Annual Journal of Hydraulic Engineering, JSCE, Vol.52, 2008, February

A MONITORING METHOD FOR POLLUTANT LOAD IN RIVERS BY USING ARTIFICIAL NEURAL NETWORK

Minghuan LIU[1], Tadaharu ISHIKAWA[2], Kenji YOSHIMI[3] and Kentaro KUDO[4]

[1]Student Member of JSCE, Doctoral Student
[2]Fellow of JSCE, Dr. of Eng., Professor
[3]Student Member of JSCE, Doctoral Student
[4]Member of JSCE, Dr. of Eng.
[1]Dept. of Civil Engineering, Tokyo Institute of Technology, Nagatsuta 4259, Yokohama 226-8502, Japan
[2, 3]Dept. of Environmental Science and Technology, Tokyo Institute of Technology, Nagatsuta 4259, Yokohama 226-8502, Japan
[4]IDEA Consultants, Inc., Hayabuchi 2-2-2, Yokohama 224-0025, Japan

This paper proposes a new method for continuous measurement of pollutant load in rivers without much cost. The basic idea is making the most of "empirical correlations which exist in the target" in order to relate what we can measure to what we want to know. In a field experiment presented here, signals from two types of optical sensors were used to estimate the loads of chemical oxygen demand (*COD*), total nitrogen (*T-N*) and total phosphorus (*T-P*), and artificial neural network (ANN) models were trained to fix "the empirical correlations" among them. The field data were collected in seven rivers located in the watershed of Lake Kasumigaura. The experimental results showed that the three items of water quality were stably estimated with good accuracy for rather long time without too much training data.

Key Words: *artificial neural network, monitoring, pollutant load, empirical correlation, optical sensor*

1. INTRODUCTION

Reduction and control of pollutant load from non-point sources is the key issue to solve the eutrophication problem in closed water bodies such as lakes, reservoirs, inner bays. In order to take effective actions for the purpose, monitoring of water quality in inflowing rivers with continuous or high frequent measurement is required because the pollutant load from non-point sources increases sharply during a rain runoff[1)].

Recently, immersed-type optical sensors of self-recording have been commercialized for continuous *in situ* measurement for some items of water quality such as *Chl-a*, *D-COD*[2)]. Because optical measurement has a strong point of getting continuous data, it has a large potential to be used in the monitoring of river water. At present, however, the measurable items are very limited, and we cannot obtain the indices of pollutant load relating to eutrophication such as *COD*, *T-P* and *T-N*.

On the other hand, many *in situ* measurements depend on empirical relations obtained "locally" to some extent: Time series of river flow rate (*Q*) is usually obtained from water level (*H*) being based on an empirical correlation between them; *Chl-a* measurement by using a fluorometer needs calibration for the conditions of algae in each water area[3)]; particulate phosphorus (*P-P*) estimation by using turbidity meter needs an empirical correlation obtained in each river[4)]; *L-Q* method for pollutant load (*L*) depends on an empirical correlation between the two factors. However, those techniques work well when the empirical correlations are formulated properly.

Considering the above mentioned conditions, the authors proposed a technique of water quality measurement in which signals from an optical measurement device are converted to time series of *COD*, *T-P* and *T-N* being based on empirical correlations obtained in each river[5)]. The technique was extended by introducing ANN (artificial neural network) for modeling the empirical relations[6)]. In the previous two papers, due to limit of the range of

Table 1 Land-uses in the watershed of the present study

Name of river	Area(km^2)	Land-use (%)					Measurement period
		Paddy	Cropland	Forest	Urban use	Other use*	
Koise	144.6	19.5	18.6	***49.4*** †	7.0	5.5	6/1/2005 - 12/1/2005& 5/18/2006 - 8/1/2007
Sonobe	71.6	15.3	41.9	23.8	13.1	5.9	
Sakura	333.0	***28.1*** †	16.9	34.6	11.7	8.7	9/15/2006 - 8/1/2007
Seimei	25.0	14.0	32.8	18.2	***21.5*** †	***13.5*** †	
Ono	144.8	20.0	29.1	22.7	18.5	9.7	
Tomoe	113.2	15.9	46.8	21.6	11.0	4.7	
Hokota	39.5	11.3	***53.9*** †	24.2	7.5	3.1	

* Other use includes golf field, wild land and water surface.
† The largest for each land-use among the seven river basins.

