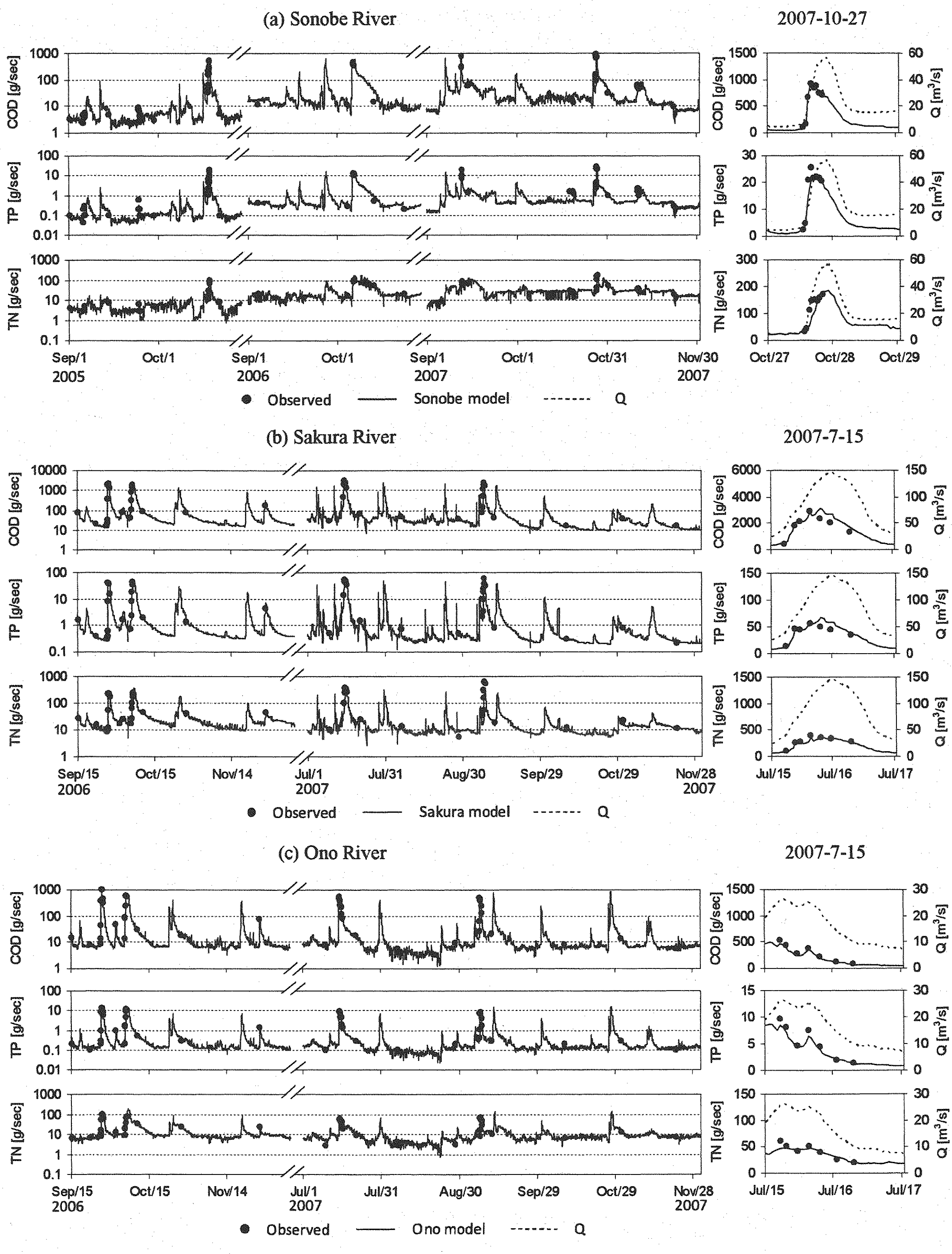

Fig. 9 Time series of pollutant fluxes of three rivers estimated by the ANN models and enlarged figures of one flood event. (a) Sonobe River; (b) Sakura River; (c) Ono River

space only results of the Sonobe River, the Sakura River and the Ono River are shown here. The detailed variations of the same data during one flood event are enlarged. The solid lines show the results estimated by the ANN models, and