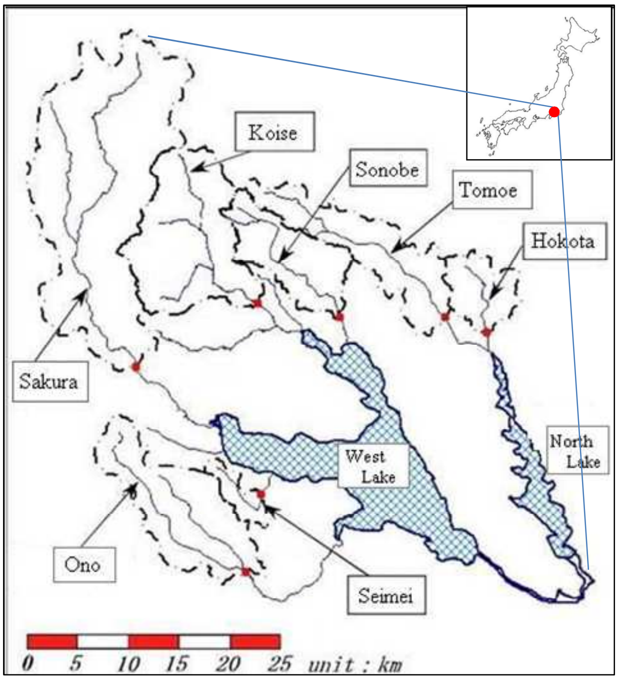

Fig. 1 Location of measurement sites

data, so its general practicality was not well defined.

In this paper, the following points are discussed being based on a plenty of data obtained recently: (1) A long term application of an empirical model constructed by ANN. (2) Practicality of the ANN modeling in river basins with different land uses. (3) Comparing an ANN model for estimation of pollutant concentration with another ANN model for direct estimation of pollutant load.

2. ARTIFICIAL NEURAL NETWORK

Recently, ANN is widely used for empirical modeling of hidden dynamics in the environment[7-10]. It is said that ANN belongs to a class of data driven approach whereas conventional statistical methods are model driven[11]. In other words, ANN is more flexible than conventional statistic methods to catch complex relations among environmental data, or, in a sense, ANN is able to simplify the procedure of statistical analysis of the complex relations[12].

In this study, the software named "Predict" supplied by the Neuralware Company was adopted to model the correlation of pollutant load L (COD, $T\text{-}P$, $T\text{-}N$) with two signals from an optical measurement device (Compact-CLW: ALEC Electronics) X_1, X_2 and a time series of river flow rate Q.

$$L = f\,(X_1, X_2, Q) \qquad (1)$$

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[13]. One of the benefits of this method is that the network retains the structure even if the training set changes, and it requires no back-propagation of error signals through the connection of the network. In our problems, therefore, even if the condition of river basin changes gradually in time, the model can be easily improved by adding training data obtained from recent measurement.

3. FIELD EXPERIMENT

Field experiment was carried out in seven rivers flowing into Lake Kasumigaura. **Fig. 1** shows the river basins and the locations of measurement sites. An immersed-type optical measurement device (Compact-CLW: ALEC Electronics) was placed in the low water channel of each river. Measurement interval was 10 minutes. Surface water was sampled from a bridge during floods as well as the time of normal condition. The items and the methods of water analysis were identical with those described in previous papers[5), 6)]. **Table 1** shows the area and the land-uses of the river basins upstream from the measurement sites with the period of experiment in

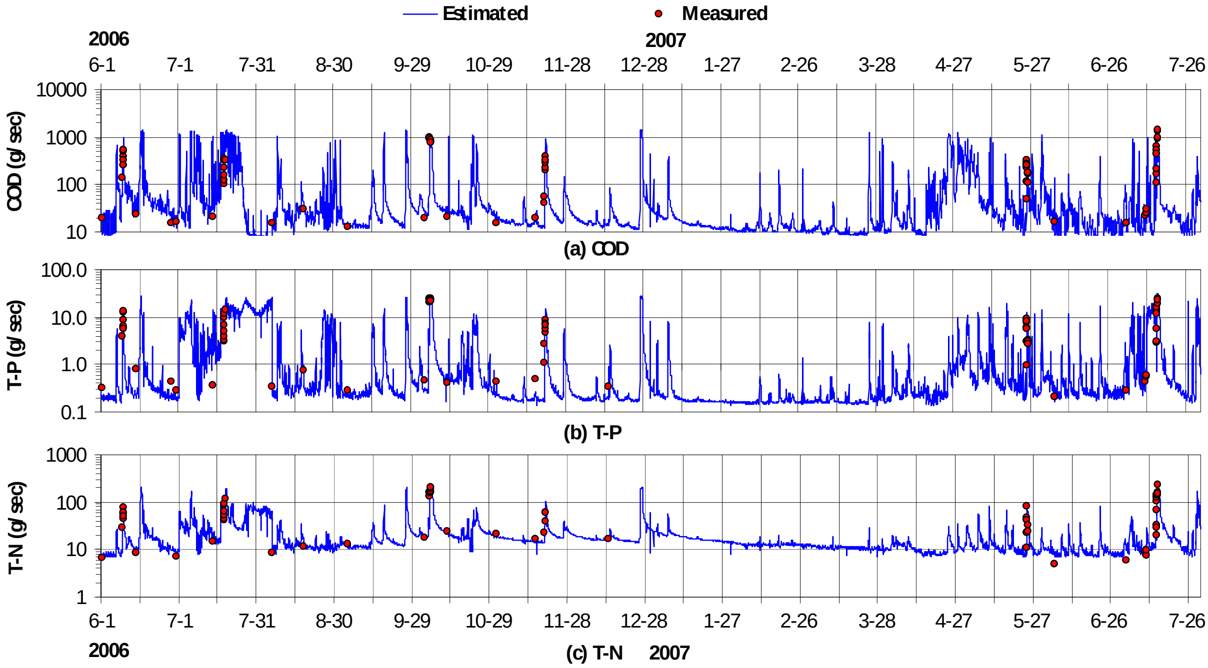

Fig. 2 Time series of *COD*, *T-P* and *T-N* load of the Koise River from the year 2006 through the year 2007