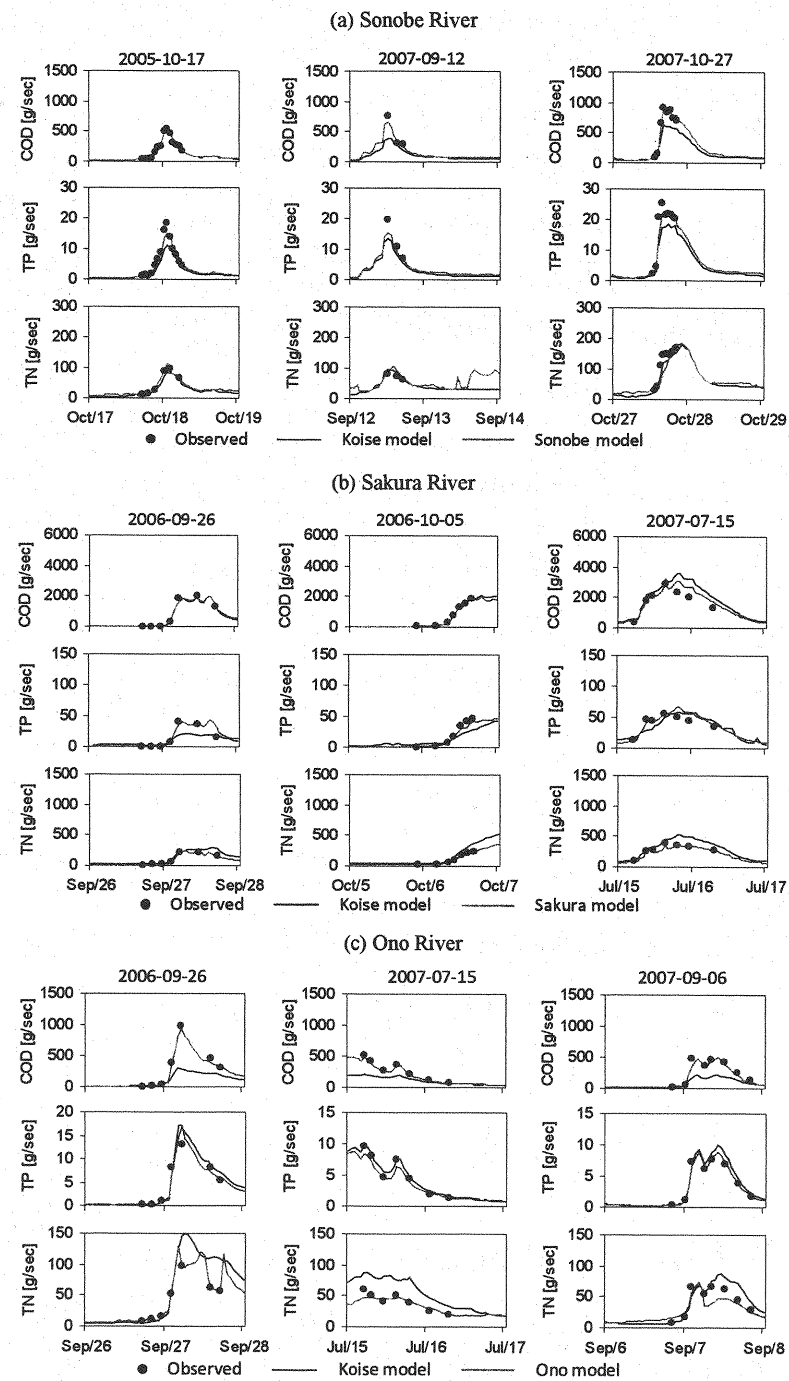

Fig. 10 Enlarged figures of pollutant fluxes of three flood events which were estimated by different ANN models (a) Sonobe River; (b) Sakura River; (c) Ono River

the dots are observed results of chemical analysis. The close agreements of estimated and observed results in the figures suggest that the ANN model can learn the peculiar correlations in river basins where the land use conditions are different from the Koise River when the field data are supplied for its training.