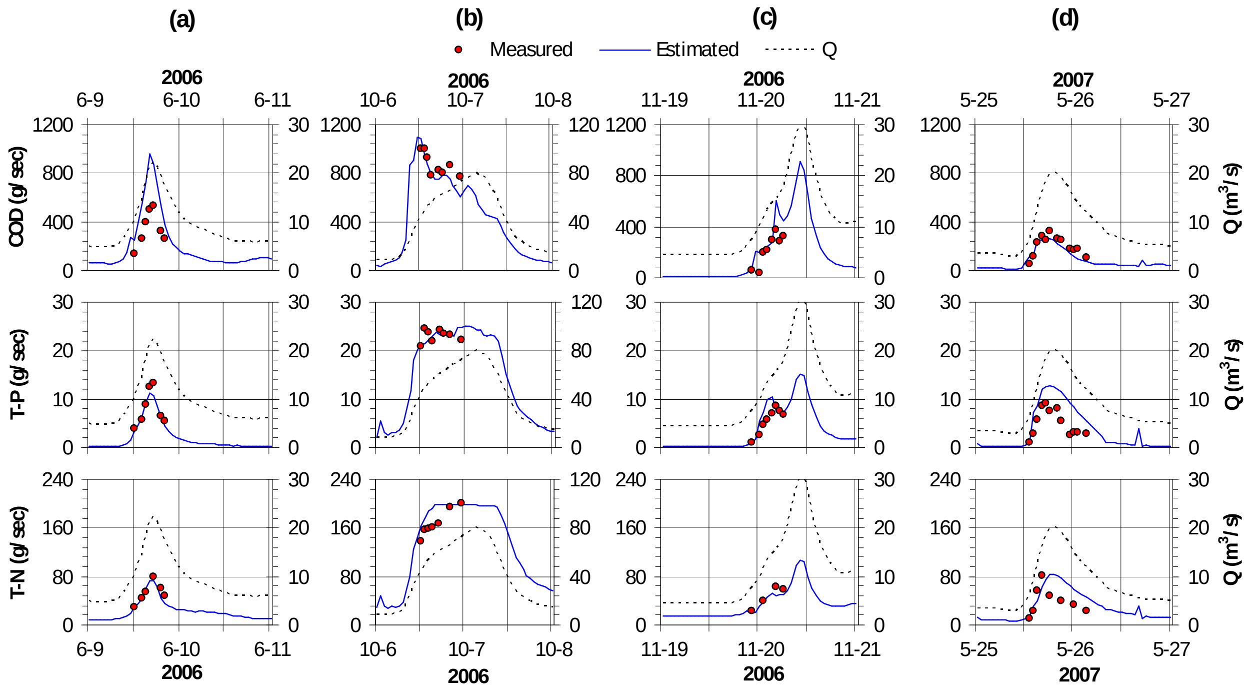

Fig. 3 Enlarged figures of four storm events of pollutant load of the Koise River in the years 2006 and 2007
(a) 2006.6.9-10; (b) 2006.10.6-7; (c) 2006.11.19-20; (d) 2007.5.25-26

each river. The land-uses of the river basins are very different from one to another.

4. RESULTS AND DISCUSSION

(1) Practicality for long term application

Data of the Koise River, in which the experimental period was the longest and the volume of data was the largest among the seven rivers, was used to discuss the long term application of an empirical model constructed by the ANN. The data was divided into two parts: the data obtained in the year 2005 for training the model and the data in the years 2006 and 2007 for evaluating the performance of the model. The number of training data is 53 for floods and 14 for normal condition, while the number of verification data is 54 for floods and 19 for normal condition.

Fig. 2 shows the verification results. The blue solid lines in the figures show the time series of