Probability of applying the Koise River model to other rivers

In principle, an ANN model is applicable only to rivers where the training data were collected because the correlation that ANN produces is peculiar to each river. In practice, however, we want to apply the model to different rivers where much field data has not been collected in order to reduce costs and labor for data collection. It might be possible if the land use conditions are similar between the river watersheds. In this section, the ANN model for the Koise River is applied to some nearby rivers in order to test the possibility. The results of applications to three rivers for some flood events are shown in Fig. 10. The black lines indicate the prediction by ANN models of the Koise River and the gray lines represent the estimation by ANN models constructed for each of the three target rivers, respectively. The dots show the results of the chemical analysis of water samples obtained from each river.

The performance of the ANN models trained with the data of the target river is better than the ANN models of the Koise River, of course. However, the difference between the two kind results was not very significant for the Sonobe River and the Sakura River. On the other hand, the agreement was worse in the case of the Ono River, especially for *COD* and *TN*. Reasons for this difference are not clear but it might be explained by looking at the data of land use. As is shown in Table 1, the percentage of "urban areas" and "other use (including wasteland)" is the largest in the Ono River among the seven rivers; the total of these two in the river basin is almost as twice large as others. In this study, the ANN model is used to express peculiar correlations between optical signals and pollutant concentration. The pollutants from the above mentioned types of lands might have different optical characteristics from those from other kinds of lands with vegetation and surface soils such as forests and farm lands.

This fact suggests the possibility that an ANN model constructed in a river basin can be tentatively applied to other river basins where data of water sampling analysis are not enough to calibrate the ANN model if the land use conditions are similar to some extent. It might be useful to make groups of land use in a view point of ANN modeling in order to reduce costs and labor for water sample analysis in flood events.

CONCLUSIONS

The correlations between the optical characteristics of river water and the eutrophication indices which are peculiar to each river are formulated by using the Artificial Neural Network in order to estimate the pollutant loads. The major conclusions are listed below:

(1) The ANN model can reproduce the peculiar correlation between the optical characteristics of river water and the eutrophication indices such as *COD*, *TP* and *TN*.

(2) The ANN model trained by the data obtained in one year can estimate the eutrophication indices in the successive two years with a high accuracy in the Koise River.

(3) The monitoring method introduced in this work is applicable to other six river basins in the watershed of Lake Kasumigaura where the land use conditions are different from the Koise River.

(4) Findings in this research indicate that an ANN model constructed in a river basin can be applied tentatively to other river basins where data of water sampling analysis are not enough to calibrate the ANN model if the land use conditions are similar to some extent.

REFERENCES

1. Räike, A., Pietiläinen,O. P., Rekolainen, S., Kauppila, P., Pitkänen, H., Niemi, J., Raateland, A. and Vuorenmaa, J. : Trends of phosphorus, nitrogen and chlorophyll a concentrations in Finnish rivers and lakes in 1975–2000, The Science of Total Environment, Vol.310, pp.47-59, 2003.
2. Richards, R.P. : Estimation of pollutant loads in rivers and streams: a guidance document fro NPS programs. Project report prepared under Grant X998397-01-0, US Environmental Protection Agency, Region VIII, Denver, pp.7-12, 1998.
3. Takahashi, M., Chida, H. and Matsuo, S. : Examinations of turbidity and water quality at floods in the catchment area of a reservoir, Journal of Hydroscience and Hydraulic Engineering, Vol.43, pp.593-598, 1999.
4. Grayson, R. B., Finlayson, B. L., Gippel, C. J. and Hart, B. T. : The potential of field turbidity measurement for the computation of total phosphorus and suspended solids loads, Journal of Environmental Management, Vol.47, pp.257-267, 1996.
5. Yamamoto, K., Futamura, T. and Sakano, A. : Estimation of particulate nutrients loads using turbidity meter, Advances in River Engineering, Vol.9, pp.515-520, 2003 (in Japanese).
6. Nihei, Y. and Hanayama, T. : Attempt to measure a periphyton concentration in river water with a fluorometer, Journal of Hydraulic, Coastal and Environmental Engineering, No.782/II-70, pp.143-148, 2005 (in Japanese).
7. Kato, S., Sato, K. and Yoshimura, H. : River water quality monitoring by portable ultraviolet photometer case study: Carmona, Cavite Province, Philippines, Annual Report of Mie Prefectural Science and Technology Promotion Center of Public Health and Environment Research Division, Vol.48, No.5, pp.49-54, 2003 (in Japanese).
8. Nakasone, H. : Runoff water quality characteristics in a small agriculture watershed, Paddy and Water Environment, Vol.1, pp.183-188, 2003.
9. Kawara, O., Hiryama, K. and Kunimatsu, T. : A study on pollution loads from the forest and rice paddy fields, Water Science and Technology, Vol.33, No.4-5, pp.159-165, 1996.
10. Liu, M., Ishikawa, T., Yoshimi, K. and Kunihiro, E. : A practical approach to the estimation of pollutant load from optical sensor measurement, Annual Journal of Hydraulic Engineering, Vol.51, pp.193-198, 2007.
11. Yoshimi, K., Liu, M. and Ishikawa, T. : A new practical measurement of pollutant load in rivers by using optical sensors, Advances in River Engineering, Vol.13, pp.219-224, 2007 (in Japanese).
12. Liu, M., Ishikawa, T., Yoshimi, K. and Kudo, K. : A monitoring method for pollutant load in rivers by using an artificial neural network, Annual Journal of Hydraulic Engineering, Vol.52, pp.127-132, 2008.
13. Charef, A., Ghauch, A., Baussand, P. and Bouyer, M. M. : Water quality monitoring using a smart sensing system, Measurement, Vol.28, pp.219-224, 2000.
14. Hsu, K. L., Gupta, H. V. and Sorooshia, S. : Artificial neural network modeling of rainfall-runoff process, Water Resources Research, Vol. 31, pp.2517-2530, 1995.
15. Wei, B., Sugiura, N. and Maekawa, T. : Use of artificial neural network in the prediction of algal blooms, Water Resources Research, Vol.25, pp.2022-2028, 2001.
16. Ahmad, S. and Simonovic, S. P. : An artificial neural network model for generating hydrograph from hydro-meteorological parameters, Journal of Hydrology, Vol.315, pp.236-251, 2005.
17. Chakraborty, K., Mehrotra, K., Mohan, C. K. and Ranka, S. : Forecasting the behavior of multivariate time series using neural networks, Neural Networks, Vol.5, pp.961-970, 1992.
18. Maier, H. R. and Dandy, G. C. : Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, Environmental Modeling & Software, Vol.15, pp.101-124, 2000.
19. Fahlman, S. E. and Lebiere, C. : The cascade-correlation learning architecture, Article created for National Science Foundation under Contract Number EET-8716324 and Defense Advanced Research Projects Agency (DOD), ARPA

Order No.4976 under Contract F33615-87-C-1499, 1991.

20. Fujita, K., Ito, H., Oro T. and Anma, T. : A study on evaluation of water environment policy through watershed-scale hydrological & material cycle simulation models to Kasumigaura Lake and its watershed. Technical note of National Institute for Land and Infrastructure Management. No. 299. pp. 11, chapter 2, 2006.

APPENDIX – NOTATION

The following symbols are used in this paper:

C	= concentration;
Chl-a	= chlorophyll-a;
L	= load;
Q	= river flow rate;
Tb	= turbidity;
X	= input index; and
α and β	= water quality items.

(Received Mar 03, 2008 ; revised Jul 23, 2008)