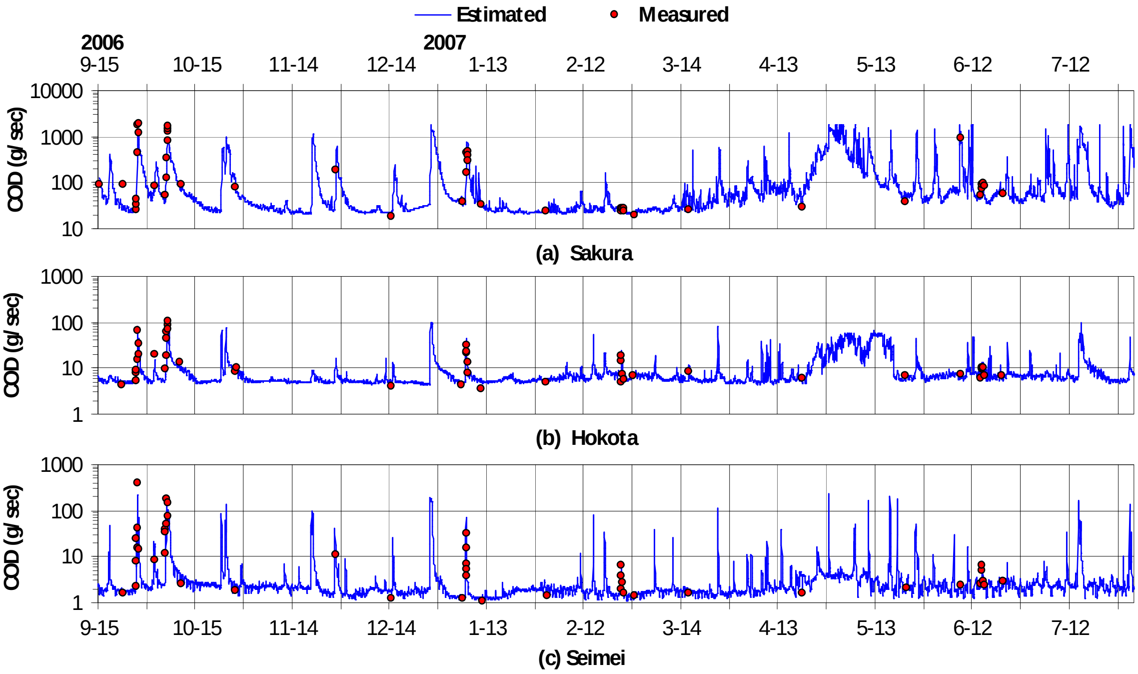

Fig. 4 Time series of *COD* load of other rivers from the year 2006 through the year 2007
(a) the Sakura River; (b) the Hokota River; (c) the Seimei River

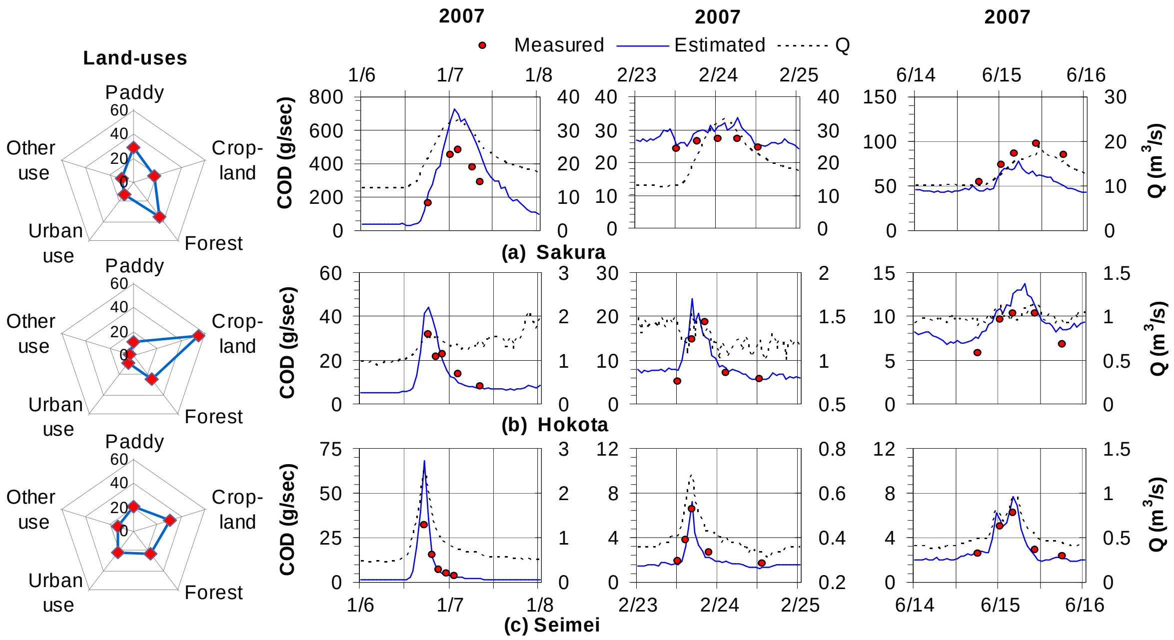

Fig. 5 Land-uses and enlarged figures of three storm events for each river in the year 2007
(a) the Sakura River; (b) the Hokota River; (c) the Seimei River

(a) *COD*, (b) *T-P* and (c) *T-N* for the years 2006 and 2007 produced by the ANN model that was trained with the data of the year 2005. Red dots show the results of water analysis. The detailed variations of the same data during four flood events are shown in **Figs. 3**. Dotted lines in the figures show the river flow rate. There are small discrepancies in some parts, but the general agreement seems good although they were deduced from one model that was calibrated in a different year.

The scales of the hydrograph and pollute-graph shown in **Figs. 3** vary from one flood to another flood. From the **Fig. 3(b)**, the peak discharge of the flood is not synchronous with the peak loads of *COD*, *T-P* and *T-N*. This kind of phenomena cannot be described by the conventional *L-Q* method. The fact shows the possibility that introduction of the proposed method will improve the estimation accuracy of pollutant load remarkably.

Needless to say, every field measurement based on empirical relations will fail if site condition changes. Therefore, the ANN model should be

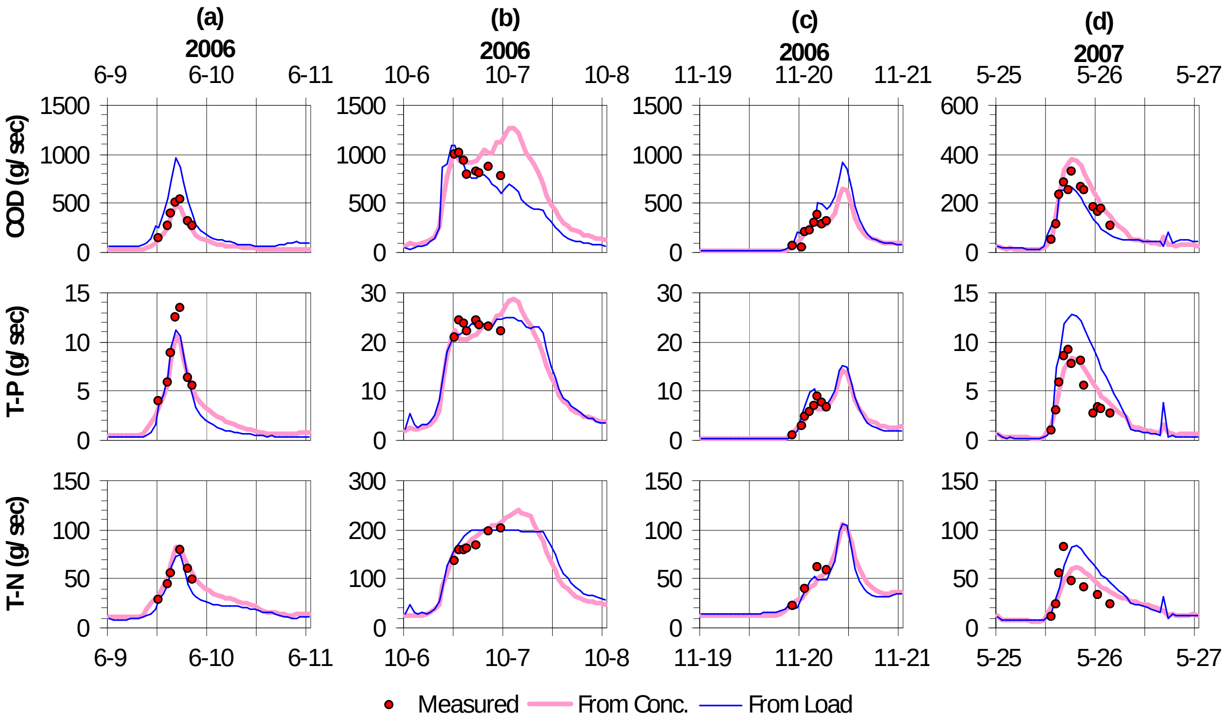

Fig. 6 Comparison of load estimation of the Koise River by different ANN models
(a) 2006.6.9-10; (b) 2006.10.6-7; (c) 2006.11.19-20; (d) 2007.5.25-26

trained with new data at some intervals. As **Fig. 2** showed that the training with the data obtained in the year 2005 was still effective for the years 2006 and 2007, very frequent calibration may not be necessary. However, the criterion of its frequency is a subject to be studied in the future.

(2) Practicality for application to river basins of different characteristics

ANN models were constructed for the seven rivers listed in **Table 1** in order to discuss the model practicality to the river basins of different characteristics. Because of the restriction of space, the results of the Sakura River, the Hokota River and the Seimei River are presented in this paper. The land uses of the river basins are very different from one to another.

The field experiment in those rivers was carried out from September 15, 2006 to August 1, 2007. The data was divided into two parts: the data obtained in the year 2006 for training the ANN model and the data in the year 2007 for evaluating the performance of it. The number of training data is 14 for floods and 7 for normal condition, while the number of verification data is 15 for floods and 9 for normal condition on the average.

Figs. 4 show the results for *COD* in the same way as **Figs. 2**. **Figs. 5** show the detailed variations of the same data during three storm events, which are different in scale of flood to each other. The pentagonal graphs on the left hand side show the land use of each river basin. The general agreements are worse than the case of the Koise River shown in **Figs. 2** and **Fig. 3**. The reasons may be: 1) the period of training data was short; 2) the season of training data is different from that of verification data. In this case, storm events included in the training data occurred in autumn (September. 26 and October 5, 2006), while storm events in the verification data occurred from winter to early summer. However, in spite of the problems of training data, the ANN model estimates *COD* in rivers of different characteristics fairly well.

(3) Comparison of two different methods of load estimation

Because pollutant load is a product of pollutant concentration and flow rate, there can be another use of ANN based on the following equation:

$$L = C(X_1, X_2)\times Q \qquad (2)$$

where Q is separated from other variables, and ANN is used to estimate $C(X_1, X_2)$. In this case, the number of input variables is reduced from three in Eq.(1) to two in Eq.(2), which reduces the freedom of the ANN model a little bit: If the dependency of C on X_1 and X_2 is changed by some unknown factors and if the factors have some local correlation with flow rate, then Eq.(1) will estimate L better than Eq.(2). But, if the correlation between the variables is absent or small, the scheme of

Eq.(1) may mislead the ANN model because of much freedom.

Fig. 6 shows the comparison of the two methods for four flood events. The scheme of Eq.(2) seems slightly better than that of Eq.(1). However, because the difference between them is not very clear, we cannot make decision at this moment.

5. CONCLUSIONS

The environment is very complex. In order to acquire more information from the environment, we should make full use of existing knowledge including empirical and local ones rather than simply pursuing the perfect or universal knowledge of science. In the previous study[5], the authors proposed an idea of utilizing empirical and local correlations to estimate time series of pollutant loads in rivers. Artificial Neural Network must be a power tool to model the correlations because we behave successfully in daily life by utilizing the empirical and local correlations stored in our human neural networks. Major conclusions of this study are as follows.

(1) ANN software "Predict" based on Cascade-Correlation Learning Architecture which was calibrated with the data of the year 2005 successfully estimated the time series of *COD*, *T-P* and *T-N* load in the Koise River in the year 2006 and 2007.

(2) The results of application to other rivers showed that the ANN model is applicable to river basins of different characteristics fairly well though the estimation accuracy was slightly lower than the case of the Koise River because of the lack of the data amount.

(3) An ANN model for estimation of pollutant concentration based on Eq.(2) was compared to another ANN model for load estimation based on Eq.(1). The results from the former seemed slightly better than the latter, but the difference was so small that the conclusion on this point could not be made.

This fact suggests that introduction of the proposed method will improve the estimation accuracy of pollutant load remarkably.

ACKNOWLEDGEMENT: Part of data presented in this paper was supplied by the Kasumigaura River Office, Ministry of Land, Infrastructure and Transport of Japan. We would like to thank them for the help. The authors are also grateful to the stuff of IDEA Consultants, Inc. and all members in Ishikawa Laboratory for their assistance in the field collection.

REFERENCES

1) Cohn, T. A., Caulder, D. L., Gilroy, E. J., Zynjuk, L. D. and R.M. Summers, The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay, *Water Resources Research*, vol.28(9), pp. 2353-2363, 1992.

2) Kato, S., Sato, K., 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(5), pp.49-54, 2003.

3) Nihei, Y. and Hanayama, T., Attempt to measure a periphyton concentration in river water with a fluorometer (in Japanese), *Journal of Hydraulic, Coastal and Environmental Engineering*, No.782/II-70, pp.143-148, 2005.

4) Yamamoto, K., Futamura, T. and Sakano, A., Estimation of particulate nutrients loads using turbidity meter (in Japanese). *Advances in River Engineering* vol.9, pp.515-520, 2003.

5) 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.

6) 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.

7) 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.

8) 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.

9) 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.

10) 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.

11) 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.

12) Maier, H. R., 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.

13) 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.

(Received September 30, 2007)